ÜBER DIE IN DER INTERVALLRECHNUNG
AUFTRETENDEN RÄUME
UND EINIGE ANWENDUNGEN

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Naturwissenschaften I
der Universität Karlsruhe
genehmigte
Dissertation

von
Diplom-Mathematiker
Otto Mayer
aus München

Tag der mündlichen Prüfung: 25. November 1968
Referent: Prof. Dr. U. Kulisch
Korreferent: Prof. Dr. K. Nickel

Inhalt

		Seite
	Einleitung	3
1.	Rechnen mit Intervallen; die Menge $I(R)$	8
2.	Der quasilineare Raum	15
2.1	Zur Definition des quasilinearen Raumes	15
2.2	Norm und Metrik in quasilinearen Räumen	17
2.3	Ein Beispiel für eine supermetrische und nicht translationsinvariante Metrik in $I(R)$	20
3.	Die Räume der Intervallrechnung	23
3.1	Intervallrechnung für Vektoren und Matrizen	23
3.2	Ordnungsstrukturen in $I(R)$, $V_n(I(R))$ und $M_{n\times m}(I(R))$	26
3.3	Die Halbordnungen in $V_n(I(R))$ als Bilder von Halbordnungen in R^{2n}	30
3.4	Intervallrechnung über allgemeinen halbgeordneten linearen Räumen	33
4.	Normen in $V_n(I(R))$ und $M_n(I(R))$	37
4.1	Normen und Abstandsnormen	37
4.2	Monotone Normen	41
4.3	Verträgliche und zugeordnete Normen	45
4.4	Über den Zusammenhang der multiplikativen monotonen Normen für Matrizen $\mathfrak{A} \in M_n(I(R))$ bzw. $M_n(R)$ mit dem Spektralradius $\varrho(\|\mathfrak{A}\|)$ des Absolutbetrags $\|\mathfrak{A}\|$	53
5.	Metriken in den Räumen $V_n(I(R))$ und $M_n(I(R))$	57
5.1	Monotone Normen und zugeordnete Metriken	57
5.2	Über die Beziehungen zwischen Normen und Metriken in $V_n(I(R))$ bzw. $M_n(I(R))$ und die Existenz verträglicher Normen und Metriken	62
5.3	Durchmesser für Elemente aus $I(R)$, $V_n(I(R))$ bzw. $M_n(I(R))$	64

		Seite
6.	Die Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$	68
6.1	Das Konvergenzverhalten der Iteration $\varphi_{\nu+1} = \mathfrak{A}\varphi_\nu + \mathfrak{b}$	68
6.2	Gesamt- und Einzelschrittverfahren zur Lösung der Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$	73
7.	Über die Bestimmung von Einschließungsmengen für die Lösungen reeller linearer Gleichungssysteme mit nichtkonstanten Koeffizienten	83
7.1	Allgemeine Überlegungen	
7.2	Abschätzungen für die Normen und Durchmesser der Lösung φ^* der Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$, sowie der Hülle φ_T^* zur Lösung φ_T^* der Gleichung $\varphi_T = \mathfrak{A}_T\varphi_T + \mathfrak{b}_T$	93
8.	Iteration mit Einschließungsmengen bei der Lösung der Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$	97
8.1	Allgemeine Überlegungen	97
8.2	Ein Verfahren zur Bestimmung einer Einschließungsmenge für die Lösung der Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$	108
9.	Literaturverzeichnis	123

Einleitung

In der vorliegenden Arbeit werden Struktur und Eigenschaften der in der Intervallrechnung auftretenden Räume untersucht und darauf aufbauend einige für die Intervallrechnung typische Anwendungsmöglichkeiten aufgezeigt.

Die Aufgabe, bei der numerischen Behandlung mathematischer Probleme auf digitalen Rechenautomaten die Rundungsfehler mit automatischen Methoden zu erfassen, führt zwangsläufig auf eine Intervallrechnung. Man verfolgt nämlich dabei das Ziel, als Ergebnis einer numerischen Rechnung nicht mehr nur irgendeinen Näherungswert, sondern sichere Schranken für die exakte Lösung des gegebenen Problems anzugeben; dies aber setzt ein Rechnen mit abgeschlossenen Intervallen reeller Zahlen voraus, das noch auf der Maschine geeignet approximiert werden muß. Weil auch alle Algorithmen zunächst in der exakten über dem Körper der reellen Zahlen aufgebauten Intervallrechnung hergeleitet werden müssen, ist eine genaue Kenntnis der Struktur und Eigenschaften der dabei auftretenden Räume unerläßlich. Diese Strukturen und Eigenschaften werden in der vorliegenden Arbeit untersucht. Fragen der Approximation der über dem Körper der reellen Zahlen aufgebauten Intervallrechnung auf der Rechenanlage werden nur bei der Besprechung von Anwendungen kurz gestreift; sie wurden bereits in [1], [10], [12], [15] weitgehend behandelt.

Übersicht:

Bereits R. C. Young und R. E. Moore definieren eine Arithmetik in der Menge $I(R)$ der abgeschlossenen Intervalle über dem Körper R der reellen Zahlen. Die dadurch in $I(R)$ erklärte algebraische Struktur unterscheidet sich von bekannten Strukturen wie Körpern oder linearen Räumen zunächst durch das Fehlen der inversen Elemente. Sie wird hier im Anschluß an die Arbeiten von N. Apostolatos und U. Kulisch weiter untersucht; als kennzeichnend erweist sich dabei vor allem ein eingeschränktes Distributivgesetz. Die wesentlichen Eigenschaften dieser Struktur werden dann axiomatisch erfaßt in der abstrakten Definition des quasilinearen Raumes. Diese Verallgemeinerung des linearen Raumes ist nun grundlegend für die gesamte Intervallrechnung. Zunächst erweisen sich nämlich alle bisher in der Intervallrechnung bekannten Räume als quasilinear; darüberhinaus wird hier allgemein eine Intervallrechnung über beliebigen halbgeordneten linearen Räumen definiert und gezeigt, daß diese stets auf quasilineare Räume führt. Ausgehend von der abstrakten Definition werden nun ihre algebraischen Eigenschaften allgemein untersucht.

Normen werden in quasilinearen Räumen durch dieselben Forderungen wie in linearen Räumen definiert; anders als in linearen normierten Räumen gibt aber hier (wegen des Fehlens der inversen Elemente) die Norm der Differenz zweier Elemente keine Metrik.

Im Hinblick auf Anwendungen benötigt man nun einen Überblick über die in quasilinearen Räumen möglichen metrischen Strukturen und daraus folgende Eigenschaften dieser Räume; vor allem interessiert man sich für metrische Strukturen, die gewisse Verträglichkeitseigenschaften in bezug auf die algebraische Struktur besitzen.

Für die diesbezüglichen Untersuchungen wird hier eine sogenannte supermetrische Eigenschaft definiert; in dieser kommt wie in den bekannten Eigenschaften der Homogenität und Translationsinvarianz eine Verträglichkeit der Metrik mit der algebraischen Struktur zum Ausdruck; dabei ist diese supermetrische Eigenschaft in quasilinearen Räumen eine Abschwächung der Translationsinvarianz einer Metrik.

Quasilineare Räume, die vollständig sind bezüglich einer homogenen supermetrischen Metrik sind nun Verallgemeinerungen von Banachräumen, sie werden hier Quasi-Banachräume genannt.
Für alle bisher in der Intervallrechnung bekannten Räume werden Metriken angegeben, mit denen sich diese als Quasi-Banachräume erweisen; darüberhinaus wird gezeigt, welche allgemeinen Voraussetzungen ein halbgeordneter linearer Raum erfüllen muß, damit die darauf aufgebaute Intervallrechnung einen Quasi-Banachraum ergibt.

Neben diesen allgemeinen Überlegungen werden auch spezielle Eigenschaften von Normen und Metriken in den für die Anwendungen besonders wichtigen Räumen der Intervallrechnung untersucht. Betrachtet werden hier der schon erwähnte Raum $I(R)$ der abgeschlossenen Intervalle über dem Körper R der reellen Zahlen, sowie die Räume $V_n(I(R))$ und $M_n(I(R))$ der n-tupel (Vektoren) und $n \times n$-Matrizen mit Komponenten aus $I(R)$, denen die linearen Räume $V_n(R)$ und $M_n(R)$ der n-tupel und $n \times n$-Matrizen mit Komponenten aus R zugrundeliegen.
Dabei ist $R \subset I(R)$, $V_n(R) \subset V_n(I(R))$ und $M_n(R) \subset M_n(I(R))$.
Die Untersuchungen über die Eigenschaften von Normen für Vektoren aus $V_n(I(R))$ und Matrizen aus $M_n(I(R))$ zeigen, daß monotone Normen für die Intervallrechnung besonders geeignet sind.
In $I(R)$, $V_n(I(R))$ und $M_n(I(R))$ fehlt wie in allen quasilinearen Räumen der von linearen normierten Räumen bekannte Zusammenhang zwischen Norm und Metrik; deshalb liegt ein wichtiges Problem darin, hier Metriken anzugeben, die auch für praktische Anwendungen, d. h. für Abschätzungen und Konvergenzuntersuchungen geeignet sind.
Im Anschluß an die Untersuchungen über die algebraische Struktur im Raum $I(R)$ werden hier für eine schon von R. E. Moore in $I(R)$ angegebene Metrik Eigenschaften bewiesen, die erst diese Metrik für Abschätzungen geeignet machen.

Aufbauend auf den Ergebnissen über die Eigenschaften monotoner Normen und die Eigenschaften der oben erwähnten Metrik in $I(R)$ wird nun gezeigt, daß sich alle durch verträgliche monotone Normen in $V_n(R)$ und $M_n(R)$ gegebene Metriken gleichzeitig mit diesen Normen so auf die umfassenderen quasilinearen Räume $V_n(I(R))$ und $M_n(I(R))$ fortsetzen lassen, daß auch die sich dort ergebenden Metriken und Normen Verträglichkeitseigenschaften wie die Ausgangsnormen besitzen und damit in gleicher Weise für Abschätzungen geeignet sind.

Am Beispiel des Gleichungssystems $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ (mit $\mathfrak{A} \in M_n(I(R))$ und $\mathfrak{b} \in V_n(I(R))$) wird gezeigt, wie diese Hilfsmittel angewendet werden können. Zunächst wird eine hinreichende Bedingung für die Existenz einer eindeutigen Lösung, sowie ein notwendiges und hinreichendes Kriterium für die Konvergenz der Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ zur Bestimmung dieser Lösung hergeleitet. Dann werden verschiedene Möglichkeiten der iterativen Lösung untersucht und miteinander verglichen. Anwendungen, die auf die Lösung eines solchen Gleichungssystems mit Intervallkoeffizienten führen, werden besprochen; vor allem wird gezeigt, daß die Lösung dieses Gleichungssystems die Lösungen aller reellen linearen Gleichungssysteme $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in \mathfrak{A}$ und $\mathfrak{b} \in \mathfrak{b}$ enthält und in einigen wichtigen Fällen die bestmögliche Einschließungsmenge für die Gesamtheit aller dieser Lösungen gibt.

Abschließend wird die praktische Bestimmung der Lösung einer Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ auf der Rechenanlage behandelt.

Die diesbezüglichen Überlegungen zeigen, daß die in einer Maschinenintervallarithmetik durch Iteration berechnete Approximation für die exakte Lösung diese nur dann mit Sicherheit enthält, wenn die Iteration mit Einschließungsmengen durchgeführt wird. Einige Sätze über die Iteration mit Einschließungsmengen werden bewiesen; darauf aufbauend wird ein Verfahren zur Bestimmung eines für die Iteration mit Einschließungsmengen geeigneten Anfangsvektors entwickelt und gezeigt, wie dieses Verfahren anzulegen ist, damit der gesamte Rechenaufwand für die Bestimmung dieses Anfangsvektors und die anschließende Iteration minimal wird.

Aus den Überlegungen zu diesen Fragen ergibt sich dann auch als neue Erkenntnis, daß die Maschinenintervallarithmetik die (von

der exakten Intervallrechnung über dem Körper der reellen Zahlen her bekannte) Teilmengeneigenschaft besitzen muß, wenn in ihr für die Inklusion Monotonieeigenschaften wie in der exakten Intervallrechnung gelten sollen.

1. Rechnen mit Intervallen; die Menge $\mathbb{I}(\mathbb{R})$

Im Anschluß an [1], [8] und [17] definieren wir zunächst eine Arithmetik für abgeschlossene Intervalle:

$\mathbb{R}$ sei der Körper der reellen Zahlen; Elemente aus $\mathbb{R}$ bezeichnen wir mit a, b, c, ...
$\mathbb{I}(\mathbb{R})$ sei die Menge der abgeschlossenen Intervalle über R; für $a_1, a_2 \in R$ mit $a_1 \le a_2$ ist dann $[a_1,a_2] := \{a \mid a_1 \le a \le a_2\} \in \mathbb{I}(\mathbb{R})$;
Elemente aus $\mathbb{I}(\mathbb{R})$ bezeichnen wir mit A, B, C, ...

Definition 1.1: Zwei Intervalle $A = [a_1,a_2]$ und $B = [b_1,b_2]$ heißen gleich,
i. Z.: $A = B$, genau dann, wenn $a_1 = b_1$ und $a_2 = b_2$ ist.

Dieser Gleichheitsbegriff ist reflexiv, symmetrisch und transitiv.

Definition 1.2: Bezeichnet $*$ eine der vier Grundrechenoperationen $+, -, \cdot, :$, so sei für $A = [a_1,a_2]$, $B = [b_1,b_2]$:
$A*B = [a_1,a_2] * [b_1,b_2] := \{c = a*b \mid a_1 \le a \le a_2,\ b_1 \le b \le b_2\}$.
Im Falle der Division setzen wir dabei $0 \notin B$ voraus.

Diese Verknüpfungen führen bekanntlich nicht aus $\mathbb{I}(\mathbb{R})$ hinaus; $\mathbb{I}(\mathbb{R})$ erhält damit eine algebraische Struktur. Die Teilmenge der Intervalle $[a,a]$ ist dabei isomorph dem Körper der reellen Zahlen. Durch die Identifizierung von $[a,a] \in I(R)$ mit $a \in R$ wird $R \subset I(R)$; Elemente aus der Teilmenge $R \subset I(R)$ bezeichnen wir zur Unterscheidung von beliebigen Elementen aus $I(R)$ auch mit $\underline{A}, \underline{B}, \underline{C}, \dots$

Addition und Multiplikation sind kommutativ und assoziativ:

$$A+B=B+A, \qquad A+(B+C)=(A+B)+C,$$
$$AB=BA, \qquad (AB)C=A(BC).$$

Addition bzw. Multiplikation haben je ein neutrales Element, nämlich die Zahlen 0 bzw. 1; es ist $0+A=A+0=A$ bzw. $1A=A1=A$.

$\mathbb{I}(\mathbb{R})$ ist jedoch bezüglich dieser Arithmetik kein Körper; dazu wird in [1] gezeigt:

a) Ein beliebiges Element $A=[a_1,a_2] \in \mathbb{I}(\mathbb{R})$ mit $a_1 \neq a_2$ besitzt keine Inversen bezüglich Addition und Multiplikation.

Beweis: Es seien $A=[a_1,a_2]$, $B=[b_1,b_2] \in \mathbb{I}(\mathbb{R})$.
Aus $A+B=[a_1+b_1, a_2+b_2]=0$ folgt $a_1+b_1=a_2+b_2=0$;
mit $a_1 \le a_2$ und $b_1 \le b_2$ erhält man $a_1=a_2$, $b_1=b_2$ und $a_1+b_1=0$.
Aus $AB=1$ folgt $\min(a_i b_j)=\max(a_i b_j)=1$, d. h.
$a_1b_1=a_1b_2=a_2b_1=a_2b_2=1$; also ist $a_1=a_2$, $b_1=b_2$ und $a_1b_1=1$.

b) Das distributive Gesetz ist nicht erfüllt, es gilt aber das sogenannte subdistributive Gesetz:

$$A(B+C) \subset AB+AC\,.$$

Beweis: Nach Definition der Verknüpfungen in $\mathbb{I}(\mathbb{R})$ ist
$A(B+C) = \{a(b+c) \mid a \in A,\ b \in B,\ c \in C\} \subset \{ab+\tilde{a}c \mid a,\tilde{a} \in A,\ b \in B,\ c \in C\} = AB+AC$

Wir erwähnen noch einen für die folgenden Überlegungen wichtigen Spezialfall:
Für $a, b \in \mathbb{R}$ mit $ab \ge 0$ und $C := [c_1,c_2] \in \mathbb{I}(\mathbb{R})$ gilt:

$$(a+b)C = aC+bC$$

Beweis:

$$(a+b)C = \begin{cases} [(a+b)c_1,\ (a+b)c_2] & \text{für } a,\ b \ge 0 \\ [(a+b)c_2,\ (a+b)c_1] & \text{für } a,\ b \le 0 \end{cases}$$

$$aC+bC = \begin{cases} [ac_1+bc_1,\ ac_2+bc_2] & \text{für } a,\ b \ge 0 \\ [ac_2+bc_2,\ ac_1+bc_1] & \text{für } a,\ b \le 0 \end{cases}$$

In $\mathbb{R}$ gilt das distributive Gesetz; mit $ac_1+bc_1=(a+b)c_1$ usw. folgt die Behauptung.

Für $ab \le 0$, $\mathbb{R} \not\ni C \in \mathbb{I}(\mathbb{R})$ gilt dagegen nur $(a+b)C \subset aC+bC$
Beispiel: $0=(1-1)C \subset C-C \neq 0$.

Wegen dieser besonderen Eigenschaften muß in einem Intervallausdruck die Reihenfolge der auszuführenden Rechenoperationen etwa durch Klammerstruktur eindeutig festgelegt sein. Abänderung der Reihenfolge und Umformungen wie in einem Körper sind hier nicht zulässig.

Eine wichtige Eigenschaft der Intervallarithmetik ist noch die sogenannte Teilmengeneigenschaft (siehe wieder [1]):

Bedeutet $*$ eine der Verknüpfungen $+$, $-$, $\cdot$, $:$, so folgt aus $A_1 \subset B_1$ und $A_2 \subset B_2$ stets $A_1 * A_2 \subset B_1 * B_2$.

Dies ergibt sich sofort aus der Definition der Verknüpfungen. Durch vollständige Induktion zeigt man ihre Gültigkeit für beliebige rationale Intervallausdrücke, d. h. aus Größen $A_1, \ldots A_n$ mittels der vier Grundoperationen aufgebaute Ausdrücke $f(A_1 \ldots A_n)$, in denen die Reihenfolge der Rechenoperationen eindeutig festgelegt ist.

Im weiteren verstehen wir unter $I(R)$ immer die Menge der abgeschlossenen Intervalle, ausgestattet mit der oben definierten algebraischen Struktur. Einige für die weiteren Überlegungen wichtige Aussagen fassen wir in einem Satz zusammen:

<u>Satz 1.1:</u> Zu je zwei Elementen A, $B \in I(R)$ gibt es genau ein Element $A+B \in I(R)$; für jedes $A \in I(R)$ und jedes $a \in R$ gibt es genau ein Element $aA \in I(R)$.
Dabei gilt für die Addition in $I(R)$ und die skalare Multiplikation der Elemente aus $I(R)$ mit Elementen aus R:

1) $A+B=B+A$,
2) $A+(B+C)=(A+B)+C$,
3) $a(B+C)=aB+aC$,
4) $(a+b)C=aC+bC$ für $ab \geq 0$,
5) $a(bC)=(ab)C$,
6) $0A=0$ für alle $A \in I(R)$,
7) $1A=A$ für alle $A \in I(R)$.

Auf die Bedeutung der Aussagen von Satz 1.1 gehen wir in Abschnitt 2.1 näher ein. Sie bilden die Grundlage für eine axiomatische Erfassung der dem Raum $I(R)$ und ähnlich gebildeten Räumen zugrunde liegenden algebraischen Struktur.

In $I(R)$ kann auf vielfältige Weise eine Metrik definiert werden; wir verwenden hier eine schon in [9] angegebene Metrik:

<u>Definition 1.3:</u> Als Abstand $q(A,B)$ von zwei Intervallen $A=[a_1,a_2]$ und $B=[b_1,b_2]$ aus $I(R)$ erklären wir nach [9]:

$$q(A,B):=\max\,(|a_1-b_1|,\ |a_2-b_2|).$$

<u>Satz 1.2:</u> $q(A,B)$ ist eine homogene translationsinvariante Metrik, d. h. es gilt:

1) $q(A,B)=0 \Leftrightarrow A=B$
2) $q(A,B) \leq q(A,C)+q(B,C)$
3) $q(cA,cB)=|c|\,q(A,B)$
4) $q(A+C,B+C)=q(A,B)$

<u>Bemerkung:</u> Die Eigenschaften 1) und 2) sind kennzeichnend für eine Metrik; eine Metrik mit der Eigenschaft 3) bzw. 4) nennen wir homogen bzw. translationsinvariant.

<u>Beweis:</u> 1) $\max(|a_i-b_i|)=0 \Leftrightarrow a_i=b_i$ für $i=1,2$

2) $\max(|a_i-b_i|)=\max(|a_i-c_i+c_i-b_i|) \leq \max(|a_i-c_i|+|c_i-b_i|) \leq$
$\leq \max(|a_i-c_i|)+\max(|b_i-c_i|)$;

3) $\max(|ca_i-cb_i|)=\max(|c|\,|a_i-b_i|)=|c|\max(|a_i-b_i|)$;

4) $\max(|a_i+c_i-b_i-c_i|)=\max(|a_i-b_i|)$.

<u>Bemerkung:</u> Ist M ein beliebiger metrischer Raum, so ist die Menge der abgeschlossenen Punktmengen von M ein metrischer Raum bezüglich einer von Hausdorff angegebenen Metrik (siehe dazu etwa Alexandroff-Hopf, Topologie I.). Einschränkung dieser Hausdorff-Metrik für die Menge der abgeschlossenen Punktmengen über R auf die Teilmenge $I(R)$ gibt dort die Metrik $q(A,B)$.

<u>Satz 1.3:</u> $I(R)$ ist vollständig bezüglich der Metrik $q(A,B)$.
Zum Beweis bilden wir $I(R)$ auf eine vollständige Teilmenge des R^2 isometrisch ab.
Für Vektoren $\vec{x}=\binom{x_1}{x_2} \in R^2$ verwenden wir die Norm $\|\vec{x}\|=\max\limits_i |x_i|$.
Bezüglich der dadurch gegebenen Metrik ist die Teilmenge

$$R^2_{\leq} = \left\{ \binom{a_1}{a_2} \,\middle|\, a_1 \leq a_2 \right\} \subset R^2$$

abgeschlossen, also vollständig.

Die Abbildung T: $I(R) \ni A = [a_1, a_2] \longrightarrow \hat{a} = \binom{a_1}{a_2} \in R^2_*$ ist isometrisch;

es gilt nämlich: $q(A,B) = \max(|a_i - b_i|) = \|\hat{a} - \hat{b}\| = \max(|a_i - b_i|)$.
Damit überträgt sich die Vollständigkeit des R^2_* auf $I(R)$.
Für $A = a$, $B = b \in R \subset I(R)$ wird $q(A,B) = |a-b|$, d. h. die Metrik $q(A,B)$ in $I(R)$ ist eine Fortsetzung der im Teilraum R durch $|a-b|$ gegebenen Metrik.

<u>Bemerkung:</u> Durch die umkehrbar eindeutige Abbildung T der Menge $I(R)$ auf $R^2_* \subset R^2$ wird jeder Metrik in R^2 eine Metrik in $I(R)$ zugeordnet.
Es ist $T([a_1, a_2]) = \binom{a_1}{a_2} \in R^2$.
Ist $q(\hat{a}, \hat{b})$ irgendeine Metrik in R^2, so ergibt die Definition $q(A,B) := q(T(A), T(B))$ eine Metrik in $I(R)$.

<u>Definition 1.4:</u> Es sei $A = [a_1, a_2] \in I(R)$;
die reelle Zahl $|A| := q(A,0) = \max(|a_1|, |a_2|)$
heißt Absolutbetrag von A,

das Intervall $\mathrm{abs}(A) := \begin{cases} A & a_1 \geq 0 \\ -A & \text{für } a_2 \leq 0 \\ [0, \max(|a_1|, |a_2|)] & a_1 \leq 0 \leq a_2 \end{cases}$

heißt Intervallbetrag von A.

Für den Absolutbetrag $|A| := q(A,0)$ gelten wegen der Homogenität und Translationsinvarianz der Metrik $q(A,B)$ in $I(R)$ die folgenden Regeln (Beweis in Satz 2.4):

(A1) $A \neq 0 \Rightarrow |A| > 0$,
(A2) $|A+B| \leq |A| + |B|$,
(A3) $|cA| = |c||A|$.

Entsprechende Aussagen für den Intervallbetrag abs(A) können wir ableiten, sobald wir eine Halbordnung für Intervalle definiert haben.
Bei Einschränkungen auf R ergeben $|A|$ und abs(A) den gewöhnlichen Absolutbetrag $|a|$ in R.

<u>Satz 1.4:</u> Für alle A, B, $C \in I(R)$ gilt: $q(AB, AC) \leq |A|\, q(B,C)$.

<u>Beweis:</u> Zu $D = [d_1, d_2] \in I(R)$ bezeichnen wir die untere Intervallgrenze d_1 auch mit inf(D) und entsprechend die obere Intervallgrenze d_2 mit sup (D). Damit lautet die Behauptung des Satzes ausführlicher geschrieben:

$\max\{|\inf(AB) - \inf(AC)|, |\sup(AB) - \sup(AC)|\} \leq |A|\, q(B,C)$.
Nach Satz 1.2/3 gilt für $a \in R$: $q(aB, aC) = |a|\, q(B,C)$, also
$\max\{|\inf(aB) - \inf(aC)|, |\sup(aB) - \sup(aC)|\} = |a|\, q(B,C)$. (1)
Wir zeigen nun $|\inf(AB) - \inf(AC)| \leq |A|\, q(B,C)$
(die entsprechende Aussage für das Supremum ergibt sich analog).

O. B. d. A. sei $\inf(AB) \geq \inf(AC)$.
Da Infimum (und Supremum) jeweils angenommen werden, existiert ein $a \in A$, so daß $\inf(AC) = \inf(aC)$
wegen $\inf(aB) \geq \inf(AB)$ erhalten wir
$\inf(aB) - \inf(aC) \geq \inf(AB) - \inf(AC) \geq 0$, also
$|\inf(AB) - \inf(AC)| = \inf(AB) - \inf(AC) \leq \inf(aB) - \inf(aC) =$
$= |\inf(aB) - \inf(aC)| \leq |a|\, q(B,C) \leq |A|\, q(B,C)$.
(Die vorletzte Ungleichung folgt aus (1); für die letzte Ungleichung wird benutzt, daß $|a| \leq |A|$ für $a \in A$.)

Entsprechend schließt man $|\sup(AB) - \sup(AC)| \leq |A|\, q(B,C)$.

Zusammenfassen ergibt die Behauptung.

Eine wesentliche Voraussetzung für die in dieser Arbeit erzielten Fortschritte ist die konsequente Verwendung der schon in [9] angegebenen Metrik $q(A,B)$ unter Berücksichtigung ihrer hier in Satz 1.2/3,4 und Satz 1.4 anscheinend zum ersten Mal bewiesenen speziellen Eigenschaften.
In der Homogenität und Translationsinvarianz (Satz 1.2/3,4) kommt eine gewisse Verträglichkeit der Metrik $q(A,B)$ mit der algebraischen Struktur von $I(R)$ zum Ausdruck; Satz 1.4 sagt aus, daß die Metrik $q(A,B)$ mit dem Absolutbetrag in $I(R)$ verträglich ist.

Erst auf Grund dieser Verträglichkeitseigenschaften wird die Metrik q(A,B) für praktische Anwendungen, insbesondere für Abschätzungen brauchbar. Solche Anwendungen werden in den Abschnitten 5 und 6 dieser Arbeit behandelt.

2. Der quasilineare Raum

2.1 Zur Definition des quasilinearen Raumes

Definition 2.1: Der quasilineare Raum:

Es sei QL eine Menge mit Elementen A, B, C, ...,
R der Körper der reellen Zahlen mit Elementen a, b, c...

Wir nehmen an, es sei eine Abbildung $(A,B) \longrightarrow A+B$ von $QL \times QL$ in QL definiert, die wir Addition nennen, und eine Abbildung $(a,A) \rightarrow aA$ von $R \times QL$ in QL, die wir skalare Multiplikation nennen, so daß die folgenden Axiome erfüllt sind:

(QL1)	$A+B=B+A$	
(QL2)	$(A+B)+C=A+(B+C)$	
(QL3)	$a(B+C)=aB+aC$	
(QL4)	$(a+b)C=aC+bC$	für $ab \geq 0$
(QL5)	$a(bC)=(ab)C$	
(QL6)	Es existiert ein sogenanntes Nullelement $\Theta \in QL$ mit der Eigenschaft: $0A=\Theta$	für alle $A \in QL$
(QL7)	$1A=A$	für alle $A \in QL$.

Ausgestaltet mit dieser Struktur heißt QL quasilinearer Raum über dem Zahlkörper R.

Bemerkung: Der quasilineare Raum ist eine Verallgemeinerung des linearen Raumes. Ersetzt man in der obigen Definition das Axiom (QL4) durch die schärfere Forderung

(L4) $(a+b)C=aC+bC$ für alle $a, b \in R$

so erhält man eine Definition für den linearen Raum (vergleiche etwa die Definition des linearen Raumes in [16]).

Beispiele für quasilineare Räume:

1) Jeder lineare Raum ist quasilinear.
2) I(R) ist ein (nicht linearer) quasilinearer Raum (vergleiche Abschnitt 1.).

Weitere Beispiele nicht linearer, quasilinearer Räume werden wir noch kennenlernen.

Satz 2.1: In einem quasilinearen Raum QL ist nur das in Axiom (QL6) postulierte Nullelement θ ein neutrales Element der Addition.

Beweis: Für beliebiges $X \in QL$ gilt nach Axiom (QL4), (QL6) und (QL7) $X+\theta=1X+0X=(1+0)X=X$.
θ ist also ein neutrales Element der Addition; ist nun $\hat{\theta}$ irgendein neutrales Element der Addition, dann gilt $\hat{\theta}+\theta=\hat{\theta}$ und $\theta+\hat{\theta}=\theta$ mit (QL1) folgt daraus $\hat{\theta}=\theta$.

Von weiteren Eigenschaften quasilinearer Räume, die aus dem Axiomensystem folgen, erwähnen wir noch:

$a\theta=\theta$ für alle $a \in R$.

Beweis: $a(0B)=(a0)B=0B=\theta$.
$aX=\theta \Rightarrow a=0$ oder $X=\theta$.
Beweis: Annahme $a \neq 0 \Rightarrow X = \frac{1}{a}(aX) = \frac{1}{a}\theta = \theta$.
Für $(-1)X$ schreiben wir $-X$, für $Y+(-X)$ schreiben wir $Y-X$ und es ist möglich, daß $X+(-X)=X-X\neq 0$ ist.
Beispiel: Für $A \in I(R)$ und $A \notin R$ ist stets $A-A \neq 0$.

Die Existenz inverser Elemente kann in quasilinearen Räumen nicht bewiesen werden (tatsächlich stellten wir auch in Abschnitt 1 fest, daß $I(R)$ nicht zu allen Elementen Inverse enthält.), gilt dagegen das Distributivgesetz in der strengen Form (L4), so lassen sich Existenz und Eindeutigkeit inverser Elemente beweisen:
Aus (QL6), (QL7) und (L4) folgt für beliebige $A \in QL$:
$A+(-A)=1A+(-1)A=(1-1)A=0A=\theta$, also ist $(-A)$ invers zu A.
Existiert zu A noch ein weiteres inverses Element C, dann gilt
$C=C+\theta=C+A+(-A)=-A$,
also ist $(-A)$ das einzige inverse Element zu A.

Wir halten fest:
Quasilineare Räume unterscheiden sich von linearen Räumen durch die schwächere Forderung im Distributivgesetz; dies bedingt das Fehlen der inversen Elemente.

Konvexe Teilmengen werden in quasilinearen Räumen wie in linearen Räumen definiert:

Definition 2.2: Eine Teilmenge M von QL heißt konvex, wenn mit zwei Elementen A und B auch alle Elemente $(1-c)A+cB$ mit $0\leq c\leq 1$ zu M gehören.
Als konvexe Hülle H einer Teilmenge N von QL bezeichnen wir den Durchschnitt aller konvexen Mengen in QL, die N enthalten.

2.2 Norm und Metrik in quasilinearen Räumen

Definition 2.3: Ein reelles Funktional $p(A)$, definiert auf einem quasilinearen Raum QL heißt Norm, wenn

1) $A\neq 0 \Rightarrow p(A) > 0$,
2) $p(A+B) \leq p(A)+p(B)$,
3) $p(cA) = |c|\cdot p(A)$.

Wir sprechen von einem quasilinearen normierten Raum.

In linearen normierten Räumen ist bekanntlich die Norm $p(A-B)$ der Differenz von zwei Elementen A und B eine Metrik; dies gilt aber nicht in quasilinearen Räumen; in diesen ist im allgemeinen $A-A\neq 0$, also $p(A-A)\neq 0$. Deshalb ist $p(A-B)$ keine Metrik.

Wir werden jedoch sehen, daß in den quasilinearen Räumen der Intervallrechnung, die hier in dieser Arbeit untersucht werden, jeder vorgegebenen Norm eine Metrik zugeordnet werden kann (vgl. Abschnitt 3.4 und 5).

Definition 2.4: Die in einem quasilinearen Raum QL erklärte Metrik $q(A,B)$ heißt

homogen,	wenn	(H)	$q(cA,cB)= \|c\|\, q(A,B)$,
translationsinvariant,	wenn	(TI)	$q(A+C,B+C)= q(A,B)$,
supermetrisch,	wenn	(S)	$q(A+B,C+D) \leq q(A,C)+ q(B,D)$,

jeweils für alle $c \in R$ und alle $A, B, C, D \in QL$.
Ein Raum mit supermetrischer Metrik heißt auch selbst supermetrisch.

In der Homogenität und Translationsinvarianz oder supermetrischen Eigenschaft kommt eine Verträglichkeit der Metrik mit der algebraischen Struktur zum Ausdruck.

Eine Definition der Eigenschaft "supermetrisch" wurde - soweit uns bekannt ist - in der Literatur bisher nur für lineare Räume gegeben. So nennt Collatz in [5] einen linearen Raum L mit der Metrik $q(f_1,f_2)$ supermetrisch, wenn

(LS) $q(f_1,f_2+f_3)= q(f_1-f_3,f_2)$ für alle $f_1,f_2,f_3 \in L$.

Wir zeigen nun, daß unsere Definition bei Einschränkung auf lineare Räume der von Collatz gegebenen äquivalent ist. In einem linearen metrischen Raum L ist zunächst die Eigenschaft (LS) äquivalent zur oben definierten Translationsinvarianz (TI); man erkennt das sofort, wenn man $A+C:=f_1$, $B:=f_2$ und $C:=f_3$ setzt.

Jetzt zeigen wir noch die Äquivalenz der Eigenschaft (TI) mit der Eigenschaft (S), die hier den supermetrischen Raum definiert:

Es gelte zunächst (TI); dann ist nach der Dreiecksungleichung
$q(A+B,C+D) \leq q(A+B,C+B)+ q(C+B,C+D)= q(A,C)+ q(B,D)$.

Setzt man (S) voraus, so ergibt sich $q(A+B,C+B)\leq q(A,C)$
und entsprechend $q(A,C)= q((A+B)+(-B),(C+B)+(-B))\leq q(A+B,C+B)$
also $q(A+B,C+B)= q(A,C)$.

Damit erhalten wir also folgenden Satz:

<u>Satz 2.2:</u> L sei ein linearer Raum mit der Metrik $q(A,B)$, dann sind die folgenden Eigenschaften äquivalent:

(TI) $q(A+C,B+C)= q(A,B)$
(LS) $q(E,F+G)= q(E-G,F)$
(S) $q(A+B,C+D)\leq q(A,C)+ q(B,D)$

d. h. die Metrik eines linearen Raumes ist genau dann translationsvariant, wenn sie supermetrisch ist.

Für quasilineare Räume gilt nur:

<u>Satz 2.3:</u> Ein quasilinearer Raum mit translationsinvarianter Metrik ist auch supermetrisch.

Der Schluß von der Eigenschaft (TI) auf die Eigenschaft (S) ergibt sich wie in Satz 2.2.

Dagegen folgt in quasilinearen Räumen aus der supermetrischen Eigenschaft nicht die Translationsinvarianz der Metrik.
In Abschnitt 2.3 wird ein Beispiel für eine supermetrische und nicht translationsinvariante Metrik in $I(R)$ angegeben.

<u>Satz 2.4:</u> In einem quasilinearen Raum mit homogener supermetrischer Metrik $q(A,B)$ ist der Abstand $q(A,\Theta)$ eines Elementes A vom neutralen Element der Addition Θ eine Norm.

<u>Beweis:</u> 1) $A\neq\Theta \succ q(A,\Theta) > 0$,
2) $q(A+B,\Theta)= q(A+B,\Theta+\Theta)\leq q(A,\Theta)+ q(B,\Theta)$,
3) $q(cA,\Theta)= |c|\, q(A,\Theta)$.

<u>Beispiel:</u> In $I(R)$ ist der Absolutbetrag $|A|= q(A,0)$ eine Norm.

Ein linearer Raum mit homogener supermetrischer Metrik ist bekanntlich ein normierter Raum (vergleiche etwa [5]; daß in linearen Räumen unsere Definition der Eigenschaft "supermetrisch" der in [5] gegebenen äquivalent ist, wurde oben gezeigt).

Ein quasilinearer Raum mit homogener supermetrischer Metrik unterscheidet sich in den definierenden Eigenschaften vom linearen normierten Raum einzig durch die Abschwächung von Axiom (L4) zu (QL4).

<u>Definition 2.5:</u> Ein quasilinearer Raum, der bezüglich einer homogenen supermetrischen Metrik vollständig ist, heißt Quasi-Banachraum.

<u>Beispiel:</u> $I(R)$ ist ein Quasi-Banachraum, denn die Metrik $q(A,B)$ in $I(R)$ ist homogen und translationsinvariant, also nach Satz 2.3 auch supermetrisch.

2.3 Ein Beispiel für eine supermetrische und nicht translationsinvariante Metrik in $I(R)$

Für Metriken in linearen Räumen sind nach Satz 2.2 die Eigenschaften "translationsinvariant" und "supermetrisch" äquivalent. Für Metriken in quasilinearen Räumen besteht diese Äquivalenz nicht:
Hier folgt nur die "supermetrische Eigenschaft" aus der "Translationsinvarianz" der Metrik (siehe Satz 2.3); dagegen folgt die "Translationsinvarianz" nicht mehr aus der "supermetrischen Eigenschaft"; dies zeigt das folgende Beispiel für eine supermetrische und nicht translationsinvariante Metrik im quasilinearen Raum $I(R)$.

Wir erklären zur Vorbereitung noch den Begriff des Durchmessers für Elemente aus $I(R)$.

Definition 2.7: Es sei $A = [a_1,a_2] \in I(R)$; dann bezeichnen wir die reelle Zahl $d(A) := (a_2 - a_1)$ als Durchmesser von A.

Ersichtlich gilt für alle $A, B \in I(R)$:

$$\begin{aligned} & d(A) \geq 0 \quad \text{mit} \quad d(A) = 0 \Leftrightarrow A \in R \\ \text{und} \quad & d(A+B) = d(A) + d(B). \end{aligned} \tag{1}$$

$q(A,B)$ sei die in Abschnitt 1 eingeführte Metrik in $I(R)$ (siehe dazu Satz 1.2).

Damit erklären wir auf $I(R) \times I(R)$ das reelle Funktional

$$qd(A,B) := \begin{cases} q(A,B) & \text{für } A, B \in R \subset I(R)) \\ q(A,B) + \dfrac{|d(A) - d(B)|}{\max\{d(A),d(B)\}} & \text{sonst} \end{cases}$$

Es gilt nun

Satz 2.6: $qd(A,B)$ ist eine supermetrische und nicht translationsinvariante Metrik in $I(R)$.

Beweis: Wir zeigen zunächst, daß $qd(A,B)$ eine Metrik ist:
Aus der Definition ergibt sich sofort:

$$\begin{aligned} & qd(A,B) \geq 0 \quad \text{mit} \quad qd(A,B) = 0 \Leftrightarrow A = B \\ \text{und} \quad & qd(A,B) = qd(B,A). \end{aligned}$$

Zum Nachweis der Dreiecksungleichung für $qd(A,B)$ genügt es (weil die Metrik $q(A,B)$ die Dreiecksungleichung erfüllt, siehe Satz 1.2) diese für das reelle Funktional

$$d(A,B) := \begin{cases} 0 & \text{für } A, B \in R \subset I(R) \\ \dfrac{|d(A) - d(B)|}{\max\{d(A),d(B)\}} & \text{sonst} \end{cases}$$

nachzuweisen.
Wegen $d(A) \geq 0$ für alle $A \in I(R)$ ist aber dieser Nachweis äquivalent mit dem Nachweis der Dreiecksungleichung für das auf $R^+ x R^+$ (R^+ sei die Menge der nichtnegativen reellen Zahlen) erklärte Funktional

$$d(a,b) := \begin{cases} 0 & a = b = 0 \\ \dfrac{|a - b|}{\max\{a,b\}} \quad \text{für} & a + b > 0 \end{cases}$$

Für $d(a,b)$ beweist man die Dreiecksungleichung $d(a,b) \leq d(a,c) + d(b,c)$ am einfachsten durch Fallunterscheidung:
In den Fällen $a = b = 0,\ c \geq 0$ und $a > 0,\ b = 0,\ c \geq 0$ erkennt man ihre Gültigkeit unmittelbar.

Im Falle $a, b > 0$ kann man o. B. d. A. $a \geq b$ voraussetzen und unterscheidet dann die Unterfälle

1) $a \geq b \geq c \geq 0$, 2) $a \geq c \geq b > 0$, 3) $c \geq a \geq b > 0$;

die Gültigkeit der Dreiecksungleichung ergibt sich in jedem dieser Fälle wieder unmittelbar.

$qd(A,B)$ ist also eine Metrik in $I(R)$.

Die Metrik $qd(A,B)$ ist nicht translationsinvariant; etwa für $A=[2,3]$, $B=[4,6]$, $C=[1,2]$ wird nämlich

$$qd(A,B) = 3\tfrac{1}{2} \neq qd(A+C,B+C) = 3\tfrac{1}{3}.$$

Wir zeigen aber jetzt, daß $qd(A,B)$ supermetrisch ist:

Nach Definition von $d(A,B)$ gilt
für alle $A, B \in \mathbb{R} \subset \mathbf{I}(R)$, $C \in \mathbf{I}(R)$: $\quad d(A+C,B+C) = 0 = d(A,B)$;
in allen anderen Fällen folgt aus (1):

$$d(A+C,B+C) = \frac{|d(A+B) - d(B+C)|}{\max\{d(A+C),d(B+C)\}} \leq \frac{|d(A) - d(B)|}{\max\{d(A),d(B)\}} = d(A,B).$$

Damit erhalten wir insgesamt:
Für alle $A, B, C \in \mathbf{I}(R)$ ist $\quad d(A+C,B+C) \leq d(A,B)$. (2)

Für die Metrik $q(A,B)$ in $\mathbf{I}(R)$ gilt nach Satz 1.2/3
für alle $A, B, C \in \mathbf{I}(R)$: $\quad q(A+C,B+C) = q(A,B)$. (3)

Aus (2) und (3) folgt
für alle $A, B, C \in \mathbf{I}(R)$: $\quad qd(A+C,B+C) \leq qd(A,B)$

Daraus folgt jetzt mit Hilfe der Dreiecksungleichung
für alle $A, B, C, D \in \mathbf{I}(R)$:

$$qd(A+B,C+D) \leq qd(A+B,C+B) + qd(C+B,C+D) \leq qd(A,C) + qd(B,D),$$

d. h. die Metrik $qd(A,B)$ ist supermetrisch.

3. Die Räume der Intervallrechnung

3.1 Intervallrechnung für Vektoren und Matrizen

Wir trieben bisher Intervallrechnung in der Menge $\mathbf{I}(R)$, deren Struktur uns zur Definition des sogenannten Quasi-Banachraumes veranlaßte. Nun führen wir im Anschluß an [3] eine Intervallrechnung für Vektoren und Matrizen mit Elementen aus $I(R)$ ein; dabei lernen wir weitere für die Praxis besonders wichtige Beispiele von Quasi-Banachräumen kennen.

Es sei $V_n(\mathbb{R}) := \mathbb{R}^n$ der euklidische Raum den n-Tupel reeller Zahlen

$V_n(\mathbb{I}(\mathbb{R}))$ die Menge der n-Tupel von Intervallen aus $\mathbf{I}(\mathbb{R})$

und entsprechend

$M_{n\times m}(\mathbb{R})$ bzw. $M_{n\times m}(\mathbf{I}(\mathbb{R}))$ die Menge der $n\times m$-Matrizen mit Elementen aus $\mathbb{R}$ bzw. $\mathbf{I}(\mathbb{R})$.

Für quadratische $n\times n$-Matrizen verwenden wir nur einen Index; z. B. $M_n(\mathbb{R})$.

Die Elemente von $V_n(\mathbf{I}(\mathbb{R}))$ bzw. $M_{n\times m}(\mathbf{I}(\mathbb{R}))$ nennen wir Intervallvektoren bzw. Intervallmatrizen.

Es ist $R \subset \mathbb{I}(R)$, also entsprechend $V_n(R) \subset V_n(\mathbf{I}(R))$,
bzw. $M_{n\times m}(R) \subset M_{n\times m}(\mathbf{I}(R))$.

Elemente aus $V_n(\mathbf{I}(R))$ bezeichnen wir mit kleinen deutschen Buchstaben

$$\mathfrak{a} = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}, \ \mathfrak{b} = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} \quad \text{mit } A_i, B_i \in \mathbf{I}(R)$$

Elemente aus $M_n(\mathbb{I}(\mathbb{R}))$ mit großen deutschen Buchstaben, also

$$\mathfrak{A} = (A_{ij}),\ \mathfrak{B} = (B_{ij}) \quad \text{mit } A_{ij}, B_{ij} \in \mathbf{I}(R).$$

Wollen wir besonders hervorheben, daß ein Element aus $V_n(\mathbf{I}(R))$ bzw. $M_n(\mathbf{I}(R))$ schon in $V_n(\mathbb{R})$ bzw. $M_n(\mathbb{R})$ liegt, so sprechen wir auch von reellen Vektoren bzw. reellen Matrizen und schreiben $\dot{\mathfrak{a}}$ bzw. $\dot{\mathfrak{A}}$.

Für Vektoren $\mathfrak{a}=(A_i)$, $\mathfrak{b}=(B_i)$ aus $V_n(\mathbb{I}(\mathbb{R}))$ bzw. Matrizen $\mathfrak{A}=(A_{ij})$, $\mathfrak{B}=(B_{ij})$ aus M $(\mathbb{I}(\mathbb{R}))$ erklären wir Gleichheit, Addition (Substraktion) und eine skalare Multiplikation mit reellen Zahlen $a\in\mathbb{R}$; dabei definieren wir wie üblich komponentenweise:

$\mathfrak{A}=\mathfrak{B}:\bigwedge\ A_{ij}=B_{ij}$ für alle i, j, bzw. $\mathfrak{a}=\mathfrak{b}:\bigwedge\ A_i=B_i$ für alle i,

$\mathfrak{A}\pm\mathfrak{B}:=(A_{ij}\pm B_{ij})$, $\qquad \mathfrak{a}\pm\mathfrak{b}:=(A_i\pm B_i)$,

$a\ \mathfrak{B}:=(a\ B_{ij})$ $\qquad a\cdot\mathfrak{b}=(a\ B_i)$.

Ausgestattet mit dieser algebraischen Struktur sind $V_n(\mathbb{I}(\mathbb{R}))$ und $M_n(\mathbb{I}(\mathbb{R}))$ quasilineare Räume.

Die Axiome des quasilinearen Raumes QL sind nämlich jeweils erfüllt, denn alle in den Axiomen auftretenden Relationen zwischen Elementen aus $V_n(\mathbb{I}(\mathbb{R}))$ bzw. $M_n(\mathbb{I}(\mathbb{R}))$ sind komponentenweise definiert (bestehen also genau dann, wenn sie zwischen jeweils entsprechenden Komponenten bestehen), die Komponenten aber sind Elemente des quasilinearen Raumes $\mathbb{I}(\mathbb{R})$.
Nullelement in $V_n(\mathbb{I}(\mathbb{R}))$ bzw. $M_n(\mathbb{I}(\mathbb{R}))$ ist der Nullvektor bzw. die Nullmatrix.

Darüberhinaus definieren wir noch wie üblich eine Multiplikation der Vektoren bzw. Matrizen mit Intervallen aus $\mathbb{I}(\mathbb{R})$, und eine Multiplikation von Matrizen mit Matrizen bzw. Vektoren:

$A\cdot\mathfrak{B}:=(A\cdot B_{ij})$ $\qquad A\ \mathfrak{b}=(A\ B_i)$

$\mathfrak{A}\mathfrak{B}:=(\sum_j A_{ij}\ B_{jk})$ $\qquad \mathfrak{A}\ \mathfrak{b}=(\sum_j A_{ij}\ B_j)$

Vorausgesetzt wird dabei natürlich, daß die Zeilen- bzw. Spaltenzahl der Matrizen und Vektoren jeweils aufeinanderpaßt.

In $V_n(\mathbb{I}(\mathbb{R}))$ bzw. $M_n(\mathbb{I}(\mathbb{R}))$ lassen sich auf mannigfache Art und Weise Metriken einführen; wir erwähnen nur je eine Möglichkeit:

Die Definition $q(\mathfrak{a},\mathfrak{b}):=\max_i q(A_i,B_i)$

bzw. $q(\mathfrak{A},\mathfrak{B}):=\max_i \sum_k q(A_{ik},B_{ik})$

ergibt eine homogene und translationsinvariante Metrik in $V_n(\mathbb{I}(\mathbb{R}))$ bzw. $M_n(\mathbb{I}(\mathbb{R}))$; dies folgt aus der Homogenität und Translationsinvarianz der Metrik $q(A,B)$ in $\mathbb{I}(\mathbb{R})$.
$V_n(\mathbb{I}(\mathbb{R}))$ und $M_n(\mathbb{I}(\mathbb{R}))$ sind beide vollständig bezüglich ihrer Metrik; dies folgt aus der Vollständigkeit von $\mathbb{I}(\mathbb{R})$, denn Konvergenz in $V_n(\mathbb{I}(\mathbb{R}))$ bzw. $M_n(\mathbb{I}(\mathbb{R}))$ ist jeweils komponentenweise Konvergenz. Insgesamt erwiesen sich also $V_n(\mathbb{I}(\mathbb{R}))$ und $M_n(\mathbb{I}(\mathbb{R}))$ als Quasi-Banachräume.

Diese Überlegungen lassen sich auch sofort übertragen auf Intervallvektoren mit unendlich vielen Komponenten (oder entsprechende Intervallmatrizen):

$V_\infty(\mathbb{I}(\mathbb{R}))$ sei der Raum der beschränkten Intervallfolgen mit Elementen

$$\{X\}:=\{X_1,X_2,X_3,\ldots\ldots\},$$

d. h. zu jedem Element $\{X\}$ existiere eine Konstante K_X, so daß für alle i

$$q(X_i,0)=|X_i|\le K_X.$$

Gleichheit, Addition und Multiplikation mit Intervallen (also auch reellen Zahlen) werden komponentenweise definiert; damit wird $V_\infty(\mathbb{I}(\mathbb{R}))$ ein quasilinearer Raum.

Für zwei Elemente $\{X\}, \{Y\}\in V_\infty(\mathbb{I}(\mathbb{R}))$ definieren wir als Abstand

$$q(\{X\},\{Y\})=\sup_i q(X_i,Y_i).$$

Dadurch erhalten wir eine homogene translationsinvariante Metrik in $V_\infty(\mathbb{I}(\mathbb{R}))$, denn durch diese Definition von $q(\{X\},\{Y\})$ als maximaler Komponentenabstand $q(X_i,Y_i)$ überträgt sich diese Eigenschaft der Metrik $q(X_i,Y_i)$ auf den $V_\infty(\mathbb{I}(\mathbb{R}))$. Die Konvergenz in $V_\infty(\mathbb{I}(\mathbb{R}))$ ist die gleichgradig komponentenweise Konvergenz; deshalb ergibt sich die Vollständigkeit von $V_\infty(\mathbb{I}(\mathbb{R}))$ im wesentlichen aus der von $\mathbb{I}(\mathbb{R})$; dabei ist nur noch gesondert zu zeigen, daß das Grenzelement wieder eine beschränkte Intervallfolge ist. Der Beweis dafür erfolgt wie im Reellen. Damit ist auch $V_\infty(\mathbb{I}(\mathbb{R}))$ ein Quasi-Banachraum.

3.2 Ordnungsstrukturen in $I(R)$, $V_n(I(R))$ und $M_{n\times m}(I(R))$

In diesem Abschnitt untersuchen wir Ordnungsstrukturen in den Räumen $I(R)$, $V_n(I(R))$ und $M_{n\times m}(I(R))$; Anwendungen dafür werden wir in Abschnitt 8 kennenlernen.

Definition 3.1: Kleinergleichrelation "$\leq$" in $I(R)$:
Für $A = [a_1,a_2]$, $B = [b_1,b_2] \in I(R)$ sei

$$A \leq B \;:\Longleftrightarrow\; a_1 \leq b_1 \quad \text{und} \quad a_2 \leq b_2$$

Bemerkung: Bei Einschränkung auf $R \subset I(R)$ erhält man die übliche Bedeutung von "$\leq$" in R.

Bezüglich dieser Relation "$\leq$" wird $I(R)$ zu einer Halbordnung.

Es gilt nämlich für beliebige $A,B,C \in I(R)$:

(01) $A \leq A$

(02) $A \leq B \wedge B \leq C \Rightarrow A \leq C$

(03) $A \leq B \wedge B \leq A \Rightarrow A = B$

$I(R)$ ist sogar ein Verband, denn es gilt:

(04) Zu je zwei Elementen $A,B \in I(R)$ existieren $\sup(A,B)$ und $\inf(A,B)$ in $I(R)$.

Für $A = [a_1,a_2]$, $B = [b_1,b_2] \in I(R)$ wird

$$\sup(A,B) = [\max(a_1,b_1), \max(a_2,b_2)],$$
$$\inf(A,B) = [\min(a_1,b_1), \min(a_2,b_2)].$$

Man erkennt auch, daß der Verband ordnungsvollständig ist (siehe dazu etwa [13]), d. h. es gilt:

(05) Jede nicht leere beschränkte Teilmenge von $I(R)$ hat ein Supremum in $I(R)$.

Dies ergibt sich sofort aus der entsprechenden Eigenschaft für reelle Zahlen und Anwendung auf beide Intervallgrenzen.

Die Relation "$\leq$" ist auch mit der algebraischen Struktur des Raumes $I(R)$ verträglich, d. h. die folgende Axiome sind erfüllt:

(A01) $A \leq B \Rightarrow A+B \leq B+C$ für alle $A,B,C \in I(R)$,

(A02) $A \leq B \Rightarrow \lambda A \leq \lambda B$ für alle $A,B \in I(R)$, $\lambda \in R$ und $\lambda \geq 0$.

(A01) drückt die Translationsinvarianz der Ordnung aus,
(A02) bedeutet die Invarianz der Ordnung bei der Abbildung $A \rightarrow \lambda A$ mit reellem $\lambda \geq 0$.

Sind zwei Intervalle $A,B \in I(R)$ in der Halbordnung $\{I(R),\leq\}$ nicht vergleichbar, so ist eines im anderen "enthalten", denn es ist $A \subset B \Longleftrightarrow a_1 \geq b_1 \wedge a_2 \leq b_2$.
Ist $A \subset B \in I(R)$ und $A = \underset{.}{A} \in R$, so schreiben wir auch $\underset{.}{A} \in B$.

Bezüglich der Relation ist nun $I(R)$ wieder eine Halbordnung. Die Menge aller Teilmengen von $I(R)$, dazu gehört jetzt auch die leere Menge, die in der algebraischen Struktur $I(R)$ nicht enthalten ist, bildet einen Verband bezüglich der Relation "$\subset$". Das Infimum (bzw. Supremum) von zwei Intervallen bezüglich der Relation "$\subset$" ist dann ihr Durchschnitt (bzw. ihre konvexe Hülle).

In Definition 1.4 erklärten wir zu $A \in I(R)$ Absolutbetrag und Intervallbetrag; eine äquivalente Definition ist

Definition 3.2: Es sei $A \in I(R)$, dann heißt

$|A| := \sup\limits_{\underset{.}{A} \in A} |\underset{.}{A}| = \sup\limits_{\underset{.}{A} \in A} \sup(\underset{.}{A},-\underset{.}{A})$ Absolutbetrag von A,

$\mathrm{abs}(A) := \sup(0,A,-A)$ Intervallbetrag von A

(0 ist hier die Null: $0 \in R \subset I(R)$).

Beide Suprema existieren, da $I(R)$ ein ordnungsvollständiger Verband ist; es ist $|A| \in R$, $\mathrm{abs}(A) \in I(R)$.

Die stetige Funktion $|\underset{.}{A}|$ nimmt ihr Supremum auf der abgeschlossenen beschränkten Teilmenge A von R an, d. h. es existiert ein $\underset{.}{A}^* \in R$ mit der Eigenschaft $\underset{.}{A}^* \in A$ und $|\underset{.}{A}^*| = \sup\limits_{\underset{.}{A} \in A} |\underset{.}{A}| = |A|$.

<u>Satz 3.1:</u> Für den Absolutbetrag $|A|$ bzw. Intervallbetrag $\mathrm{abs}(A)$ von $A \in \mathbb{I}(\mathbb{R})$ gilt:

(A1)	$A \neq 0 \Rightarrow \lvert A\rvert > 0$	bzw.:	(IA1)	$\mathrm{abs}(A) \geq 0$, mit $\mathrm{abs}(A)=0 \Leftrightarrow A=0$
(A2)	$\lvert A+B\rvert \leq \lvert A\rvert + \lvert B\rvert$		(IA2)	$\mathrm{abs}(A+B) \leq \mathrm{abs}(A)+\mathrm{abs}(B)$
(A3)	$\lvert cA\rvert = \lvert c\rvert \lvert A\rvert$		(IA3)	$\mathrm{abs}(cA) = \lvert c\rvert\, \mathrm{abs}(A)$

<u>Beweis:</u> Die Regeln für $|A|$ ergeben sich aus Satz 2.4.

Aus $\mathrm{abs}(A)=[\min_{\mathsf{A}\in A}|\mathsf{A}|,\ \max_{\mathsf{A}\in A}|\mathsf{A}|]$ folgen sofort (IA1) und (IA3).
Für (IA2) müssen wir zeigen:

$$\max_{\substack{\mathsf{A}\in A\\ \mathsf{B}\in B}} |\mathsf{A}+\mathsf{B}| \leq \max_{\mathsf{A}\in A} |\mathsf{A}| + \max_{\mathsf{B}\in B} |\mathsf{B}| \quad \text{und} \quad \min_{\substack{\mathsf{A}\in A\\ \mathsf{B}\in B}} |\mathsf{A}+\mathsf{B}| \leq \min_{\mathsf{A}\in A} |\mathsf{A}| + \min_{\mathsf{B}\in B} |\mathsf{B}|;$$

Die erste Aussage für das Maximum ist trivial; die zweite für das Minimum beweisen wir durch Fallunterscheidung:
Für $0\in A \wedge 0\in B$ ist die Aussage trivial; also sei etwa $0\notin B$:
Für $0\leq B$ verifiziert man die Behauptung in allen Fällen $0\leq A$, $0\in A$, $0\geq A$; Vorzeichenumkehr bei B ergibt jetzt auch die restlichen Fälle.

<u>Definition 3.3:</u> Kleinergleichrelation "$\leq$" in $V_n(\mathbb{I}(\mathbb{R}))$:

für $\mathfrak{a}=(A_i)$, $\mathfrak{b}=(B_i) \in V_n(\mathbb{I}(\mathbb{R}))$ sei

$$\mathfrak{a}\leq\mathfrak{b} \;:\Leftrightarrow\; A_i \leq B_i \quad \text{für alle } i.$$

Bezüglich der Relation "$\leq$" ist $V_n(\mathbb{I}(\mathbb{R}))$ eine Halbordnung. Zu $\mathfrak{a}=(A_i)$, $\mathfrak{b}=(B_i) \in V_n(\mathbb{I}(\mathbb{R}))$ existieren in $V_n(\mathbb{I}(\mathbb{R}))$ auch $\sup(\mathfrak{a},\mathfrak{b}):=(\sup(A_i,B_i))$ und $\inf(\mathfrak{a},\mathfrak{b}):=(\inf(A_i,B_i))$,

ebenso zu jeder nicht leeren beschränkten Teilmenge ein Supremum.

$V_n(\mathbb{I}(\mathbb{R}))$ ist also ein ordnungsvollständiger Verband und es gelten die Axiome (O1) bis (O5). Dies ergibt sich wegen der komponentenweisen Definition der Relation "$\leq$" sofort aus den entsprechenden Eigenschaften von $\mathbb{I}(\mathbb{R})$. Aus demselben Grund ist die Halbordnung auch hier mit der quasilinearen Struktur verträglich, d. h. es gelten (AO1) und (AO2) (siehe oben).

Für die Inklusion "$\subset$" gilt: $\mathfrak{a}\subset\mathfrak{b} \Leftrightarrow A_i \subset B_i$ **für alle** i.
Im Fall $\mathfrak{a}\subset\mathfrak{b}$ **und** $\mathfrak{a} = a \in V_n(\mathbb{R})$ **schreiben wir auch** $a \in \mathfrak{b}$
Bezüglich der Relation "$\subset$" ist $V_n(\mathbb{I}(\mathbb{R}))$ **eine Halbordnung.**

<u>Definition 3.4:</u> Es sei $\mathfrak{a}=(A_i)\in V_n(\mathbb{I}(\mathbb{R}))$; dann heißen die Vektoren

$$|\mathfrak{a}| := \sup_{a\in\mathfrak{a}} |a| := \sup_{a\in\mathfrak{a}} \sup(a,-a) \qquad \text{Absolutbetrag von } \mathfrak{a}$$

$$\mathrm{abs}(\mathfrak{a}) := \sup(\mathfrak{o},\mathfrak{a},-\mathfrak{a}) \qquad \text{Intervallbetrag von } \mathfrak{a}$$

($\mathfrak{o}$ ist der Nullvektor)

Beide Suprema existieren, da $V_n(\mathbb{I}(\mathbb{R}))$ ein ordnungsvollständiger Verband ist; wegen der komponentenweisen Definition der Halbordnung und damit des Supremums gilt:

$$|\mathfrak{a}| = (|A_i|) \in V_n(\mathbb{R}) \quad \text{bzw.}$$
$$\mathrm{abs}(\mathfrak{a}) = (\mathrm{abs}(A_i)) \in V_n(\mathbb{I}(\mathbb{R}));$$

damit gelten die in $\mathbb{I}(\mathbb{R})$ für $|A_i|$ bzw. $\mathrm{abs}(A_i)$ bewiesenen Regeln (A1), (A2), (A3) bzw. (IA1), (IA2), (IA3) (siehe Satz 3.1) jetzt auch für den Absolutbetrag $|\mathfrak{a}|$ bzw. Intervallbetrag $\mathrm{abs}(\mathfrak{a})$.

Weiter ergibt sich für jeden Vektor $\mathfrak{a}=(A_i) \in V_n(\mathbb{I}(\mathbb{R}))$ aus der entsprechenden Eigenschaft für die Komponenten:
Es existiert ein Vektor $a'\in V_n(\mathbb{R})$, mit der Eigenschaft

$$a'\in\mathfrak{a} \quad \text{und} \quad |a'| = \sup_{a\in\mathfrak{a}} |a| = |\mathfrak{a}|.$$

<u>Bemerkung:</u> Alle hier gegebenen Definitionen und sämtliche Überlesungen lassen sich wörtlich auf den $V_\infty(\mathbb{I}(\mathbb{R}))$ übertragen.

Für Intervallmatrizen aus $M_{n\times m}(\mathbb{I}(\mathbb{R}))$ definiert man die Relation "$\leq$" wie für Intervallvektoren komponentenweise, ebenso die zugehörigen Infima und Suprema; so erhält man wieder einen ordnungsvollständigen Verband (es gelten (O1) bis (O5)), und die Halbordnung ist mit der quasilinearen Struktur verträglich ((AO1) und (AO2)). Die Relationen "$\subset$" bzw. "$\in$" sind auch klar.

Zu $\mathfrak{A}=(A_{ik}) \in M_{n\times m}(\mathbb{I}(\mathbb{R}))$ nennen wir wie oben die Matrizen

$$|\mathfrak{A}| := \sup_{A\in\mathfrak{A}} |A| = \sup_{A\in\mathfrak{A}} \sup(A,-A) \qquad \text{Absolutbetrag von } \mathfrak{A}$$

$$\mathrm{abs}(\mathfrak{A}) := \sup(\mathfrak{O},\mathfrak{A},-\mathfrak{A}) \qquad \text{Intervallbetrag von } \mathfrak{A}$$

($\mathfrak{O}$ bedeute die Nullmatrix)

Es ist $\mathfrak{A}=(|A_{ik}|)\in M_{n\times m}(R)$ bzw. $\mathrm{abs}(\mathfrak{A})=(\mathrm{abs}(A_{ik}))\in M_{n\times m}(I(R))$.

Wie oben in $V_n(I(R))$ gelten die Regeln (A1), (A2), (A3) bzw. (IA1), (IA2), (IA3) von Satz 3.1 auch hier für den Absolutbetrag $|\mathfrak{A}|$ bzw. Intervallbetrag $\mathrm{abs}(\mathfrak{A})$
ebenso gilt für jede Matrix $\mathfrak{A}\in M_{n\times m}(I(R))$:

Es existiert eine Matrix $\mathfrak{A}^*\in M_{n\times m}(R)$ mit der Eigenschaft

$$\mathfrak{A}^*\in\mathfrak{A} \quad\text{und}\quad |\mathfrak{A}^*| = \sup_{\alpha\in\mathfrak{A}}|\alpha| = |\mathfrak{A}|.$$

3.3 Die Halbordnungen in $V_n(I(R))$ als Bilder von Halbordnungen in R^{2n}

Die oben im $V_n(I(R))$ definierte Halbordnung steht in engem Zusammenhang mit der natürlichen Halbordnung im R^{2n}; betrachten wir nämlich mit den Vektoren

$$\mathfrak{a}=([a_i^{(1)},a_i^{(2)}]),\ \mathfrak{b}=([b_i^{(1)},b_i^{(2)}])\in V_n(I(R))$$

zugleich die Vektoren

$$\vec{a}=\begin{pmatrix} a_1^{(1)}\\ a_1^{(2)}\\ \vdots\\ a_n^{(1)}\\ a_n^{(2)}\end{pmatrix},\quad \vec{b}=\begin{pmatrix} b_1^{(1)}\\ b_1^{(2)}\\ \vdots\\ b_n^{(1)}\\ b_n^{(2)}\end{pmatrix}\in R^{2n}=V_{2n}(R),$$

dann gilt nach unserer Definition der Relation "$\leq$" in $V_n(I(R))$ und der "natürlichen" Relation "$\leq$" in R^{2n}:

$$\mathfrak{a}\leq\mathfrak{b} \Leftrightarrow \vec{a}\leq\vec{b}.$$

Wir wollen den Zusammenhang zwischen den Halbordnungen in $V_n(I(R))$ und in R^{2n} untersuchen und erklären:
Dazu betrachten wir die oben schon benutzte Abbildung T des $V_n(I(R))$ auf die Teilmenge

$$R^{2n}_*:=\left\{\begin{pmatrix}a_1\\ \vdots\\ a_{2n}\end{pmatrix}\ \middle|\ a_{2i-1}\leq a_{2i}\ \text{für } i=1,..n\right\}\subset R^{2n}$$

die dem Vektor $\mathfrak{a}=([a_i^{(1)},a_i^{(2)}])\in V_n(I(R))$

den Vektor $T(\mathfrak{a}):=\vec{a}=\begin{pmatrix}a_1^{(1)}\\ \vdots\\ a_n^{(2)}\end{pmatrix}\in R^{2n}$ zuordnet.

Diese Abbildung T von $V_n(I(R))$ auf R^{2n}_* ist ein-eindeutig; deshalb kann man mit ihr auch jeder Halbordnung, die in einem der beiden Räume $V_n(I(R))$ bzw. R^{2n}_* vorgegeben ist, eine Halbordnung im zweiten Raum zuordnen. Man braucht nur die Bilder im zweiten Raum den Urbildern im ersten entsprechend zu ordnen.

Ist nun in R^{2n} und damit auch in $R^{2n}_*\subset R^{2n}$ eine Halbordnung vorgegeben, so definiert die Festsetzung

$\mathfrak{a}=T^{-1}(\vec{a})\leq\mathfrak{b}=T^{-1}(\vec{b}) :\Leftrightarrow \vec{a}\leq\vec{b}$ eine Halbordnung in $V_n(I(R))$.

Genauso kann in umgekehrter Richtung zu einer in $V_n(I(R))$ vorgegebenen Halbordnung eine solche in R^{2n}_* definiert werden; jede Halbordnung in der Teilmenge R^{2n}_* läßt sich auf den ganzen Raum R^{2n} fortsetzen; diese Fortsetzung ist jedoch im allgemeinen nicht eindeutig. Man hat also insgesamt eine Abbildung der Menge der Halbordnungen in R^{2n} auf die Menge der Halbordnungen in $V_n(I(R))$.

Die ein-eindeutige Abbildung T von $V_n(I(R))$ auf R^{2n}_* kann jedoch nicht strukturtreu sein, denn R^{2n} hat eine lineare, $V_n(I(R))$ aber eine quasilineare Struktur.

Dennoch gilt: $T(\mathfrak{a}+\mathfrak{b})=T(\mathfrak{a})+T(\mathfrak{b})$,
$T(\lambda\mathfrak{a})=\lambda T(\mathfrak{a})$ für $\lambda\geq 0$,
(für $\lambda<0$ ist $T(\lambda\mathfrak{a})\neq\lambda T(\mathfrak{a})$).

Wir nennen die Abbildung bedingt strukturtreu; dabei verwenden wir die folgende

Definition 3.5: Eine umkehrbar eindeutige Abbildung T eines quasilinearen Raumes QL mit Elementen α, $\mathfrak{b}$, $\mathfrak{t}$, ... auf eine Teilmenge eines linearen Raumes L mit Elementen $a:=T(\alpha)$, $b:=T(\mathfrak{b})$, $c:=T(\mathfrak{t})$, ... heißt bedingt strukturtreu, wenn

$$T(\alpha+\mathfrak{b})=T(\alpha)+T(\mathfrak{b})$$

und $$T(\lambda\alpha)=\lambda T(\alpha) \quad \text{für } \lambda\geq 0.$$

Von Interesse sind nun in $V_n(I(R))$ wie in R^{2n} vor allem Halbordnungen, die mit der jeweils zugrundeliegenden algebraischen Struktur verträglich sind; die diesbezüglichen Bedingungen waren:

(A01) $\alpha\leq\mathfrak{b} \succ \alpha+\mathfrak{t}\leq\mathfrak{b}+\mathfrak{t}$ für alle Elemente $\mathfrak{t}$ des betrachteten Raumes,

(A02) $\alpha\leq\mathfrak{b} \succ \lambda\alpha\leq\lambda\mathfrak{b}$ für alle $\lambda\geq 0$.

Wir betrachten wieder die einer gegebenen Halbordnung in R^{2n} durch die Definition $\alpha\leq\mathfrak{b} :\Leftrightarrow T(\alpha)\leq T(\mathfrak{b})$ zugeordnete Halbordnung in $V_n(I(R))$: Ist nämlich die vorgegebene Halbordnung in R^{2n} mit der linearen Struktur von R^{2n} verträglich, so wird die ihr zugeordnete Halbordnung in $V_n(I(R))$ mit der quasilinearen Struktur von $V_n(I(R))$ verträglich.

Beweis: Die Verträglichkeitseigenschaften (A01) und (A02) für die Halbordnung in R^{2n} gelten auch bei Einschränkung auf R^{2n}_{*}; da $V_n(I(R))$ durch T umkehrbar eindeutig auf R^{2n} abgebildet wird, können wir dafür auch schreiben:

$$T(\alpha)\leq T(\mathfrak{b}) \succ \begin{cases} T(\alpha)+T(\mathfrak{t})\leq T(\mathfrak{b})+T(\mathfrak{t}) & \text{für alle } \mathfrak{t}\in V_n(I(R)) \\ \lambda T(\alpha)\leq\lambda T(\mathfrak{b}) & \text{für alle } \lambda\geq 0 \end{cases}$$

Wegen der bedingten Strukturtreue der Abbildung T und der Definition der Halbordnung in $V_n(I(R))$ folgt daraus:

$$\alpha\leq\mathfrak{b} \succ \begin{cases} \alpha+\mathfrak{t}\leq\mathfrak{b}+\mathfrak{t} & \text{für alle } \mathfrak{t}\in V_n(I(R)), \\ \lambda\alpha\leq\lambda\mathfrak{b} & \text{für alle } \lambda\geq 0; \end{cases}$$

d. h. die in $V_n(I(R))$ erzeugte Halbordnung ist hier mit quasilinearer Struktur verträglich. □

Die Ergebnisse dieses Abschnittes fassen wir zusammen im folgenden

Satz 3.2:

a) Durch die Abbildung T wird jeder Halbordnung in R^{2n} eine Halbordnung in $V_n(I(R))$ zugeordnet und jede Halbordnung in $V_n(I(R))$ läßt sich auf diese Weise erzeugen.

b) Jede mit der linearen Struktur verträgliche Halbordnung in R^{2n} erzeugt dabei eine Halbordnung in $V_n(I(R))$, die dort mit der quasilinearen Struktur verträglich ist.

3.4 Intervallrechnung über allgemeinen halbgeordneten linearen Räumen

Im folgenden wird allgemein eine Intervallrechnung über einem beliebigen halbgeordneten linearen Raum B definiert und gezeigt, daß man dadurch stets einen quasilinearen Raum erhält; darüberhinaus wird geklärt, welche Voraussetzungen B erfüllen muß, damit die darauf aufgebaute Intervallrechnung einen Quasi-Banachraum ergibt. Alle bisher eingeführten Räume der Intervallrechnung lassen sich auf diese Weise erzeugen.

Sei also B (mit Elementen a, b, c ...) ein halbgeordneter linearer Raum über R; die Halbordnung von B sei mit der algebraischen Struktur verträglich, d. h. es gelte

(A01) $a\leq b \succ a+c\leq b+c$ für alle $a, b, c\in B$

(A02) $a\leq b \succ \lambda a\leq\lambda b$ für alle $a, b\in B$ und alle reellen Zahlen $\lambda\geq 0$.

Zu $a_1, a_2\in B$ mit $a_1\leq a_2$ bezeichnen wir die Menge der Elemente $A:=[a_1,a_2]:=\{a\in B \mid a_1\leq a\leq a_2\}$ als Intervall von a_1 bis a_2. $I(B)$ bezeichne die Menge aller dieser Intervalle über B, dann ist $I(B)\supset B$.

Wie in $\mathbf{I}(\mathbb{R})$ definieren wir für zwei Intervalle $A=[a_1,a_2]$ und $B=[b_1,b_2]$:

Gleichheit: $A=B \;:\Leftrightarrow\; a_1=b_1 \wedge a_2=b_2$,

Addition bzw. Subtraktion: $A\pm B \;:=\{a\pm b \mid a\in A \wedge b\in B\}$,

Multiplikation mit reellen Zahlen λ: $\lambda A \;:=\{\lambda a \mid a\in A\} = \begin{cases}[\lambda a_1,\lambda a_2] & \text{für } \lambda\geq 0\\ [\lambda a_2,\lambda a_1] & \text{für } \lambda<0\end{cases}$

Die Verträglichkeit der Halbordnung mit der algebraischen Struktur in $\mathbb{B}$ garantiert, daß mit $A,B\in\mathbf{I}(\mathbb{B})$ und $\lambda\in\mathbb{R}$ auch $A\pm B$ und λA in $\mathbf{I}(\mathbb{B})$ liegen.
Damit haben wir auch in $\mathbf{I}(\mathbb{B})$ eine algebraische Struktur erklärt; $\mathbf{I}(\mathbb{B})$ bezeichne ab jetzt die Menge der Intervalle über $\mathbb{B}$ mitsamt dieser algebraischen Struktur.
Genau wie früher für $\mathbf{I}(\mathbb{R})$ ergibt sich auch hier, daß $\mathbf{I}(\mathbb{B})$ alle Axiome des quasilinearen Raumes erfüllt, aber nicht linear ist, weil das Distributivgesetz nicht in voller Allgemeinheit gilt, sondern nur in der dem quasilinearen Raum gemäßen Form:

$$(\lambda+\mu)C = \lambda C+\mu C \quad \text{für alle } C\in\mathbf{I}(\mathbb{B}) \text{ und alle } \lambda,\mu\in\mathbb{R} \text{ mit } \lambda\mu>0,$$

$$\text{aber } (\lambda+\mu)C \neq \lambda C+\mu C \quad \text{für } \mathbb{B}\not\ni C\in\mathbf{I}(\mathbb{B}) \text{ und } \lambda\mu<0.$$

Allgemein gilt das subdistributive Gesetz:

$$(\lambda+\mu)C \subset \lambda C+\mu C \quad \text{für alle } C\in\mathbf{I}(\mathbb{B}) \text{ und alle } \lambda,\mu\in\mathbb{R}.$$

Ist $\mathbb{B}$ ein metrischer Raum, so läßt sich in $\mathbf{I}(\mathbb{B})$ auf mannigfache Art und Weise eine Metrik einführen; die Bemerkung im Anschluß an Satz 1.3 zeigt, wie man dabei vorzugehen hat. Wir betrachten nun gleich den Spezialfall, daß $\mathbb{B}$ ein Banachraum ist mit der Norm $\|a\|$. Wir fordern noch zusätzlich, daß die Metrik in $\mathbf{I}(\mathbb{B})$ eine Fortsetzung der in $\mathbb{B}$ durch die Norm $\|a\|$ gegebenen Metrik $\|a-b\|$ ist. Von den vielen dann noch möglichen Definitionen einer Metrik in $\mathbf{I}(\mathbb{B})$ erwähnen wir nur eine spezielle; wir definieren als Abstand zweier Intervalle

$$A=[a_1,a_2],\quad B=[b_1,b_2] \quad \text{aus } \mathbf{I}(\mathbb{B})$$
$$q(A,B):=\max(\|a_1-b_1\|,\|a_2-b_2\|).$$

$q(A,B)$ ist nun eine homogene und translationsinvariante Metrik in $\mathbf{I}(\mathbb{B})$ (man zeigt dies wie für die Metrik in $\mathbf{I}(\mathbb{R})$);
für $A=a$, $B=b\in\mathbb{B}$ wird $q(A,B)=\|a-b\|$.
Der Abstand $q(A,\theta)$ eines Elementes A vom neutralen Element der Addition ist nach Satz 2.4 eine Norm.
Bis jetzt ist $\mathbb{B}$ ein Banachraum mit strukturverträglicher Halbordnung (d. h. es gelten (A01) und (A02)). Wir benötigen für die folgenden Überlegungen noch den Zusammenhang dieser strukturverträglichen Halbordnung mit ihrem sogenannten Positivitätskegel.
Nach [13] gilt:
In einem linearen Raum $\mathbb{L}$ mit strukturverträglicher Halbordnung bildet die Menge $\{x \mid x\geq 0\}$ der positiven Elemente einen konvexen Kegel H mit den Eigenschaften

(H1) $a,\ b\in H \Rightarrow a+b\in H$,

(H2) $a\in H \wedge \lambda\geq 0 \Rightarrow \lambda a\in H$,

(H3) $H\cap(-H) = \{0\}$.

H heißt Positivitätskegel der gegebenen Halbordnung. Umgekehrt definiert jeder konvexe Kegel $H\subset\mathbb{L}$ mit den Eigenschaften (H1), (H2) und (H3) durch die Festsetzung $x\leq y \;:\Leftrightarrow\; y-x\in H$ eine strukturverträgliche Halbordnung in $\mathbb{L}$.
Jede strukturverträgliche Halbordnung in $\mathbb{L}$ ist so durch ihren Positivitätskegel H eindeutig bestimmt.

Ist jetzt der zur Halbordnung in $\mathbb{B}$ gehörige Positivitätskegel H in $\mathbb{B}$ abgeschlossen, so wird $\mathbf{I}(\mathbb{B})$ bezüglich der Metrik $q(A,B)$ auch vollständig, also ein Quasi-Banachraum.

Zum Beweis betrachten wir eine Cauchy-Folge von Intervallen $[a_\nu,b_\nu]\in\mathbf{I}(\mathbb{B})$; die Cauchy-Konvergenz dieser Folge ist nach Definition der Metrik in $\mathbf{I}(\mathbb{B})$ äquivalent mit der (gleichgradigen) Cauchy-Konvergenz der beiden Folgen $\{a_\nu\}$ und $\{b_\nu\}$ in $\mathbb{B}$;

da B vollständig ist, existieren in B $\lim_{\nu\to\infty} a_\nu$ und $\lim_{\nu\to\infty} b_\nu$; $[a_\nu, b_\nu] \in I(B)$ ist gleichbedeutend mit $a_\nu \leq b_\nu$, also $b_\nu - a_\nu \in H$.
Aus der Abgeschlossenheit des Positivitätskegels H folgt nun: $\lim_{\nu\to\infty} (b_\nu - a_\nu) = (\lim_{\nu\to\infty} b_\nu - \lim_{\nu\to\infty} a_\nu) \in H$; das wieder ist äquivalent mit $\lim_{\nu\to\infty} a_\nu \leq \lim_{\nu\to\infty} b_\nu$.
$I(B)$ enthält also ein Element $[\lim_{\nu\to\infty} a_\nu, \lim_{\nu\to\infty} b_\nu]$, und es ist $\lim_{\nu\to\infty}[a_\nu, b_\nu] = [\lim_{\nu\to\infty} a_\nu, \lim_{\nu\to\infty} b_\nu]$, d. h. $I(B)$ ist vollständig.

Bemerkung: Die angegebene allgemeine Konstruktion des Intervallraumes $I(B)$ über dem linearen halbgeordneten Raum B führt in den Spezialfällen $B = R$ bzw. $V_n(R)$ bzw. $V_\infty(R)$ oder $M_{n\times m}(R)$ (zugrundegelegt sei jeweils die "natürliche" Halbordnung) auf die oben schon eingeführten Räume $I(R)$ bzw. $V_n(I(R))$ bzw. $V_\infty(I(R))$ und $M_{n\times m}(I(R))$.

Wie in Abschnitt 3.2 läßt sich mit Hilfe der Halbordnung von B eine Halbordnung in $I(B)$ erklären; ebenso lassen sich alle Überlegungen von Abschnitt 3.3 übertragen, es ergibt sich ein entsprechender Zusammenhang der Ordnungsstrukturen des linearen Raumes $B \times B$ mit denjenigen des quasilinearen Raumes $I(B)$.

4. Normen in $V_n(I(R))$ und $M_n(I(R))$

In diesem Abschnitt untersuchen wir die Eigenschaften von Normen für Intervallvektoren und Intervallmatrizen. Wichtig sind hier vor allem die sogenannten Abstandsnormen und unter diesen wieder besonders die monotonen Normen. Wie im Reellen gehört zu jeder Vektornorm eine Klasse von Matrixnormen, die mit ihr verträglich sind; von diesen ist auch hier eine durch eine Minimaleigenschaft eindeutig ausgezeichnet, und in besonderer Weise der gegebenen Vektornorm zugeordnet. Und diese Normen haben Eigenschaften, die denen verträglicher bzw. zugeordneter Normen für reelle bzw. komplexe Vektoren und Matrizen ähnlich sind. Schließlich beweisen wir auch noch einen wichtigen Satz über den Zusammenhang der monotonen Normen einer Matrix mit dem Spektralradius ihrer Absolutbetrages.

4.1 Normen und Abstandsnormen

Definition 4.1: Ein für alle $\mathfrak{x} \in V_n(I(R))$ bzw. $\mathfrak{A} \in M_n(I(R))$ definiertes reelles Funktional $\|\mathfrak{x}\|$ bzw. $\|\mathfrak{A}\|$ heißt Norm, wenn es folgende Bedingungen erfüllt:

1)	$\mathfrak{x} \neq \mathfrak{o} \Rightarrow \|\mathfrak{x}\| > 0$	bzw.	1')	$\mathfrak{A} \neq \mathfrak{o} \Rightarrow \|\mathfrak{A}\| > 0$
2)	$\|\mathfrak{x} + \mathfrak{y}\| \leq \|\mathfrak{x}\| + \|\mathfrak{y}\|$		2')	$\|\mathfrak{A} + \mathfrak{B}\| \leq \|\mathfrak{A}\| + \|\mathfrak{B}\|$
3)	$\|c\,\mathfrak{x}\| = \|c\|\,\|\mathfrak{x}\|$		3')	$\|c\,\mathfrak{A}\| = \|c\|\,\|\mathfrak{A}\|$

jeweils für alle $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$ bzw. alle $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$ und alle $c \in R$; $\mathfrak{o}$ steht hier für den Nullvektor bzw. die Nullmatrix.

Eine Matrixnorm heißt multiplikativ, wenn für alle $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$ gilt: 4') $\|\mathfrak{A}\mathfrak{B}\| \leq \|\mathfrak{A}\| \cdot \|\mathfrak{B}\|$

Normen in $V_n(R)$ bzw. $M_n(R)$ werden durch dieselben Forderungen definiert; die Einschränkung einer Norm in $V_n(I(R))$ bzw. $M_n(I(R))$

auf $V_n(R)$ bzw. $M_n(R)$ gibt also eine Norm in $V_n(R)$ bzw. $M_n(R)$.

Um zwischen der Norm als auf $V_n(I(R))$ bzw. $M_n(I(R))$ erklärtem Funktional und dem Wert der Norm für ein bestimmtes Element sicher unterscheiden zu können, reservieren wir die Buchstaben $\mathfrak{w}$ und $\mathfrak{m}$ für die folgende Sprechweise: "Norm $\|\mathfrak{w}\|$ in $V_n(I(R))$" bzw. "Norm $\|\mathfrak{m}\|$ in $M_n(I(R))$"; damit sei die Norm als das für alle Elemente aus $V_n(I(R))$ bzw. $M_n(I(R))$ erklärte Funktional gemeint.

Im folgenden betrachten wir vor allem die sogenannten Abstandsnormen; bei diesen kommt nämlich gerade die Auffassung des Rechnens mit Intervallen als Verallgemeinerung des Rechnens mit reellen Zahlen besonders zum Ausdruck. Wir besprechen dabei Vektor- und Matrixnormen nebeneinander.

Der folgende Satz ergibt sich aus entsprechenden Sätzen in [8].

Satz 4.1: Sind $\|\underset{\cdot}{\mathfrak{x}}\|$ bzw. $\|\underset{\cdot}{\mathfrak{A}}\|$ Normen für reelle Vektoren $\underset{\cdot}{\mathfrak{x}} \in V_n(R)$ bzw. Matrizen $\underset{\cdot}{\mathfrak{A}} \in M_n(R)$, dann sind

$\|\mathfrak{x}\| := \sup\limits_{\underset{\cdot}{\mathfrak{x}} \in \mathfrak{x}} \|\underset{\cdot}{\mathfrak{x}}\|$ bzw. $\|\mathfrak{A}\| := \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|\underset{\cdot}{\mathfrak{A}}\|$ Normen für Intervallvektoren $\mathfrak{x} \in V_n(I(R))$ bzw. Intervallmatrizen $\mathfrak{A} \in M_n(I(R))$.

Zum Beweis (nach [8]) verifiziert man die Eigenschaften 1), 2), 3) bzw. 1'), 2'), 3'). Wir zeigen dies etwa für die Matrixnorm:

1') $\|\mathfrak{A}\| = \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|\underset{\cdot}{\mathfrak{A}}\| \geq 0$, mit $\|\mathfrak{A}\| = 0 \iff \forall \underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}: \underset{\cdot}{\mathfrak{A}} = \underset{\cdot}{\mathfrak{o}} \iff \mathfrak{A} = \mathfrak{o}$;

2') $\|\mathfrak{A} + \mathfrak{B}\| = \sup\limits_{\substack{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A} \\ \underset{\cdot}{\mathfrak{B}} \in \mathfrak{B}}} \|\underset{\cdot}{\mathfrak{A}} + \underset{\cdot}{\mathfrak{B}}\| \leq \sup\limits_{\substack{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A} \\ \underset{\cdot}{\mathfrak{B}} \in \mathfrak{B}}} (\|\underset{\cdot}{\mathfrak{A}}\| + \|\underset{\cdot}{\mathfrak{B}}\|) = \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|\underset{\cdot}{\mathfrak{A}}\| + \sup\limits_{\underset{\cdot}{\mathfrak{B}} \in \mathfrak{B}} \|\underset{\cdot}{\mathfrak{B}}\| = \|\mathfrak{A}\| + \|\mathfrak{B}\|$;

3') $\|c\mathfrak{A}\| = \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|c\underset{\cdot}{\mathfrak{A}}\| = |c| \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|\underset{\cdot}{\mathfrak{A}}\| = |c| \|\mathfrak{A}\|$. □

Die Normen $\|\mathfrak{x}\| = \sup\limits_{\underset{\cdot}{\mathfrak{x}} \in \mathfrak{x}} \|\underset{\cdot}{\mathfrak{x}}\|$ in $V_n(I(R))$ und $\|\mathfrak{A}\| = \sup\limits_{\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}} \|\underset{\cdot}{\mathfrak{A}}\|$ in $M_n(I(R))$ sind jeweils Fortsetzungen der in den linearen Teilräumen $V_n(R)$ und $M_n(R)$ gegebenen Normen $\|\underset{\cdot}{\mathfrak{x}}\|$ und $\|\underset{\cdot}{\mathfrak{A}}\|$.

Definition 4.2: Solche Normen für Intervallvektoren bzw. Intervallmatrizen, die sich auf diese Weise durch Supremumsbildung aus Normen für reelle Vektoren bzw. Matrizen erzeugen lassen, nennen wir Abstandsnormen.

Bemerkung: Nicht alle Normen in $V_n(I(R))$ bzw. $M_n(I(R))$ sind Abstandsnormen, wie das folgende Beispiel in $I(R) = V_1(I(R))$ zeigt: Es sei $\|([a_1, a_2])\| := |a_1| + |a_2|$.
Dies ist eine Norm, aber es ist etwa für $A := [2,10]$
$\|(A)\| := \|([2,10])\| = 12 < \|([10,10])\| = 20 = \sup\limits_{\underset{\cdot}{A} \in A} \|(\underset{\cdot}{A})\|$.
Hier ist $\|(A)\| \neq \sup\limits_{\underset{\cdot}{A} \in A} \|(\underset{\cdot}{A})\|$
Aus $B \subset A$ kann hier nicht $\|(B)\| \leq \|(A)\|$ gefolgert werden (z. B. ist $\|([2,10])\| = 12 < \|([4,10])\| = 14$).

Für Abstandsnormen gilt jedoch

Satz 4.2: Es seien $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$ und $\|\mathfrak{w}\|$ eine Abstandsnorm in $V_n(I(R))$; dann gilt:

a) $\mathfrak{x} \subset \mathfrak{y} \Rightarrow \|\mathfrak{x}\| \leq \|\mathfrak{y}\|$

b) $\|\mathfrak{y}\| > \|\mathfrak{x}\| \Rightarrow \mathfrak{y} \not\subset \mathfrak{x}$

Beweis: a) Das Supremum über die Teilmenge $\mathfrak{x}$ von $\mathfrak{y}$ ist nicht größer als das Supremum über $\mathfrak{y}$

b) Die Annahme $\mathfrak{y} \subset \mathfrak{x}$ führt wegen a) auf den Widerspruch $\|\mathfrak{y}\| \leq \|\mathfrak{x}\|$. □

Bemerkung: Für Abstandsnormen in $M_n(I(R))$ gilt die entsprechende Aussage.

Vektornormen in $V_n(R)$ lassen sich bekanntlich geometrisch mit Hilfe gewisser konvexer Körper definieren. Wir skizzieren jetzt kurz nach [7] diese Darstellung einer Vektornorm in $V_n(R)$ durch ihren sogenannten Eichkörper und verallgemeinern sie dann für die zugehörige Abstandsnorm.

Eine abgeschlossene, beschränkte, konvexe Punktmenge in $V_n(R)$ heißt konvexer Körper. Ist K ein konvexer Körper in $V_n(R)$, $\underset{\cdot}{\mathfrak{A}}$ eine Matrix aus $M_n(R)$, so bedeute $\underset{\cdot}{\mathfrak{A}}K$ den Körper

$$\underset{\cdot}{\mathfrak{A}}K := \{\underset{\cdot}{\mathfrak{A}}\underset{\cdot}{\mathfrak{x}} \mid \underset{\cdot}{\mathfrak{x}} \in K\};$$

für $(a\mathcal{E})K$ ($\mathcal{E}$ ist die Einheitsmatrix, $a \in R$) schreiben wir auch aK.
Ein konvexer Körper heißt ausbalanziert, wenn aus $\underset{\cdot}{\varphi} \in K$ und $|a| \le 1$ folgt $a\underset{\cdot}{\varphi} \in K$.

Ist in $V_n(R)$ irgendein ausbalanzierter, konvexer Körper vorgegeben, der den Nullpunkt als inneren Punkt enthält, so gibt die Definition

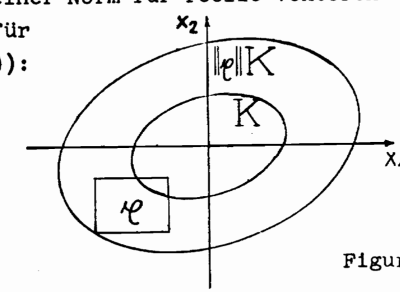
$$\|\underset{\cdot}{\varphi}\| := \inf\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in K\}$$

eine Vektornorm in $V_n(R)$.
Geht man umgekehrt aus von einer Vektornorm $\|\underset{\cdot}{\varphi}\|$ in $V_n(R)$, so ist die Menge $K := \{\underset{\cdot}{\varphi} \mid \|\underset{\cdot}{\varphi}\| \le 1\}$ ein ausbalanzierter, konvexer Körper, der den Nullpunkt als inneren Punkt enthält, und für die gegebene Norm gilt die Darstellung:

(RN) $$\|\underset{\cdot}{\varphi}\| = \inf\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in rK\}.$$

K heißt deshalb Eichkörper der Vektornorm

Jetzt betrachten wir die zu einer Norm für reelle Vektoren aus $V_n(R)$ gehörige Abstandsnorm für Intervallvektoren aus $V_n(I(R))$:
Ein Intervallvektor $\varphi = (X_i)$ aus $V_n(I(R))$ wird in $V_n(R)$ durch einen abgeschlossenen beschränkten Quader mit den Kanten X_i dargestellt (siehe Figur 1).



Figur 1

Die Definition der Abstandsnorm in $V_n(I(R))$ und der Norm in $V_n(R)$ ergeben nun

$$\|\varphi\| := \sup_{\underset{\cdot}{\varphi} \in \varphi} \|\underset{\cdot}{\varphi}\| = \sup_{\underset{\cdot}{\varphi} \in \varphi} \inf\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in rK\}$$

Daraus folgt:

(AN) $$\|\varphi\| = \inf\{r \mid r \ge 0 \wedge \varphi \subset rK\}$$

Beweis: $\forall \underset{\cdot}{\varphi} \in \varphi$: $\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in rK\} \supset \{r \mid r \ge 0 \wedge \varphi \in rK\}$

$\Rightarrow$ $\forall \underset{\cdot}{\varphi} \in \varphi$: $\|\underset{\cdot}{\varphi}\| = \inf\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in rK\} \le \inf\{r \mid r \ge 0 \wedge \varphi \in rK\}$

$\Rightarrow$ $r^* := \sup_{\underset{\cdot}{\varphi} \in \varphi} \|\underset{\cdot}{\varphi}\| = \sup_{\underset{\cdot}{\varphi} \in \varphi} \inf\{r \mid r \ge 0 \wedge \underset{\cdot}{\varphi} \in rK\} \le \inf\{r \mid r \ge 0 \wedge \varphi \in rK\}$

Für r^* gilt: $\forall \underset{\cdot}{\varphi} \in \varphi : \underset{\cdot}{\varphi} \in r^*K \Rightarrow \varphi \subset r^*K \Rightarrow r^* \ge \inf\{r \mid r \ge 0 \wedge \varphi \subset rK\}$

Aus den beiden Ungleichungen für r^* folgt:

$$r^* = \sup_{\underset{\cdot}{\varphi} \in \varphi} \|\underset{\cdot}{\varphi}\| = \inf\{r \mid r \ge 0 \wedge \varphi \subset rK\}. \square$$

Die Darstellung (AN) für die Abstandsnorm in $V_n(I(R))$ mit Hilfe des Eichkörpers K verallgemeinert gerade die Darstellung (RN) für die Norm in $V_n(R)$, von der wir ausgegangen sind.
Die Zuordnung der Abstandsnorm in $V_n(I(R))$ zur Ausgangsnorm in $V_n(R)$ und damit zu deren Eichkörper K in $V_n(R)$ ist umkehrbar eindeutig; die Darstellung (AN) für die Abstandsnorm berechtigt uns, von K auch als dem Eichkörper der zugehörigen Abstandsnorm zu sprechen.

4.2 Monotone Normen

Definition 4.2: a) Normen für Vektoren aus $V_n(I(R))$ bzw. $V_n(R)$ heißen monoton, wenn (M) $|\varphi| \le |\eta| \Rightarrow \|\varphi\| \le \|\eta\|$.

b) Normen für Matrizen aus $M_n(I(R))$ bzw. $M_n(R)$ heißen monoton, wenn (M) $|\mathfrak{A}| \le |\mathfrak{B}| \Rightarrow \|\mathfrak{A}\| \le \|\mathfrak{B}\|$

Eine äquivalente Definition erhalten wir jeweils durch die Forderung der Eigenschaft

a) (A) $\|\varphi\| = \||\varphi|\|$ für alle $\varphi \in V_n(I(R))$ bzw. alle $\varphi \in V_n(R)$
b) (A) $\|\mathfrak{A}\| = \||\mathfrak{A}|\|$ für alle $\mathfrak{A} \in M_n(I(R))$ bzw. alle $\mathfrak{A} \in M_n(R)$

Die Eigenschaften (A) und (M) sind nämlich für alle Normen in den oben angegebenen Räumen äquivalent.

Beweis: Wir beweisen zuerst die Äquivalenz der Eigenschaften (A) und (M) für Vektornormen in $V_n(R)$. K sei der Eichkörper der betrachteten Vektornorm .
Zunächst werde (M) vorausgesetzt:
Für $\varphi \in V_n(R)$ haben φ und $|\varphi|$ denselben Absolutbetrag $|\varphi| = ||\varphi||$
Aus $|\varphi| \le ||\varphi||$ bzw. $|\varphi| \ge ||\varphi||$ folgt mit (M)
$\|\varphi\| \le \||\varphi|\|$ bzw. $\|\varphi\| \ge \||\varphi|\|$; damit gilt: (A) $\|\varphi\| = \||\varphi|\|$. $\square$

Der folgende Schluß von (A) auf (M) wird aus [7], Seite 47, übernommen; dort wird nämlich eine entsprechende Äquivalenz für Vektornormen in $V_n(C)$ bewiesen ($V_n(C)$ ist der lineare Raum der n-tupel komplexen Zahlen).
Wir betrachten zwei Vektoren $\mathfrak{x}, \mathfrak{y} \in V_n(R)$ mit $|\mathfrak{x}| \le |\mathfrak{y}|$;
o. B. d. A. sei $\|\mathfrak{y}\| = 1$. Für jede Diagonalmatrix $\vartheta \in M_n(R)$ mit $|\vartheta| = \mathfrak{E}$ ($\mathfrak{E}$ ist die Einheitsmatrix) ist $|\mathfrak{y}| = |\vartheta\mathfrak{y}|$, also $\|\mathfrak{y}\| = \||\mathfrak{y}|\| = \|\vartheta\mathfrak{y}\|$; folglich liegt $\vartheta\mathfrak{y}$ auf dem Rand von K. Wegen der Konvexität von K kann jetzt $\mathfrak{x}$ nur noch im Innern oder auf dem Rand von K liegen; also gilt: (M) $\|\mathfrak{x}\| \le \|\mathfrak{y}\|$.

Für Vektornormen in $V_n(I(R))$ folgt der Schluß von (M) auf (A) wie oben für Normen in $V_n(R)$.
Jetzt werde (A) vorausgesetzt:
Für $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$ gilt $|\mathfrak{x}|, |\mathfrak{y}| \in V_n(R)$.
Für Vektornormen in $V_n(R)$ gilt aber der Schluß von (A) auf (M) (siehe oben); damit folgt aus $|\mathfrak{x}| \le |\mathfrak{y}|$ $\quad \||\mathfrak{x}|\| \le \||\mathfrak{y}|\|$;
mit (A) erhalten wir $\quad \|\mathfrak{x}\| = \||\mathfrak{x}|\| \le \||\mathfrak{y}|\| = \|\mathfrak{y}\|$,
also gilt insgesamt (M) $\quad |\mathfrak{x}| \le |\mathfrak{y}| \Rightarrow \|\mathfrak{x}\| \le \|\mathfrak{y}\|$.

Damit haben wir die Äquivalenz der Eigenschaften (A) und (M) für Vektornormen in $V_n(R)$ bzw. $V_n(I(R))$ bewiesen.
Für Normen von Matrizen aus $M_n(R)$ bzw. $M_n(I(R))$ folgt sie jetzt einfach daraus, daß man jede solche Matrixnorm als Vektornorm in $V_{n^2}(R)$ bzw. $V_{n^2}(I(R))$ auffassen kann. □

Eine unserer Definition äquivalente Definition für monotone Normen in $V_n(I(R))$ bzw. $M_n(I(R))$ wurde bereits in [8] gegeben; aus [8] übernehmen wir auch den folgenden

Satz 4.3: Ist eine Norm für Vektoren aus $V_n(R)$ bzw. Matrizen aus $M_n(R)$ monoton, dann ist auch die durch sie definierte Abstandsnorm für Intervallvektoren aus $V_n(I(R))$ bzw. Intervallmatrizen aus $M_n(I(R))$ monoton.

Beweis: (Der Beweis wird hier anders als in [8] geführt.)
Wir geben den Beweis nur im Falle der Matrixnorm; im Falle der Vektornorm verläuft er analog.

Für alle $A \in \mathfrak{A}$ ist $|A| \le \sup_{A\in\mathfrak{A}} |A| = |\mathfrak{A}|$, also wegen der Monotonie der Norm $\quad \|A\| \le \||\mathfrak{A}|\|$, daraus folgt

$$\|\mathfrak{A}\| = \sup_{A\in\mathfrak{A}} \|A\| \le \||\mathfrak{A}|\| . \qquad (1)$$

Nach Abschnitt 3.2 existiert ein $A^* \in \mathfrak{A}$ mit der Eigenschaft $|A^*| = |\mathfrak{A}|$, also wegen der Monotonie der Norm

$$\|A^*\| = \||\mathfrak{A}|\| . \qquad (2)$$

(1) und (2) ergeben zusammen $\quad \|\mathfrak{A}\| = \sup_{A\in\mathfrak{A}} \|A\| = \||\mathfrak{A}|\|$. □

Satz 4.4: Jede monotone Norm für Intervallvektoren bzw. Intervallmatrizen ist eine Abstandsnorm.

Beweis: Wir führen den Beweis nur für die Matrixnorm, für die Vektornorm verläuft er analog.

Für alle $A \in \mathfrak{A}$ ist $|A| \le \sup_{A\in\mathfrak{A}} |A| = |\mathfrak{A}|$, also wegen der Monotonie der Norm

$$\|A\| \le \|\mathfrak{A}\| . \qquad (1)$$

Nach Abschnitt 3.2 existiert ein $A^* \in \mathfrak{A}$ mit der Eigenschaft $|A^*| = |\mathfrak{A}|$, also wegen der Monotonie der Norm

$$\|A^*\| = \|\mathfrak{A}\| . \qquad (2)$$

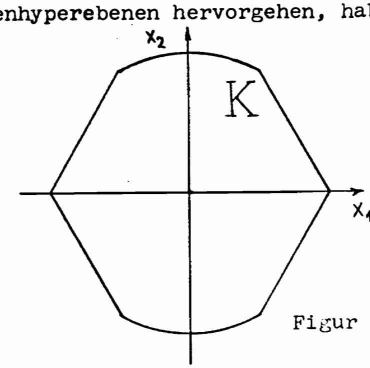
(1) und (2) ergeben zusammen $\quad \|\mathfrak{A}\| = \sup_{A\in\mathfrak{A}} \|A\|$. □

In der definierenden Eigenschaft (M) für die monotonen Normen kommt eine gewisse Verträglichkeit dieser Normen mit der vorgegebenen "natürlichen" Halbordnung des betreffenden Raumes zum Ausdruck. Wegen dieser Monotonieeigenschaft sind monotone Normen auch besonders handlich und für Abschätzungen gut geeignet. Dies gilt schon für Vektornormen in $V_n(R)$ bzw. $V_n(C)$ und entsprechende Matrixnormen. In der Intervallrechnung sind nun die Gründe für die Verwendung monotoner Normen noch zwingender.
Der Definition von Intervallen, Intervallvektoren und Intervallmatrizen und damit der Definition der Räume $I(R) = V_1(I(R))$, $V_n(I(R))$ bzw. $M_n(I(R))$ liegt die "natürliche" Halbordnung in den Ausgangsräumen R, $V_n(R)$ bzw. $M_n(R)$ zugrunde (siehe Abschnitt 1 und 3.4). Die Eigenschaften der Räume $V_n(I(R))$ und

$M_n(I(R))$ sind damit wesentlich durch die "natürliche" Halbordnung in $V_n(R)$ bzw. $M_n(R)$ bestimmt. Aber nur monotone Normen in $V_n(I(R))$ bzw. $M_n(I(R))$ sind auf Grund ihrer Eigenschaft (M) mit der Halbordnung in den Teilräumen $V_n(R)$ bzw. $M_n(R)$ - die Absolutbeträge von Elementen aus $V_n(I(R))$ bzw. $M_n(I(R))$ liegen in den Teilräumen $V_n(R)$ bzw. $M_n(R)$ - verträglich. Wegen dieses Zusammenhangs ist zu erwarten, daß für die Intervallrechnung nur die Verwendung monotoner Normen sinnvoll ist.

Wir gehen noch kurz auf die Gestalt des Eichkörpers monotoner Vektornormen ein.
Monotone Normen für reelle Vektoren $\varphi \in V_n(R)$ bzw. Intervallvektoren $\varphi \in V_n(I(R))$ sind durch die Eigenschaft
(A) $\|\varphi\| = \| |\varphi| \|$ bzw. $\|\varphi\| = \| |\varphi| \|$ gekennzeichnet. Der Zusammenhang zwischen beiden ist gegeben durch $|\varphi| = \sup\limits_{\varphi \in \varphi} |\varphi|$.

Alle Vektoren φ', die aus einem gegebenen Vektor φ durch Spiegelung an beliebigen Koordinatenhyperebenen hervorgehen, haben denselben Betrag $|\varphi|$. Bei Zugrundelegung einer monotonen Norm sind die Normen dieser Vektoren alle gleich, deshalb muß der zu einer monotonen Norm gehörige Eichkörper invariant sein gegenüber Spiegelungen an beliebigen Koordinatenhyperebenen (siehe Figur 2).



Figur 2

Umgekehrt definiert auch jeder solche, gegenüber Spiegelungen an beliebigen Koordinatenebenen invariante konvexe Körper eine Norm mit der Eigenschaft (A) $\|\varphi\| = \| |\varphi| \|$, d. h. eine monotone Norm.
Interessant ist hier, daß in der Intervallrechnung selbst auch Elemente mit entsprechenden Symmetrieeigenschaften auftreten.

<u>Definition 4.3:</u> $A \in I(R)$ bzw. $\varphi \in V_n(I(R))$ bzw. $\mathfrak{A} \in M_n(I(R))$ heißt symmetrisch, wenn $A = -A$ bzw. $\varphi = -\varphi$ bzw. $\mathfrak{A} = -\mathfrak{A}$.

Dieses Auftreten symmetrischer Elemente ungleich Null ist kennzeichnend für die Intervallrechnung und die Nichtlinearität der ihr zugrundeliegenden Räume.

Etwa ein Vektor $\varphi \in V_n(I(R))$ ist genau dann symmetrisch, wenn er bei Spiegelung an beliebigen Koordinatenhyperebenen in sich übergeht. Durch die entsprechende Invarianzeigenschaft bei der Norm, nämlich Invarianz des Eichkörpers bei Spiegelung an beliebigen Koordinatenhyperebenen sind gerade die monotonen Normen gekennzeichnet.

<u>4.3 Verträgliche und zugeordnete Normen</u>

<u>Definition 4.4:</u> Eine Vektornorm $\|\mathfrak{w}\|$ in $V_n(I(R))$ und eine Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$ heißen verträglich, wenn die Matrixnorm multiplikativ ist, und für alle $\varphi \in V_n(I(R))$ und alle $\mathfrak{A} \in M_n(I(R))$ gilt: $\|\mathfrak{A}\varphi\| \le \|\mathfrak{A}\| \cdot \|\varphi\|$

bzw. zugeordnet, wenn sie verträglich sind und zu jeder Matrix $\mathfrak{A} \in M_n(I(R))$ ein Vektor $\mathfrak{y} \in V_n(I(R))$ (ungleich dem Nullvektor) existiert, so daß $\|\mathfrak{A}\mathfrak{y}\| = \|\mathfrak{A}\| \, \|\mathfrak{y}\|$

Zu einer gegebenen Vektornorm $\|\mathfrak{w}\|$ in $V_n(I(R))$ gibt es, wie wir sehen werden, immer eine Klasse von mit ihr verträglichen Matrixnormen in $M_n(I(R))$. Die zugeordnete Matrixnorm ist dagegen eindeutig. Dies ergibt sich aus dem folgenden Satz über die Minimaleigenschaft der zugeordneten Matrixnorm. (Der folgende Satz, das anschließende Korollar und die zugehörigen Beweise stimmen mit den entsprechenden bekannten Aussagen für Normen reeller Vektoren und Matrizen überein. Sie werden aber hier für Normen von Intervallvektoren und Intervallmatrizen ausgesprochen.)

Satz 4.5: Es seien $\|\mathfrak{x}\|$ eine Vektornorm in $V_n(I(R))$, $\|\mathfrak{M}\|_V$ eine mit $\|\mathfrak{x}\|$ verträgliche, $\|\mathfrak{M}\|_Z$ eine der Norm $\|\mathfrak{x}\|$ zugeordnete Matrixnorm in $M_n(I(R))$, dann gilt für jede Matrix $\mathfrak{A} \in M_n(I(R))$: $\quad \|\mathfrak{A}\|_Z \leq \|\mathfrak{A}\|_V$

Beweis: Zu jedem $\mathfrak{A} \in M_n(I(R))$ existiert ein $\mathfrak{y} \neq \mathfrak{o}$ aus $V_n(I(R))$, derart, daß $\|\mathfrak{A}\mathfrak{y}\| = \|\mathfrak{A}\|_Z \|\mathfrak{y}\|$; zugleich ist $\|\mathfrak{A}\mathfrak{y}\| \leq \|\mathfrak{A}\|_V \|\mathfrak{y}\|$; wegen $\|\mathfrak{y}\| \neq 0$ folgt daraus die Behauptung. □

Korollar: Die einer gegebenen Vektornorm $\|\mathfrak{x}\|$ in $V_n(I(R))$ zugeordnete Matrixnorm $\|\mathfrak{M}\|_Z$ in $M_n(I(R))$ ist eindeutig.

Beweis: Sind $\|\mathfrak{M}\|_{Z1}$ und $\|\mathfrak{M}\|_{Z2}$ der gegebenen Vektornorm $\|\mathfrak{x}\|$ zugeordnet, so sind beide mit ihr verträglich; deshalb gilt für jede Matrix $\mathfrak{A} \in M_n(I(R))$: $\|\mathfrak{A}\|_{Z1} \leq \|\mathfrak{A}\|_{Z2}$ und $\|\mathfrak{A}\|_{Z1} \geq \|\mathfrak{A}\|_{Z2}$,
also $\|\mathfrak{A}\|_{Z1} = \|\mathfrak{A}\|_{Z2}$. □

Die folgenden Sätze geben Auskunft über das Verhalten der Eigenschaften "verträglich" und "zugeordnet" beim Übergang von Normen für reelle Vektoren bzw. Matrizen zu den daraus abgeleiteten Abstandsnormen; zugleich führen sie die Existenz verträglicher bzw. zugeordneter Normen für Intervallvektoren und Intervallmatrizen auf die Existenz verträglicher bzw. zugeordneter Normen für reelle Vektoren bzw. Matrizen zurück.

Satz 4.6: a) Ist eine monotone Norm in $M_n(R)$ multiplikativ, so gilt dies auch für die damit gebildete Abstandsnorm in $M_n(I(R))$.

b) Ist eine monotone Norm in $V_n(R)$ verträglich mit einer monotonen Norm in $M_n(R)$, so gilt dies auch für die damit gebildeten Abstandsnormen in $V_n(I(R))$ und $M_n(I(R))$.

(Beide Aussagen sind bereits in entsprechenden Sätzen in [8] enthalten.)

Beweis: a) Für alle $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$ gilt $|\mathfrak{A}|, |\mathfrak{B}| \in M_n(R)$; aus $|\mathfrak{A}\mathfrak{B}| \leq |\mathfrak{A}||\mathfrak{B}|$ und der Monotonie der Norm folgt jetzt

$$\|\mathfrak{A}\mathfrak{B}\| = \||\mathfrak{A}\mathfrak{B}|\| \leq \||\mathfrak{A}||\mathfrak{B}|\| \leq \||\mathfrak{A}|\| \, \||\mathfrak{B}|\| = \|\mathfrak{A}\| \, \|\mathfrak{B}\|.$$

b) Die Multiplikativität der Abstandsnorm in $M_n(I(R))$ wurde bereits unter a) nachgewiesen. Auch für alle $\mathfrak{x} \in V_n(I(R)$ ist $|\mathfrak{x}| \in V_n(R)$; aus $|\mathfrak{A}\mathfrak{x}| \leq |\mathfrak{A}||\mathfrak{x}|$ und der Monotonie der Normen folgt wie oben: $\|\mathfrak{A}\mathfrak{x}\| = \||\mathfrak{A}\mathfrak{x}|\| \leq \||\mathfrak{A}||\mathfrak{x}|\| = \||\mathfrak{A}|\| \, \||\mathfrak{x}|\| = \|\mathfrak{A}\| \cdot \|\mathfrak{x}\|$. □

Hilfssatz 4.7: Es seien $\mathfrak{A} = (A_{ik}) \in M_n(I(R))$, $\mathfrak{x} = (X_i) \in V_n(I(R))$ und dazu $\mathfrak{x}$ symmetrisch, dann ist auch $\mathfrak{A}\mathfrak{x}$ symmetrisch; dabei ist $\mathfrak{A}\mathfrak{x} = |\mathfrak{A}|\mathfrak{x}$ und $|\mathfrak{A}\mathfrak{x}| = |\mathfrak{A}| \cdot |\mathfrak{x}|$.

Beweis: Zu A, $X \in I(R)$, A beliebig, X symmetrisch ist auch das Produkt AX symmetrisch; dabei ist $AX = |A|X$ und $|AX| = |A||B|$. Die Summe symmetrischer Intervalle ist wieder symmetrisch. Anwendung auf $\mathfrak{A}\mathfrak{x} = (\sum_k A_{ik}X_k)$ gibt die Behauptung. □

Satz 4.8: Sind zwei monotone Normen für reelle Vektoren aus $V_n(R)$ und reelle Matrizen aus $M_n(R)$ einander zugeordnet, so gilt dies auch für die daraus abgeleiteten Abstandsnormen für Intervallvektoren und Intervallmatrizen. Zu jeder Matrix $\mathfrak{A} \in M_n(I(R))$ kann dabei immer ein symmetrischer Vektor $\mathfrak{x} \in V_n(I(R))$ angegeben werden, der die Zuordnung verifiziert.

Beweis: Diese Abstandsnormen für $\mathfrak{x} \in V_n(I(R))$ und $\mathfrak{A} \in M_n(I(R))$ sind nach Satz 4.3 ebenfalls monoton, also ist $\||\mathfrak{A}|\| = \|\mathfrak{A}\|$ und $\||\mathfrak{x}|\| = \|\mathfrak{x}\|$.

Wegen der Zuordnung der Normen im Reellen existiert zu $|\mathfrak{A}|$ ein $\mathfrak{x}$, so daß $\||\mathfrak{A}|\mathfrak{x}\| = \||\mathfrak{A}|\| \cdot \|\mathfrak{x}\|$.

Damit wird

$$\||\mathfrak{A}||\mathfrak{x}|\| \leq \||\mathfrak{A}|\| \cdot \||\mathfrak{x}|\| = \||\mathfrak{A}|\| \cdot \|\mathfrak{x}\| = \||\mathfrak{A}|\mathfrak{x}\| \leq \||\mathfrak{A}||\mathfrak{x}|\|$$

Die letzte Ungleichung folgt aus der Monotonie der Norm und der Beziehung $\quad ||\mathfrak{A}|\mathfrak{x}| \leq |\mathfrak{A}| \cdot |\mathfrak{x}|)$

Es steht also überall das Gleichheitszeichen und wir erhalten $\||\mathfrak{A}||\mathfrak{x}|\| = \||\mathfrak{A}|\| \cdot \||\mathfrak{x}|\|$.

Es sei $\mathfrak{x}$ die konvexe Hülle von $|\mathfrak{x}|$ und $-|\mathfrak{x}|$ in $V_n(I(R))$, dann ist $\mathfrak{x}$ symmetrisch mit $|\mathfrak{x}| = |\mathfrak{x}|$; nach Hilfssatz 4.7 folgt daraus $|\mathfrak{A}\mathfrak{x}| = |\mathfrak{A}| \cdot |\mathfrak{x}|$ und wir erhalten insgesamt

$$\|\mathfrak{A}\mathfrak{x}\| = \||\mathfrak{A}\mathfrak{x}|\| = \||\mathfrak{A}||\mathfrak{x}|\| = \||\mathfrak{A}|\| \cdot \||\mathfrak{x}|\| = \|\mathfrak{A}\| \cdot \|\mathfrak{x}\|. \quad □$$

Der folgende Satz verallgemeinert einen bekannten Satz für Vektornormen in $V_n(R)$ auf Abstandsnormen in $V_n(I(R))$; für den Beweis sind nur geringe zusätzliche Überlegungen nötig.

<u>Satz 4.9:</u> Zu jeder Abstandsnorm $\|\omega\|$ in $V_n(I(R))$ gibt es genau eine zugeordnete Matrixnorm in $M_n(I(R))$; diese ist für jedes $\mathfrak{A} \in M_n(I(R))$ gegeben durch:

$$\|\mathfrak{A}\| := \sup_{\varphi \neq \mathfrak{o}} \frac{\|\mathfrak{A}\varphi\|}{\|\varphi\|} = \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\|.$$

<u>Beweis:</u> Die Eindeutigkeit der zugeordneten Matrixnorm zeigten wir schon im Korollar zu Satz 4.5. Wir müssen also jetzt nur noch beweisen, daß $\|\mathfrak{A}\| := \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\|$ tatsächlich eine zugeordnete Matrixnorm ist. Zuerst zeigen wir, daß überhaupt eine Matrixnorm voliegt und daß diese multiplikativ ist:

1') $\mathfrak{A} \neq 0 \succ \|\mathfrak{A}\| = \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\| > 0$.

2') $\|\mathfrak{A} + \mathfrak{B}\| = \sup_{\|\varphi\|=1} \|(\mathfrak{A} + \mathfrak{B})\varphi\| \leq \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi + \mathfrak{B}\varphi\| \leq \sup_{\|\varphi\|=1} (\|\mathfrak{A}\varphi\| + \|\mathfrak{B}\varphi\|) \leq$

$\leq \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\| + \sup_{\|\varphi\|=1} \|\mathfrak{B}\varphi\| = \|\mathfrak{A}\| + \|\mathfrak{B}\|$.

Für die erste Ungleichung wird das subdistributive Gesetz und Satz 4.2 benutzt; aus $(\mathfrak{A} + \mathfrak{B})\varphi \subset \mathfrak{A}\varphi + \mathfrak{B}\varphi$ folgt $\|(\mathfrak{A} + \mathfrak{B})\varphi\| \leq \|\mathfrak{A}\varphi + \mathfrak{B}\varphi\|$

3') $\|c\mathfrak{A}\| = |c|\,\|\mathfrak{A}\|$ ergibt sich sofort aus der entsprechenden Eigenschaft der Vektornorm.

4') $\|\mathfrak{A}\mathfrak{B}\| = \sup_{\varphi \neq \mathfrak{o}} \dfrac{\|\mathfrak{A}\mathfrak{B}\varphi\| \, \|\mathfrak{B}\varphi\|}{\|\varphi\| \, \|\mathfrak{B}\varphi\|} \leq \sup_{\varphi \neq \mathfrak{o}} \dfrac{\|\mathfrak{A}\mathfrak{B}\varphi\|}{\|\mathfrak{B}\varphi\|} \cdot \sup_{\varphi \neq \mathfrak{o}} \dfrac{\|\mathfrak{B}\varphi\|}{\|\varphi\|} \leq \|\mathfrak{A}\| \, \|\mathfrak{B}\|$.

Daß die Matrixnorm $\|\mathfrak{A}\| = \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\|$ der gegebenen Vektornorm zugeordnet ist, zeigt man jetzt wie üblich; zunächst gilt für alle $\mathfrak{y} \in V_n(I(R))$ und alle $\mathfrak{A} \in M_n(I(R))$: $\|\mathfrak{A}\mathfrak{y}\| \leq \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\| \, \|\mathfrak{y}\|$ d. h. die Normen sind verträglich.
In Abschnitt 4.1 haben wir gezeigt, daß jede Abstandsnorm in $V_n(I(R))$ durch den Eichkörper K der ihr zugrundeliegenden

Norm in $V_n(R)$ dargestellt werden kann; Intervallvektoren aus $V_n(I(R))$ werden dabei in $V_n(R)$ als achsenparallele, abgeschlossene Quader gedeutet. Aus dieser Darstellung ergibt sich: Die Menge $\{\varphi \in V_n(I(R)) \mid \|\varphi\| = 1\}$ ist im Eichkörper K der Norm in $V_n(R)$ enthalten, also beschränkt und abgeschlossen. Die reellwertige, stetige Funktion $\|\mathfrak{A}\varphi\|$ nimmt auf dieser Menge ihr Supremum an; jeder Vektor $\mathfrak{y}$, für den das Supremum angenommen wird, verifiziert die Zuordnung der Matrixnorm. □

<u>Bemerkung:</u> Aus der eben bewiesenen Darstellung für die zugeordnete Matrixnorm folgt, daß auch jede einer Abstandsnorm in $V_n(I(R))$ zugeordnete Norm in $M_n(I(R))$ für die Einheitsmatrix $\mathfrak{E}$ den Wert $\|\mathfrak{E}\| = 1$ annimmt.
<u>Bemerkung:</u> Ist eine Norm $\|\mathfrak{A}\|$ für Matrizen $\mathfrak{A} \in M_n(I(R))$ verträglich mit einer Vektornorm $\|\omega\|$ in $V_n(I(R))$, so ergibt für jedes $a \geq 1$ die Definition $\|\mathfrak{A}\|_a := a\|\mathfrak{A}\|$ eine mit der Vektornorm verträgliche Matrixnorm. Damit existieren zu jeder Abstandsnorm in $V_n(I(R))$ genau eine zugeordnete und unendlich viele verträgliche Matrixnormen in $M_n(I(R))$.

<u>Satz 4.10:</u> Jede einer Abstandsnorm in $V_n(I(R))$ zugeordnete Matrixnorm in $M_n(I(R))$ ist eine Abstandsnorm, d. h. für alle $\mathfrak{A} \in M_n(I(R))$ gilt:

$$\sup_{\dot{\mathfrak{A}} \in \mathfrak{A}} \sup_{\|\varphi\|=1} \|\dot{\mathfrak{A}}\varphi\| = \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\|.$$

<u>Beweis:</u> Wir betrachten ein festes $\mathfrak{A} \in M_n(I(R))$; für alle $\dot{\mathfrak{A}} \in \mathfrak{A}$ und alle $\varphi \in V_n(I(R))$ ist nach der Teilmengeneigenschaft $\dot{\mathfrak{A}}\varphi \subset \mathfrak{A}\varphi$, also nach Satz 4.2 $\|\dot{\mathfrak{A}}\varphi\| \leq \|\mathfrak{A}\varphi\|$

Damit gilt für alle $\dot{\mathfrak{A}} \in \mathfrak{A}$ $\qquad \sup_{\|\varphi\|=1} \|\dot{\mathfrak{A}}\varphi\| \leq \sup_{\|\varphi\|=1} \|\mathfrak{A}\varphi\| \qquad (1)$

Für jedes $\varphi \in V_n(I(R))$ ist $\|\mathfrak{A}\varphi\| = \sup_{\dot{\mathfrak{y}} \in \mathfrak{A}\varphi} \|\dot{\mathfrak{y}}\|$.

Also existiert zu jedem $\varepsilon > 0$ ein $\dot{\mathfrak{y}} \in \mathfrak{A}\varphi$ mit der Eigenschaft

$$\|\dot{\mathfrak{y}}\| > \|\mathfrak{A}\varphi\| - \varepsilon. \qquad (2)$$

Für alle $\dot{\mathfrak{y}} \in \mathfrak{A}\varphi$ existiert ein $\dot{\mathfrak{A}} \in \mathfrak{A}$ mit der Eigenschaft

$\mathfrak{y} \in \underset{\cdot}{\mathfrak{A}}\mathfrak{x}$, also $\|\underset{\cdot}{\mathfrak{A}}\mathfrak{x}\| \geq \|\mathfrak{y}\|$

Mit (2) folgt daraus: $\|\underset{\cdot}{\mathfrak{A}}\mathfrak{x}\| > \|\mathfrak{A}\mathfrak{x}\| - \varepsilon$;

daraus folgt $\sup\limits_{\|\mathfrak{x}\|=1} \|\underset{\cdot}{\mathfrak{A}}\mathfrak{x}\| \geq \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| - \varepsilon$.

Damit gilt: Zu jedem $\varepsilon > 0$ existiert ein $\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}$ mit der Eigenschaft $\sup\limits_{\|\mathfrak{x}\|=1} \|\underset{\cdot}{\mathfrak{A}}\mathfrak{x}\| = \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| - \varepsilon$.

(1) und (3) ergeben zusammen die Behauptung.

<u>Satz 4.11:</u> Ist die Norm $\|\mathfrak{x}\|$ für Vektoren $\mathfrak{x}$ aus $V_n(I(R))$ monoton, so ist auch die zugeordnete Norm $\|\mathfrak{A}\| = \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|$ für Matrizen $\mathfrak{A}$ aus $M_n(I(R))$ monoton.

<u>Beweis:</u> Nach Satz 4.4 ist jede monotone Norm in $V_n(I(R))$ eine Abstandsnorm; nach Abschnitt 4.2 ist der Eichkörper K einer monotonen Norm invariant gegenüber Spiegelungen an beliebigen Koordinatenhyperebenen.

Es ist $\|\mathfrak{A}\| = \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| = \sup\limits_{\|\mathfrak{x}\|\leq 1} \|\mathfrak{A}\mathfrak{x}\| = \sup\limits_{\mathfrak{x}\subset K} \|\mathfrak{A}\mathfrak{x}\|$

Es sei $\mathfrak{x} \in V_n(I(R))$ mit $\mathfrak{x} \subset K$ und $\mathfrak{y}$ die konvexe Hülle von $\mathfrak{x}$ und $(-\mathfrak{x})$ in $V_n(I(R))$;

dann ist $\mathfrak{x} \subset \mathfrak{y}$ und $\mathfrak{y} = -\mathfrak{y}$ (d. h. $\mathfrak{y}$ ist symmetrisch); aus der Spiegelungsinvarianz von K folgt nun $\mathfrak{y} \subset K$.

Aus $\mathfrak{x} \subset \mathfrak{y}$ folgt nach der Teilmengeneigenschaft $\mathfrak{A}\mathfrak{x} \subset \mathfrak{A}\mathfrak{y}$ mit Satz 4.2 folgt daraus $\|\mathfrak{A}\mathfrak{x}\| \leq \|\mathfrak{A}\mathfrak{y}\|$; man kann sich also bei der Supremumsbildung in $\sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|$ auf symmetrische Vektoren beschränken.

Für symmetrische $\mathfrak{x}$ ist aber nach Hilfssatz 4.8: $|\mathfrak{A}\mathfrak{x}| = |\mathfrak{A}||\mathfrak{x}|$;

damit wird $\sup\limits_{\mathfrak{x}\subset K} \|\mathfrak{A}\mathfrak{x}\| = \sup\limits_{\mathfrak{x}\subset K} \||\mathfrak{A}\mathfrak{x}|\| = \sup\limits_{\mathfrak{x}\subset K} \||\mathfrak{A}||\mathfrak{x}|\|$

also ist $\|\mathfrak{A}\| = \||\mathfrak{A}|\|$, d. h. die Matrixnorm ist monoton.□

<u>Bemerkung:</u> Für Vektornormen in $V_n(R)$ und Matrixnormen in $M_n(R)$ gilt keine entsprechende Aussage; ist $\|\mathfrak{x}\|$ eine monotone Norm für Vektoren $\mathfrak{x} \in V_n(R)$, so ist die zugeordnete Norm $\|\mathfrak{A}\| = \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|$ für Matrizen $\mathfrak{A} \in M_n(R)$ nicht notwendig monoton.

<u>Beispiel:</u> In $V_n(R)$ mit Elementen $\mathfrak{x} = (x_i)$ ist die euklidische Norm $\|\mathfrak{x}\|_E := \sqrt{\sum_i |x_i|^2}$ monoton, die zugeordnete Norm in $M_n(R)$ ist die sogenannte Spektralnorm; diese ist nicht monoton.
Das erkennt man folgendermaßen:
Für hermitesche Matrizen $\mathfrak{A} \in M_n(R)$ wird $\sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|_E = \varrho(\mathfrak{A})$ (siehe etwa [14], Theorem 1.3); im allgemeinen ist aber $\varrho(\mathfrak{A}) \neq \varrho(|\mathfrak{A}|)$.

Die Ursache für das andersartige Verhalten der Normen für Intervallvektoren $\mathfrak{x} \in V_n(I(R))$ und Intervallmatrizen $\mathfrak{A} \in M_n(I(R))$ ist, daß hier das Supremum $\sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|$ über die umfassendere Menge $\{\mathfrak{x} \in V_n(I(R)) \mid \|\mathfrak{x}\| = 1\}$ gebildet wird; in dieser Menge sind jetzt auch symmetrische Vektoren $\mathfrak{x} \neq 0$ mit $\mathfrak{x} = -\mathfrak{x}$ enthalten, und das Supremum wird auch gerade für symmetrische Vektoren angenommen.

Die einer Abstandsnorm $\|\mathfrak{x}\|$ für Vektoren $\mathfrak{x} \in V_n(I(R))$ zugeordnete Norm $\|\mathfrak{A}\| := \sup\limits_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\|$ für Matrizen $\mathfrak{A} \in M_n(I(R))$ läßt sich nun auch mit Hilfe des Eichkörpers K der gegebenen Vektornorm darstellen.

Zunächst gilt:

$$\|\mathfrak{A}\| = \sup_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| = \sup_{\|\mathfrak{x}\|\leq 1} \|\mathfrak{A}\mathfrak{x}\| = \sup_{\mathfrak{x}\subset K} \|\mathfrak{A}\mathfrak{x}\| = \sup_{\mathfrak{x}\subset K} \inf\{r \mid r \geq 0 \wedge \mathfrak{A}\mathfrak{x} \subset rK\};$$

Daraus folgt:

$$\|\mathfrak{A}\| = \sup_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| = \inf\{r \mid r \geq 0 \wedge \mathfrak{A}K \subset rK\}.$$

(Dies ist eine Verallgemeinerung der für zugeordnete Normen in $M_n(R)$ bekannten Darstellung; siehe dazu etwa [7].)

Man zeigt dies genauso wie für die Darstellung (AN) von Abstandsnormen in $V_n(I(R))$ (siehe Abschnitt 4.1); der dort dafür gegebene Beweis läßt sich dabei vollständig übernehmen, wenn man die darin auftretenden Größen entsprechend ersetzt.

Satz 4.12: Zu jeder multiplikativen Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$ gibt es beliebig viele verträgliche Vektornormen in $V_n(I(R))$.
Es seien $k_1, k_2, \dots . k_n \in R$, mindestens ein $k_i \neq 0$ und $\mathfrak{x} = (X_i) \in V_n(I(R))$, dann ist

$$\|\mathfrak{x}\| := \left\| \begin{pmatrix} k_1X_1 & \cdots\cdots & k_nX_1 \\ & & \\ k_1X_n & & k_nX_n \end{pmatrix} \right\|$$

eine mit der gegebenen Matrixnorm verträgliche Vektornorm.

(Dies ist eine Verallgemeinerung eines bekannten Satzes für Normen von Matrizen und Vektoren, deren Komponenten reelle oder komplexe Zahlen sind; der Beweis läßt sich hier unverändert übernehmen.)
Beweis: Die Eigenschaften 1), 2), 3) der Vektornorm ergeben sich aus den entsprechenden Eigenschaften 1'), 2'), 3') der Matrixnorm; aus der Multiplikativität der Matrixnorm folgt die Verträglichkeit. □

Zusatz: Ist die gegebene Matixnorm in $M_n(I(R))$ monoton, also $\|\mathfrak{A}\| = \||\mathfrak{A}|\|$ für alle $\mathfrak{A} \in M_n(I(R))$, so ergibt die obige Definition $\|\mathfrak{x}\| = \||\mathfrak{x}|\|$ für alle $\mathfrak{x} \in V_n(I(R))$, also eine monotone verträgliche Vektornorm in $V_n(I(R))$.

Zum Abschluß bringen wir noch einige Beispiele für zugeordnete monotone Normen:
Es seien $\mathfrak{x} = (X_i) \in V_n(I(R))$, $\mathfrak{A} = (A_{ik}) \in M_n(I(R))$.

1) $\|\mathfrak{x}\| := \sum_i |X_i|$ und $\|\mathfrak{A}\| := \max_k \sum_i |A_{ik}|$ sind zugeordnete monotone Normen; das Maximum in $\|\mathfrak{A}\|$ werde angenommen für $k = k_j$; dann verifiziert der Vektor $\mathfrak{y} = (Y_i)$ mit den Komponenten

$$Y_i = \left\{ \begin{matrix} 0 \\ [-1,1] \end{matrix} \right. \text{ für } \left. \begin{matrix} k \neq k_j \\ k = k_j \end{matrix} \right\}$$

die Zuordnung der beiden Normen.

2) $k_1, k_2, \dots\dots k_n$ seien beliebige positive Zahlen;
dann sind $\|\mathfrak{x}\| := \max_i \frac{1}{k_i} |X_i|$ und $\|\mathfrak{A}\| := \max_i \frac{1}{k_i} \sum_j |A_{ij}| k_j$
zugeordnete monotone Normen; ihre Zuordnung wird verifiziert durch den Vektor

$$\mathfrak{y} = \begin{pmatrix} [-k_1, k_1] \\ \vdots \\ [-k_n, k_n] \end{pmatrix}.$$

4.4 Über den Zusammenhang der multiplikativen monotonen Normen für Matrizen $\mathfrak{A} \in M_n(I(R))$ bzw. $M_n(R)$ mit dem Spektralradius $\varrho(|\mathfrak{A}|)$ des Absolutbetrags $|\mathfrak{A}|$.

Wir beweisen einen wichtigen Satz über den Zusammenhang der multiplikativen monotonen Normen für Matrizen $\mathfrak{A} \in M_n(I(R))$ bzw. $M_n(R)$ mit dem Spektralradius $\varrho(|\mathfrak{A}|)$ des Absolutbetrags $|\mathfrak{A}|$. Zur Vorbereitung erinnern wir an eine entsprechende Beziehung für Matrizen $\mathfrak{A} \in M_n(R)$ und multiplikative Normen in $M_n(R)$.

Satz 4.13: Es sei $\mathfrak{A} \in M_n(R)$, dann gilt (siehe etwa [7]):

(Ra) Für jede multiplikative Norm $\|\mathfrak{M}\|$ in $M_n(R)$ ist
$$\|\mathfrak{A}\| \geq \varrho(\mathfrak{A}).$$

(Rb) Zu jedem $\varepsilon > 0$ existiert eine multiplikative Matrixnorm $\|\mathfrak{M}\|$ in $M_n(R)$, für die $\|\mathfrak{A}\| < \varrho(\mathfrak{A}) + \varepsilon$.

(Ra) und (Rb) zusammen sind gleichbedeutend mit:
(R) $\varrho(\mathfrak{A}) = \inf \left\{ \|\mathfrak{A}\| \,\middle|\, \|\ \| \text{ eine multiplikative Norm in } M_n(R) \right\}$

Die entsprechenden Aussagen für Matrizen $\mathfrak{A} \in M_n(I(R))$ bzw. $M_n(R)$ und ihre multiplikativen monotonen Normen enthält der folgende

Satz 4.14: Es sei $\mathfrak{A} = (A_{ij}) \in M_n(I(R))$ bzw. $M_n(R)$; dann gilt:

(Ia) Für jede multiplikative monotone Norm $\|\mathfrak{M}\|_m$ in $M_n(I(R))$ bzw. $M_n(R)$ ist $\|\mathfrak{A}\|_m \geq \varrho(|\mathfrak{A}|)$.

(Ib) Zu jedem $\varepsilon > 0$ existiert eine multiplikative monotone Matrixnorm $\|\mathfrak{M}\|_m$ in $M_n(I(R))$ bzw. $M_n(R)$, für die $\|\mathfrak{A}\|_m < \varrho(|\mathfrak{A}|) + \varepsilon$.

(Ia) und (Ib) zusammen sind wieder gleichbedeutend mit:

(I) $\varrho(|\mathfrak{A}|) = \inf\left\{ \|\mathfrak{A}\| \;\middle|\; \|\ \| \text{ eine } \begin{array}{r}\text{monotone multiplikative Norm}\\ \text{in } M_n(\mathbb{I}(\mathbb{R})) \text{ bzw. } M_n(\mathbb{R})\end{array} \right\}$

Beweis: Monotone Normen sind gekennzeichnet durch die Eigenschaft (A) $\|\mathfrak{A}\|_m = \||\mathfrak{A}|\|_m$. Wegen $|\mathfrak{A}| \in M_n(R)$ ergibt sich (Ia) sofort aus (Ra); ein entsprechender Schluß von (Rb) auf (Ib) oder von (R) auf (I) ist dagegen nicht möglich. Für eine beliebige Norm $\|\mathfrak{M}\|_b$ in $M_n(R)$ gibt nämlich das Funktional

$\||\mathfrak{A}|\|_b$ nicht immer eine Norm für Matrizen $\mathfrak{A} \in M_n(R)$
und entsprechend
$\||\mathfrak{A}|\|_b$ nicht immer eine Norm für Matrizen $\mathfrak{A} \in M_n(\mathbb{I}(\mathbb{R}))$.

Zum Beweis von (Ib) können wir uns nun sogar auf mit Gewichten $k_1, k_2, \ldots, k_n > 0$ gebildete monotone Normen der Gestalt

$$\|\mathfrak{A}\|_k := \max_i \frac{1}{k_i} \sum_j |A_{ij}|\, k_j$$

beschränken; diese Normen haben wir bereits bei den oben angegebenen Beispielen für monotone zugeordnete Normen erwähnt; sie sind also auch multiplikativ. Wir zeigen jetzt:

$$\text{(K)} \quad \varrho(|\mathfrak{A}|) = \text{Infimum aller Normen } \|\mathfrak{A}\|_k = $$
$$= \inf_{k_1,\ldots,k_n>0} \left\{ \max_i \frac{1}{k_i} \sum_j |A_{ij}|\, k_j \right\}.$$

Für irreduzible Matrizen $\mathfrak{A}$ ergibt sich (K) sofort aus der Perron-Frobenius-Theorie für nichtnegative Matrizen (siehe etwa [14], insbesondere Theorem 2.1 und 2.2):
In diesem Falle ist nämlich $\varrho(|\mathfrak{A}|)$ Eigenwert von $|\mathfrak{A}|$ mit einem reellen Eigenvektor $\varphi = (x_i) > 0$, und es ist

$$\varrho(|\mathfrak{A}|) = \frac{1}{x_i} \sum_j |A_{ij}|\, k_j \qquad \text{für alle } i;$$

das Infimum in (K) wird also angenommen.

Ist nun $\mathfrak{A}$ reduzibel, dann bilden wir zu $|\mathfrak{A}| =: (a_{ij})$ eine

irreduzible Matrix $\mathfrak{A}' =: (a'_{ij})$ nach folgender Vorschrift:

Für $a_{ij} > 0$ sei $a'_{ij} := a_{ij}$,
für $a_{ij} = 0$ sei $a'_{ij} := \alpha > 0$;

dann ist $\mathfrak{A}'$ irreduzibel, $|\mathfrak{A}'| \geq |\mathfrak{A}| \geq \sigma$ und $\varrho(\mathfrak{A}') \geq \varrho(|\mathfrak{A}|)$ (siehe etwa [14], Theeorem 2.7).

Jetzt ist aber $\varrho(\mathfrak{A}')$ Eigenwert von $\mathfrak{A}'$ mit einem positiven Eigenvektor $\varphi = (x_i) > 0$; damit gilt:

$$\varrho(\mathfrak{A}') = \frac{1}{k_i} \sum_j a'_{ij}\, x_j \qquad \text{für alle } i.$$

Nach Definition der Matrix $\mathfrak{A}'$ ist $a'_{ij} \geq a_{ij}$ für alle i, j.
damit wird für alle i:

$$\varrho(\mathfrak{A}') = \frac{1}{x_i} \sum_j a'_{ij} x_j \geq \max_i \frac{1}{x_i} \sum_j a_{ij} x_j \geq \varrho(|\mathfrak{A}|). \qquad (1)$$

(Die letzte Ungleichung in (1) folgt daraus, daß keine Norm von $|\mathfrak{A}|$ kleiner als $\varrho(|\mathfrak{A}|)$ist.)

Wegen der stetigen Abhängigkeit des Spektralradius einer Matrix $\mathfrak{A} \in M_n(R)$ von ihren Komponenten gibt es zu jedem $\varepsilon > 0$ ein $\delta(\varepsilon)$ derart, daß

$$0 \leq \varrho(\mathfrak{A}') - \varrho(|\mathfrak{A}|) < \varepsilon \quad \text{für alle } \alpha < \delta(\varepsilon). \qquad (2)$$

Aus (1) und (2) folgt jetzt:

$$0 \leq \max_i \frac{1}{x_i} \sum_j a_{ij} x_j - \varrho(|\mathfrak{A}|) < \varepsilon.$$

Zu jedem $\varepsilon > 0$ existieren also positive Zahlen $k_1 := x_1, \ldots\ldots .k_n := x_n$, so daß für die damit gebildete monotone Norm

$$\|\mathfrak{A}\|_k = \max_i \frac{1}{k_i} \sum_j |A_{ij}|\, k_j$$

gilt

$$\|\mathfrak{A}\|_k \leq \varrho(|\mathfrak{A}|) + \varepsilon. \qquad (3)$$

(3) enthält die zu beweisende Aussage (Ib); zugleich folgt jetzt aus (Ia) und (3) auch Aussage (K). □

Eine wichtige Folgerung aus Satz 4.14 enthält das folgende

Korollar: Es sei $\mathfrak{A} \in M_n(I(R))$ bzw. $M_n(R)$; dann ist $\rho(|\mathfrak{A}|) < a$ dann und nur dann, wenn in $M_n(I(R))$ bzw. $M_n(R)$ eine multiplikative monotone Matrixnorm $\|\mathfrak{M}\|_m$ existiert, für die $\|\mathfrak{A}\|_m < a$.

Beweis: Der Schluß von $\|\mathfrak{A}\|_m < a$ auf $\rho(|\mathfrak{A}|) < a$ folgt unmittelbar aus Aussage (Ia) von Satz 4.14.
Der Schluß von $\rho(|\mathfrak{A}|) < a$ auf die Existenz einer Matrixnorm $\|\mathfrak{A}\|_m < a$ ergibt sich aus Aussage (Ib) mit $0 < \varepsilon < a - \rho(|\mathfrak{A}|)$. $\square$

Bemerkung: Für Matrixnormen in $M_n(I(R))$ kann die Aussage von Satz 4.14 und des anschließenden Korollars noch wesentlich verschärft werden; in Abschnitt 6 wird gezeigt, daß der Satz für Normen in $M_n(I(R))$ auch dann noch richtig bleibt, wenn man die Beschränkung auf monotone Normen fallen läßt und nur noch ihre Multiplikativität voraussetzt (siehe Satz 6.3).

5. Metriken in den Räumen $V_n(I(R))$ und $M_n(I(R))$

Im vorhergehenden Abschnitt 4 untersuchten wir die Eigenschaften von Normen in $V_n(I(R))$ und $M_n(I(R))$; wegen des Fehlens inverser Elemente gibt aber in diesen quasilinearen Räumen die Norm der Differenz zweier Elemente keine Metrik. Für Abschätzungen, insbesondere bei Konvergenzuntersuchungen benötigt man deshalb nicht nur Normen sondern auch eigene Metriken. Für Abschätzungen eignen sich aber nur Metriken, die einmal mit der algebraischen Struktur verträglich, d. h. homogen und translationsinvariant sind und darüberhinaus mit bekannten Matrixnormen verträglich sind ($M_n(I(R))$ ist ja ein Raum von Operatoren über $V_n(I(R))$ bzw. $M_n(I(R))$). Wir verwenden dabei folgende

Definition 5.1: Eine Metrik $q(\mathfrak{x},\mathfrak{y})$ in $V_n(I(R))$ bzw. $q(\mathfrak{A},\mathfrak{B})$ in $M_n(I(R))$ und eine Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$ heißen miteinander verträglich, wenn die Matrixnorm multiplikativ ist, und für alle $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$ und alle $\mathfrak{A}, \mathfrak{B}, \mathfrak{C} \in M_n(I(R))$ gilt:

$$q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \|\mathfrak{A}\|\, q(\mathfrak{x},\mathfrak{y}) \quad \text{bzw.} \quad q(\mathfrak{A}\mathfrak{B},\mathfrak{A}\mathfrak{C}) \leq \|\mathfrak{A}\|\, q(\mathfrak{B},\mathfrak{C})$$

Beispiel: Die in Abschnitt 1 eingeführte Metrik $q(A,B)$ in $I(R) = M_1(I(R))$ ist nach Satz 1.4 mit dem in $I(R)$ erklärten Absolutbetrag $|A|$ (dieser ist eine Norm) verträglich.

Auch bei Einschränkung auf die linearen Teilräume $V_n(R)$ und $M_n(R)$ sind die Verhältnisse einfach. Jede als Norm der Differenz zweier Elemente darstellbare Metrik ist dort homogen und translationsinvariant. $M_n(R)$ ist der Raum der linearen Operatoren über $V_n(R)$; bei Verwendung zugeordneter bzw. verträglicher Normen $\|\mathfrak{x}\|$ und $\|\mathfrak{M}\|$ ist die Metrik $q(\mathfrak{x},\mathfrak{y}) := \|\mathfrak{x} - \mathfrak{y}\|$ bzw. $q(\mathfrak{A},\mathfrak{B}) := \|\mathfrak{A} - \mathfrak{B}\|$ von selbst mit der Matrixnorm $\|\mathfrak{M}\|$ verträglich.

5.1 Monotone Normen und zugeordnete Metriken

In $V_n(I(R))$ und $M_n(I(R))$ führen wir nun zu beliebig vorgegebenen monotonen Normen sogenannte zugeordnete Metriken ein und zeigen

anschließend, daß diese Verträglichkeitseigenschaften besitzen, wie wir sie eben für $V_n(R)$ und $M_n(R)$ beschrieben.
Wir verwenden dabei die in Abschnitt 1. eingeführte Metrik $q(A,B)$ in $I(R)$; für $A = [a_1,a_2]$, $B = [b_1,b_2] \in I(R)$ ist $q(A,B) := \max(|a_1-b_1|,|a_2-b_2|)$. Nach Satz 1.2 ist diese Metrik homogen und translationsinvariant.

a) In $V_n(I(R))$ sei eine monotone Vektornorm $\|\mathfrak{w}\|_m$ vorgegeben; zu $\mathfrak{x} = (X_i)$, $\mathfrak{y} = (Y_i) \in V_n(I(R))$ betrachten wir das reelle Funktional

$$q_m(\mathfrak{x},\mathfrak{y}) := \left\| \begin{pmatrix} q(X_1,Y_1) \\ q(X_n,Y_n) \end{pmatrix} \right\|_m$$

b) In $M_n(I(R))$ sei entsprechend eine monotone Matrixnorm $\|\mathfrak{M}\|_m$ vorgegeben; zu $\mathfrak{A} = (A_{ik})$, $\mathfrak{B} = (B_{ik})$ bilden wir

$$q_m(\mathfrak{A},\mathfrak{B}) := \left\| \left(q(A_{ik},B_{ik}) \right) \right\|_m$$

<u>Bemerkung:</u> Für die Definition von $q_m(\mathfrak{x},\mathfrak{y})$ bzw. $q_m(\mathfrak{A},\mathfrak{B})$ wird tatsächlich nur die Einschränkung der in $V_n(I(R))$ bzw. $M_n(I(R))$ gegebenen monotonen Norm auf $V_n(R)$ bzw. $M_n(R)$ benötigt.

<u>Satz 5.1:</u> $q_m(\mathfrak{x},\mathfrak{y})$ bzw. $q_m(\mathfrak{A},\mathfrak{B})$ ist eine homogene translationsinvariante Metrik in $V_n(I(R))$ bzw. $M_n(I(R))$, d. h. es gilt für $q_m(\mathfrak{x},\mathfrak{y})$ bzw. für $q_m(\mathfrak{A},\mathfrak{B})$ entsprechend:

1) $q_m(\mathfrak{x},\mathfrak{y}) = 0 \Leftrightarrow \mathfrak{x} = \mathfrak{y}$
2) $q_m(\mathfrak{x},\mathfrak{y}) \le q_m(\mathfrak{x},\mathfrak{z}) + q_m(\mathfrak{y},\mathfrak{z})$
3) $q_m(c\mathfrak{x},c\mathfrak{y}) = |c|\, q_m(\mathfrak{x},\mathfrak{z})$
4) $q_m(\mathfrak{x}+\mathfrak{z},\mathfrak{y}+\mathfrak{z}) = q_m(\mathfrak{x},\mathfrak{z})$

<u>Beweis:</u> Wir führen den Beweis stellvertretend für $q_m(\mathfrak{x},\mathfrak{y})$; für $q_m(\mathfrak{A},\mathfrak{B})$ läuft er ganz analog.
Es seien $\mathfrak{x} = (X_i)$, $\mathfrak{y} = (Y_i)$, $\mathfrak{z} = (Z_i) \in V_n(I(R))$.
Die Komponenten $q(X_i,Y_i)$ des Vektors $\left(q(X_i,Y_i)\right)$ sind gerade die Abstände der entsprechenden Komponenten X_i und Y_i von $\mathfrak{x}$ und $\mathfrak{y}$.

Der Abstand $q(X_i,Y_i)$ hat aber schon die Eigenschaft 1) bis 4) einer homogenen translationsinvarianten Metrik. Bei der Bildung der monotonen Norm übertragen sich dann diese Eigenschaften auf $q_m(\mathfrak{x},\mathfrak{y})$:

1) $q_m(\mathfrak{x},\mathfrak{y}) = \|(q(X_i,Y_i))\|_m = 0 \Leftrightarrow \forall i:\ q(X_i,Y_i) = 0 \Leftrightarrow \forall i:\ X_i = Y_i \Leftrightarrow \mathfrak{x} = \mathfrak{y}$.

2) $0 \le q(X_i,Y_i) \le q(X_i,Z_i) + q(Y_i,Z_i)$ $\Rightarrow$

$$q_m(\mathfrak{x},\mathfrak{y}) = \|(q(X_i,Y_i))\|_m \le \|(q(X_i,Z_i) + q(Y_i,Z_i))\|_m \le$$
$$\le \|(q(X_i,Z_i))\|_m + \|(q(Y_i,Z_i))\|_m = q_m(\mathfrak{x},\mathfrak{z}) + q_m(\mathfrak{y},\mathfrak{z}).$$

(Bei der Bildung der ersten Ungleichung wird hier die Monotonie der Norm benutzt; für beliebige Normen ist dieser Schluß nicht zulässig.)

3) $q_m(c\mathfrak{x},c\mathfrak{y}) = \|(q(cX_i,cY_i))\|_m = \||c|(q(X_i,Y_i))\|_m = |c|\,\|(q(X_i,Y_i))\|_m =$
$= |c|\, q_m(\mathfrak{x},\mathfrak{y})$.

4) $q_m(\mathfrak{x}+\mathfrak{y},\mathfrak{z}+\mathfrak{y}) = \|(q(X_i+Y_i,Z_i+Y_i))\|_m = \|(q(X_i,Z_i))\|_m = q_m(\mathfrak{x},\mathfrak{z})$. □

<u>Definition 5.1:</u> $q_m(\mathfrak{x},\mathfrak{y})$ bzw. $q_m(\mathfrak{A},\mathfrak{B})$ heißt die der gegebenen monotonen Vektornorm in $V_n(I(R))$ bzw. der gegebenen monotonen Matrixnorm in $M_n(I(R))$ zugeordnete Metrik.

$V_n(I(R))$ und $M_n(I(R))$ sind jeweils vollständig bezüglich der Metrik $q_m(\mathfrak{x},\mathfrak{y})$ bzw. $q_m(\mathfrak{A},\mathfrak{B})$; denn die Konvergenz ist jeweils komponentenweise Konvergenz bezüglich der Metrik $q(A,B)$ des Raumes $I(R)$ und $I(R)$ ist vollständig.

Für Vektoren $\mathfrak{x} = \mathfrak{x}$, $\mathfrak{y} = \mathfrak{y} \in V_n(R)$ wird $q_m(\mathfrak{x},\mathfrak{y}) = \|\mathfrak{x} - \mathfrak{y}\|_m$; die Metrik $q_m(\mathfrak{x},\mathfrak{y})$ in $V_n(I(R))$ ist also eine Fortsetzung der im linearen Teilraum $V_n(R)$ durch die Vektornorm $\|\mathfrak{w}\|_m$ festgelegten Metrik $\|\mathfrak{x}-\mathfrak{y}\|_m$ auf den umfassenderen Raum $V_n(I(R))$.
Entsprechend ist die Metrik $q_m(\mathfrak{A},\mathfrak{B})$ in $M_n(I(R))$ eine Fortsetzung der im linearen Teilraum $M_n(R)$ durch die Matrixnorm $\|\mathfrak{M}\|_m$ festgelegten Metrik $\|\mathfrak{A}-\mathfrak{B}\|_m$ auf den Raum $M_n(I(R))$.

Satz 5.2: Es sei $A \in I(R)$, $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$, $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$; für die einer monotonen Norm in $V_n(I(R))$ bzw. $M_n(I(R))$ zugeordnete Metrik $q_m(\mathfrak{x},\mathfrak{y})$ bzw. $q_m(\mathfrak{A},\mathfrak{B})$ gilt:

$$q_m(A\mathfrak{x},A\mathfrak{y}) \leq |A|\, q_m(\mathfrak{x},\mathfrak{y}) \quad \text{bzw.} \quad q_m(A\mathfrak{A},A\mathfrak{B}) \leq |A|\, q_m(\mathfrak{A},\mathfrak{B}).$$

Beweis: Es sei $\mathfrak{x} = (X_i)$, $\mathfrak{y} = (Y_i)$; mit Satz 1.4 und der Monotonie der Norm folgt:

$$q_m(A\mathfrak{x},A\mathfrak{y}) = \|(q(AX_i,AY_i))\|_m \leq \|(|A|\, q(X_i,Y_i))\|_m = |A|\,\|(q(X_i,Y_i))\|_m = \\ = |A|\, q_m(\mathfrak{x},\mathfrak{y}).$$

Für $q_m(A\mathfrak{A},A\mathfrak{B})$ schließt man entsprechend. □

Satz 5.3: Es seien $\mathfrak{A} = (A_{ik})$, $\mathfrak{B} = (B_{ik})$, $\mathfrak{C} = (C_{ik}) \in M_n(I(R))$, $\mathfrak{x} = (X_i)$, $\mathfrak{y} = (Y_i) \in V_n(I(R))$; weiter seien $\|\mathfrak{v}\|_m$ und $\|\mathfrak{M}\|_m$ verträgliche monotone Normen in $V_n(I(R))$ bzw. $M_n(I(R))$, sowie $q_m(\mathfrak{x},\mathfrak{y})$ und $q_m(\mathfrak{A},\mathfrak{B})$ die zugeordneten Metriken.

Dann gilt:

a) $q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \|\mathfrak{A}\|_m q_m(\mathfrak{x},\mathfrak{y})$,

b) $q_m(\mathfrak{A}\mathfrak{B},\mathfrak{A}\mathfrak{C}) \leq \|\mathfrak{A}\|_m q_m(\mathfrak{B},\mathfrak{C})$.

Beweis: a) Zunächst gilt für alle i:

$$q\Big(\sum_k A_{ik}X_k, \sum_k A_{ik}Y_k\Big) \leq \sum_k q(A_{ik}X_k, A_{ik}Y_k) \leq \sum_k |A_{ik}|\, q(X_k,Y_k).$$

(Hier wird zunächst die supermetrische Eigenschaft $q(A+B,C+D) \leq q(A,C) + q(C,D)$ der Metrik $q(A,B)$ in $I(R)$ und dann Satz 1.4 benutzt.)

Zusammenfassen in Matrixschreibweise ergibt (für die i-te Komponente eines Vektors $\mathfrak{z}$ schreiben wir hier auch $(\mathfrak{z})_i$):

$$\big(q((\mathfrak{A}\mathfrak{x})_i,(\mathfrak{A}\mathfrak{y})_i)\big) \leq |\mathfrak{A}|\,\big(q(X_i,Y_i)\big)$$

In den zugrundegelegten verträglichen monotonen Vektor- und Matrixnormen erhalten wir nun

$$q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) = \|(q((\mathfrak{A}\mathfrak{x})_i,(\mathfrak{A}\mathfrak{y})_i))\|_m \leq \|\mathfrak{A}\|_m \|(q(X_i,Y_i))\|_m = \|\mathfrak{A}\|_m q_m(\mathfrak{x},\mathfrak{y}).$$

Aussage b) beweist man ganz entsprechend; nur verwendet man hier anstatt der Verträglichkeit der Normen die Multiplikativität der Matrixnorm. □

Satz 5.3 drückt aus, daß die Metriken $q_m(\mathfrak{x},\mathfrak{y})$ und $q_m(\mathfrak{A},\mathfrak{B})$ mit der Matrixnorm $\|\mathfrak{M}\|_m$ verträglich sind.

Korollar: Für beliebige $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ ist die Abbildung: $\mathfrak{x} \longrightarrow \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ von $V_n(I(R))$ in sich gleichmäßig stetig bezüglich der Metrik $q_m(\mathfrak{x},\mathfrak{y})$ in $V_n(I(R))$ und es gilt:

$$q_m(\mathfrak{A}\mathfrak{x}+\mathfrak{b},\mathfrak{A}\mathfrak{y}+\mathfrak{b}) \leq \|\mathfrak{A}\|_m q_m(\mathfrak{x},\mathfrak{y}) < \varepsilon \quad \text{für } q_m(\mathfrak{x},\mathfrak{y}) \leq \frac{\varepsilon}{\|\mathfrak{A}\|_m}$$

Beweis: Wegen der Translationsinvarianz (Satz 5.1) ist

$q_m(\mathfrak{A}\mathfrak{x}+\mathfrak{b},\mathfrak{A}\mathfrak{y}+\mathfrak{b}) = q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y})$;

die Behauptung folgt jetzt aus Satz 5.3.

Entsprechend zeigt man: Für $\mathfrak{A}$, $\mathfrak{B} \in M_n(I(R))$ ist die Abbildung: $\mathfrak{X} \longrightarrow \mathfrak{A}\mathfrak{X} + \mathfrak{B}$ von $M_n(I(R))$ in sich gleichmäßig stetig bezüglich der Metrik $q_m(\mathfrak{X},\mathfrak{Y})$ in $M_n(I(R))$. □

Bemerkung: Zu jeder multiplikativen monotonen Matrixnorm $\|\mathfrak{M}\|_m$ existiert nach Satz 4.12 eine verträgliche monotone Vektornorm $\|\mathfrak{v}\|_m$; die den Normen $\|\mathfrak{v}\|_m$ und $\|\mathfrak{M}\|_m$ jeweils zugeordneten Metriken $q_m(\mathfrak{x},\mathfrak{y})$ und $q_m(\mathfrak{A},\mathfrak{B})$ sind nun mit der Matrixnorm $\|\mathfrak{M}\|_m$ verträglich.

Für $\|\mathfrak{A}\|_m < 1$ ist die Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ von $V_n(I(R))$ in sich bezüglich der Metrik $q_m(\mathfrak{x},\mathfrak{y})$ kontrahierend. Dies wird in Abschnitt 6 eine wichtige Rolle spielen.

5.2 Über die Beziehungen zwischen Normen und Metriken in $V_n(I(R))$ bzw. $M_n(I(R))$ und die Existenz verträglicher Normen und Metriken

In einem kurzen Rückblick fassen wir zunächst die Beziehungen zwischen Normen und Metriken in $V_n(R)$ und $V_n(I(R))$ zusammen (für $M_n(R)$ und $M_n(I(R))$ liegen die Verhältnisse entsprechend): Der lineare Raum $V_n(R)$ ist eingebettet in den umfassenderen quasilinearen Raum $V_n(I(R))$. In $V_n(R)$ gehört zu jeder Norm $\|\mathfrak{y}\|$ eine Metrik der Gestalt $\|\mathfrak{x}-\mathfrak{y}\|$; bei Fortsetzung der Norm $\|\mathfrak{y}\|$ und der Metrik $\|\mathfrak{x}-\mathfrak{y}\|$ auf den quasilinearen Raum $V_n(I(R))$ kann dieser im linearen Raum bestehende Zusammenhang nicht in dieser Form bestehen bleiben; denn in $V_n(I(R))$ kann wegen des Fehlens inverser Elemente keine Metrik als Norm der Differenz zweier Elemente dargestellt werden. In Abschnitt 4.2 haben wir gezeigt, wie beliebige Normen von $V_n(R)$ als Abstandsnormen auf $V_n(I(R))$ fortgesetzt werden können. In Abschnitt 3.4 haben wir eine allgemeine Methode angegeben, mit der jede durch eine Norm darstellbare Metrik $\|\mathfrak{x}-\mathfrak{y}\|$ in $V_n(R)$ zu einer homogenen und translationsinvarianten Metrik in $V_n(I(R))$ fortgesetzt werden kann. Das eben Gesagte gilt entsprechend auch für die Beziehungen zwischen $M_n(R)$ und $M_n(I(R))$.

Die zu verträglichen monotonen Normen $\|\mathfrak{y}\|_m$ in $V_n(R)$ und $\|\mathfrak{M}\|_m$ in $M_n(R)$ gehörigen Metriken $\|\mathfrak{x}-\mathfrak{y}\|_m$ bzw. $\|\mathfrak{A}-\mathfrak{B}\|_m$ haben wir nun oben so auf den $V_n(I(R))$ bzw. $M_n(I(R))$ fortgesetzt, daß die sich dort ergebenden (den monotonen Normen zugeordneten) Metriken $q_m(\mathfrak{x},\mathfrak{y})$ und $q_m(\mathfrak{A},\mathfrak{B})$ auch für Abschätzungen und Konvergenzuntersuchungen geeignet sind; die dafür nötigen Verträglichkeitseigenschaften, nämlich die Homogenität und Translationsinvarianz und die Verträglichkeit mit der durch $\|\mathfrak{M}\|_m$ erklärten Abstandsnorm $\|\mathfrak{M}\|_m$ in $M_n(I(R))$ haben wir in Satz 5.1 und 5.3 nachgewiesen.

Für etwa nach der Methode von Abschnitt 3.4 auf $V_n(I(R))$ und $M_n(I(R))$ fortgesetzte Metriken sind Verträglichkeitseigenschaften wie in Satz 5.3 auch bei Zugrundelegung monotoner Normen nicht

immer erfüllt. Wir zeigen dies an einem Gegenbeispiel: Für jede Vektornorm $\|\mathfrak{y}\|$ in $V_n(I(R))$ (oder $V_n(R)$) ist

$$q(\mathfrak{x},\mathfrak{y}) := \max\left(\left\|\inf_{\xi\in\mathfrak{x}}\xi - \inf_{\eta\in\mathfrak{y}}\eta\right\|, \left\|\sup_{\xi\in\mathfrak{x}}\xi - \sup_{\eta\in\mathfrak{y}}\eta\right\|\right)$$

eine homogene und translationsinvariante Metrik in $V_n(I(R))$. Wir legen jetzt in $V_2(I(R))$ und $M_2(I(R))$ mit Elementen $\mathfrak{x} = (X_i)$ bzw. $\mathfrak{A} = (A_{ik})$ die zugeordneten monotonen Normen $\|\mathfrak{x}\| := \sum_i |X_i|$ und $\|\mathfrak{A}\| := \max_l \sum_i |A_{il}|$ zugrunde. Dann erhält man für

$$\mathfrak{x} := \begin{pmatrix} [1,3] \\ [2,4] \end{pmatrix}, \quad \mathfrak{y} := \begin{pmatrix} [3,4] \\ [1,2] \end{pmatrix} \in V_2(I(R)), \quad \mathfrak{A} := \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \in M_2(I(R)):$$

$$q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) = \max\left(\left\|\inf_{\mathfrak{z}\in\mathfrak{A}\mathfrak{x}}\mathfrak{z} - \inf_{\mathfrak{z}\in\mathfrak{A}\mathfrak{y}}\mathfrak{z}\right\|, \left\|\sup_{\mathfrak{z}\in\mathfrak{A}\mathfrak{x}}\mathfrak{z} - \sup_{\mathfrak{z}\in\mathfrak{A}\mathfrak{y}}\mathfrak{z}\right\|\right) = 8,$$

$$q(\mathfrak{x},\mathfrak{y}) = \max\left(\left\|\inf_{\xi\in\mathfrak{x}}\xi - \inf_{\eta\in\mathfrak{y}}\eta, \sup_{\xi\in\mathfrak{x}}\xi - \sup_{\eta\in\mathfrak{y}}\eta\right)\right) = 3,$$

$$\|\mathfrak{A}\| = 2; \qquad \text{also} \quad q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) > \|\mathfrak{A}\|\, q(\mathfrak{x},\mathfrak{y}). \ \square$$

Die Frage, ob zu jeder homogenen und translationsinvarianten Metrik in $V_n(I(R))$ oder $M_n(I(R))$ eine verträgliche Matrixnorm in $M_n(I(R))$ existiert, scheint nicht auf Anhieb beantwortbar. Die folgenden Bemerkungen sollen die hier auftretenden Schwierigkeiten zeigen. Wir betrachten stellvertretend den Fall einer Metrik in $V_n(I(R))$, für eine Metrik in $M_n(I(R))$ gelten die entsprechenden Überlegungen.

Ist $q(\mathfrak{x},\mathfrak{y})$ eine homogene und translationsinvariante Metrik in $V_n(I(R))$, so gilt für jede damit verträgliche Matrixnorm $\|\mathfrak{A}\|$ in $M_n(I(R))$:

$$\sup_{q(\mathfrak{x},\mathfrak{y})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \|\mathfrak{A}\| \qquad \text{für alle } \mathfrak{A}\in M_n(I(R))$$

Existiert nun $\sup\limits_{q(\mathfrak{x},\mathfrak{y})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y})$ nicht für alle $\mathfrak{A}\in M_n(I(R))$, so gibt es keine mit $q(\mathfrak{x},\mathfrak{y})$ verträgliche Matrixnorm.

Existiert aber das Supremum, so gilt:

$$q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \sup_{q(\mathfrak{x},\mathfrak{y})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y})\; q(\mathfrak{x},\mathfrak{y})$$

$$\text{für alle } \mathfrak{x},\mathfrak{y}\in V_n(I(R)),\ \mathfrak{A}\in M_n(I(R));$$

das ist gerade die Verträglichkeitseigenschaft von Definition 5.1 und es erhebt sich die Frage, ob etwa dieses Supremum selbst schon eine Matrixnorm ist. Die Eigenschaften 1') und 3'), sowie die Multiplikativität 4') läßt sich leicht verifizieren, aber die Eigenschaft 2') kann man wegen der Subdistributivität nicht allgemein nachweisen.

Wir zeigen aber jetzt noch, daß für die einer monotonen Vektornorm $\|\mathfrak{w}\|_m$ zugeordnete Metrik $q_m(\mathfrak{x},\mathfrak{y})$ der Ausdruck $\sup_{q(\mathfrak{x},\mathfrak{y})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y})$ tatsächlich eine verträgliche Matrixnorm gibt:

Nach Definition ist hier $\quad q_m(\mathfrak{x},\mathfrak{o}) = \||\mathfrak{x}|\|_m = \|\mathfrak{x}\|_m$;

$\|\mathfrak{A}\|_m := \sup_{\|\mathfrak{x}\|=1} \|\mathfrak{A}\mathfrak{x}\| = \sup_{q_m(\mathfrak{x},\mathfrak{o})=1} q_m(\mathfrak{A}\mathfrak{x},\mathfrak{o})$ ist die der Vektornorm $\|\mathfrak{x}\|_m$ zugeordnete Matrixnorm (Satz 4.9). Nach Satz 5.3 gilt für diese:

$q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \|\mathfrak{A}\|_m\, q_m(\mathfrak{x},\mathfrak{y}) \qquad$ für alle $\mathfrak{x},\mathfrak{y} \in V_n(\mathbf{I}(R))$, $\mathfrak{A} \in M_n(\mathbf{I}(R))$;

daraus folgt

$$\sup_{q_m(\mathfrak{x},\mathfrak{y})=1} q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \leq \|\mathfrak{A}\|_m = \sup_{q_m(\mathfrak{x},\mathfrak{o})=1} q_m(\mathfrak{A}\mathfrak{x},\mathfrak{o}). \tag{1}$$

Die umgekehrte Ungleichung

$$\sup_{q_m(\mathfrak{x},\mathfrak{y})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}) \geq \sup_{q_m(\mathfrak{x},\mathfrak{o})=1} q(\mathfrak{A}\mathfrak{x},\mathfrak{o}) \tag{2}$$

gilt für jede Metrik $q(\mathfrak{x},\mathfrak{y})$ in $V_n(\mathbf{I}(R))$, weil links das Supremum über die umfassendere Menge gebildet wird.

(1) und (2) ergeben zusammen:

$$\|\mathfrak{A}\|_m = \sup_{q_m(\mathfrak{x},\mathfrak{o})=1} q_m(\mathfrak{A}\mathfrak{x},\mathfrak{o}) = \sup_{q_m(\mathfrak{x},\mathfrak{y})=1} q_m(\mathfrak{A}\mathfrak{x},\mathfrak{A}\mathfrak{y}). \; \square$$

5.3 Durchmesser für Elemente aus $\mathbf{I}(R)$, $V_n(\mathbf{I}(R))$ bzw. $M_n(\mathbf{I}(R))$

Alle Elemente aus den Räumen $\mathbf{I}(R)$, $V_n(\mathbf{I}(R))$ bzw. $M_n(\mathbf{I}(R))$ sind auf Grund ihrer Definition jeweils spezielle Teilmengen der linearen Räume R, $V_n(R)$ bzw. $M_n(R)$; für diese Teilmengen werden zu jeder in R, $V_n(R)$ bzw. $M_n(R)$ vorgegebenen Metrik wie

üblich Durchmesser erklärt. (Wir halten uns hier im folgenden an die in [3] und [8] gegebene Darstellung.)

Für reelle Zahlen $a, b \in R$ sei die natürliche Metrik mit dem Abstand $|a - b|$ zugrundegelegt; damit erhalten wir für den Durchmesser eines Intervalls $A = [a_1, a_2] \in \mathbf{I}(R)$:

$$d(A) = \sup_{\substack{\bar{a} \in A \\ a \in A}} |\bar{a} - a| = |a_1 - a_2|;$$

Man erkennt: Für alle $A \in \mathbf{I}(R)$ ist $\; d(A) = |A - A|$.

Satz 5.4: Für $A, B \in \mathbf{I}(R)$ gilt:

a) $\quad d(A \pm B) = d(A) + d(B)$;

b) $\quad |A|\, d(B) \leq d(AB) \leq |A|\, d(B) + |B|\, d(A)$.

Beweis: a) $d(A \pm B) = |A \pm B - A \mp B| = |A - A| + |B - B| = d(A) + d(B)$;

b1) $\displaystyle |A|\, d(B) = \sup_{\substack{a \in A \\ b,\bar{b} \in B}} |a(b - \bar{b})| = \sup_{\substack{a \in A \\ b,\bar{b} \in B}} |ab - a\bar{b}| \leq \sup_{\substack{a,\bar{a} \in A \\ b,\bar{b} \in B}} |ab - \bar{a}\bar{b}| = d(AB)$,

b2) $\displaystyle d(AB) = \sup_{\substack{a,\bar{a} \in A \\ b,\bar{b} \in B}} |ab - \bar{a}\bar{b}| = \sup_{\substack{a,\bar{a} \in A \\ b,\bar{b} \in B}} |a(b - \bar{b}) + \bar{b}(a - \bar{a})| \leq$

$$\leq \sup_{\substack{a \in A \\ b,\bar{b} \in B}} |a(b - \bar{b})| + \sup_{\substack{a,\bar{a} \in A \\ \bar{b} \in B}} |\bar{b}(a - \bar{a})| = |A|\, d(B) + |B|\, d(A). \; \square$$

Die linke Ungleichung in Satz 5.4/b ist neu gegenüber [3] und [8]; sie spielt im Beweis von Satz 6.2 eine wichtige Rolle.

In den linearen Räumen $V_n(R)$ bzw. $M_n(R)$ definiert jede Norm $\|\mathfrak{w}\|$ bzw. $\|\mathfrak{M}\|$ eine Metrik; in dieser erhalten wir als Durchmesser für Elemente $\mathfrak{x} \in V_n(\mathbf{I}(R))$ bzw. $\mathfrak{A} \in M_n(\mathbf{I}(R))$:

$$d(\mathfrak{x}) = \sup_{\mathfrak{x}_1,\mathfrak{x}_2 \in \mathfrak{x}} \|\mathfrak{x}_1 - \mathfrak{x}_2\| \qquad \text{bzw.} \qquad d(\mathfrak{A}) = \sup_{\mathfrak{A}_1,\mathfrak{A}_2 \in \mathfrak{A}} \|\mathfrak{A}_1 - \mathfrak{A}_2\|.$$

Bei Verwendung der zu $\|\mathfrak{w}\|$ bzw. $\|\mathfrak{M}\|$ gehörigen Abstandsnormen $\|\mathfrak{w}\|$ bzw. $\|\mathfrak{M}\|$ in $V_n(\mathbf{I}(R))$ bzw. $M_n(\mathbf{I}(R))$ wird nun

$$d(\mathfrak{x}) = \|\mathfrak{x} - \mathfrak{x}\| \qquad \text{bzw.} \qquad d(\mathfrak{A}) = \|\mathfrak{A} - \mathfrak{A}\|.$$

Für $\mathfrak{x}, \mathfrak{y} \in V_n(I(R))$ bzw. $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$ folgt daraus durch Anwendung der Dreiecksungleichung für die Normen

$$d(\mathfrak{x}+\mathfrak{y}) \le d(\mathfrak{x}) + d(\mathfrak{y}) \quad \text{bzw.} \quad d(\mathfrak{A}+\mathfrak{B}) \le d(\mathfrak{A}) + d(\mathfrak{B})$$

Satz 5.5: Für $\mathfrak{x}=(X_i) \in V_n(I(R))$ bzw. $\mathfrak{A}=(A_{ik}) \in M_n(I(R))$ gilt bei Zugrundelegung monotoner Vektor- bzw. Matrixnormen $\|\mathfrak{w}\|_m$ bzw. $\|\mathfrak{M}\|_m$:

$$d(\mathfrak{x}) = \|(d(X_i))\|_m \quad \text{bzw.} \quad d(\mathfrak{A}) = \|(d(A_{ik}))\|_m$$

Beweis: $d(\mathfrak{x}) = \|\mathfrak{x}-\mathfrak{x}\|_m = \|\,|\mathfrak{x}-\mathfrak{x}|\,\|_m = \|(|X_i - X_i|)\|_m = \|(d(X_i))\|_m$.
Für $d(\mathfrak{A})$ schließt man entsprechend. □

Satz 5.6: a) Für $\mathfrak{A}, \mathfrak{B} \in M_n(I(R))$ gilt bei Zugrundelegeng monotoner multiplikativer Matrixnormen $\|\mathfrak{M}\|_m$:

$$d(\mathfrak{A}\mathfrak{B}) \le \|\mathfrak{A}\|_m d(\mathfrak{B}) + d(\mathfrak{A})\|\mathfrak{B}\|_m.$$

b) Für $\mathfrak{x} \in V_n(I(R))$, $\mathfrak{A}\in M_n(I(R))$ gilt bei Zugrundelegung verträglicher monotoner Vektor- und Matrixnormen $\|\mathfrak{w}\|_m$ und $\|\mathfrak{M}\|_m$:

$$d(\mathfrak{A}\mathfrak{x}) \le \|\mathfrak{A}\|_m d(\mathfrak{x}) + d(\mathfrak{A})\|\mathfrak{x}\|_m.$$

Beweis:

a) Für alle i, k, l gilt nach Satz 5.4:
$d(A_{ik}B_{kl}) \le d(A_{ik})|B_{kl}| + |A_{ik}|d(B_{kl})$;
Summation über k und Zusammenfassen in Matrixschreibweise ergibt:

$$\left(d\left(\sum_k A_{ik}B_{kl}\right)\right) \le (d(A_{ik}))(|B_{kl}|) + (|A_{ik}|)(d(B_{kl})).$$

In der multiplikativen monotonen Matrixnorm erhält man:

$$\left\|\left(d\left(\sum_k A_{ik}B_{kl}\right)\right)\right\|_m \le \|(d(A_{ik}))\|_m \|\mathfrak{B}\|_m + \|\mathfrak{A}\|_m \|(d(B_{kl}))\|_m$$

also $$d(\mathfrak{A}\mathfrak{B}) \le d(\mathfrak{A})\|\mathfrak{B}\|_m + \|\mathfrak{A}\|_m d(\mathfrak{B}).$$

b) Aussage b) beweist man analog; nur an Stelle der in a) verwendeten Multiplikativität der Norm benützt man jetzt die Verträglichkeit der Normen. □

Bemerkung: In der Beziehung $d(AB) \le |A|d(B) + |B|d(A)$ von Satz 5.4 steht das Gleichheitszeichen, sobald A oder B eine reelle Zahl ist; in den entsprechenden Aussagen von Satz 5.6 a) bzw. b) kann das Gleichheitszeichen gelten, wenn ein Faktor aus $M_n(R)$ bzw. $V_n(R)$ ist; trivialerweise gilt es, wenn beide Faktoren reell sind.

Eine Beziehung zwischen $\|\mathfrak{A}\|_m d(\mathfrak{B})$ und $d(\mathfrak{A}\mathfrak{B})$,
bzw. $\|\mathfrak{A}\|_m d(\mathfrak{x})$ und $d(\mathfrak{A}\mathfrak{x})$,
die der Beziehung $|A|d(B) \le d(AB)$ entspräche, existiert nicht; es sind vielmehr jeweils alle Relationen $<$, $=$ oder $>$ möglich. Wir zeigen dies an einfachen Beispielen für das Prudukt $\mathfrak{A}\mathfrak{x}$:

Für $\mathfrak{A}=(A_{ik})\in M_2(I(R))$ bzw. $\mathfrak{x}=(X_i)\in V_2(I(R))$ verwenden wir die zugeordneten monotonen Normen

$$\|\mathfrak{A}\| := \max_i \sum_k |A_{ik}| \quad \text{bzw.} \quad \|\mathfrak{x}\| := \max_i |X_i|$$

Für $\mathfrak{A} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$, $\mathfrak{x}_1 = \begin{pmatrix} [-2,5] \\ [-2,5] \end{pmatrix}$, $\mathfrak{x}_2 = \begin{pmatrix} [-2,5] \\ [-2,4] \end{pmatrix}$ wird

$\mathfrak{A}\mathfrak{x}_1 = \begin{pmatrix} [-7,7] \\ [-7,7] \end{pmatrix}$, $\mathfrak{A}\mathfrak{x}_2 = \begin{pmatrix} [-7,6] \\ [-7,6] \end{pmatrix}$, $\|\mathfrak{A}\| = 2$, $d(\mathfrak{x}_1) = d(\mathfrak{x}_2) = 7$,

$d(\mathfrak{A}\mathfrak{x}_1) = 14 = \|\mathfrak{A}\| d(\mathfrak{x}_1)$ und $d(\mathfrak{A}\mathfrak{x}_2) = 13 < \|\mathfrak{A}\| d(\mathfrak{x}_2)$.

Für $\mathfrak{A} = \begin{pmatrix} [3,4] & [3,4] \\ [3,4] & [3,4] \end{pmatrix}$, $\mathfrak{x} = \begin{pmatrix} [1,2] \\ [1,2] \end{pmatrix}$ wird $\mathfrak{A}\mathfrak{x} = \begin{pmatrix} [6,16] \\ [6,16] \end{pmatrix}$ und

$\|\mathfrak{A}\| = 8$, $d(\mathfrak{x}) = 1$, $d(\mathfrak{A}\mathfrak{x}) = 10 > \|\mathfrak{A}\| d(\mathfrak{x})$.

6. Die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$

Definition 6.1: Gegeben seien $\mathfrak{A} \in M_n(I(R))$ und $\mathfrak{b} \in V_n(I(R))$. Ein Intervallvektor $\mathfrak{x}^* \in V_n(I(R))$ mit der Eigenschaft $\mathfrak{x}^* = \mathfrak{A}\mathfrak{x}^* + \mathfrak{b}$ heißt Fixpunkt der Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ und Lösung der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$. Hat die Abbildung keinen Fixpunkt, so heißt die Gleichung unlösbar.

Wir sprechen gelegentlich ohne Unterschied in der Bedeutung auch von einem Gleichungssystem.

Auf den Zusammenhang zwischen Gleichungssystemen der Form $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ und reellen linearen Gleichungssystemen $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ derselben Gestalt, aber mit $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$ werden wir in Abschnitt 7 näher eingehen. Zum Beispiel gilt im Fall $\mathfrak{A} \in \mathfrak{A}$, $\mathfrak{b} \in \mathfrak{b}$ auch für die Lösungen $\mathfrak{x}^*$ bzw. $\mathfrak{x}^*$ dieser Gleichungssysteme die Beziehung $\mathfrak{x}^* \in \mathfrak{x}^*$.

Ist die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ lösbar, dann kann die Lösung eventuell durch sukzessive Approximation bestimmt werden; diese Frage wollen wir jetzt untersuchen.

6.1 Das Konvergenzverhalten der Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$

Die beiden folgenden Sätze ergeben zusammen ein notwendiges und hinreichendes Kriterium für die Konvergenz des Iterationsverfahrens $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit einer Matrix $\mathfrak{A} \in M_n(I(R))$ und beliebigem $\mathfrak{b} \in V_n(I(R))$.

Satz 6.1: Es sei $\mathfrak{A} = (A_{il}) \in M_n(I(R))$ und $\varrho(|\mathfrak{A}|) < 1$, dann hat die Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ für beliebige $\mathfrak{b} \in V_n(I(R))$ jeweils genau einen Fixpunkt $\mathfrak{x}^*$, d. h. die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ genau eine Lösung $\mathfrak{x}^*$ und das Iterationsverfahren $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ konvergiert für jeden Anfangsvektor $\mathfrak{x}_0 \in V_n(I(R))$ gegen $\mathfrak{x}^*$.

Beweis: Wegen $\varrho(|\mathfrak{A}|) < 1$ existiert nach dem Korollar zu Satz 4.14 in $M_n(I(R))$ eine monotone multiplikative Matrixnorm $\|\mathfrak{M}\|_m$, für die $\|\mathfrak{A}\|_m < 1$.

Zu dieser Matrixnorm existiert nach Satz 4.12 eine verträgliche monotone Vektornorm $\|\mathfrak{w}\|_m$ in $V_n(I(R))$; nach Satz 5.3 ist nun auch die dieser Vektornorm zugeordnete Metrik $q_m(\mathfrak{x},\mathfrak{y})$ in $V_n(I(R))$ mit der Matrixnorm $\|\mathfrak{M}\|_m$ verträglich; damit gilt

$$q_m(\mathfrak{A}\mathfrak{x} + \mathfrak{b}, \mathfrak{A}\mathfrak{y} + \mathfrak{b}) \leq \|\mathfrak{A}\|_m \, q_m(\mathfrak{x},\mathfrak{y}) \quad \text{für alle } \mathfrak{x}, \mathfrak{y} \in V_n(I(R)).$$

Wegen $\|\mathfrak{A}\|_m < 1$ ist die Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ in der Metrik $q_m(\mathfrak{x},\mathfrak{y})$ kontrahierend; der Raum $V_n(I(R))$ ist vollständig; der Fixpunktsatz von Banach über Kontraktionsoperatoren in vollständigen metrischen Räumen (siehe etwa [16]) gibt jetzt die Aussage des Satzes. □

Der folgende Satz enthält die Umkehrung von Satz 6.1.

Satz 6.2: Gegeben seien $\mathfrak{A} \in M_n(I(R))$ und $\mathfrak{b} \in V_n(I(R))$; konvergiert das Iterationsverfahren $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit jedem Anfangsvektor $\mathfrak{x}_0 \in V_n(I(R))$ gegen denselben Vektor $\mathfrak{x}^* \in V_n(I(R))$, dann ist $\varrho(|\mathfrak{A}|) < 1$.

Beweis: Für die Komponentendurchmesser $d(X_{i\nu})$ der Iterationsvektoren $\mathfrak{x}_\nu = (X_{i\nu})$ gilt nach Satz 5.4:

$$d(X_{i\,\nu+1}) = d(\sum_l A_{il} X_{l\nu}) + d(B_i) \geq \sum_l d(A_{il} X_{l\nu}) \geq \sum_l |A_{il}| \, d(X_{l\nu})$$

Wir fassen die Komponentendurchmesser $d(X_{i\nu})$ von $\mathfrak{x}_\nu = (X_{i\nu})$ zusammen im reellen Vektor $(d(X_{i\nu}))$; dann gilt

$$(d(X_{i\nu})) \geq |\mathfrak{A}| \, (d(X_{i\,\nu-1})) \geq \ldots\ldots \geq |\mathfrak{A}|^\nu (d(X_{i0})). \qquad (1)$$

Nach Voraussetzung konvergiert für jeden Anfangsvektor $\mathfrak{x}_0 \in V_n(I(R))$ die Folge der Iterationsvektoren $\mathfrak{x}_\nu$ gegen $\mathfrak{x}^* = (X_i^*)$; dies beinhaltet aber die Konvergenz der Vektorfolge $(d(X_{i\nu}))$ gegen $(d(X_i^*))$. Deshalb gilt:

Für alle $\mathfrak{x}_0 \in V\,(I(R))$ und alle $\varepsilon > 0$ existiert eine natürliche

Zahl $n(\varepsilon,\mathscr{X}_o)$ mit der Eigenschaft, daß für alle $\nu > n(\varepsilon,\mathscr{X}_o)$ gilt:

$$d(X_{i\nu}) \leq d(X_i^*) + \varepsilon \qquad \text{für alle } 1 \leq i \leq n,$$

also auch $d(X_{i\nu}) \leq \max_i d(X_i^*) + \varepsilon$ für alle $1 \leq i \leq n$. (2)

$D_\varepsilon := \max_i d(X_i^*) + \varepsilon$ ist jeweils durch $\mathfrak{A}$, $\mathfrak{b}$ und ε bestimmt und endlich.

Für beliebige $c > 0$ und ein festes $\bar{\varepsilon} > 0$ betrachten wir jetzt den Anfangsvektor

$\mathscr{X}_o(c,\bar{\varepsilon}) := (X_{io})$ mit $X_{io} = [0, cD_{\bar{\varepsilon}}]$, also $d(X_{io}) = cD_{\bar{\varepsilon}}$ für alle $1 \leq i \leq n$.

Nach (1) und (2) gilt jetzt für alle $\nu \geq n(\bar{\varepsilon},\mathscr{X}_o(c,\bar{\varepsilon}))$

$$\begin{pmatrix} D_{\bar{\varepsilon}} \\ \vdots \\ D_{\bar{\varepsilon}} \end{pmatrix} \geq \begin{pmatrix} d(X_1^*)+\bar{\varepsilon} \\ \vdots \\ d(X_n^*)+\bar{\varepsilon} \end{pmatrix} \geq \begin{pmatrix} d(X_{1\nu}) \\ \vdots \\ d(X_{n\nu}) \end{pmatrix} \geq |\mathfrak{A}|^\nu \begin{pmatrix} cD_{\bar{\varepsilon}} \\ \vdots \\ cD_{\bar{\varepsilon}} \end{pmatrix}, \text{ also}$$

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \geq c\,|\mathfrak{A}|^\nu \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Damit gilt für die Potenzen $|\mathfrak{A}|^\nu =: (a_{ij}^{(\nu)})$:
Für jedes $c > 0$ existiert eine natürliche Zahl $n(c)$, nämlich $n(c) := n(\bar{\varepsilon},\mathscr{X}_o(c,\bar{\varepsilon}))$ mit der Eigenschaft, daß für alle $\nu > n(c)$ gilt:

$$\sum_i a_{ij}^{(\nu)} \leq \frac{1}{c} \qquad \text{für alle } 1 \leq j \leq n,$$

also auch $a_{ij}^{(\nu)} \leq \frac{1}{c}$ für alle $1 \leq i,j \leq n$.

Folglich ist $\lim_{\nu\to\infty} a_{ij}^{(\nu)} = 0$ für alle $1 \leq i,j \leq n$,

d. h. $\lim_{\nu\to\infty} |\mathfrak{A}|^\nu = \mathfrak{O}$; dies aber ist (siehe etwa [14], Theorem 1.4) äquivalent mit $\varrho(|\mathfrak{A}|) < 1$. □

Zusammenfassen der Sätze 6.1 und 6.2 liefert <u>ein notwendiges und hinreichendes Konvergenzkriterium für die Iteration</u>
<u>$\mathscr{X}_{\nu+1} = \mathfrak{A}\mathscr{X}_\nu + \mathfrak{b}$:</u>

Es seien $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$; dann sind folgende Aussagen äquivalent:

1) Die Iteration $\mathscr{X}_{\nu+1} = \mathfrak{A}\mathscr{X}_\nu + \mathfrak{b}$ konvergiert mit jedem Anfangsvektor $\mathscr{X}_o \in V_n(I(R))$.

2) $\varrho(|\mathfrak{A}|) < 1$.

Berücksichtigung von Satz 4.14 und des anschließenden Korollars liefert noch die Äquivalenz von 1) und 2) mit

2') Es existiert eine monotone multiplikative Norm $\|\mathfrak{M}\|_m$ in $M_n(I(R))$, für die $\|\mathfrak{A}\|_m < 1$.

Ein wichtiger Spezialfall des oben behandelten Gleichungssystems ist das reelle lineare Gleichungssystem $\mathscr{X} = \mathfrak{A}\mathscr{X} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$. Die Iteration $\mathscr{X}_{\nu+1} = \mathfrak{A}\mathscr{X}_\nu + \mathfrak{b}$ zur Bestimmung der Lösung $\mathscr{X}^*$ kann hier entweder, wie eben beschrieben, mit Intervallvektoren $\mathscr{X}_o \in V_n(I(R))$ oder aber in herkömmlicher Weise mit reellen Vektoren $\mathscr{X}_o = \mathfrak{x}_o \in V_n(R)$ gestartet und durchgeführt werden. Für die Iteration mit Intervallvektoren haben wir oben ein notwendiges und zugleich hinreichendes Konvergenzkriterium abgeleitet, für die Iteration mit reellen Vektoren sind solche Konvergenzkriterien allgemein bekannt; wir zitieren sie zum Vergleich mit den Kriterien bei der intervallmäßigen Iteration:

Es seien $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$; dann sind folgende Aussagen äquivalent:

1) Die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ konvergiert mit jedem Anfangsvektor $\mathfrak{x}_o \in V_n(R)$.

2) $\varrho(\mathfrak{A}) < 1$.

2') Es existiert eine multiplikative Norm $\|\mathfrak{M}\|$ in $M_n(R)$, für die $\|\mathfrak{A}\| < 1$ (siehe Satz 4.13).

Damit erhalten wir:

Für ein reelles lineares Gleichungssystem $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$ konvergiert das Iterationsverfahren

$\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$

für alle $\mathfrak{x}_o = \mathfrak{x}_o \in V_n(R)$ dann und nur dann, wenn $\varrho(\mathfrak{A}) < 1$,

für alle $\mathfrak{x}_o \in V_n(I(R))$ dann und nur dann, wenn $\varrho(|\mathfrak{A}|) < 1$.

Für jede Matrix $\mathfrak{A} \in M_n(R)$ ist bekanntlich $\varrho(\mathfrak{A}) \le \varrho(|\mathfrak{A}|)$ (vergleiche etwa [14], Lemma 2.3; die Aussage wird dort für irreduzible Matrizen bewiesen; für reduzible folgt sie dann aus der stetigen Abhängigkeit des Spektralradius von den Matrixkomponenten).

Der Konvergenz für die umfassendere Menge von Anfangsvektoren (es ist $V_n(I(R)) \supset V_n(R)$) entspricht also die schärfere Bedingung.

Eine wichtige Folgerung aus Satz 6.2 enthält das folgende

Korollar: Ist eine Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$ multiplikativ, so gilt für alle $\mathfrak{A} \in M_n(I(R))$: $\|\mathfrak{A}\| \ge \varrho(|\mathfrak{A}|)$.

Beweis: Nach Satz 4.12 existiert zur Matrixnorm $\|\mathfrak{M}\|$ eine verträgliche Vektornorm $\|\mathfrak{w}\|$ in $V_n(I(R))$.

Wir betrachten nun ein festes $\mathfrak{A} \in M_n(I(R))$; es sei $a := \|\mathfrak{A}\|$ und $0 < \varepsilon \in R$, dann hat die Matrix $\frac{1}{a+\varepsilon}\mathfrak{A}$ die Norm $\left\|\frac{1}{a+\varepsilon}\mathfrak{A}\right\| = \frac{a}{a+\varepsilon} < 1$.

Die Iteration $\mathfrak{x}_{\nu+1} = \frac{1}{a+\varepsilon}\mathfrak{A}\mathfrak{x}_\nu$ konvergiert jetzt auf Grund der Normabschätzung $\|\mathfrak{x}_\nu\| \le \left\|\frac{1}{a+\varepsilon}\mathfrak{A}\right\|^\nu \|\mathfrak{x}_o\|$ für jeden Anfangsvektor $\mathfrak{x}_o$ aus $V_n(I(R))$ gegen den Nullvektor.

Mit Satz 6.2 folgt daraus

$$\varrho\left(\left|\frac{1}{a+\varepsilon}\mathfrak{A}\right|\right) = \frac{1}{a+\varepsilon}\varrho(|\mathfrak{A}|) < 1;$$

also ist $\varrho(|\mathfrak{A}|) < a+\varepsilon$ für alle $\varepsilon > 0$; das bedeutet $\varrho(|\mathfrak{A}|) \le a = \|\mathfrak{A}\|$. □

Mit diesem Korollar und dem früher bewiesenen Satz 4.14 er-

halten wir jetzt den folgenden

Satz 6.3: Es sei $\mathfrak{A} \in M_n(I(R))$; dann gilt:

(Pa) Für jede multiplikative Norm $\|\mathfrak{M}\|$ in $M_n(I(R))$ ist $\|\mathfrak{A}\| \ge \varrho(|\mathfrak{A}|)$.

(Pb) Zu jedem $\varepsilon > 0$ existiert eine multiplikative Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$, für die $\|\mathfrak{A}\| < \varrho(|\mathfrak{A}|) + \varepsilon$.

(Pa) und (Pb) zusammen sind gleichbedeutend mit:

(P) $\varrho(|\mathfrak{A}|) = \inf\left\{\|\mathfrak{A}\| \;\middle|\; \|\;\| \text{ eine multiplikative Norm in } M_n(I(R))\right\}$

Beweis: (Pa) folgt aus dem eben bewiesenen Korollar;
(Pb) folgt aus Aussage (Ib) von Satz 4.14.

Satz 6.3 macht gerade die Satz 4.13 entsprechende Aussage für multiplikative Normen in $M_n(I(R))$. □

6.2 Gesamt- und Einzelschrittverfahren zur Lösung der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$

Wir betrachten wieder das Gleichungssystem $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$ und $\mathfrak{b} \in V_n(I(R))$, und zerlegen $\mathfrak{A}$ in die Summe $\mathfrak{A} = \mathfrak{L} + \mathfrak{D} + \mathfrak{R}$; dabei sei $\mathfrak{L}$ eine strenge untere, $\mathfrak{R}$ eine strenge obere Dreiecksmatrix, $\mathfrak{D}$ eine Diagonalmatrix.

Wie bei Gleichungssystemen üblich, bezeichnen wir die Iterationsvorschrift

$\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ als Gesamtschrittverfahren;

$\mathfrak{x}_{\nu+1} = \mathfrak{L}\mathfrak{x}_{\nu+1} + (\mathfrak{D} + \mathfrak{R})\mathfrak{x}_\nu + \mathfrak{b}$ als Einzelschrittverfahren.

Nach Satz 6.1 hat die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ für $\varrho(|\mathfrak{A}|) < 1$ genau eine Lösung $\mathfrak{x}^*$ und das Gesamtschrittverfahren konvergiert gegen diese Lösung.

Aus $\mathfrak{A}\mathfrak{x} = (\mathfrak{L} + \mathfrak{D} + \mathfrak{R})\mathfrak{x} = \mathfrak{L}\mathfrak{x} + (\mathfrak{D} + \mathfrak{R})\mathfrak{x}$ folgt, daß die Gleichungen $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ und $\mathfrak{x} = \mathfrak{L}\mathfrak{x} + (\mathfrak{D} + \mathfrak{R})\mathfrak{x} + \mathfrak{b}$ dieselbe

Lösung $\mathfrak{x}^*$ haben; das bedeutet, daß das Einzelschrittverfahren im Falle der Konvergenz auch gegen die Lösung $\mathfrak{x}^*$ konvergiert. Wir werden zeigen, daß das Einzelschrittverfahren genau dann konvergiert bzw. divergiert, wenn auch das zugehörige Gesamtschrittverfahren konvergiert oder divergiert.
Zur Vorbereitung darauf skizzieren wir zunächst, wie analoge Vergleichsaussagen bei reellen linearen Gleichungssystemen und Iteration mit Vektoren aus $V_n(R)$ gewonnen werden. Bei der intervallmäßigen Iteration werden wir dann ein Verhalten finden, das dem bei reellen linearen Gleichungssystemen $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ $(\mathfrak{A} \in M_n(R),\ \mathfrak{b} \in V_n(R))$ mit nichtnegativem $\mathfrak{A}$ entspricht.

$\mathfrak{x} = \mathfrak{L}\mathfrak{x} + \mathfrak{b}$ sie irgendeine äquivalente Darstellung eines Gleichungssystems $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$; die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ konvergiert dann und nur dann für alle $\mathfrak{x}_0 \in V_n(R)$ wenn $\varrho(\mathfrak{L}) < 1$.
Für die Abweichung $(\mathfrak{x}_\nu - \mathfrak{x}^*)$ der ν-ten Näherung $\mathfrak{x}_\nu$ von der Lösung $\mathfrak{x}^*$ gilt nämlich

$$(\mathfrak{x}_\nu - \mathfrak{x}^*) = \mathfrak{L}^\nu (\mathfrak{x}_0 - \mathfrak{x}^*) \qquad (1R)$$

und nach [14], Theorem 1.4 ist $\lim\limits_{\nu\to\infty} \mathfrak{L}^\nu = \mathfrak{O} \iff \varrho(\mathfrak{L}) < 1.$
Mit verträglichen Vektor- und Matrixnormen erhält man aus (1R)

$$\|\mathfrak{x}_\nu - \mathfrak{x}^*\| \le \|\mathfrak{L}^\nu\|\ \|\mathfrak{x}_0 - \mathfrak{x}^*\|.$$

Im Falle $\|\mathfrak{x}_0 - \mathfrak{x}^*\| \neq 0$ gibt $\|\mathfrak{L}^\nu\|$ eine obere Schranke für das Verhältnis

$$\frac{\|\mathfrak{x}_\nu - \mathfrak{x}^*\|}{\|\mathfrak{x}_0 - \mathfrak{x}^*\|}.$$

Um dafür möglichst gute obere Schranken zu erhalten, verwendet man zweckmäßig zugeordnete Matrixnormen (in Satz 4.5 haben wir deren Minimaleigenschaft bewiesen).
An Hand von $\|\mathfrak{L}\|^\nu$ können nun verschiedene Iterationsverfahren verglichen werden. Bekanntlich (siehe [14]) heißt
im Falle $\|\mathfrak{L}^m\| < 1$
$R(\mathfrak{L}^m) := -\ln \|\mathfrak{L}^m\|^{\frac{1}{m}}$ durchschnittliche Konvergenzgeschwindigkeit des Iterationsverfahrens mit der Matrix $\mathfrak{L}$ bei m-Iterationen.

$R_\infty(\mathfrak{L}) := \lim\limits_{m\to\infty} R(\mathfrak{L}^m)$ asymptotische Konvergenzgeschwindigkeit des Iterationsverfahrens mit der Matrix $\mathfrak{L}$

Für jede zugeordnete Matrixnorm ist nun

$$R_\infty(\mathfrak{L}) := \lim_{m\to\infty} R(\mathfrak{L}^m) = \lim \frac{-\ln \|\mathfrak{L}^m\|}{m} = -\ln \varrho(\mathfrak{L}).$$

Dies wird für die Spektralnorm (diese ist der euklidischen Vektornorm zugeordnet) in [14] bewiesen; der dort angegebene Beweis läßt sich jedoch leicht auf beliebige zugeordnete Matrixnormen übertragen.

Bei zwei Iterationsverfahren (I) und (II) zur Bestimmung der Lösung $\mathfrak{x}^*$ einer Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$, bei denen jeweils $(\mathfrak{x}_\nu - \mathfrak{x}^*) = \mathfrak{L}_I^\nu (\mathfrak{x}_0 - \mathfrak{x}^*)$ bzw. $(\mathfrak{x}_\nu - \mathfrak{x}^*) = \mathfrak{L}_{II}^\nu (\mathfrak{x}_0 - \mathfrak{x}^*)$ gilt, vergleicht man üblicherweise das Konvergenzverhalten an Hand der oben definierten Konvergenzgeschwindigkeiten; die asymptotische Konvergenzgeschwindigkeit, gegeben durch den Spektralradius, liefert dabei das am einfachsten anwendbare Maß. Das Verfahren, dessen Iterationsmatrix den kleineren Spektralradius hat, heißt asymptotisch schneller. Bei solchen Vergleichen sind jedoch die folgenden Vorbehalte zu beachten:
Grundsätzlich können ja nur endlich viele Iterationsschritte durchgeführt werden; die asymptotische Konvergenzgeschwindigkeit liefert jeweils nur Abschätzungen der durchschnittlichen Konvergenzgeschwindigkeit bei endlich vielen Iterationsschritten, aber auch wenn diese exakt bekannt sind, ist wegen der Ungleichheitszeichen in den Normbeziehungen $\|\mathfrak{x}_\nu - \mathfrak{x}^*\| \le \|\mathfrak{L}_i^\nu\|\ \|\mathfrak{x}_0 - \mathfrak{x}^*\|$, $i = I, II$, ein sicherer Vergleich dieser Iterationsverfahren für beliebige Anfangsvektoren $\mathfrak{x}_0$ nicht möglich.

Speziell für das Gesamt- und Einzelschrittverfahren zur Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ gelten die folgenden Beziehungen:

$$(\mathfrak{x}_\nu - \mathfrak{x}^*) = \mathfrak{A}^\nu (\mathfrak{x}_0 - \mathfrak{x}^*) \qquad (2R)$$
$$(\mathfrak{x}_\nu - \mathfrak{x}^*) = \left((\mathfrak{E} - \mathfrak{L})^{-1} (\mathfrak{D} + \mathfrak{R})\right)^\nu (\mathfrak{x}_0 - \mathfrak{x}^*) \qquad (3R)$$

Für beliebige Matrizen $\mathfrak{A} \in M_n(R)$ ist hier sowohl $\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})) \le \varrho(\mathfrak{A})$ also auch $\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})) > \varrho(\mathfrak{A})$

möglich (siehe dazu [14]), das Einzelschrittverfahren kann deshalb auch bei konvergentem Gesamtschrittverfahren divergieren und umgekehrt. Für nichtnegative Matrizen $\mathcal{A} \in M_n(R)$ jedoch konvergieren Gesamt- und Einzelschrittverfahren stets gleichzeitig; dies folgt aus einem in [4] bewiesenen Satz. Wir zitieren diesen Satz so, daß wir ihn auch zur Herleitung entsprechender Aussagen bei der intervallmäßigen Iteration verwenden können.

Satz 6.4: Ist $\mathcal{A} = \mathcal{L} + \mathcal{D} + \mathcal{R}$ eine reelle nichtnegative Matrix aus $M_n(R)$, (mit der oben beschriebenen Zerlegung), dann gilt für

a) $\varrho(\mathcal{A}) < 1$: $\varrho((\mathcal{E} - \mathcal{L})^{-1}(\mathcal{D} + \mathcal{R})) \le \varrho(\mathcal{A}) < 1$, mit strenger Ungleichheit bei irreduziblem $\mathcal{A}$.

b) $\varrho(\mathcal{A}) \ge 1$: $\varrho((\mathcal{E} - \mathcal{L})^{-1}(\mathcal{D} + \mathcal{R})) \ge \varrho(\mathcal{A}) \ge 1$.

Dieser Satz ist eine Verallgemeinerung des bekannten Satzes von Stein-Rosenberg (siehe etwa [14]) für den Fall $\mathcal{D} \neq \mathcal{O}$.

Für reelle lineare Gleichungssysteme $\mathfrak{x} = \mathcal{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathcal{A} \ge 0$ folgt aus Satz 6.4 sofort, daß mit dem Gesamtschrittverfahren stets auch das Einzelschrittverfahren konvergiert bzw. divergiert; im Falle der Konvergenz ist das Einzelschrittverfahren asymptotisch mindestens ebenso schnell wie das Gesamtschrittverfahren, bei irreduzibler Matrix $\mathcal{A}$ sogar asymptotisch schneller.

Jetzt vergleichen wir Gesamt- und Einzelschrittverfahren beim Gleichungssystem $\mathfrak{x} = \mathcal{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathcal{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$. $\mathcal{A} = \mathcal{L} + \mathcal{D} + \mathcal{R}$ sei die oben beschriebene Zerlegung von $\mathcal{A}$ und es gelte $\varrho(|\mathcal{A}|) < 1$; nach Satz 6.1 hat dann das Gleichungssystem ganau eine Lösung und das Gesamtschrittverfahren konvergiert gegen diese Lösung.

$\mathfrak{x}^* = (X_i^*)$ bezeichne die eindeutige Lösung des Gleichungssystems,

$\mathfrak{x}_\nu = (X_{i\nu})$ den ν-ten Iterationsvektor.

Beim Gesamtschrittverfahren $\mathfrak{x}_{\nu+1} = \mathcal{A}\mathfrak{x}_\nu + \mathfrak{b}$ ist

$X_{i\nu} = \sum_k A_{ik}X_{k\nu-1} + B_i$ und $X_i^* = \sum_k A_{ik}X_k^* + B_i$ jeweils für alle i.

Mit der Metrik $q(A,B)$ in $I(R)$ gilt jetzt für alle i:

$$\begin{aligned} q(X_{i\nu}, X_i^*) &= q\Big(\sum_k A_{ik}X_{k\nu-1} + B_i, \sum_k A_{ik}X_k^* + B_i\Big) = \\ &= q\Big(\sum_k A_{ik}X_{k\nu-1}, \sum_k A_{ik}X_k^*\Big) \le \sum_k q(A_{ik}X_{k\nu-1}, A_{ik}X_k^*) \\ &\le \sum_k |A_{ik}|\, q(X_{k\nu-1}, X_k^*). \end{aligned}$$

(Hier wird zunächst die Translationsinvarianz und die daraus folgende supermetrische Eigenschaft der Metrik $q(A,B)$ und schließlich Satz 1.4 benutzt.)

Zusammenfassen in Matrixschreibweise ergibt nun für das Gesamtschrittverfahren:

$$\big(q(X_{i\nu}, X_i^*)\big) \le |\mathcal{A}| \big(q(X_{i\nu-1}, X_i)\big) \le |\mathcal{A}|^\nu \big(q(X_{i0}, X_i)\big) \qquad (2)$$

Beim Einzelschrittverfahren $\mathfrak{x}_{\nu+1} = \mathcal{L}\mathfrak{x}_{\nu+1} + (\mathcal{D} + \mathcal{R})\mathfrak{x}_\nu + \mathfrak{b}$ gilt zunächst komponentenweise für alle i:

$X_{i\nu} = \sum_k L_{ik}X_{k\nu} + D_{ii}X_{i\nu-1} + \sum_k R_{ik}X_{k\nu-1} + B_i$ und

$X_i^* = \sum_k L_{ik}X_k^* + D_{ii}X_i^* + \sum_k R_{ik}X_k^* + B_i$.

Daraus folgt wie oben beim Gesamtschrittverfahren für alle i:

$$q(X_{i\nu}, X_i^*) \le \sum_k |L_{ik}|\, q(X_{k\nu}, X_k) + |D_{ii}|\, q(X_{i\nu-1}, X_i^*) + \sum_k |R_{ik}|\, q(X_{k\nu-1}, X_k^*).$$

Zusammenfassen in Matrixschreibweise ergibt:

$$\big(q(X_{i\nu}, X_i^*)\big) \le |\mathcal{L}| \big(q(X_{i\nu}, X_i^*)\big) + (|\mathcal{D}| + |\mathcal{R}|)\big(q(X_{i\nu-1}, X_i^*)\big),$$

also $(\mathcal{E} - |\mathcal{L}|)^{-1}\big(q(X_{i\nu}, X_i^*)\big) \le (|\mathcal{D}| + |\mathcal{R}|)\big(q(X_{i\nu-1}, X_i^*)\big)$.

$\mathcal{L}$ ist eine strenge untere Dreiecksmatrix, also ist $|\mathcal{L}|^\nu = \mathcal{O}$ für alle $\nu > n$; damit wird $(\mathcal{E} - |\mathcal{L}|)^{-1} = \sum_{\nu=0}^{n} |\mathcal{L}|^\nu > 0$.

Damit gilt für das Einzelschrittverfahren

$$\begin{aligned} \big(q(X_{i\nu}, X_i^*)\big) &\le (\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)\big(q(X_{i\nu-1}, X_i^*)\big) \le \\ &\le \big((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)\big)^\nu \big(q(X_{i0}, X_i^*)\big). \end{aligned} \qquad (3)$$

Die Beziehungen (2) und (3) haben die allgemeine Gestalt

$$(q(X_{i\nu},X_i^*)) \leq |\mathcal{B}|^\nu (q(X_{io},X_i)); \qquad (1)$$

Konvergenz ergibt sich jeweils für $\varrho(|\mathcal{B}|) < 1$.

(2) und (3) entsprechen den Beziehungen (2R) und (3R) mit der allgemeinen Gestalt (1R) $(\varphi_\nu - \varphi^*) = \mathcal{B}^\nu (\varphi_o - \varphi^*)$ bei reellen linearen Gleichungssystemen und Iteration mit Vektoren aus $V_n(\mathbb{R})$. Wie für diesen Fall beschrieben gehen wir jetzt auch in (1) zu zugeordneten Normen über; dann ergibt sich entsprechend

$$\frac{\|(q(X_{i\nu},X_i))\|}{\|(q(X_{io},X_i))\|} \leq \||\mathcal{L}|\|^\nu$$

und das asymptotische Konvergenzverhalten der durch (1) beschriebenen Iteration läßt sich jetzt mit dem Spektralradius

$\varrho(|\mathcal{L}|) = \lim\limits_{n\to\infty} \||\mathcal{L}|^n\|^{\frac{1}{n}}$ abschätzen.

Wenden wir jetzt Satz 6.4 auf die Matrizen $|\mathfrak{A}|$ und $(\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)$ an den Beziehungen (2) und (3) an, so ergibt sich im Falle $\varrho(|\mathfrak{A}|) < 1$:

Konvergiert das Gesamtschrittverfahren $\varphi_{\nu+1} = \mathfrak{A}\varphi_\nu + \mathfrak{b}$ für die Matrix $\mathfrak{A} \in M_n(\mathbb{I}(\mathbb{R}))$, so konvergiert auch das zugehörige Einzelschrittverfahren $\varphi_{\nu+1} = \mathcal{L}\varphi_{\nu+1} + (\mathcal{D} + \mathcal{R})\varphi_\nu + \mathfrak{b}$ und zwar asymptotisch mindestens genauso schnell wie das Einzelschrittverfahren, bei irreduzibler Matrix $\mathfrak{A}$ sogar schneller.

Daß bei divergentem Gesamtschrittverfahren auch das Einzelschrittverfahren divergiert, oder gleichbedeutend bei konvergentem Einzelschrittverfahren auch das Gesamtschrittverfahren konvergiert, kann wegen der Ungleichheitszeichen nicht aus den Beziehungen (2) und (3) abgeleitet werden. Um dies zu zeigen gehen wir ähnlich vor wie im Beweis von Satz 6.2; dort wurde für die Komponentendurchmesser der Iterationsvektoren $\varphi_\nu = (X_{i\nu})$ beim Gesamtschrittverfahren die Beziehung

$$(d(X_{i\nu})) \geq |\mathfrak{A}|^\nu (d(X_{io})) \qquad (4)$$

abgeleitet.

Die entsprechende Beziehung für das Einzelschrittverfahren lautet:

$$(d(X_{i\nu})) \geq ((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|))^\nu (d(X_{io})) \qquad (5)$$

<u>Beweis:</u> Mit $\mathcal{L} = (L_{ik})$, $\mathcal{D} = (D_{ik})$, $\mathcal{R} = (R_{ik})$ ist für alle $1 \leq i \leq n$:

$$X_{i\nu+1} = \sum L_{ik}X_{k\nu+1} + D_{ii}X_{i\nu} + \sum R_{ik}X_{k\nu} + B_i$$

Mit Satz 5.4 folgt daraus

$$d(X_{i\nu+1}) \geq \sum_k |L_{ik}|\, d(X_{k\nu+1}) + |D_{ii}|\, d(X_{i\nu}) + \sum_k |R_{ik}|\, d(X_{k\nu});$$

Zusammenfassen in Matrixschreibweise ergibt

$(d(X_{i\nu+1})) \geq |\mathcal{L}|\,(d(X_{i\nu+1})) + (|\mathcal{D}| + |\mathcal{R}|)\,(d(X_{i\nu}))$, also

$(\mathcal{E} - |\mathcal{L}|)\,(d(X_{i\nu+1})) \geq (|\mathcal{D}| + |\mathcal{R}|)\,(d(X_{i\nu}))$.

Wegen $(\mathcal{E} - |\mathcal{L}|)^{-1} = \sum\limits_{\nu=1}^{n} |\mathcal{L}|^\nu \geq \sigma$ (siehe oben) erhalten wir

$$(d(X_{i\nu+1})) \quad ((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|))\,(d(X_{i\nu}));$$

daraus folgt jetzt sofort (5).

Aus (5) folgt mit derselben Schlußweise wie im Beweis von Satz 6.2:

Konvergiert das Einzelschrittverfahren für alle $\varphi_o \in V_n(\mathbb{I}(\mathbb{R}))$, so ist $\varrho((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)) < 1$. □

<u>Wir fassen</u> nochmals <u>zusammen:</u>

Aus (3) und (5) folgt:

Das Einzelschrittverfahren $\varphi_{\nu+1} = \mathcal{L}\varphi_{\nu+1} + \mathcal{D}\varphi_\nu + \mathcal{R}\varphi_\nu + \mathfrak{b}$ konvergiert dann und nur dann, wenn $\varrho((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)) < 1$.

Aus den Sätzen 6.1 und 6.2 (ebenso aus (2) und (4)) folgt:

Das Gesamtschrittverfahren $\varphi_{\nu+1} = \mathfrak{A}\varphi_\nu + \mathfrak{b}$ konvergiert dann und nur dann, wenn $\varrho(|\mathfrak{A}|) < 1$.

Nach Satz 6.4 ist

entweder $\varrho((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)) \leq \varrho(|\mathfrak{A}|) < 1$

mit strenger Ungleichheit für irreduzible $\mathfrak{A}$,

oder $\varrho((\mathcal{E} - |\mathcal{L}|)^{-1}(|\mathcal{D}| + |\mathcal{R}|)) \geq \varrho(|\mathfrak{A}|) \geq 1$.

Daraus entnehmen wir den folgenden

Satz 6.5: Konvergiert das Gesamtschrittverfahren $\varphi_{\nu+1} = \mathfrak{A}\varphi_\nu + \mathfrak{b}$ mit einer Matrix $\mathfrak{A} \in M_n(I(R))$, so konvergiert auch das zugehörige Einzelschrittverfahren $\varphi_{\nu+1} = \mathfrak{L}\varphi_{\nu+1} + (\mathfrak{D}+\mathfrak{R})\varphi_\nu + \mathfrak{b}$ und zwar asymptotisch mindestens genauso schnell wie das Gesamtschrittverfahren, bei irreduziblem $\mathfrak{A}$ asymptotisch schneller. Divergiert das Gesamtschrittverfahren, so divergiert auch stets das zugehörige Einzelschrittverfahren.

Diese Aussage ist hier im Gegensatz zu beliebigen reellen linearen Gleichungssystemen allgemein möglich, weil das Konvergenzverhalten beim Gesamt- bzw. Einzelschrittverfahren wegen der von vorneherein wesentlich schärferen Konvergenzbedingungen $\varrho(|\mathfrak{A}|) < 1$ bzw. $\varrho((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|)) < 1$ nur vom Absolutbetrag $|\mathfrak{A}|$ der Matrix $\mathfrak{A}$ abhängt. Das sinngemäße reelle Analogon geben also gerade die reellen linearen Gleichungssysteme $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$ mit nichtnegativem $\mathfrak{A}$.

Wir schließen diesen Abschnitt mit einigen kritischen Bemerkungen zur Beurteilung und zum Vergleich des asymptotischen Konvergenzverhaltens bei den oben untersuchten intervallmäßigen Iterationsverfahren; die Beziehungen (2) und (3) beim Gesamt- und Einzelschrittverfahren haben die allgemeine Gestalt

$$(q(X_{i\nu}, X_i^*)) \leq |\mathfrak{B}|^\nu (q(X_{i0}, X_i)), \tag{1}$$

und die darin enthaltenen Abschätzungen gelten für alle Anfangsvektoren $\varphi_0 \in V_n(I(R))$. Diese Abschätzungen sind deshalb im Einzelfall oft recht grob; trotzdem kann man sie global nicht verbessern. Das asymptotische Konvergenzverhalten kann auf Grund einer solchen Beziehung zwar mit dem Spektralradius $\varrho(|\mathfrak{B}|)$ global für alle $\varphi_0 \in V_n(I(R))$ abgeschätzt werden, aber $\varrho(|\mathfrak{B}|)$ beschreibt das asymptotische Konvergenzverhalten im Einzelfall nicht immer genau.

Als Beispiel dafür betrachten wir etwa den Spezialfall eines reellen linearen Gleichungssystems $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$ und $\varrho(|\mathfrak{A}|) < 1$. Für Iteration mit Intervallvektoren muß wegen $\varrho((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|)) \leq \varrho(|\mathfrak{A}|)$ (siehe oben) das Einzelschrittverfahren als günstiger angesehen werden als das Gesamtschrittverfahren; trotzdem kann für Iteration mit Vektoren aus der Teilmenge $V_n(R) \subset V_n(I(R))$ gerade das Gegenteil der Fall sein.

Bei Iteration mit Vektoren aus $V_n(R)$ gelten für das Gesamt- bzw. Einzelschrittverfahren die Beziehungen

$$(\varphi_\nu - \varphi^*) = \mathfrak{A}^\nu(\varphi_0 - \varphi^*) \quad \text{bzw.} \tag{2R}$$

$$(\varphi_\nu - \varphi^*) = ((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R}))^\nu (\varphi_0 - \varphi^*), \tag{3R}$$

und die Spektralradien $\varrho(\mathfrak{A})$ bzw. $\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R}))$ beschreiben das jeweilige asymptotische Konvergenzverhalten.

Aus den Beziehungen (2) und (3), die global das Konvergenzverhalten bei Iteration mit Vektoren aus $V_n(I(R))$ beschreiben, erhalten wir für Iteration mit Vektoren aus der Teilmenge $V_n(R)$ (für $A = a$, $B = b \in R$ wird $q(A,B) = |a - b|$):

$$|\varphi_\nu - \varphi^*| \leq |\mathfrak{A}|^\nu |\varphi_0 - \varphi^*| \quad \text{bzw.} \tag{2'}$$

$$|\varphi_\nu - \varphi^*| \leq ((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|))^\nu |\varphi_0 - \varphi^*|. \tag{3'}$$

(2') bzw. (3') erhält man aber auch aus (2R) bzw. (3R), wenn man dort alle Komponenten durch ihre Absolutbeträge abschätzt. Das Konvergenzverhalten beim Gesamt- und Einzelschrittverfahren wird also durch die Beziehungen (2') und (3') im Vergleich zu (2R) und (3R) nur grob abgeschätzt. Entsprechend wird auch das asymptotische Konvergenzverhalten durch die Spektralradien $\varrho(|\mathfrak{A}|)$ und $\varrho((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|))$ jeweils nur grob abgeschätzt, und erst durch $\varrho(\mathfrak{A})$ bzw. $\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R}))$ genauer beschrieben.

Es ist $\qquad \varrho(\mathfrak{A}) \leq \varrho(|\mathfrak{A}|) \qquad (6)$
(siehe Ende von Abschnitt 6.1)

und $\qquad \varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})) \leq \varrho((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|)); \qquad (7)$

Ungleichung (7) ergibt sich folgendermaßen:

Zunächst ist nach (6) $\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})) \leq \varrho(|(\mathfrak{E} - \mathfrak{L})^{-1}((\mathfrak{D} + \mathfrak{R})|)$.
Wegen $|(\mathfrak{E} - \mathfrak{L})^{-1}| = |\sum_{\nu=0}^{n} \mathfrak{L}^{\nu}| \leq \sum_{\nu=0}^{n} |\mathfrak{L}|^{\nu} = (\mathfrak{E} - |\mathfrak{L}|)^{-1}$
ist $\quad |(\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})| \leq ((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|);$

Nach der Perron-Frobenius-Theorie (siehe etwa [14]) folgt daraus $\quad \varrho(|(\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})|) \leq \varrho((\mathfrak{E} - |\mathfrak{L}|)^{-1}(|\mathfrak{D}| + |\mathfrak{R}|))$.

(7) folgt jetzt aus den beiden Ungleichungen für die Spektralradien.

Auf Grund von (6) und (7) schließen nun die Ungleichungen

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

und
$$\varrho((\mathfrak{E} - \mathfrak{L})^{-1}(\mathfrak{D} + \mathfrak{R})) > \varrho(\mathfrak{A})$$

einander nicht aus. Das bedeutet: Obwohl für Iteration mit Vektoren aus $V_n(I(R))$ global das Einzelschrittverfahren als günstiger anzusehen ist, ist nicht auszuschließen, daß für Iteration mit Vektoren aus der Teilmenge $V_n(R)$ das Gesamtschrittverfahren günstiger ist.

<u>Bemerkung:</u> Divergiert bei Iteration mit Vektoren aus $V_n(R)$ das Gesamt- oder Einzelschrittverfahren, so divergieren bei Iteration mit Vektoren aus $V_n(I(R))$ beide Verfahren; denn nach Satz 6.4 konvergieren bzw divergieren beide Verfahren bei Iteration mit Vektoren aus $V_n(I(R))$ stets gleichzeitig, und ein Verfahren kann nicht bei Iteration mit Vektoren aus $V_n(I(R))$ konvergieren, wenn es schon bei Iteration mit Vektoren aus der Teilmenge $V_n(R)$ divergiert.

7. Über die Bestimmung von Einschließungsmengen für die Lösungen reeller linearer Gleichungssysteme mit nichtkonstanten Koeffizienten

7.1 Allgemeine Überlegungen

Die Intervallrechnung entstand aus dem Bedürfnis, die bei der numerischen Behandlung mathematischer Probleme auf digitalen Rechenanlagen unvermeidbaren Rundungsfehler unter Kontrolle zu bringen, und als Ergebnis einer numerischen Rechnung nicht nur irgendwelche Näherungswerte, sondern sichere Scranken für die exakte Lösung des gegebenen Problems zu bestimmen. Dies setzt ein Rechnen mit speziellen Mengen reeller Zahlen, nämlich abgeschlossenen Intervallen voraus, das auf der Maschine geeignet zu approximieren ist (siehe dazu [1], [12], [15]).

Die Intervallrechnung bietet nun darüberhinaus die Möglichkeit bei Problemen, deren Ausgangsdaten nicht exakt vorliegen, sondern innerhalb vorgegebener Grenzen streuen, Einschließungsschranken für die Ergebnisse zu bestimmen. Wir wollen dies an dem von uns betrachteten Problem der Gleichungsauflösung deutlich machen (vergleiche dazu [8]):

Wir gehen aus von der Aufgabe ein reelles lineares Gleichungssystem $\mathfrak{C}\mathfrak{x} = \mathfrak{b}$ mit $\mathfrak{C} = (c_{ij}) \in M_n(R)$, $\mathfrak{b} = (b_i) \in V_n(R)$ zu lösen und nehmen an, daß die Größen c_{ij} und b_i aus irgendwelchen Gründen nicht exakt vorliegen oder genau bekannt sind, sondern beliebige Werte aus vorgegebenen Intervallen C_{ij} und B_i annehmen können. Wir suchen einen Vektor $\mathfrak{x}^* = (X_i) \in V_n(I(R))$ mit minimalen Komponentenlängen, der alle Lösungsvektoren der Gleichungen $\mathfrak{C}\mathfrak{x} = \mathfrak{b}$ für $\mathfrak{C} = (c_{ij}) \in \boldsymbol{\mathfrak{C}} := (C_{ij})$ und $\mathfrak{b} = (b_i) \in \boldsymbol{\mathfrak{b}} := (B_i)$ enthält.

Wir wollen das Problem noch etwas allgemeiner stellen:
Im Gleichungssystem $\mathfrak{C}\mathfrak{x} = \mathfrak{b}$ seien die Komponenten der Matrix $\mathfrak{C}$ und des Vektors $\mathfrak{b}$ Funktionen einer Variablen t, die in einem gewissen Bereich T variiert, also $\mathfrak{C}(t) = (c_{ij}(t))$, $\mathfrak{b}(t) = (b_i(t))$.

Für jedes $t \in T$ sei das reelle lineare Gleichungssystem $\mathfrak{A}(t)\mathfrak{x} = \mathfrak{b}(t)$ lösbar. Mit vernünftigem Aufwand ist es jedoch oft nicht möglich, numerisch die Lösung des Gleichungssystems als Funktion von t zu ermitteln; deshalb suchen wir wieder nach einem Vektor $\mathfrak{x}^* = (X_i) \in V_n(I(R))$ mit minimalen Komponentenlängen, der für jedes $t \in T$ die Lösung enthält.

Da wir diesen Vektor $\mathfrak{x}^*$ durch Iteration bestimmen wollen, gehen wir im folgenden immer von einer iterationsfähigen Gestalt $\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ des Gleichungssystems aus.

Die grundlegenden Überlegungen über den Zusammenhang des Rechnens mit Vektoren und Matrizen, deren Komponenten Funktionen eines Parameters t sind, und der Intervallrechnung für Vektoren und Matrizen findet man in [3] und [8]; wir erläutern hier kurz die wesentlichen Begriffe soweit, wie sie für den Zusammenhang mit den in dieser Arbeit gewonnenen Ergebnissen von Bedeutung sind.

T sei eine abgeschlossene, beschränkte, zusammenhängende Teilmenge des R^m (m beliebig), die von einer Variablen t durchlaufen wird.

$(T,R)_c$ bezeichne die Menge der stetigen Abbildungen von T in R; Elemente aus $(T,R)_c$ bezeichnen wir mit $a_T, b_T, \ldots\ldots$.

$V_n(T,R)_c$ bezeichne die Menge der n-tupel (Vektoren) mit Komponenten aus $(T,R)_c$; für Elemente aus $V_n(T,R)_c$ schreiben wir

$$\mathfrak{x}_T = (x_T^{\,i}) \ .$$

$M_n(T,R)_c$ bezeichne die Menge der quadratischen $n \times n$-Matrizen mit Komponenten aus $(T,R)_c$; für Elemente aus $M_n(T,R)_c$ schreiben wir

$$\mathfrak{A}_T = (a_T^{\,ij}) \ .$$

Wir <u>definieren</u>: Zwischen Elementen $a_T, b_T, \ldots.$ aus $(T,R)_c$ bestehe Gleichheit oder irgendeine von R her bekannte Relation (etwa "$\leq$"), wenn diese Relation für jedes $t \in T$ zwischen

den reellen Zahlen $a_T(t), b_T(t), \ldots\ldots$ besteht. Entsprechend werden auch die Verknüpfungen "$+,-,\cdot,:$" für Elemente aus $(T,R)_c$ erklärt.

Damit erhält man z.B.:

$$\begin{array}{llll} x_T = y_T : & \bigwedge \; x_T(t) = y_T(t) & \text{für alle} & t \in T \\ x_T \leq y_T : & \bigwedge \; x_T(t) \leq y_T(t) & \text{für alle} & t \in T \\ (x_T + y_T)(t) & := x_T(t) + y_T(t) & \text{für alle} & t \in T \end{array}$$

Die Verknüpfungen für Vektoren und Matrizen aus $V_n(T,R)_c$ und $M_n(I,R)_c$ und alle Relationen werden wie üblich über die Komponenten definiert und so auf Verknüpfungen bzw. Relationen zurückgeführt.

Es sei $f_T \in (T,R)_c$;
die Menge $\{f_T\} := \{f_T(t) \mid t \in T\}$ bezeichnen wir als Komplex von f_T .

Nach Voraussetzung ist $f_T(t)$ stetig auf der abgeschlossenen, beschränkten, zusammenhängenden Teilmenge $T \subset R^m$; deshalb ist $\{f_T\} \in I(R)$.

Für Vektoren $\mathfrak{x}_T = (x_T^{\,i}) \in V_n(T,R)_c$ heißt

$\{\mathfrak{x}_T\} := \{(x_T^{\,i}(t)) \mid t \in T\}$ Komplex von $\mathfrak{x}_T$;

$\overline{\mathfrak{x}_T} := (\{x_T^{\,i}(t)\})$ Hülle von $\mathfrak{x}_T$;

$\{\mathfrak{x}_T\}$ ist eine Menge von Vektoren aus $V_n(R)$;

$\overline{\mathfrak{x}_T}$ ist ein Intervallvektor aus $V_n(I(R))$; $\overline{\mathfrak{x}_T}$ ist Durchschnitt aller Intervallvektoren $\mathfrak{x} \in V_n(I(R))$ mit der Eigenschaft $\mathfrak{x} \supset \{\mathfrak{x}_T\}$.

Für Matrizen $\mathfrak{A}_T = (a_T^{\,ij}) \in M_n(T,R)_c$ heißt entsprechend

$\{\mathfrak{A}_T\} := \{(a_T^{ij}(t)) |\ t \in T\}$ Komplex von $\mathfrak{A}_T$;
$\overline{\mathfrak{A}_T} := (\{a_T^{ij}(t)\})$ Hülle von $\mathfrak{A}_T$.

$\{\mathfrak{A}_T\}$ ist eine Menge von Matrizen aus $M_n(R)$;
$\overline{\mathfrak{A}_T}$ ist eine Intervallmatrix aus $M_n(I(R))$; $\overline{\mathfrak{A}_T}$ ist Durchschnitt aller Intervallmatrizen $\mathfrak{A} \in M_n(I(R))$ mit der Eigenschaft $\mathfrak{A} \supset \{\mathfrak{A}_T\}$.

Zur vorgegebenen $\mathfrak{A}_T \in M_n(T,R)_c$, $\mathfrak{b}_T \in V_n(T,R)_c$ betrachten wir die Gleichung

$$\varphi_T = \mathfrak{A}_T \cdot \varphi_T + \mathfrak{b}_T \ .$$

Gesucht ist die Lösung $\varphi_T^* \in V_n(T,R)_c$.

Ist nun $\mathfrak{E} - \mathfrak{A}_T(t)$ nicht singulär für alle $t \in T$, so hat die Gleichung die eindeutige Lösung $\varphi_T^* = (\mathfrak{E} - \mathfrak{A}_T)^{-1} \cdot \mathfrak{b}_T$. Diese Bedingung für die Existenz einer eindeutigen Lösung φ_T^* ist nun natürlich in der Praxis i.a. schwer nachprüfbar; sie ist jedoch erfüllt, wenn eine multiplikative Matrixnorm $\|\mathfrak{M}\|$ in $M_n(I(R))$ existiert, für die $\|\overline{\mathfrak{A}_T}\| < 1$.

Dann ist nämlich nach 6.3 auch $\rho(|\overline{\mathfrak{A}_T}|) < 1$, und für alle $t \in T$ gilt jetzt:

$$\rho(\mathfrak{A}_T(t)) \le \rho(|\mathfrak{A}_T(t)|) \le \rho(|\overline{\mathfrak{A}_T}|) < 1 \ .$$

(Die erste Ungleichung hier wurde schon in Abschnitt 6.1 benutzt, sie folgt aus [14], Lemma 2.3; die zweite Ungleichung folgt wegen $\mathfrak{A}_T(t) \in \overline{\mathfrak{A}_T}$ also $|\mathfrak{A}_T(t)| \le |\overline{\mathfrak{A}_T}|$ aus [14], Theorem 2.7 .)

Mit $\rho(\mathfrak{A}_T(t)) < 1$ ist jetzt aber $\mathfrak{E} - \mathfrak{A}_T(t)$ nicht singulär.

Nur in Ausnahmefällen läßt sich nun die Lösung φ_T^* unter vertretbarem Aufwand als Funktion von t bestimmen, deshalb muß man sich i.a. damit begnügen, in $V_n(I(R))$ Abschätzungen für den Komplex $\{\varphi_T\}$ und die Hülle $\overline{\varphi_T}$ anzugeben. Dabei sind aber nur

Abschätzungen interessant, die die gesuchte Lösung φ_T^* mit Sicherheit einschließen. Eine solche Abschätzung in $V_n(I(R))$ liefert nun - falls vorhanden - die Lösung φ^* der Gleichung $\varphi = \overline{\mathfrak{A}_T} \cdot \varphi + \overline{\mathfrak{b}_T}$; dies ergibt sich aus dem folgenden

Satz 7.1: Gegeben seien $\mathfrak{A}_T \in M_n(T,R)_c$, $\mathfrak{b}_T \in V_n(T,R)_c$ und es sei $\mathfrak{E} - \mathfrak{A}_T(t)$ nichtsingulär für alle $t \in T$; die Gleichung $\varphi = \overline{\mathfrak{A}_T}\varphi + \overline{\mathfrak{b}_T}$ habe die Lösung $\varphi^* \in V_n(I,R)$; dann hat die Gleichung $\varphi_T = \mathfrak{A}_T \varphi_T + \mathfrak{b}_T$ die eindeutige Lösung $\varphi_T = (\mathfrak{E} - \mathfrak{A}_T)^{-1}\mathfrak{b}_T$ und für diese gilt:

$$\{\varphi_T^*\} \subset \varphi^* \quad \text{und} \quad \overline{\varphi_T^*} \subset \varphi^* .$$

(Ähnliche Aussagen wurden unter der schärferen Voraussetzung einer konvergenten Iterationsfolge $\varphi_{\nu+1} = \overline{\mathfrak{A}_T}\varphi_\nu + \overline{\mathfrak{b}_T}$ und auf andere Weise schon in [8] bewiesen.)

Beweis: Daß φ_T die angegebene Gestalt hat, wurde bereits gezeigt. Wir betrachten jetzt ein festes $t \in T$ und zeigen dafür $\varphi_T(t) \in \varphi^*$.

$\mathfrak{A}_T(t)$ ist eine Matrix aus $M_n(R)$ mit $\mathfrak{A}_T(t) \in \overline{\mathfrak{A}_T}$,
$\mathfrak{b}_T(t)$ ist ein Vektor aus $V_n(R)$ mit $\mathfrak{b}_T(t) \in \overline{\mathfrak{b}_T}$,
φ sei ein Vektor aus $V_n(R)$ mit $\varphi \in \varphi^*$.

Wegen $\varphi^* = \overline{\mathfrak{A}_T} \cdot \varphi^* + \overline{\mathfrak{b}_T}$ folgt aus der Teilmengeneigenschaft

$$\varphi^* \ni \mathfrak{A}_T(t)\varphi + \mathfrak{b}_T(t) \ .$$

$\{\varphi \in V_n(R) | \varphi \in \varphi^*\}$ ist eine kompakte, konvexe Teilmenge des $V_n(R)$, die durch die stetige Abbildung $\varphi \to \mathfrak{A}_T(t)\varphi + \mathfrak{b}_T(t)$ in sich abgebildet wird; nach dem Brouwerschen Fixpunktsatz enthält also die Menge $\{\varphi \in V_n(R) | \varphi \in \varphi^*\}$ mindestens einen Fixpunkt dieser Abbildung, d.h. mindestens eine Lösung des reellen linearen Gleichungssystems

$$\text{(RL)} \qquad \varphi = \mathfrak{A}_T(t)\varphi + \mathfrak{b}_T(t) \ .$$

Nach Voraussetzung ist aber $\mathfrak{E} - \mathfrak{A}_T(t)$ nichtsingulär, deshalb hat (RL) genau eine Lösung $\mathfrak{x}_T^*(t)$; für diese gilt deshalb

$$\mathfrak{x}_T^*(t) \in \{\mathfrak{x} \in V_n(R) \mid \mathfrak{x} \in \mathfrak{x}^*\};$$

also ist $\mathfrak{x}_T^*(t) \in \mathfrak{x}^*$ für alle $t \in T$,

d.h. $\{\mathfrak{x}_T^*\} = \{\mathfrak{x}_T^*(t) \mid t \in T\} \subset \mathfrak{x}^*$.

Die Hülle $\overline{\mathfrak{x}_T}$ ist Durchschnitt aller Intervallvektoren $\mathfrak{x}$ mit der Eigenschaft $\mathfrak{x} \supset \{\mathfrak{x}_T^*\}$; deshalb ist auch $\overline{\mathfrak{x}_T^*} \subset \mathfrak{x}^*$. □

<u>Bemerkung</u>: Im allgemeinen wird $\mathfrak{x}^*$ eine echte Obermenge von $\overline{\mathfrak{x}_T}$.

Wir betrachten dazu ein einfaches Beispiel im Fall $n = 1$:

<u>Beispiel 1</u>: In der Gleichung $x_T = a_T \cdot x_T + b_T$ sei

$T := ([-0.5,-0.25] \times [1.5,3])$, $t := (t_1,t_2) \in T$ und

$a_T(t) := t_1$, $b_T(t) := t_2$;

damit wird $x_T(t) = \dfrac{t_2}{1-t_1}$, also

$$\{x_T\} = \overline{x_T} = \left\{\frac{t_2}{1-t_1} \;\middle|\; t_1 \in [-0.5,-0.25] \wedge t_2 \in [1.5,3]\right\} = [1,2.4] .$$

Es ist $\overline{a_T} = [-0.5,-0.25]$, $\overline{b_T} = [1.5,3]$, und die Gleichung

$X = \overline{a_T} \cdot X + \overline{b_T}$ hat die Lösung $X^* = [0,3]$.

(Die Eindeutigkeit dieser Lösung folgt wegen $|\overline{a_T}| = \frac{1}{2} < 1$ aus Satz 6.1 .)

X^* ist also echte Obermenge von $\overline{x_T}$.

Es gibt jedoch auch einige wichtige Fälle mit spezieller Gestalt von $\mathfrak{A}_T$ und $\mathfrak{b}_T$, für die $\overline{\mathfrak{x}_T} = \mathfrak{x}^*$. Die Lösung der Gleichung $\mathfrak{x} = \overline{\mathfrak{A}_T}\,\mathfrak{x} + \overline{\mathfrak{b}_T}$ gibt in diesen Fällen die bestmöglichen Einschließungsschranken für die Lösung $\mathfrak{x}_T$ der Gleichung $\mathfrak{x}_T = \mathfrak{A}_T \cdot \mathfrak{x}_T + \mathfrak{b}_T$.

Eine erste Klasse von Paaren solcher $\mathfrak{A}_T$ und $\mathfrak{b}_T$ gibt uns der folgende

<u>Satz 7.2:</u> Es seien $\mathfrak{A} \in M_n(I(R))$ mit $\mathfrak{A} = (A_{ik}) \geq 0$ und $\rho(|\mathfrak{A}|) < 1$,
$\mathfrak{b} \in V_n(I(R))$ mit $\mathfrak{b} = (B_i) \geq 0$,

$T := A_{11} \times A_{12} \times \ldots \times A_{nn} \times B_1 \times B_2 \times \ldots \times B_n$,

$t := (t_{11}, t_{12}, \quad , t_{nn}, t_1, t_2, \ldots, t_n) \in T$,

$\mathfrak{A}_T \in M_n(T,R)_c$, $\mathfrak{b}_T \in V_n(T,R)_c$ mit

$\mathfrak{A}_T(t) := (t_{ij})$, $\mathfrak{b}_T(t) := (t_i)$,

$\mathfrak{x}^* \in V_n(I(R))$ Lösung von $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$,

$\mathfrak{x}_T^* \in V_n(T,R)_c$ Lösung von $\mathfrak{x}_T = \mathfrak{A}_T \cdot \mathfrak{x}_T + \mathfrak{b}_T$;

dann ist $\overline{\mathfrak{x}_T} = \mathfrak{x}^*$.

Dieser Satz ergibt sich unmittelbar aus dem Satz 8.1 der Arbeit [3], wenn man den hier eingeführten Begriff der Lösung $\mathfrak{x}^*$ von $\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ benutzt.

Wir erklären zunächst einige für die Formulierung des Beweises nötige Begriffe in der folgenden

<u>Definition 7.1</u>: Zu $D := [d_1,d_2] \in I(R)$ sei

$$\inf(D) := d_1 , \qquad \sup(D) := d_2 ;$$

entsprechend sei für $\mathfrak{x} = (X_i) \in V_n(I(R))$ bzw. $\mathfrak{A} = (A_{ik}) \in M_n(I(R))$

$$\inf(\mathfrak{x}) := (\inf(X_i)) , \qquad \sup(\mathfrak{x}) := (\sup(X_i))$$

bzw. $\inf(\mathfrak{A}) := (\inf(A_{ik}))$, $\sup(\mathfrak{A}) := (\sup(A_{ik}))$.

Es ist $\inf(D), \sup(D) \in R$,

$\inf(\mathfrak{x}), \sup(\mathfrak{x}) \in V_n(R)$ bzw. $\inf(\mathfrak{A}), \sup(\mathfrak{A}) \in M_n(R)$.

Bei Zugrundelegung der natürlichen Halbordnung in $V_n(R)$ ist

$$\inf(\mathfrak{x}) = \inf_{\mathfrak{y} \in \mathfrak{x}} \mathfrak{y} , \qquad \sup(\mathfrak{x}) = \sup_{\mathfrak{y} \in \mathfrak{x}} \mathfrak{y} .$$

Entsprechendes gilt für Matrizen.

Für $\mathfrak{x} \in V_n(I(R))$ bzw. $\mathfrak{A} \in M_n(I(R))$ schreiben wir in Verallgemeinerung der Intervallschreibweise bzw. im Anschluß an Abschnitt 3.4:

$$\mathfrak{x} = [\inf(\mathfrak{x}),\ \sup(\mathfrak{x})] \quad \text{bzw.} \quad \mathfrak{A} = [\inf(\mathfrak{A}),\ \sup(\mathfrak{A})] \ .$$

Beweis von Satz 7.2: Für $\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ schreiben wir

$$[\inf(\mathfrak{x}),\ \sup(\mathfrak{x})] = [\inf(\mathfrak{A}),\ \sup(\mathfrak{A})]\ [\inf(\mathfrak{x}),\ \sup(\mathfrak{x})] + [\inf(\mathfrak{b}),\ \sup(\mathfrak{b})] \ ;$$

daraus erhalten wir wegen $\mathfrak{A} \geq 0$, $\mathfrak{b} \geq 0$ und nach Definition der auftretenden Verknüpfungen "$\cdot$" und "+" :

$$\inf(\mathfrak{x}) = \inf(\mathfrak{A})\cdot\inf(\mathfrak{x}) + \inf(\mathfrak{b}) \qquad (1)$$

und

$$\sup(\mathfrak{x}) = \sup(\mathfrak{A})\ \sup(\mathfrak{x}) + \sup(\mathfrak{b}) \ , \qquad (2)$$

d.h. der Fixpunkt $\mathfrak{x}^* = [\inf(\mathfrak{x}^*),\ \sup(\mathfrak{x}^*)]$ kann hier durch Lösen der beiden reellen linearen Gleichungssysteme (1) und (2) bestimmt werden.

Wegen $\inf(\mathfrak{A}) \in \mathfrak{A}$, $\inf(\mathfrak{b}) \in \mathfrak{b}$ und $\sup(\mathfrak{A}) \in \mathfrak{A}$, $\sup(\mathfrak{b}) \in \mathfrak{b}$ ist nun

$$\inf(\mathfrak{x}^*) \in \overline{\mathfrak{x}_T^*} \qquad \text{und} \qquad \sup(\mathfrak{x}^*) \in \overline{\mathfrak{x}_T^*} \ .$$

Wegen $\mathfrak{x}^*, \overline{\mathfrak{x}_T^*} \in V_n(I(R))$ folgt daraus $\mathfrak{x}^* \subset \overline{\mathfrak{x}_T^*}$;
nach Satz 7.1 gilt immer $\mathfrak{x}^* \supset \overline{\mathfrak{x}_T^*}$;
also ist $\mathfrak{x}^* = \overline{\mathfrak{x}_T^*}$. □

Wir geben auch für diesen in Satz 7.2 beschriebenen Fall ein einfaches Beispiel:

Beispiel 2: In der Gleichung $x_T = a_T\, x_T + b_T$ sei $T := ([0.25,0.5] \times [1.5,3])$, $t := (t_1,t_2) \in T$ und $a_T(t) := t_1$, $b_T(t) := t_2$.

damit wird $x_T^*(t) = \dfrac{t_2}{1-t_1}$, also

$$\overline{x_T^*} = \{x_T^*\} = \left\{ \frac{t_2}{1-t_1} \;\middle|\; t_1 \in [0.25,0.5] \wedge t_2 \in [1.5,3] \right\} = [2,6] \ .$$

Es ist $\overline{a_T} = [0.25,0.5]$, $\overline{b_T} = [1.5,3]$ und die Gleichung

$X = \overline{a_T}\cdot X + \overline{b_T}$ hat die Lösung $X^* = [2,6]$.

Wie nach Satz 7.2 zu erwarten, ist $\overline{x_T^*} = X^*$.

Wir untersuchen eine weitere Klasse von Paaren $\mathfrak{A}_T$ und $\mathfrak{b}_T$, für die (im Sinne von Satz 7.1) $\overline{\mathfrak{x}_T^*} = \mathfrak{x}^*$.

Dazu benötigen wir den folgenden

Hilfssatz 7.3: Es sei $\mathfrak{A} \in M_n(I(R))$ mit $\varrho(|\mathfrak{A}|) < 1$, $\mathfrak{b} \in V_n(I(R))$ und $\mathfrak{b}$ symmetrisch, dann hat die Gleichung

$\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ die symmetrische Lösung $\mathfrak{x}^* = (\mathfrak{E} - |\mathfrak{A}|)^{-1}\cdot\mathfrak{b}$.

Beweis: Aus der Voraussetzung $\varrho(|\mathfrak{A}|) < 1$ folgt nach Satz 6.1, daß die Gleichung genau eine Lösung $\mathfrak{x}^*$ besitzt, und daß die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\,\mathfrak{x}_\nu + \mathfrak{b}$ für jeden Anfangsvektor $\mathfrak{x}_0 \in V_n(I(R))$ gegen $\mathfrak{x}^*$ konvergiert. $\mathfrak{x}^*$ ist nun symmetrisch; das erkennt man folgendermaßen:

Geht man aus von einem symmetrischen Anfangsvektor $\mathfrak{x}_0$, so wird zunächst $\mathfrak{x}_1$ symmetrisch (nach Hilfssatz 4.7 ist mit $\mathfrak{x}_0$ auch $\mathfrak{A}\,\mathfrak{x}_0$ symmetrisch; mit $\mathfrak{x}_0$ und $\mathfrak{b}$ ist auch $\mathfrak{x}_1 = \mathfrak{A}\,\mathfrak{x}_0 + \mathfrak{b}$ symmetrisch;); vollständige Induktion ergibt die Symmetrie für alle Iterierten $\mathfrak{x}_\nu$; $\mathfrak{x}^*$ ist nun als Grenzwert eine Folge symmetrischer Vektoren ebenfalls symmetrisch.

Es sei $\mathfrak{b} = ([-b_i,b_i])$, also $b_i \geq 0$ für alle i;
$\mathfrak{x}^*$ bestimmen wir aus dem Ansatz

$$\mathfrak{x}^* = ([-x_i,x_i]) \quad \text{mit} \quad x_i \geq 0 \quad \text{für alle } i \ .$$

Das ergibt

$$([-x_i,x_i]) = \mathfrak{A}\,([-x_i,x_i]) + ([-b_i,b_i]) \ .$$

Daraus entnehmen wir für den nichtnegativen reellen Vektor (x_i) das folgende Gleichungssystem

$$(R) \qquad (x_i) = |\mathfrak{A}|(x_i) + (b_i) .$$

Nach [14], Theorem 3.8 gilt nun:

$$\varrho(|\mathfrak{A}|) < 1 \Leftrightarrow \mathfrak{E} - |\mathfrak{A}| \text{ nichtsingulär } \wedge (\mathfrak{E}-|\mathfrak{A}|)^{-1} \geq 0 .$$

(R) hat also die eindeutige Lösung $(x_i) = (\mathfrak{E}-|\mathfrak{A}|)^{-1}(b_i)$;
nach dem eben zitierten Theorem ist $(x_i) \geq 0$,
damit erhalten wir

$$\mathfrak{x}^* = ([-x_i,x_i]) = (\mathfrak{E}-|\mathfrak{A}|)^{-1}([-b_i,b_i]) = (\mathfrak{E}-|\mathfrak{A}|)^{-1}\cdot\mathfrak{b} \ . \ \square$$

Bemerkung: Für $\varrho(|\mathfrak{A}|) \geq 1$ ist nach dem oben zitierten Satz $(\mathfrak{E}-|\mathfrak{A}|)^{-1} \not\geq 0$. Der angegebene Ansatz führt deshalb im Falle $\varrho(|\mathfrak{A}|) \geq 1$ nur dann zu einem sinnvollen Ergebnis, wenn zufällig $(\mathfrak{E}-|\mathfrak{A}|)^{-1}(b_i)$ ein nichtnegativer Vektor wird. Wir geben dafür ein einfaches Beispiel:

Das Gleichungssystem $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ hat für

$$\mathfrak{A} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \text{ (hier ist } \varrho(|\mathfrak{A}|)=2) \text{ und } \quad \mathfrak{b} = \begin{pmatrix} [-a,a] \\ [-b,b] \\ 0 \\ 0 \end{pmatrix}$$

(mit $a,b \geq 0$) die Lösung $\mathfrak{x}^* = \mathfrak{b}$.

Satz 7.4: Es sei $\mathfrak{A} \in M_n(R)$ mit $\mathfrak{A} \geq 0$ und $\varrho(\mathfrak{A}) < 1$,
$\mathfrak{b} = (B_i) \in V_n(I(R))$ und $\mathfrak{b}$ symmetrisch ,
$T := (B_1 \times B_2 \times \ldots\ldots \times B_n)$, $t = (t_1,t_2,\ldots,t_n) \in T$,
$\mathfrak{A}_T \in M_n(T,R)_c$ mit $\mathfrak{A}_T = \mathfrak{A}$
$\mathfrak{b}_T \in V_n(T,R)_c$ mit $\mathfrak{b}_T(t) := (b^i_T(t)) := (t_i)$;
$\mathfrak{x}^*$ Lösung von $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$
$\mathfrak{x}^*_T$ Lösung von $\mathfrak{x}_T = \mathfrak{A}_T \cdot \mathfrak{x}_T + \mathfrak{b}_T$

dann ist $\overline{\mathfrak{x}^*_T} = \mathfrak{x}^*$.

Beweis: Es ist $\mathfrak{x}_T = (\mathfrak{E}-\mathfrak{A}_T)^{-1}\mathfrak{b}_T = (\mathfrak{E}-\mathfrak{A})^{-1}\mathfrak{b}_T$.
Wegen der speziellen Gestalt von $\mathfrak{b}_T$ (in der i-ten Komponente tritt nur die Variable t_i auf), wird

$$\overline{\mathfrak{x}^*_T} = \overline{(\mathfrak{E}-\mathfrak{A})^{-1}\mathfrak{b}_T} = (\mathfrak{E}-\mathfrak{A})^{-1}\overline{\mathfrak{b}_T} \qquad (1)$$

Nach Hilfssatz 7.3 ist $\mathfrak{x}^* = (\mathfrak{E}-|\mathfrak{A}|)^{-1}\mathfrak{b}$. (2)

Mit $\overline{\mathfrak{b}_T} = \mathfrak{b}$ und $\mathfrak{A} = |\mathfrak{A}|$ folgt aus (1) und (2): $\overline{\mathfrak{x}^*_T} = \mathfrak{x}^*$. $\square$

Auch für den im Satz 7.4 beschriebenen Fall geben wir ein einfaches Beispiel:

Beispiel 3: Wir betrachten zunächst die Gleichung $X=AX+B$; diese hat für $B = [-8,8]$ und $A = \pm 0,2$ die Lösung $X^* = [-10,10]$.

Die Gleichung $x_T = Ax_T + b_T$ mit $T := B$, $t \in T$ und $b_T := t$ hat die Lösung

$$x_T = \frac{t}{1 \mp 0,2}$$

Im Falle $A = 0,2$ wird $\overline{x^*_T} = [-10,10] = X^*$;
im Falle $A = -0,2$ wird $\overline{x^*_T} = \left[-6\tfrac{2}{3}, 6\tfrac{2}{3}\right] \neq X^*$.

7.2 Abschätzungen für die Normen und Durchmesser der Lösung $\mathfrak{x}^*$ der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$, sowie der Hülle $\overline{\mathfrak{x}_T}$ zur Lösung $\mathfrak{x}^*_T$ der Gleichung $\mathfrak{x}_T = \mathfrak{A}_T \cdot \mathfrak{x}_T + \mathfrak{b}_T$

Satz 7.5: Es seien $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$, $\|\mathfrak{M}\|$ und $\|\mathfrak{w}\|$ verträgliche monotone Namen in $M_n(I(R))$ bzw. $V_n(I(R))$ mit $\|\mathfrak{A}\| < 1$;
dann gilt für die Lösung $\mathfrak{x}^*$ der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$:

a) $\|\mathfrak{x}^*\| \leq \dfrac{\|\mathfrak{b}\|}{1 - \|\mathfrak{A}\|}$;

b) $d(\mathfrak{x}^*) \leq \dfrac{d(\mathfrak{A})\,\|\mathfrak{x}^*\| + d(\mathfrak{b})}{1 - \|\mathfrak{A}\|} \leq \dfrac{d(\mathfrak{A})\,\|\mathfrak{b}\|}{(1-\|\mathfrak{A}\|)^2} + \dfrac{d(\mathfrak{b})}{1-\|\mathfrak{A}\|}$.

<u>Beweis:</u> a) $\left.\mathfrak{x}^* = \mathfrak{A}\mathfrak{x}^* + \mathfrak{b}\right\} \quad \|\mathfrak{x}^*\| \le \|\mathfrak{A}\|\,\|\mathfrak{x}^*\| + \|\mathfrak{b}\|$;

wegen $\|\mathfrak{A}\| < 1$ folgt daraus $\|\mathfrak{x}^*\| \le \dfrac{\|\mathfrak{b}\|}{1-\|\mathfrak{A}\|}$.

b) $d(\mathfrak{x}^*) \le d(\mathfrak{A}\mathfrak{x}^*+\mathfrak{b}) = \|\mathfrak{A}\mathfrak{x}^*+\mathfrak{b} - \mathfrak{A}\mathfrak{x}^* - \mathfrak{b}\| \le \|\mathfrak{A}\mathfrak{x}^* - \mathfrak{A}\mathfrak{x}^*\| + \|\mathfrak{b}-\mathfrak{b}\| =$

$= d(\mathfrak{A}\mathfrak{x}^*) + d(\mathfrak{b}) \le d(\mathfrak{A})\|\mathfrak{x}^*\| + \|\mathfrak{A}\|\,d(\mathfrak{x}^*) + d(\mathfrak{b})$ (nach Satz 5.6);

wegen $\|\mathfrak{A}\| < 1$ folgt daraus $d(\mathfrak{x}^*) \le \dfrac{d(\mathfrak{A})\|\mathfrak{x}^*\| + d(\mathfrak{b})}{1-\|\mathfrak{A}\|}$;

setzt man hier die unter a) gewonnene Abschätzung für $\|\mathfrak{x}^*\|$ ein, so erhält man die zweite Ungleichung in b) . □

<u>Bemerkung:</u> a) $\|\mathfrak{x}^*\|$ kann den Wert $\dfrac{\|\mathfrak{b}\|}{1-\|\mathfrak{A}\|}$ annehmen.

<u>Beispiel:</u> Es sei $\mathfrak{A} = (a_{ik})$ mit $0 \le a_{ik} := \alpha < \frac{1}{n}$ für alle i, k.
$\mathfrak{b} = (B_i)$ mit $B_i = [-b, b]$ und $b \ge 0$ für alle i.

dann hat $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$

die Lösung $\mathfrak{x}^* = (X_i^*)$ mit $X_i^* = \left[\dfrac{-b}{1-n\alpha}, \dfrac{b}{1-n\alpha}\right]$ für alle i .

Für die zugeordneten Namen $\|(A_{ie})\| := \max\limits_i \sum\limits_e |A_{ie}|$ und
$\|(X_i)\| = \max\limits_i |X_i|$

wird $\|\mathfrak{x}^*\| = \dfrac{b}{1-n\alpha}$, $\|\mathfrak{b}\| = b$ und $\|\mathfrak{A}\| = n\alpha$.

b) Das oben angegebene Beispiel zeigt auch, daß $d(\mathfrak{x}^*)$ für $\mathfrak{A} \in M_n(R)$, also $d(\mathfrak{A}) = 0$ den Wert
$$d(\mathfrak{x}^*) = \frac{d(\mathfrak{b})}{1-\|\mathfrak{A}\|}$$
annehmen kann.
Die Abschätzung für $d(\mathfrak{x}^*)$ ist nun scharf in dem Sinne, daß sie für $d(\mathfrak{A}) = d(\mathfrak{b}) = 0$ auch $d(\mathfrak{x}^*) = 0$ liefert.

<u>Korollar:</u> Es seien $\mathfrak{A}_T \in M_n(T,R)_c$, $\mathfrak{b}_T \in V_n(T,R)_c$
$\mathfrak{x}_T^*$ Lösung der Gleichung $\mathfrak{x}_T = \mathfrak{A}_T\mathfrak{x}_T + \mathfrak{b}_T$
$\mathfrak{x}^*$ Lösung der Gleichung $\mathfrak{x} = \overline{\mathfrak{A}_T}\cdot\mathfrak{x} + \overline{\mathfrak{b}_T}$.

$\|\mathfrak{m}\|$ und $\|\mathfrak{v}\|$ seien seien verträgliche monotone Normen in $M_n(I(R))$ bzw. $V_n(I(R))$ mit $\|\overline{\mathfrak{A}_T}\| < 1$; dann gilt:

a) $\|\overline{\mathfrak{x}_T^*}\| \le \|\mathfrak{x}^*\| \le \dfrac{\|\overline{\mathfrak{b}_T}\|}{1-\|\overline{\mathfrak{A}_T}\|}$

b) $d(\overline{\mathfrak{x}_T^*}) \le d(\mathfrak{x}^*) \le \dfrac{d(\overline{\mathfrak{A}_T})\|\overline{\mathfrak{b}_T}\|}{1-\|\overline{\mathfrak{A}_T}\|^2} + \dfrac{d(\overline{\mathfrak{b}_T})}{1-\|\overline{\mathfrak{A}_T}\|}$

c) $\|\overline{\mathfrak{x}_T^*} - \mathfrak{x}^*\| \le \dfrac{d(\overline{\mathfrak{A}_T})\|\overline{\mathfrak{b}_T}\|}{(1-\|\overline{\mathfrak{A}_T}\|)^2} + \dfrac{d(\overline{\mathfrak{b}_T})}{1-\|\overline{\mathfrak{A}_T}\|}$.

<u>Bemerkung:</u> Für $\|\overline{\mathfrak{x}_T^*}\|$ und $d(\overline{\mathfrak{x}_T^*})$ werden dieselben Abschätzungen jedoch ohne Zwischenschaltung von $\|\mathfrak{x}^*\|$ und $d(\mathfrak{x}^*)$ schon in [8] angegeben. Wir haben hier in Satz 7.5 gezeigt, daß auch Norm und Durchmesser der umfassenderen Menge $\mathfrak{x}^*$ (nach Satz 7.1 ist ja $\mathfrak{x}^* \supset \overline{\mathfrak{x}_T^*}$) durch dieselben Schranken abgeschätzt werden können.

Unserer Abschätzung für $\|\overline{\mathfrak{x}_T^*} - \mathfrak{x}^*\|$ entspricht in [8] eine Abschätzung für $\limsup\limits_{\nu\to\infty} \|\overline{\mathfrak{x}_T^*} - \mathfrak{y}_\nu\|$ durch dieselbe Schranke; dabei sind die $\mathfrak{y}_\nu$ definiert durch die Iteration $\mathfrak{y}_{\nu+1} = \overline{\mathfrak{A}_T}\cdot\mathfrak{y}_\nu + \overline{\mathfrak{b}_T}$ bei beliebigem $\mathfrak{y}_0 \in V_n(I(R))$. In Satz 6.1 haben wir hier bewiesen, daß diese Iteration für beliebige $\mathfrak{y}_0 \in V_n(I(R))$ gegen $\mathfrak{x}^*$ konvergiert; mit $\lim\limits_{\nu\to\infty} \mathfrak{y}_\nu = \mathfrak{x}^*$ wird aber $\limsup\limits_{\nu\to\infty} \|\overline{\mathfrak{x}_T^*} - \mathfrak{y}_\nu\| = \|\overline{\mathfrak{x}_T^*} - \mathfrak{x}^*\|$.

<u>Beweis:</u> Nach Satz 7.1 ist $\overline{\mathfrak{x}_T^*} \subset \mathfrak{x}^*$, damit wird nach der Teilmengeneigenschaft auch $\overline{\mathfrak{x}_T^*} - \overline{\mathfrak{x}_T^*} \subset \mathfrak{x}^* - \mathfrak{x}^*$ und $\overline{\mathfrak{x}_T^*} - \mathfrak{x}^* \subset \mathfrak{x}^* - \mathfrak{x}^*$.

Monotone Normen sind nach Satz 4.4 auch Abstandsnormen; mit Satz 4.2 ergibt sich jetzt:

a) $\|\overline{\mathfrak{x}_T^*}\| \le \|\mathfrak{x}^*\|$

b) $d(\overline{\mathfrak{x}_T^*}) = \|\overline{\mathfrak{x}_T^*} - \overline{\mathfrak{x}_T^*}\| \le \|\mathfrak{x}^* - \mathfrak{x}^*\| = d(\mathfrak{x}^*)$

c) $\|\overline{\mathfrak{x}_T^*} - \mathfrak{x}^*\| \le \|\mathfrak{x}^* - \mathfrak{x}^*\| = d(\mathfrak{x}^*)$.

Für $\|\mathfrak{x}^*\|$ und $d(\mathfrak{x}^*)$ setzen wir jetzt die Abschätzungen aus dem eben bewiesenen Satz ein. □

Bemerkungen: $\|\overline{\varphi_T^*}\|$ kann den Wert $\dfrac{\|\overline{b_T}\|}{1-\|\overline{\mathfrak{A}_T}\|}$ annehmen;

für $d(\overline{\mathfrak{A}_T}) = 0$ können $d(\overline{\varphi_T^*})$ und $\|\overline{\varphi_T^*} - \varphi^*\|$ den Wert $\dfrac{d(\overline{b_T})}{1-\|\overline{\mathfrak{A}_T}\|}$ annehmen.

Beispiel: Für $\mathfrak{A} = (a_{ik}) \in M_n(R)$ mit $0 \leq a_{ik} = \alpha < \frac{1}{n}$ für alle i,k und $b = (B_i) \in V_n(I(R))$ mit $B_i = [-b,b]$ für alle i wurden in der Bemerkung zu Satz 7.5 schon Normen angegeben, für die

$$\|\varphi^*\| = \frac{\|b\|}{1-\|\mathfrak{A}\|} \quad \text{und} \quad d(\varphi^*) = \|\varphi^* - \varphi^*\| = \frac{d(b)}{1-\|\mathfrak{A}\|} \quad .$$

Jetzt sei $T := B_1 \times B_2 \ \ldots \times B_n$, $t := (t_1, t_2, \ldots t_n) \in T$,

$\mathfrak{A}_T := \mathfrak{A}$ und $b_T := (t_i)$; dann ist

$\overline{\mathfrak{A}_T} = \mathfrak{A}$, $\overline{b_T} = b$ und nach Satz 7.4 gilt hier $\overline{\varphi_T^*} = \varphi^*$.

8. Iteration mit Einschließungsmengen bei der Lösung des Gleichungssystems $\varphi = \mathfrak{A}\varphi + b$

8.1 Allgemeine Überlegungen

Im vorhergehenden Abschnitt haben wir zum Gleichungssystem $\varphi_T = \mathfrak{A}_T \varphi_T + b_T$ (gegeben: $\mathfrak{A}_T \in M_n(T,R)_c$ und $b_T \in V_n(T,R)_c$, gesucht: Lösung $\varphi_T \in V_n(T,R)_c$) noch das Gleichungssystem $\varphi = \overline{\mathfrak{A}_T}\,\varphi + \overline{b_T}$ mit der Lösung $\varphi^* \in V_n(I(R))$ betrachtet.

Sind nämlich beide Systeme lösbar (für $\varrho(|\overline{\mathfrak{A}_T}|) < 1$ sind beide sogar eindeutig lösbar) so liefert φ^* eine Abschätzung für den Komplex und die Hülle von φ_T; nach Satz 7.1 gilt nämlich die Einschließungseigenschaft $\{\varphi_T^*\} \subset \varphi^*$ und $\overline{\varphi_T^*} \subset \varphi^*$; mit $\{\varphi_T^*\} \subset \varphi^*$ ist hier auch $\overline{\varphi_T^*} \subset \varphi^*$ und umgekehrt.

Im Falle $\varrho(|\overline{\mathfrak{A}_T}|) < 1$ kann φ^* nach Satz 6.1 durch die Iteration $\varphi_{\nu+1} = \overline{\mathfrak{A}_T}\,\varphi_\nu + \overline{b_T}$ bestimmt werden.

Diese Iteration soll nun auf einer Rechenanlage durchgeführt werden; dort ist aber nur eine endliche Teilmenge R_M von reellen Zahlen, also auch nur eine endliche Teilmenge $I(R_M)$ der Intervalle aus $I(R)$ vorhanden; die Elemente von $I(R_M)$ werden durch die sogenannte Maschinenintervallarithmetik verknüpft, die die Verknüpfungen in $I(R)$ approximiert, dabei ist jedoch die Abbildung der Menge $I(R)$ auf die Menge $I(R_M)$ der Maschinenintervalle bezüglich der Verknüpfungen in $I(R)$ und $I(R_M)$ im allgemeinen nicht homomorph (siehe dazu [1]) .

Die nach der Rechenvorschrift $\varphi_{\nu+1} = \overline{\mathfrak{A}_T}\,\varphi_\nu + \overline{b_T}$ in einer vorgegebenen Maschinenintervallarithmetik berechnete Approximation φ_M^* für den theoretisch exakten Fixpunkt φ^* stimmt deshalb im allgemeinen nicht mit φ^* überein und kann auch noch vom Anfangsvektor φ_o abhängen. Dann enthält aber φ_M möglicherweise auch nicht mehr die Lösung φ_T . Dabei interessiert uns die Lösung φ^* des Gleichungssystems $\varphi = \overline{\mathfrak{A}_T}\,\varphi + \overline{b_T}$

mit Intervallkoeffizienten doch vor allem deshalb, weil sie Einschließungsmenge für die Lösungen der in ihm "enthaltenen" reellen linearen Gleichungssysteme ist (siehe Satz 7.1). Unser Ziel muß deshalb sein, die Rechnung auf der Maschine so zu führen, daß die dabei ermittelte Näherung $\mathfrak{x}^*_M$ für die exakte Lösung $\mathfrak{x}^*$ Obermenge der exakten Lösung $\mathfrak{x}^*$ wird und sie dabei noch möglichst gut approximiert.

Um die Einschließungseigenschaft $\mathfrak{x}^*_M \supset \mathfrak{x}^*$ zu erhalten, wird es im wesentlichen genügen, die Iteration mit Einschließungsmengen durchzuführen. Die Güte der Approximation von $\mathfrak{x}^*$ durch $\mathfrak{x}^*_M$ hängt natürlich weitgehend von den Eigenschaften der verwendeten Maschinenintervallarithmetik ab.

Aus den Überlegungen zu diesen Fragen ergibt sich dann auch als neue Erkenntnis, daß die Maschinenintervallarithmetik die (von der exakten Intervallrechnung über dem Körper der reellen Zahlen her bekannte) Teilmengeneigenschaft besitzen muß, wenn in ihr für die Inklusion Monotonieeigenschaften wie in der exakten Intervallrechnung gelten sollen. Wir gehen darauf weiter unten ein.

Zur Vorbereitung bringen wir zunächst zwei Sätze über die Iteration mit Einschließungsmengen:

<u>Satz 8.1</u>: Das Gleichungssystem $\mathfrak{x}_T = \mathfrak{A}_T \cdot \mathfrak{x}_T + \mathfrak{b}_T$ habe die Lösung $\mathfrak{x}^*_T$; hat nun der Anfangsvektor $\mathfrak{x}_0$ der Iteration $\mathfrak{x}_{\nu+1} = \overline{\mathfrak{A}_T} \cdot \mathfrak{x}_\nu + \overline{\mathfrak{b}_T}$ die Einschließungseigenschaft $\mathfrak{x}_0 \supset \overline{\mathfrak{x}^*_T}$, so ist $\mathfrak{x}_\nu \supset \overline{\mathfrak{x}^*_T}$ für alle ν.

Dieser Satz wird schon in [8] angegeben; der <u>Beweis</u> verwendet vollständige Induktion:

Es sei $\mathfrak{x}_n \supset \overline{\mathfrak{x}^*_T}$, dann folgt aus der Teilmengeneigenschaft

$$\mathfrak{x}_{n+1} = \overline{\mathfrak{A}_T} \cdot \mathfrak{x}_n + \overline{\mathfrak{b}_T} \supset \overline{\mathfrak{A}_T} \cdot \overline{\mathfrak{x}^*_T} + \overline{\mathfrak{b}_T} \supset \overline{\mathfrak{A}_T \mathfrak{x}^*_T + \mathfrak{b}_T} = \overline{\mathfrak{x}^*_T}$$

(die zweite Inklusion folgt unmittelbar aus der Definition der Hüllenbildung). □

Satz 8.1 drückt aus, daß die zunächst für den Anfangsvektor $\mathfrak{x}_0$ geltende Einschließungseigenschaft $\mathfrak{x}_0 \supset \overline{\mathfrak{x}^*_T}$ bei der Iteration auf alle $\mathfrak{x}_\nu$ übergeht; im Falle der Konvergenz gilt deshalb auch für das Grenzelement $\mathfrak{x}^* := \lim_{\nu \to \infty} \mathfrak{x}_\nu$: $\mathfrak{x}^* \supset \overline{\mathfrak{x}^*_T}$. Damit erhält man übrigens unter der schärferen Voraussetzung der Konvergenz der Iterationsfolge $\mathfrak{x}_{\nu+1} = \overline{\mathfrak{A}_T}\,\mathfrak{x}_\nu + \overline{\mathfrak{b}_T}$ einen anderen Beweis für Satz 7.1 .
(Die Iteration konvergiert nämlich jetzt nach Voraussetzung für jeden Anfangsvektor, also insbesondere auch für jeden $\overline{\mathfrak{x}^*_T}$ enthaltenden Anfangsvektor gegen $\mathfrak{x}^*$, deshalb ist $\mathfrak{x}^* \supset \overline{\mathfrak{x}^*_T}$.)

Existiert nun $\mathfrak{x}^* = \lim_{\nu \to \infty} \mathfrak{x}_\nu$, dann läßt sich die Einschließungseigenschaft $\mathfrak{x}_0 \supset \overline{\mathfrak{x}^*_T}$ wegen $\mathfrak{x}^* \supset \overline{\mathfrak{x}^*_T}$ (siehe Satz 7.1) immer durch die Einschließungseigenschaft $\mathfrak{x}_0 \supset \mathfrak{x}^*$ erfüllen. Der folgende Satz gibt Auskunft über das Verhalten dieser Einschließungseigenschaft bei der betrachteten Iteration; wir setzen dabei $\mathfrak{A} := \overline{\mathfrak{A}_T}$, $\mathfrak{b} := \overline{\mathfrak{b}_T}$.

<u>Satz 8.2</u>: Die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ habe genau eine Lösung $\mathfrak{x}^* = \mathfrak{A}\mathfrak{x}^* + \mathfrak{b}$; dann gilt für die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit dem Anfangsvektor $\mathfrak{x}_0 \in V_n(I(R))$:

a) Aus $\mathfrak{x}_0 \supset \mathfrak{x}^*$ folgt $\mathfrak{x}_\nu \supset \mathfrak{x}^*$ für alle ν .

b) Aus $\mathfrak{x}_0 \supset \mathfrak{x}_1$ folgt $\mathfrak{x}_\nu \supset \mathfrak{x}_{\nu+1} \supset \mathfrak{x}^*$ für alle ν , und $\lim_{\nu \to \infty} \mathfrak{x}_\nu = \mathfrak{x}^*$.

(Aussage b) von Satz 8.2 ist bereits in ähnlichen Sätzen in [8] enthalten, auch wenn dort der Begriff der Lösung für derartige Gleichungssysteme nicht verwendet wird.)

<u>Beweis</u>: a) Wir verwenden vollständige Induktion; aus $\mathfrak{x}_n \supset \mathfrak{x}^*$ folgt nach der Teilmengeneigenschaft

$$\mathfrak{x}_{n+1} = \mathfrak{A}\mathfrak{x}_n + \mathfrak{b} \supset \mathfrak{A}\mathfrak{x}^* + \mathfrak{b} = \mathfrak{x}^* .$$

b) für den Schluß von $\varphi_0 \supset \varphi_1$ auf $\varphi_\nu \supset \varphi_{\nu+1}$ für alle ν verwenden wir wieder vollständige Induktion; aus $\varphi_n \supset \varphi_{n+1}$ folgt nach der Teilmengeneigenschaft

$$\varphi_{n+1} = \mathfrak{A}\varphi_n + \mathfrak{b} \supset \mathfrak{A}\varphi_{n+1} + \mathfrak{b} = \varphi_{n+2}.$$

Also ist $\varphi_\nu \supset \varphi_{\nu+1}$ für alle ν, d.h. die φ_ν bilden eine absteigende Mengenfolge; damit existiert auch $\lim\limits_{\nu\to\infty} \varphi_\nu$ und es gilt

$$\varphi_\nu \supset \lim_{\nu\to\infty} \varphi_\nu \quad \text{für alle } \nu. \tag{1}$$

$\lim\limits_{\nu\to\infty} \varphi_\nu$ ist nun als Fixpunkt der Iteration $\varphi_{\nu+1} = \mathfrak{A}\varphi_\nu + \mathfrak{b}$ auch Lösung der Gleichung $\varphi = \mathfrak{A}\varphi + \mathfrak{b}$; wegen der Eindeutigkeit der Lösung ist jetzt $\varphi^* = \lim\limits_{\nu\to\infty} \varphi_\nu$.

und mit (1) gilt: $\varphi_\nu \supset \varphi^*$ für alle ν. □

Damit ist der Satz bewiesen.

<u>Bemerkung</u>: Die Aussagen von Satz 8.2 sind unabhängig von der speziellen Gestalt der Iterationsvorschrift; wie der Beweis zeigt, gelten die Aussagen für jedes in $V_n(I(R))$ erklärte Iterationsverfahren $\varphi_{n+1} = \bar{A}(\varphi_n)$, wenn der das Iterationsverfahren definierende Operator $\bar{A}$ genau einen Fixpunkt $\varphi^* = \bar{A}(\varphi^*)$ in $V_n(I(R))$ besitzt.

Die Aussagen der Sätze 8.1 und 8.2 gelten in der über dem Körper der reellen Zahlen aufgebauten exakten Intervallarithmetik. Wir wollen jetzt untersuchen, welche Eigenschaften eine Maschinenintervallarithmetik besitzen muß, damit diese Aussagen auch für die auf der Rechenanlage berechneten Iterationsvektoren φ_{ν_M} gelten.

Zunächst ist jede Maschinenintervallarithmetik eine Approximation der über dem Körper der reellen Zahlen aufgebauten Intervallarithmetik (diese enthält natürlich auch die Arithmetik für reelle Zahlen). Von der Aufgabe der Intervallrechnung her er-

gibt sich nun als grundsätzliche Forderung an jede Maschinenintervallarithmetik, daß alle in der Maschinenintervallarithmetik berechneten Ergebnisintervalle stets die entsprechenden der exakten Intervallarithmetik enthalten; diese Forderung erfüllt man in einfacher Weise dadurch, daß man bei jeder Approximation von Elementen aus $I(R)$ durch Maschinenintervalle aus $I(R_M)$ und bei allen Verknüpfungen in $I(R_M)$ grundsätzlich nach außen rundet (siehe dazu [1]). Damit überträgt sich auch hier jede Einschließungseigenschaft $\varphi_0 \supset \varphi^*$ bzw. $\varphi_0 \supset \overline{\varphi_T^*}$ zunächst auf die Darstellung φ_{0M} von φ_0 auf der Maschine und dann auf die von φ_{0M} ausgehend in der Maschinenintervallarithmetik nach der gegebenen Iterationsvorschrift berechneten Iterationsvektoren $\varphi_{\nu M}$; sie gilt so schließlich auch für das eventuell vorhandene Grenzelement φ_M^*.

In Satz 8.2 b folgt der Schluß von $\varphi_0 \supset \varphi_1$ auf $\varphi_\nu \supset \varphi_{\nu+1}$ für alle ν aus der Teilmengeneigenschaft der exakten Intervallarithmetik.

Ein entsprechender Schluß ist für die auf der Rechenanlage nach der obigen Vorschrift berechneten $\varphi_{\nu M}$ nur dann richtig, wenn auch die Maschinenintervallarithmetik die Teilmengeneigenschaft hat. Diese Teilmengeneigenschaft drückt aus, daß die Rundung (nach außen !) noch gewisse Monotonieeigenschaften bezüglich der Inklusion hat.

<u>Definition 8.1</u>: Wir sagen eine in der Menge $I(R_M)$ der Maschinenintervalle definierte Maschinenintervallarithmetik mit den Verknüpfungen $\oplus, \ominus, \odot, \oslash$ (diese approximieren die entsprechenden Verknüpfungen in $I(R)$) hat die Teilmengeneigenschaft, wenn folgendes gilt:

Sind A_M, B_M, C_M, D_M Maschinenintervalle aus $I(R_M)$, und bedeutet $\otimes$ eine der Verknüpfungen in $I(R_M)$, dann folgt aus $A_M \supset C_M$ und $B_M \supset D_M$

stets $A_M \otimes B_M \supset C_M \otimes D_M$.

Diese Teilmengeneigenschaft der Maschinenintervallarithmetik entspricht genau der Teilmengeneigenschaft der exakten Intervallarithmetik in der Menge $\mathbf{I}(R)$ (siehe Abschnitt 1) und gibt ihr auch entsprechende Eigenschaften. Für die praktische Anwendung der Intervallrechnung ist dies von großer Bedeutung. Jeder Algorithmus wird nämlich in der exakten Intervallarithmetik, also im Raum $\mathbf{I}(R)$ hergeleitet und alle hier im Algorithmus bestehenden Inklusionsbeziehungen bestehen nur dann auch in seiner Realisierung auf der Rechenanlage, wenn die zugrundegelegte Maschinenintervallarithmetik die Teilmengeneigenschaft hat.

Wir wenden jetzt diese Ergebnisse an auf das eingangs gestellte Problem für die exakte Lösung $\mathfrak{x}^*$ der Gleichung $\mathfrak{x} = \mathfrak{A}\,\mathfrak{x} + \mathfrak{b}$ (mit $\mathfrak{A} \in M_n(\mathbf{I}(R))$, $\mathfrak{b} \in V_n(\mathbf{I}(R))$) eine möglichst gute Approximation $\mathfrak{x}_M^*$ mit der Einschließungseigenschaft $\mathfrak{x}_M^* \supset \mathfrak{x}^*$ zu bestimmen. Die Lösung $\mathfrak{x}^*$ sei als eindeutig vorausgesetzt; $\mathfrak{x}_M^*$ soll auf der Rechenanlage durch Iteration bestimmt werden:

Wir betrachten zunächst die Iterationsvorschrift

$$\text{(I)} \qquad \mathfrak{x}_{\nu+1\,M} := \mathfrak{A}_M \odot \mathfrak{x}_{\nu M} \oplus \mathfrak{b}_M$$

($\mathfrak{A}_M$ und $\mathfrak{b}_M$ seien die Darstellungen von $\mathfrak{A}$ und $\mathfrak{b}$ in der Maschine, $\odot$ und $\oplus$ die oben in Definition 8.1 eingeführten Verknüpfungen in der Maschinenintervallarithmetik, die die entsprechenden Verknüpfungen der exakten Intervallarithmetik approximieren). Wir nehmen an, daß auf der Rechenanlage ein Anfangsvektor $\mathfrak{x}_{oM}$ mit der Einschließungseigenschaft $\mathfrak{x}_{oM} \supset \mathfrak{x}^*$ zur Verfügung steht.

Wie wir oben sahen, gilt in jeder Maschinenintervallarithmetik für die nach der Vorschrift (I) berechneten Iterierten $\mathfrak{x}_{\nu M}$ die Satz 8.2 a entsprechende Aussage, d.h.

aus $\mathfrak{x}_{oM} \supset \mathfrak{x}^*$ folgt $\mathfrak{x}_{\nu M} \supset \mathfrak{x}^*$ für alle ν . (1)

Konvergiert jetzt die von $\mathfrak{x}_{oM}$ ausgehende Iteration in der zugrundegelegten Maschinenintervallarithmetik gegen ein Element $\mathfrak{x}'_M$, so wird auch $\mathfrak{x}'_M \supset \mathfrak{x}^*$.

Durch eine einfache Abwandlung der Iterationsvorschrift läßt sich nun das Konvergenzverhalten der Iteration wesentlich verbessern und zugleich ein sicheres Kriterium für den Abbruch der Iteration gewinnen; die Iterationsvorschrift laute jetzt:

$$\text{(D)} \qquad \mathfrak{y}_{\nu+1\,M} := \mathfrak{A}_M \odot \mathfrak{y}_{\nu M} \oplus \mathfrak{b}_M \cap \mathfrak{y}_{\nu M} \;; \quad \text{dabei sei } \mathfrak{y}_{oM} := \mathfrak{x}_{oM}.$$

Aus $\mathfrak{y}_{nM} \supset \mathfrak{x}^*$ folgt zunächst nach denselben Überlegungen, die auf (1) führten $\mathfrak{A}_M \odot \mathfrak{y}_{nM} \oplus \mathfrak{b}_M \supset \mathfrak{x}^*$, und wir erhalten insgesamt $\mathfrak{y}_{n+1M} \supset \mathfrak{x}^*$.

Damit gilt jetzt auch für die Iterationsvorschrift (D):

Aus $\mathfrak{y}_{oM} \supset \mathfrak{x}^*$ folgt $\mathfrak{y}_{\nu M} \supset \mathfrak{x}^*$ für alle ν .

Zusätzlich gilt noch:

$$\mathfrak{y}_{oM} \supset \mathfrak{y}_{1M} \supset \mathfrak{y}_{2M} \supset \ldots \supset \mathfrak{y}_{\nu M} \supset \mathfrak{y}_{\nu+1M} \ldots .$$

Da nun auf jeder Rechenanlage nur eine endliche Teilmenge der reellen Zahlen zur Verfügung steht, muß diese Iteration nach endlich vielen Schritten auf der Stelle treten. Die Durchschnittsbildung im Verfahren (D) erzwingt also immer Konvergenz; zusätzlich erhält man ein einwandfreies Abbrechkriterium: Die Iteration wird abgebrochen, sobald zwei aufeinanderfolgende Iterationsvektoren gleich sind.

Den letzten Iterationsvektor bezeichnen wir mit $\mathfrak{x}_M^*$. Ersichtlich gilt $\mathfrak{x}_M^* \supset \mathfrak{x}^*$; $\mathfrak{x}_M^*$ gibt jetzt die bei vorgegebenem Anfangsvektor $\mathfrak{x}_{oM}$ in der vorgegebenen Maschinenintervallarithmetik bestmöglichen Schranken für $\mathfrak{x}^*$.

Das Hauptproblem besteht also jetzt darin, geeignete Anfangsvektoren $\mathfrak{x}_{oM} = \mathfrak{y}_{oM}$ mit der Einschließungseigenschaft $\mathfrak{x}_{oM} \supset \mathfrak{x}^*$ zu bestimmen.

Bei theoretischen Überlegungen zur Begründung von Verfahren für die Konstruktion solcher Einschließungsmengen $\mathfrak{x}_{oM}$ ist folgende Bemerkung oft nützlich:

Ist $\mathfrak{x}_o \in V_n(I(R))$ mit $\mathfrak{x}_o \supset \mathfrak{x}^*$ und $\mathfrak{x}_{oM}$ die Darstellung von $\mathfrak{x}_o$ in der Maschinenintervallarithmetik, so ist (wegen der Rundung nach außen) $\mathfrak{x}_{oM} \supset \mathfrak{x}_o \supset \mathfrak{x}^*$.

Einen wichtigen Hinweis gibt jetzt insbesondere die folgende Aussage aus Satz 8.2 b:

Gilt bei der Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ für einen Vektor $\mathfrak{x}_o \in V_n(I(R))$: $\mathfrak{x}_o \supset \mathfrak{x}_1 := \mathfrak{A}\mathfrak{x}_o + \mathfrak{b}$, so ist $\mathfrak{x}_o \supset \mathfrak{x}_1 \supset \mathfrak{x}^*$. (2)

Diese Abbildung eines Vektors $\mathfrak{x}_o$ in sich muß nun i.a. auch auf der Maschine nachgeprüft werden; hier gilt der folgende

Satz 8.3: Es sei $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b}$, $\mathfrak{x}_o \in V_n(I(R))$ und die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ habe die eindeutige Lösung $\mathfrak{x}^*$. Weiter seien $\mathfrak{A}_M$, $\mathfrak{b}_M$ und $\mathfrak{x}_{oM}$ die Darstellungen von $\mathfrak{A}$, $\mathfrak{b}$ und $\mathfrak{x}_o$ in der vorgegebenen Maschinenintervallarithmetik.

Gilt nun $\quad \mathfrak{x}_{oM} \supset \mathfrak{x}_{1M} := \mathfrak{A}_M \odot \mathfrak{x}_{oM} \oplus \mathfrak{b}_M$,

so ist $\quad \mathfrak{x}_{oM} \supset \mathfrak{x}_{1M} \supset \mathfrak{x}^*$.

Beweis: Es ist $\mathfrak{A}, \mathfrak{A}_M, \in M_n(I(R))$; $\mathfrak{b}, \mathfrak{b}_M, \mathfrak{x}_o, \mathfrak{x}_{oM} \in V_n(I(R))$; da bei jeder Darstellung reeller Intervalle durch Maschinenintervalle nach außen gerundet wird, gilt:

$$\mathfrak{A} \subset \mathfrak{A}_M \qquad \mathfrak{b} \subset \mathfrak{b}_M$$

Aus der Teilmengeneigenschaft und der Tatsache, daß auch bei jeder Verknüpfung "$\circledast$" von Maschinenintervallen gegenüber

der entsprechenden Verknüpfung "$*$" in der exakten Intervallarithmetik nach außen gerundet wird (sind $A, B \in I(R)$ und A_M, B_M die Darstellungen von A und B in der Maschinenintervallarithmetik, so sind auch $A_M, B_M \in I(R)$ und es gilt: $A \subset A_M$, $B \subset B_M$ und $A * B \subset A_M \circledast B_M$, siehe dazu [1]), folgt jetzt:

$$\mathfrak{A}\mathfrak{x}_{oM} + \mathfrak{b}_M \subset \mathfrak{A}_M \odot \mathfrak{x}_{oM} \oplus \mathfrak{b}_M = \mathfrak{x}_{1M} \subset \mathfrak{x}_{oM},$$

d.h. der Vektor $\mathfrak{x}_{oM} \in V_n(I(R))$ wird durch die Abbildung $\mathfrak{x} \to \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ in sich abgebildet.

Mit (2) folgt daraus: $\quad \mathfrak{x}_{oM} \supset \mathfrak{x}^*$

mit (1) folgt noch: $\quad \mathfrak{x}_{1M} \supset \mathfrak{x}^*$. □

Korollar: Wie oben sei $\mathfrak{x}^*$ die eindeutige Lösung der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$.

Hat die auf der Rechenanlage zugrundegelegte Maschinenintervallarithmetik die Teilmengeneigenschaft, so gilt auch für die in der Maschinenintervallarithmetik durchgeführte Iteration (I) die Satz 8.2 b entsprechende Aussage:

Aus $\mathfrak{x}_{oM} \supset \mathfrak{x}_{1M}$ folgt $\mathfrak{x}_{\nu M} \supset \mathfrak{x}_{\nu+1 M} \supset \mathfrak{x}^*$ für alle ν .

Beweis: Der Schluß von $\mathfrak{x}_{oM} \supset \mathfrak{x}_{1M}$ auf $\mathfrak{x}_{\nu M} \supset \mathfrak{x}_{\nu+1 M}$ für alle ν folgt aus der Teilmengeneigenschaft der Maschinenintervallarithmetik; der Beweis entspricht genau dem für die analoge Aussage von Satz 8.2 b . Nach Satz 8.3 gilt hier $\mathfrak{x}_{oM} \supset \mathfrak{x}^*$, nach (1) (siehe Seite 102) ist also auch $\mathfrak{x}_{\nu M} \supset \mathfrak{x}^*$ für alle ν. □

Bemerkung: Im eben betrachteten Fall bilden die Iterationsvektoren $\mathfrak{x}_{\nu M}$ schon bei Zugrundelegung der Iterationsvorschrift (I) eine absteigende Mengenfolge:

$$\mathfrak{x}_{oM} \supset \mathfrak{x}_{1M} \supset \mathfrak{x}_{2M} \supset \dots\dots$$

Die Iterationsvorschriften (I) und (D) werden also identisch; und wir finden dasselbe Konvergenzverhalten wie oben unter (D) beschrieben:
Die Iteration tritt nach endlich vielen Schritten auf der Stelle; wir brechen die Iteration wieder ab, sobald zwei aufeinanderfolgende Iterationsvektoren gleich sind.

Wir begründen und beschreiben anschließend in Abschnitt 8.2 ein Verfahren zur Bestimmung einer Einschließungsmenge für die exakte Lösung $\mathfrak{x}^*$ einer Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$, das darauf beruht, daß ein Vektor $\mathfrak{x}_o \in V_n(I(R))$ ermittelt wird, der bei der Abbildung $\mathfrak{x} \longrightarrow \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ in sich abgebildet wird.

Im folgenden werden noch kurz einige für die Herleitung und Begründung dieses Verfahrens nötige Begriffe zusammengestellt bzw. definiert.

In Definition 7.1 haben wir zu $D = [d_1, d_2] \in I(R)$ $\inf(D)$ und $\sup(D)$ erklärt durch $\inf(D) := d_1$, $\sup(D) := d_2$.

Entsprechend haben wir zu $\mathfrak{a} \in V_n(I(R))$ $\inf(\mathfrak{a})$ und $\sup(\mathfrak{a})$ erklärt; $\inf(\mathfrak{a})$ und $\sup(\mathfrak{a})$ sind Vektoren aus $V_n(R)$, deren Komponenten die unteren bzw. oberen Intervallenden der Komponenten von $\mathfrak{a}$ sind. Es ist $\inf(\mathfrak{a}) \leq \sup(\mathfrak{a})$.

Für Matrizen $\mathfrak{A} \in M_n(I(R))$ haben wir $\inf(\mathfrak{A})$ und $\sup(\mathfrak{A})$ analog erklärt.

Für $\mathfrak{a}, \mathfrak{b} \in V_n(I(R))$ gilt nach Definition der Relation "$\leq$" in $V_n(I(R))$ (siehe Abschnitt 3.2):

$$\mathfrak{a} \leq \mathfrak{b} \Leftrightarrow \inf(\mathfrak{a}) \leq \inf(\mathfrak{b}) \wedge \sup(\mathfrak{a}) \leq \sup(\mathfrak{b}).$$

Definition 8.2: Unter dem Mittelpunkt eines Intervallvektors $\mathfrak{x} = (X_i) \in V_n(I(R))$ verstehen wir den Vektor

$$\underset{\sim}{\mathfrak{x}} := \frac{1}{2}\,(\inf(\mathfrak{x}) + \sup(\mathfrak{x}))$$

Entsprechend definieren wir als Mittelpunkt einer Intervallmatrix $\mathfrak{A} = (A_{ik}) \in M_n(I(R))$ die Matrix

$$\underset{\sim}{\mathfrak{A}} := \frac{1}{2}\,(\inf(\mathfrak{A}) + \sup(\mathfrak{A})) .$$

Es ist $\underset{\sim}{\mathfrak{x}} \in V_n(R)$, $\underset{\sim}{\mathfrak{A}} \in M_n(R)$.

Definition 8.3: Zu $\mathfrak{a}, \mathfrak{b} \in V_n(I(R))$ mit $\mathfrak{a} \leq \mathfrak{b}$ bezeichne $[\mathfrak{a}, \mathfrak{b}]$ die konvexe Hülle von $\mathfrak{a}$ und $\mathfrak{b}$ in $V_n(I(R))$.

Es sei $\mathfrak{x} = (X_i)$, $\mathfrak{y} = (Y_i)$, dann ist $[\mathfrak{x}, \mathfrak{y}] = ([\inf(X_i), \sup(Y_i)])$

Definition 8.4: Zwei Matrizen $\mathfrak{A}_1 = (A^1_{ik})$ und $\mathfrak{A}_2 = (A^2_{ik})$ aus $M_n(I(R))$ heißen indexfremd, wenn für jedes Indexpaar (i,k) gilt: $A^1_{ik} \cdot A^2_{ik} = 0$, (also höchstens eines der beiden Intervalle A^1_{ik} und A^2_{ik} ungleich Null ist).

$\mathfrak{A} \in M_n(I(R))$ heißt indexfremd, wenn es als Summe $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ zweier indexfremder Matrizen $\mathfrak{A}_1 \geq 0$ und $\mathfrak{A}_2 \leq 0$ darstellbar ist.

Wir bemerken noch, daß sich jede Matrix $\mathfrak{A} \in M_n(I(R))$ eindeutig darstellen läßt als Summe $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ zweier Matrizen $\mathfrak{A}_1 \geq 0$ und $\mathfrak{A}_2 \leq 0$.

Nach den Überlegungen von Abschnitt 3.2 existieren in $M_n(I(R))$ bezüglich der natürlichen Halbordnung zu zwei Elementen stets Infimum und Supremum (d.h. $M_n(I(R))$ ist ein Verband).

Die eben beschriebene Zerlegung ist nun gegeben durch

$$\mathfrak{A} = \sup(\mathfrak{A}, \Theta) + \inf(\mathfrak{A}, \Theta) .$$

8.2 Ein Verfahren zur Bestimmung einer Einschließungsmenge für die Lösung der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$

Vorgegeben sei die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$. Ihre Lösung $\mathfrak{x}^*$ soll durch Iteration auf einer Rechenanlage bestimmt werden. Nach den in Abschnitt **8**.1 durchgeführten Überlegungen ist es zweckmäßig, diese Iteration mit Einschließungsmengen durchzuführen. Wir entwickeln ein Verfahren zur Bestimmung eines dafür geeigneten Anfangsvektors und zeigen dann, wie dieses Verfahren anzulegen ist, damit der gesamte Rechenaufwand für die Bestimmung dieses Anfangsvektors und die anschließende Iteration minimal wird. Alle Überlegungen werden zunächst in der exakten über dem Körper der reellen Zahlen aufgebauten Intervallrechnung durchgeführt; anschließend wird die Realisierung des Verfahrens auf der Rechenanlage besprochen.

Wir begründen und beschreiben jetzt ein Verfahren zur Bestimmung eines Vektors $\mathfrak{x}_0$ mit der Einschließungseigenschaft $\mathfrak{x}_0 \supset \mathfrak{x}^*$. Der Grundgedanke bei diesem Verfahren ist, zu einem beliebig vorgegebenen Vektor $\check{\mathfrak{u}} \in V_n(R)$ einen nichtnegativen Vektor $\mathfrak{z} \in V_n(R)$ (also $\mathfrak{z} \geq \mathfrak{o}$) so zu bestimmen, daß der Intervallvektor

$$\mathfrak{x}_0 := \check{\mathfrak{u}} + [\mathfrak{z}, \mathfrak{z}] = [\check{\mathfrak{u}} - \mathfrak{z}, \check{\mathfrak{u}} + \mathfrak{z}]$$

durch die Abbildung $\mathfrak{x} \longrightarrow \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ in sich abgebildet wird. Dann erfüllt nämlich $\mathfrak{x}_0$ die Voraussetzungen von Satz 8.2 b; daraus wieder folgt für die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit dem Anfangsvektor $\mathfrak{x}_0$:

$$\mathfrak{x}_0 \supset \mathfrak{x}_1 \supset \mathfrak{x}_2 \supset \ldots\ldots \supset \mathfrak{x}^*,$$

d.h. die $\mathfrak{x}_\nu$ bilden eine absteigende Mengenfolge, die gegen $\mathfrak{x}^*$ konvergiert.

Bei der Herleitung des Verfahrens wird sich dann auch ergeben, welche Voraussetzungen die Matrix $\mathfrak{A}$ erfüllen muß, damit die Bestimmung eines Vektors $\mathfrak{x}_0$ mit der angegebenen Eigenschaft möglich ist.

Die Anregung zur Entwicklung dieses Verfahrens gaben die Abhandlungen in [5] über monotone Operatoren, speziell § 21.5 . Dort wird ein Verfahren beschrieben, mit dem bei reellen linearen Gleichungssystemen geeignete Anfangsvektoren für eine spezielle Doppeliteration bestimmt werden. Dieses Verfahren läßt sich übrigens mit dem Ergebnis von Satz 4.14 noch wesentlich verallgemeinern. Die entsprechende Verallgemeinerung haben wir hier gleich mit einbezogen.

Im folgenden wird mit Ungleichungen in $V_n(I(R))$ und $M_n(I(R))$ operiert; die dabei verwendeten Regeln ergeben sich alle unmittelbar aus der Definition der Relation "$\leq$" in $V_n(I(R))$ bzw. $M_n(I(R))$ und den Regeln für das Rechnen mit Ungleichungen in R .

Die Matrix $\mathfrak{A}$ werde aufgespalten in die Summe $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ der Matrizen $\mathfrak{A}_1 \geq \mathfrak{o}$ und $\mathfrak{A}_2 \leq \mathfrak{o}$. Wegen der Subdistributivität gilt für alle $\mathfrak{x} \in V_n(I(R))$:

$$\mathfrak{A}\mathfrak{x} + \mathfrak{b} \subset \mathfrak{A}_1\mathfrak{x} + \mathfrak{A}_2\mathfrak{x} + \mathfrak{b} .$$

Für $\mathfrak{z} \geq \mathfrak{o}$ ist $\check{\mathfrak{u}} - \mathfrak{z} \leq \check{\mathfrak{u}} + \mathfrak{z}$, also

$$\mathfrak{A}_1(\check{\mathfrak{u}} - \mathfrak{z}) \leq \mathfrak{A}_1(\check{\mathfrak{u}} + \mathfrak{z}) \quad \text{und} \quad \mathfrak{A}_2(\check{\mathfrak{u}} + \mathfrak{z}) \leq \mathfrak{A}_2(\check{\mathfrak{u}} - \mathfrak{z}) .$$

Daraus folgt

$$\begin{aligned}\mathfrak{A}\,[\check{\mathfrak{u}} - \mathfrak{z}, \check{\mathfrak{u}} + \mathfrak{z}] + \mathfrak{b} \subset \mathfrak{A}_1[\check{\mathfrak{u}} - \mathfrak{z}, \check{\mathfrak{u}} + \mathfrak{z}] + \mathfrak{A}_2[\check{\mathfrak{u}} - \mathfrak{z}, \check{\mathfrak{u}} + \mathfrak{z}] + \mathfrak{b} &= \\ = [\mathfrak{A}_1(\check{\mathfrak{u}} - \mathfrak{z}), \mathfrak{A}_1(\check{\mathfrak{u}} + \mathfrak{z})] + [\mathfrak{A}_2(\check{\mathfrak{u}} + \mathfrak{z}), \mathfrak{A}_2(\check{\mathfrak{u}} - \mathfrak{z})] + \mathfrak{b} &= \\ = [\mathfrak{A}_1(\check{\mathfrak{u}} - \mathfrak{z}) + \mathfrak{A}_2(\check{\mathfrak{u}} + \mathfrak{z}), \mathfrak{A}_1(\check{\mathfrak{u}} + \mathfrak{z}) + \mathfrak{A}_2(\check{\mathfrak{u}} - \mathfrak{z})] + \mathfrak{b} & .\end{aligned}$$

Also wird $[\check{\mathfrak{x}}-\mathfrak{z},\check{\mathfrak{x}}+\mathfrak{z}]$ durch die Abbildung $\mathfrak{x}\to\mathfrak{A}\mathfrak{x}+\mathfrak{b}$ jedenfalls dann in sich abgebildet, wenn

$$\check{\mathfrak{x}}-\mathfrak{z}\le\mathfrak{A}_1(\check{\mathfrak{x}}-\mathfrak{z})+\mathfrak{A}_2(\check{\mathfrak{x}}+\mathfrak{z})+\mathfrak{b}\le\mathfrak{A}_1(\check{\mathfrak{x}}+\mathfrak{z})+\mathfrak{A}_2(\check{\mathfrak{x}}-\mathfrak{z})+\mathfrak{b}\le\check{\mathfrak{x}}+\mathfrak{z}\,. \qquad (1)$$

Die mittlere Ungleichung ist hier für $\mathfrak{z}\ge 0$ von selbst erfüllt; wir bestimmen $\mathfrak{z}\ge 0$ so, daß auch die beiden äußeren Ungleichungen gelten.

Wegen der Subdistributivität gelten die äußeren Ungleichungen in (1), wenn

$$\check{\mathfrak{x}}-\mathfrak{z}\le\mathfrak{A}_1\check{\mathfrak{x}}-\mathfrak{A}_1\mathfrak{z}+\mathfrak{A}_2\check{\mathfrak{x}}+\mathfrak{A}_2\mathfrak{z}+\mathfrak{b}\le\mathfrak{A}_1\check{\mathfrak{x}}+\mathfrak{A}_1\mathfrak{z}+\mathfrak{A}_2\check{\mathfrak{x}}-\mathfrak{A}_2\mathfrak{z}+\le\check{\mathfrak{x}}+\mathfrak{z}\,. \qquad (2)$$

Für Vektoren $\mathfrak{x}\in V_n(R)$ und Matrizen $\mathfrak{A}_1,\mathfrak{A}_2\in M_n(I(R))$ ist $\mathfrak{A}_1\mathfrak{x}+\mathfrak{A}_2\mathfrak{x}=(\mathfrak{A}_1+\mathfrak{A}_2)\mathfrak{x}$ und $\mathfrak{A}_1\mathfrak{x}-\mathfrak{A}_2\mathfrak{x}=(\mathfrak{A}_1-\mathfrak{A}_2)\mathfrak{x}$;
deshalb ist (2) äquivalent mit:

$$\check{\mathfrak{x}}-\mathfrak{z}\le(\mathfrak{A}_1+\mathfrak{A}_2)\check{\mathfrak{x}}+\mathfrak{b}-(\mathfrak{A}_1-\mathfrak{A}_2)\mathfrak{z}\le(\mathfrak{A}_1+\mathfrak{A}_2)\check{\mathfrak{x}}+\mathfrak{b}+(\mathfrak{A}_1-\mathfrak{A}_2)\mathfrak{z}\le\check{\mathfrak{x}}+\mathfrak{z} \qquad (2')$$

Es sei $\check{\mathfrak{x}}' := \mathfrak{A}\check{\mathfrak{x}}+\mathfrak{b}$,
$\hat{\mathfrak{A}} := \mathfrak{A}_1-\mathfrak{A}_2$, dann ist $\hat{\mathfrak{A}}\ge\mathfrak{o}$
und $|\hat{\mathfrak{A}}|=|\mathfrak{A}_1|+|\mathfrak{A}_2|\ge|\mathfrak{A}_1+\mathfrak{A}_2|=|\mathfrak{A}|$ (mit $|\hat{\mathfrak{A}}|=|\mathfrak{A}|$, wenn $\mathfrak{A}$ indexfremd).

Nach Voraussetzung ist $\check{\mathfrak{x}}\in V_n(R)$, also ist $\check{\mathfrak{x}}-\check{\mathfrak{x}}=\mathfrak{o}$,
deshalb ist (2') äquivalent mit

$$-\mathfrak{z}\le\check{\mathfrak{x}}'-\check{\mathfrak{x}}-\hat{\mathfrak{A}}\mathfrak{z}\le\check{\mathfrak{x}}'-\check{\mathfrak{x}}+\hat{\mathfrak{A}}\mathfrak{z}\le\mathfrak{z} \qquad (2'')$$

Die mittlere Ungleichung in (2") ist für $\mathfrak{z}\ge 0$ von selbst erfüllt, die beiden äußeren Ungleichungen in (2") sind erfüllt, wenn

$$-(\check{\mathfrak{x}}'-\check{\mathfrak{x}})\le\mathfrak{z}-\sup(\hat{\mathfrak{A}}\mathfrak{z}) \quad\text{und}\quad (\check{\mathfrak{x}}'-\check{\mathfrak{x}})\le\mathfrak{z}-\sup(\hat{\mathfrak{A}}\mathfrak{z}). \qquad (3)$$

Aus (3) folgt nämlich (wegen $\sup(\hat{\mathfrak{A}}\mathfrak{z})\in V_n(R)$ ist $\sup(\hat{\mathfrak{A}}\mathfrak{z})-\sup(\hat{\mathfrak{A}}\mathfrak{z})=\mathfrak{o}$):

$$-(\check{\mathfrak{x}}'-\check{\mathfrak{x}})+\sup(\hat{\mathfrak{A}}\mathfrak{z})\le\mathfrak{z} \quad\text{und}\quad (\check{\mathfrak{x}}'-\check{\mathfrak{x}})+\sup(\hat{\mathfrak{A}}\mathfrak{z})\le\mathfrak{z}\ ,$$

daraus wiederum folgt wegen $\hat{\mathfrak{A}}\mathfrak{z}\le\sup(\hat{\mathfrak{A}}\mathfrak{z})$:

$$-(\check{\mathfrak{x}}'-\check{\mathfrak{x}})+\hat{\mathfrak{A}}\mathfrak{z}\le\mathfrak{z} \quad\text{und}\quad (\check{\mathfrak{x}}'-\check{\mathfrak{x}})+\hat{\mathfrak{A}}\mathfrak{z}\le\mathfrak{z}\ ;$$

multiplizieren wir noch die linke Ungleichung mit -1, so haben wir gerade die beiden äußeren Ungleichungen von (2") vor uns.

Aus der Definition der Relation "$\le$" in $V_n(I(R))$ und den Beziehungen

$$\inf(\mathfrak{x})\le\sup(\mathfrak{x})\le|\mathfrak{x}| \quad\text{sowie}\quad \inf(-\mathfrak{x})\le\sup(-\mathfrak{x})\le|\mathfrak{x}|$$

folgt jetzt:

Beide Ungleichungen in (3) sind erfüllt, wenn

$$(\check{\mathfrak{x}}'-\check{\mathfrak{x}})\le\mathfrak{z}-\sup(\hat{\mathfrak{A}}\mathfrak{z})\ . \qquad (4')$$

Für $\mathfrak{A}\in M_n(I(R))$ mit $\mathfrak{A}\ge 0$ ist $\sup(\mathfrak{A})=|\mathfrak{A}|$;
mit $\mathfrak{z}\ge\mathfrak{o}$ ist dann $\sup(\hat{\mathfrak{A}}\mathfrak{z})=\sup(\hat{\mathfrak{A}})\mathfrak{z}=|\mathfrak{A}|\mathfrak{z}$.

Aus (4') wird damit:

$$|(\check{\mathfrak{x}}'-\check{\mathfrak{x}})|\le\mathfrak{z}-|\hat{\mathfrak{A}}|\mathfrak{z}\ . \qquad (4)$$

Insgesamt erhalten wir jetzt den folgenden

<u>Satz 6.4</u>: Es sei $\check{\mathfrak{x}},\mathfrak{z}\in V_n(R)$ mit $\mathfrak{z}\ge 0$, $\mathfrak{b}\in V_n(I(R))$, $\mathfrak{A}\in M_n(I(R))$, $\hat{\mathfrak{A}} := \sup(\mathfrak{A},\mathfrak{o})-\inf(\mathfrak{A},\mathfrak{o})$;

und es gelte $\qquad (\mathfrak{A}\check{\mathfrak{u}}+\mathfrak{b}-\check{\mathfrak{u}}) \leq \mathfrak{z} - |\hat{\mathfrak{A}}|\mathfrak{z}$, $\qquad$ (4)

dann wird der Intervallvektor $\mathfrak{x}_o := [\check{\mathfrak{u}}-\mathfrak{z}, \check{\mathfrak{u}}+\mathfrak{z}]$ durch die Abbildung $\mathfrak{x} \longrightarrow \mathfrak{A}\mathfrak{x}+\mathfrak{b}$ in sich abgebildet.

Der Beweis folgt unmittelbar aus den oben durchgeführten Überlegungen, wenn man berücksichtigt, daß die geforderte Darstellung von $\mathfrak{A}$ als Summe $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ mit $\mathfrak{A}_1 \geq 0$, $\mathfrak{A}_2 \leq 0$ gerade durch $\mathfrak{A} = \sup(\mathfrak{A},\mathcal{O}) + \inf(\mathfrak{A},\mathcal{O})$ gegeben ist.

Ist nun $\varrho(|\hat{\mathfrak{A}}|) < 1$, so existieren zu beliebigem $\check{\mathfrak{u}} \in V_n(R)$, $\mathfrak{b} \in V_n(I(R))$ stets Vektoren $\mathfrak{z} \in V_n(R)$ mit $\mathfrak{z} \geq 0$, die Bedingung (4) erfüllen. (Für indexfremde Matrizen $\mathfrak{A}$ ist $|\hat{\mathfrak{A}}| = |\mathfrak{A}|$ und die Forderung $\varrho(|\hat{\mathfrak{A}}|) < 1$ vereinfacht sich zu $\varrho(|\mathfrak{A}|) < 1$.)

Wie sich aus dem Beweis von Satz 4.14 ergibt, existieren nämlich dann zu $|\hat{\mathfrak{A}}| =: (|\hat{A}_{ij}|)$ positive Zahlen $k_1, \ldots k_n > 0$, für die

$$\|\hat{\mathfrak{A}}\|_k := \max \frac{1}{k_i}\sum_j |\hat{A}_{ij}|\, k_j < 1 \qquad (5)$$

Wir setzen jetzt voraus, daß solche Zahlen $k_1, \ldots k_n$ bekannt sind; in vielen praktischen Fällen kommt man mit $k_1 = k_2 = k_3 = \ldots k_n = 1$ zum Ziel; das bedeutet, daß das sogenannte "Zeilensummenkriterium" erfüllt ist.

Aus dem Beweis von Satz 4.14 ergibt sich auch, wie geeignete $k_1,..,k_n$ durch Eigenvektorbestimmung ermittelt werden können.

Geeignete Vektoren $\mathfrak{z}$, die der Bedingung (4) genügen, lassen sich nun bestimmen durch den Ansatz $\mathfrak{z} = c(k_i)$ mit $c \geq 0$. Durch Einsetzen von (4) erhalten wir folgende Bedingung für c:

$$\check{\mathfrak{u}}' - \check{\mathfrak{u}} \leq c\,(k_i - \sum_j |\hat{A}_{ij}|\, k_j) \quad \text{für alle } i .$$

Wegen (5) ist $(k_i - \sum_j |\hat{A}_{ij}|\, k_j) > 0$ für alle i .
(4) ist also erfüllt für alle $\mathfrak{z} = c(k_i)$ mit $c \geq c'$,
dabei ist:

$$c' := \max_i \frac{(|\check{\mathfrak{u}}' - \check{\mathfrak{u}}|)_i}{k_i - \sum_j |\hat{A}_{ij}|\, k_j} \qquad (6)$$

(Für die i-te Komponente eines Vektors $\mathfrak{w}$ schreiben wir hier $(\mathfrak{w})_i$.)

Damit erhalten wir insgesamt: Für $c \geq c'$ wird jeder Vektor

$$\mathfrak{x}_o(c) := \check{\mathfrak{u}} + c\,([-k_i, k_i])$$

durch die Abbildung $\mathfrak{x} \longrightarrow \mathfrak{A}\mathfrak{x}+\mathfrak{b}$ in sich abgebildet.

Daraus erfolgt nach Satz 8.2 b, daß jeder solche Vektor $\mathfrak{x}_o(c)$ Obermenge der Lösung $\mathfrak{x}^*$ der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x}+\mathfrak{b}$ ist, und daß die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit $\mathfrak{x}_o(c)$ als Anfangsvektor eine absteigende gegen $\mathfrak{x}^*$ konvergente Mengenfolge ergibt.

Für alle $c \geq c'$ ist $\quad \mathfrak{x}_o(c) \supset \mathfrak{x}_o(c') \supset \mathfrak{x}^*$.

Als Anfangsvektor für die Iteration zur Bestimmung von wählt man deshalb $\mathfrak{x}_o := \mathfrak{x}_o(c')$.
Damit sind wir grundsätzlich fertig.

In einem kurzen Rückblick weisen wir noch auf einige wichtige Gesichtspunkte hin:

Wir gingen aus von dem Problem, die Lösung einer Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ auf einer Rechenanlage durch Iteration zu bestimmen; und stellten zunächst in Abschnitt 8.1 fest, daß die auf der Rechenanlage berechnete Approximation für die exakte Lösung $\mathfrak{x}^*$ diese nur dann mit Sicherheit enthält, wenn man die Iteration mit Einschließungsmengen durchführt. Oben haben wir nun ein Verfahren zur Bestimmung eines dafür geeigneten Anfangsvektors $\mathfrak{x}_0$ entwickelt. Voraussetzung für die Anwendbarkeit dieses Verfahrens ist $\varrho(|\hat{\mathfrak{A}}|) < 1$; dabei ist $\hat{\mathfrak{A}} = \sup(\mathfrak{A},\mathfrak{O}) - \inf(\mathfrak{A},\mathfrak{O})$.

Wir möchten besonders hervorheben, daß die Bedingung $\varrho(|\hat{\mathfrak{A}}|) < 1$ wegen $|\mathfrak{A}| \leq |\hat{\mathfrak{A}}|$ also $\varrho(|\mathfrak{A}|) \leq \varrho(|\hat{\mathfrak{A}}|)$ schärfer ist, als das notwendige und hinreichende Kriterium $\varrho(|\mathfrak{A}|) < 1$ für die Konvergenz des Iterationsverfahrens $\mathfrak{x}_{v+1} = \mathfrak{A}\mathfrak{x}_v + \mathfrak{b}$ (siehe Abschnitt 6).

Für indexfremde Matrizen jedoch gilt $|\hat{\mathfrak{A}}| = |\mathfrak{A}|$; bei Konvergenz des Iterationsverfahrens ist also hier auch das angegebene Verfahren zur Bestimmung des Anfangsvektors stets durchführbar. Eine wichtige Klasse indexfremder Matrizen bilden übrigens gerade die reellen Matrizen $\mathfrak{A} \in M_n(R)$.

Wesentlich bei der Durchführung des Verfahrens ist nun die Kenntnis positiver Zahlen $k_1,\dots k_n$ mit der Eigenschaft (5). (Der Beweis von Satz 4.14 zeigt, daß unter der Voraussetzung $\varrho(|\mathfrak{A}|) < 1$ stets solche Zahlen existieren und gibt zugleich einen Hinweis, wie solche Zahlen bestimmt werden können.)

Grundsätzlich liefert jetzt das Verfahren zu jedem beliebigen Vektor $\check{\mathfrak{y}} \in V_n(R)$ einen für die Iteration geeigneten Anfangsvektor $\mathfrak{x}_0$ mit den beschriebenen Einschließungseigenschaften. Man bestimmt dazu c' aus (6), und erhält damit:

$$\mathfrak{x}_0 = \check{\mathfrak{y}} + c' \, ([-k_i,k_i]) \; .$$

Durch $\check{\mathfrak{y}}$ wird nun gerade der Mittelpunkt $\underline{\mathfrak{x}}_0$ (siehe Definition 8.2) des Vektors $\mathfrak{x}_0$ vorgegeben, deshalb hängt auch die Güte der Approximation von $\mathfrak{x}^*$ durch $\mathfrak{x}_0$ wesentlich von der Wahl des Vektors $\check{\mathfrak{y}}$ ab.

Jetzt sind wir vor allem noch daran interessiert, den gesamten Rechenaufwand zur Bestimmung von $\mathfrak{x}^*$, und zwar einschließlich der von $\mathfrak{x}_0$ ausgehenden Iteration, möglichst klein zu halten. Dazu ist es zweckmäßig, nicht von einem beliebigen $\check{\mathfrak{y}}$ auszugehen, sondern vorher durch (reelle Rechnung) ein $\check{\mathfrak{y}}$ so zu bestimmen, daß der damit bestimmte Anfangsvektor $\mathfrak{x}_0$ die exakte Lösung $\mathfrak{x}^*$ schon möglichst gut approximiert; dann führt auch die von $\mathfrak{x}_0$ ausgehende intervallmäßige Iteration in entsprechend wenig Schritten im Rahmen der erreichbaren Genauigkeit auf $\mathfrak{x}^*$.

Wie man ein solches $\check{\mathfrak{y}}$ bestimmen kann, zeigen die folgenden Überlegungen:
Da $\mathfrak{x}_0$ zu vorgegebenem Mittelpunkt $\underline{\mathfrak{x}}_0 = \check{\mathfrak{y}}$ so bestimmt wird, daß es die exakte Lösung $\mathfrak{x}^*$ enthält und möglichst approximiert, erhielte man hier optimale Einschließungsschranken für $\mathfrak{x}^*$ durch die Wahl $\check{\mathfrak{y}} = \underline{\mathfrak{x}}^*$. Der Mittelpunkt $\underline{\mathfrak{x}}^*$ von $\mathfrak{x}^*$ ist nun natürlich nicht bekannt. Deshalb werden wir $\check{\mathfrak{y}}$ so wählen, daß es Lösung eines Gleichungssystems $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in \mathfrak{A}$, $\mathfrak{b} \in \mathfrak{b}$ ist; dann ist nämlich nach Satz 7.1 schon $\check{\mathfrak{y}} \in \mathfrak{x}^*$.
(Für alle $\mathfrak{A} \in \mathfrak{A}$ ist $\varrho(\mathfrak{A}) \leq \varrho(|\mathfrak{A}|) \leq \varrho(|\mathfrak{A}|) < 1$, also $\mathfrak{E} - \mathfrak{A}$ nicht singulär und damit die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ eindeutig lösbar.) Im Durchschnitt der Fälle kommt dabei $\check{\mathfrak{y}}$ dem Mittelpunkt $\underline{\mathfrak{x}}^*$ von $\mathfrak{x}^*$ am nächsten, wenn man für $\mathfrak{A}$ und $\mathfrak{b}$ die Mittelpunkte $\underline{\mathfrak{A}}$ von $\mathfrak{A}$ und $\underline{\mathfrak{b}}$ von $\mathfrak{b}$ wählt.

Diese plausible Erklärung wollen wir noch durch folgende Überlegung unterstützen:
Unser Ziel ist ja, den Vektor $\check{\mathfrak{y}}$ so zu bestimmen, daß der dazu berechnete Anfangsvektor $\mathfrak{x}_0$ die exakte Lösung $\mathfrak{x}^*$ möglichst gut approximiert; wegen $\mathfrak{x}_0 \supset \mathfrak{x}^*$ ist das dann der Fall, wenn die Komponentendurchmesser von $\mathfrak{x}_0$ möglichst klein sind. Die Komponentendurchmesser von $\mathfrak{x}_0$ betragen $2c'k_i$, sind also im wesentlichen durch c' bestimmt. Die Abhängigkeit der Zahl c' von $\check{\mathfrak{y}}$ ergibt sich aus (6); daraus entnehmen wir:

Um ein möglichst kleines c' (und damit möglichst kleine Komponentendurchmesser für φ_0) zu erhalten, müssen wir $\mathfrak{x}$ so bestimmen, daß die Komponenten des Vektors $|\mathfrak{x}' - \mathfrak{x}| = |\mathfrak{A}\underset{\cdot}{\mathfrak{x}} + \mathfrak{b} - \underset{\cdot}{\mathfrak{x}}|$ möglichst kein werden.

Bekanntlich ist der Absolutbetrag $|A|$ eines Intervalles $A := [a, a+d(A)]$ bei vorgegebenem Durchmesser $d(A)$, und beliebigem a dann am kleinsten, wenn A symmetrisch ist. Nun sind zwar die Komponentendurchmesser von $(\mathfrak{x}' - \mathfrak{x})$ nicht konstant, aber im Durchschnitt der Fälle sind die Absolutbeträge der Komponenten des Vektors $(\mathfrak{x}' - \mathfrak{x})$ wieder dann am kleinsten, wenn wir $\underset{\cdot}{\mathfrak{x}}$ so bestimmen, daß $(\mathfrak{x}' - \underset{\cdot}{\mathfrak{x}})$ ein symmetrischer Vektor wird.

Jetzt zeigen wir noch, daß $(\mathfrak{x}' - \underset{\cdot}{\mathfrak{x}})$ dann symmetrisch wird, wenn wir - wie oben schon vorgeschlagen - für $\underset{\cdot}{\mathfrak{x}}$ die Lösung der Gleichung $\underset{\cdot}{\varphi} = \underset{\cdot}{\mathfrak{A}}\,\underset{\cdot}{\varphi} + \underset{\cdot}{\mathfrak{b}}$ mit $\underset{\cdot}{\mathfrak{A}} = \underline{\mathfrak{A}}$ (Mittelpunkt von $\mathfrak{A}$) und $\underset{\cdot}{\mathfrak{b}} = \underline{\mathfrak{b}}$ (Mittelpunkt von $\mathfrak{b}$) einsetzen.
Es sei also $\underset{\cdot}{\mathfrak{x}}$ zunächst Lösung einer Gleichung $\underset{\cdot}{\varphi} = \underset{\cdot}{\mathfrak{A}}\,\underset{\cdot}{\varphi} + \underset{\cdot}{\mathfrak{b}}$ mit $\underset{\cdot}{\mathfrak{A}} \in \mathfrak{A}$, $\underset{\cdot}{\mathfrak{b}} \in \mathfrak{b}$, dann ist

$$\underset{\cdot}{\mathfrak{x}} = (\mathfrak{E} - \underset{\cdot}{\mathfrak{A}})^{-1}\underset{\cdot}{\mathfrak{b}}, \quad \text{also}$$

$$\mathfrak{x}' - \underset{\cdot}{\mathfrak{x}} = \mathfrak{A}\underset{\cdot}{\mathfrak{x}} + \mathfrak{b} - \underset{\cdot}{\mathfrak{x}} = \mathfrak{A}(\mathfrak{E} - \underset{\cdot}{\mathfrak{A}})^{-1}\underset{\cdot}{\mathfrak{b}} + \mathfrak{b} - (\mathfrak{E} - \underset{\cdot}{\mathfrak{A}})^{-1}\underset{\cdot}{\mathfrak{b}} =$$

$$= \mathfrak{b} - (\mathfrak{A} - \mathfrak{E})(\underset{\cdot}{\mathfrak{A}} - \mathfrak{E})^{-1}\underset{\cdot}{\mathfrak{b}}.$$

Es gilt nun

<u>Hilfssatz 8.5</u>: Es seien $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ und $\underline{\mathfrak{A}}$ bzw. $\underline{\mathfrak{b}}$ die Mittelpunkte von $\mathfrak{A}$ bzw. $\mathfrak{b}$, dann ist der Vektor $\mathfrak{b} - (\mathfrak{A} - \mathfrak{E})(\underline{\mathfrak{A}} - \mathfrak{E})^{-1}\underline{\mathfrak{b}}$ symmetrisch.

<u>Beweis:</u> Wir beginnen mit einer Vorüberlegung:
Es seien $a_j, b_j \in R$, $B_j \in I(R)$ $(j = 1,2,\dots n)$, dann gilt nach dem Distributivgesetz (siehe Abschnitt 1):

$$\sum_j (b_j + B_j)\, a_j = \sum_j b_j\, a_j + \sum_j B_j\, a_j \tag{1}$$

Alle B_j werden jetzt zusätzlich als symmetrisch vorausgesetzt; nach Hilfssatz 4.7 sind dann auch die Produkte $B_j\, a_j$ symmetrisch; damit ist auch $\sum_j B_j\, a_j$ ein symmetrisches Intervall.

Jetzt sei $d := \sum_j b_j\, a_j$; dann folgt aus (1):

Das Intervall $D := \sum_j (b_j + B_j)\, a_j$ hat den Mittelpunkt $\underline{D} = d$.

Jetzt sei $\mathfrak{L} \in M_n(I(R))$, $\underline{\mathfrak{L}} = (b_{ij})$ der Mittelpunkt von $\mathfrak{L}$.

$\underline{\mathfrak{L}}$ sei nichtsingulär und $\underline{\mathfrak{L}}^{-1} = (a_{ij})$,

dann sind die Komponenten B_{ij} der Matrix $(B_{ij}) = \mathfrak{L} - \underline{\mathfrak{L}}$ symmetrisch und es ist $\mathfrak{L} = (b_{ij} + B_{ij})$.

Die Komponenten der Matrix $\mathfrak{L}\,\underline{\mathfrak{L}}^{-1} = (\sum_j (b_{ij} + B_{ij})\, a_{jl})$

sind jetzt Produktsummen der oben betrachteten Gestalt, mit $d = 0$ für $i \neq j$ und $d = 1$ für $i = j$.

$\mathfrak{L}\,\underline{\mathfrak{L}}^{-1}$ halt also die Gestalt
$$\mathfrak{L}\,\underline{\mathfrak{L}}^{-1} = \begin{pmatrix} \text{-1-} & & \text{-0-} \\ & \text{-1-} & \\ \text{-0-} & & \text{-1-} \end{pmatrix}$$

Dabei bedeutet

-0- ein symmetrisches Intervall, also mit dem Mittelpunkt 0,
-1- ein Intervall mit dem Mittelpunkt 1.

Der Mittelpunkt der Matrix $(\mathfrak{E} - \mathfrak{A})$ ist nun $(\mathfrak{E} - \underline{\mathfrak{A}})$, also hat auch $(\mathfrak{E} - \mathfrak{A})(\mathfrak{E} - \underline{\mathfrak{A}})^{-1}$ die angegebene Gestalt, damit wird aber

$$\mathfrak{b}' := (\mathfrak{E} - \mathfrak{A})(\mathfrak{E} - \underline{\mathfrak{A}})^{-1}\underline{\mathfrak{b}}$$

ein Vektor mit dem Mittelpunkt $\underline{\mathfrak{b}'} = \underline{\mathfrak{b}}$; daraus wieder folgt, daß $\mathfrak{b} - \mathfrak{b}'$ ein symmetrischer Vektor ist. □

Zum Abschluß beschreiben wir noch zusammenfassend, wie nach den bisher geführten Überlegungen die Lösung einer Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ auf der Rechenanlage durch Iteration mit Einschließungsmengen bestimmt wird.

Um Verwechslungen auszuschließen zwischen Aussagen, die im Körper der reellen Zahlen richtig sind und solchen, die in der zugrundegelegten Maschinenintervallarithmetik gelten, bezeichnen wir die maschineninterne, intervallmäßige Darstellung aller Größen durch einen Index "M". Etwa $\mathfrak{A}_M$ oder k_M sind dann die Darstellungen der Matrix $\mathfrak{A}$ bzw. der reellen Zahl k in der Maschinenintervallarithmetik; die Verknüpfungen der Maschinenintervallarithmetik bezeichnen wir wieder mit "$\oplus$, $\ominus$, $\odot$, $\oslash$"; die zugehörige Summenbildung mit $\boxed{\Sigma}$.

Vorgelegt sei jetzt die Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$; $\mathfrak{A}$ sei aufgespalten in $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ mit $\mathfrak{A}_1 \geq 0$, $\mathfrak{A}_2 \leq 0$ und für die Matrix $\hat{\mathfrak{A}} = (\hat{A}_{ij}) := \mathfrak{A}_1 - \mathfrak{A}_2$ gelte $\rho(|\hat{\mathfrak{A}}|) < 1$.

Dazu seien positive Zahlen $k_1, \ldots k_n$ bekannt, für die

$$\boxed{\Sigma}_j \, |A_{ij}|_M \odot k_{jM} < k_{iM} \quad \text{für alle } i. \tag{K}$$

Bei der Bestimmung der Lösung gehen wir in drei Schritten vor:

1. Schritt: Wir bestimmen durch gewöhnliche reelle Rechnung mit herkömmlicher Rundung eine Approximation für die Lösung des reellen linearen Gleichungssystems

$$\underset{\cdot}{\mathfrak{x}} = \underset{\cdot}{\mathfrak{A}}\,\underset{\cdot}{\mathfrak{x}} + \underset{\cdot}{\mathfrak{b}}$$

Dafür kann jedes herkömmliche Verfahren verwendet werden. Bei Iterationsverfahren können dann auch alle bekannten Maßnahmen zur Konvergenzbeschleunigung, also auch sukzessive Überrelaxation verwendet werden.

Das Kriterium für den Abbruch der Iteration wird man dabei so wählen, daß für Ausgangssysteme $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(R)$, $\mathfrak{b} \in V_n(R)$ die Lösung wirklich so genau berechnet wird, wie auf der Maschine überhaupt möglich, während für $\mathfrak{A} \in M_n(I(R))$, $\mathfrak{b} \in V_n(I(R))$ die erstrebte Genauigkeit einem vorgegebenen Bruchteil des maximalen Komponentendurchmessers in $\mathfrak{A}$ und $\mathfrak{b}$ entsprechen kann.

Die so berechnete Approximation für die Lösung der Gleichung $\underset{\cdot}{\mathfrak{x}} = \underset{\cdot}{\mathfrak{A}}\,\underset{\cdot}{\mathfrak{x}} + \underset{\cdot}{\mathfrak{b}}$ bezeichnen wir mit $\check{\mathfrak{x}} = \check{\mathfrak{x}}_M$ und geben sie als Mittelpunkt $\underset{\cdot}{\mathfrak{x}}_0 = \check{\mathfrak{x}}$ der zu bestimmenden Einschließungsmenge $\mathfrak{X}_0$ vor.

2. Schritt: Hier soll der Anfangsvektor $\mathfrak{X}_0$ bestimmt werden; dieser ist theoretisch bestimmt durch

$$\mathfrak{X}_0 := \check{\mathfrak{x}} + c' \, ([-k_i, k_i]) \quad \text{mit} \quad c' := \max_i \frac{|(\mathfrak{A}\check{\mathfrak{x}} + \mathfrak{b} - \check{\mathfrak{x}})|_i}{k_i - \sum_j |\hat{A}_{ij}| \cdot k_j}$$

Alle Rechnungen in diesem zweiten Abschnitt werden aber jetzt auf der Maschine intervallmäßig durchgeführt; die Auswertung des Ausdrucks für c' in der Maschinenintervallarithmetik ergibt dann:

$$c'_M := \max_i \left(|(\mathfrak{A}_M \odot \check{\mathfrak{x}}_M \oplus \mathfrak{b}_M \ominus \check{\mathfrak{x}}_M)|_i \right) \oslash \left(k_{iM} \ominus \boxed{\Sigma}_j |\hat{A}_{ij}|_M \odot k_{jM}\right);$$

Nach Voraussetzung (K) sind alle Divisionen auch auf der Rechenanlage durchführbar.
Es gilt $c'_M \geq 0$. Mit c'_M bilden wir schließlich

$$\mathfrak{X}_{0M} := \check{\mathfrak{x}}_M \oplus c'_M \odot ([-k_i, k_i]_M)$$

Bemerkung: Es würde hier genügen, allein die obere Intervallgrenze von c'_M zu bestimmen, denn $\mathfrak{X}_{0M}$ hängt nur von dieser ab. (In der exakten Intervallarithmetik gilt für $C, K \in I(R)$ mit $C \geq 0$ und K symmetrisch: $CK = |C|K = \sup(C)K$; in der Maschinenintervallarithmetik gilt natürlich die entsprechende Aussage.)

3. Schritt: Von $\mathfrak{x}_{oM}$ ausgehend iterieren wir intervallmässig nach der Iterationsvorschrift

$$\text{(D)} \qquad \mathfrak{x}_{\nu+1M} := \{\mathfrak{A}_M \odot \mathfrak{x}_{\nu M} \oplus \mathfrak{b}_M\} \cap \mathfrak{x}_{\nu M}$$

Wie wir oben zeigten, gilt in der exakten Intervallarithmetik für die Iteration $\mathfrak{x}_{\nu+1} = \mathfrak{A}\mathfrak{x}_\nu + \mathfrak{b}$ mit dem Anfangsvektor $\mathfrak{x}_o$ stets $\mathfrak{x}_o \supset \mathfrak{x}_1 \supset \mathfrak{x}_2 \supset \mathfrak{x}_3 \supset \ldots\ldots \supset \mathfrak{x}^*$, die Durchschnittsbildung in der Iterationsvorschrift (D) ist als theoretisch bedeutungslos.

In der Maschinenintervallarithmetik mit ihrer beschränkten Genauigkeit geht zwar wegen der Rundung nach außen auch bei der Rechenvorschrift

$$\text{(I)} \qquad \mathfrak{y}_{\nu+1M} := \mathfrak{A}_M \odot \mathfrak{y}_{\nu M} \oplus \mathfrak{b}_M$$

jede Einschließungseigenschaft $\mathfrak{y}_{oM} := \mathfrak{x}_{oM} \supset \mathfrak{x}^*$ auf alle Iterierten $\mathfrak{y}_{\nu M}$ über, aber die Einschließungseigenschaft " $\mathfrak{y}_{\nu M} \supset \mathfrak{y}_{\nu+1M}$ für alle ν " kann verlorengehen. Dies hätte eine Verlangsamung der Konvergenz zur Folge und brächte Schwierigkeiten bei der Formulierung eines geeigneten Abbruchkriteriums (siehe dazu Abschnitt 8.1); all das vermeidet man durch die in Verfahren (D) einbezogene Durchschnittsbildung.

Hat aber die zugrundegelegte Maschinenintervallarithmetik die Teilmengeneigenschaft, so ist es zweckmäßig, zunächst doch einen Iterationsschritt nach der Vorschrift (I), also ohne Durchschnittsbildung durchzuführen und nachzuprüfen, ob $\mathfrak{y}_{1M}$ in $\mathfrak{y}_{oM}$ enthalten ist.
Ist nämlich $\mathfrak{y}_{oM} \supset \mathfrak{y}_{1M}$, so gilt nach dem Korollar zu Satz 8.3 auch bei Zugrundelegung der Iterationsvorschrift (I)

$$\text{(E)} \qquad \mathfrak{y}_{\nu M} \supset \mathfrak{y}_{\nu+1M} \quad \text{für alle } \nu.$$

In diesem Falle sind also die Iterationsvorschriften (D) und (I) in ihrer Wirkung identisch, nur ist das Verfahren (D) mit der Nachprüfung der sowieso erfüllten Einschließungseigenschaft (E) belastet; diese wird deshalb hier besser weggelassen, d. h. man arbeitet mit der Vorschrift (I).

Die Iteration wird jeweils abgebrochen, sobald zwei aufeinanderfolgende Iterationsvektoren gleich sind. Dieser Fall tritt wie in Abschnitt 8.1 schon gezeigt wurde, nach endlich vielen Schritten ein. Wir erhalten so in der zugrundegelegten Maschinenintervallarithmetik für die Iterationsvorschrift (D) einen Fixpunkt $\mathfrak{x}^*_M$.
$\mathfrak{x}^*_M$ enthält nun mit Sicherheit die exakte Lösung $\mathfrak{x}^*$ der Gleichung $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ und approximiert sie im Rahmen der Maschinengenauigkeit.

Nach Satz 7.1 enthält $\mathfrak{x}^*_M$ nun auch zu jeder Gleichung $\mathfrak{x}_T = \mathfrak{A}_T\mathfrak{x}_T + \mathfrak{b}_T$ mit $\overline{\mathfrak{A}}_T \subset \mathfrak{A}$ und $\overline{\mathfrak{b}}_T \subset \mathfrak{b}$ den Komplex $\{\mathfrak{x}^*_T\}$ und die Hülle $\overline{\mathfrak{x}^*_T}$ der Lösung $\mathfrak{x}^*_T$.

Bemerkung 1: Natürlich kann man für die abschließende intervallmäßige Iteration statt des Gesamtschrittverfahrens auch das Einzelschrittverfahren verwenden; auf Grund der in Abschnitt 6.2 gewonnenen Ergebnisse muß das Einzelschrittverfahren sogar als günstiger angesehen werden.
Alle für das Gesamtschrittverfahren durchgeführten Überlegungen gelten für das Einzelschrittverfahren analog:

In der exakten Intervallarithmetik gilt für die Iterationsvorschrift $\mathfrak{x}_{\nu+1} = \mathfrak{L}\mathfrak{x}_{\nu+1} + (\mathfrak{D} + \mathfrak{R})\mathfrak{x}_\nu + \mathfrak{b}$ bei Verwendung des oben definierten Anfangsvektors $\mathfrak{x}_o$ stets

$$\mathfrak{x}_o \supset \mathfrak{x}_1 \supset \mathfrak{x}_2 \supset \mathfrak{x}_3 \supset \ldots\ldots \supset \mathfrak{x}^*.$$

Dies folgt aus der entsprechenden Eigenschaft des Gesamtschrittverfahrens und der Teilmengeneigenschaft.

In die Iterationsvorschrift für die Rechenanlage bezieht man aber aus denselben Gründen wie oben beim Gesamtschrittverfahren zweckmäßig eine Durchschnittsbildung mit ein. Die

Iterationsvorschrift lautet dann:

$$(DE)\quad \mathfrak{x}_{\nu+1\,M} = \{\mathfrak{L}_M \odot \mathfrak{x}_{\nu+1\,M} \oplus (\vartheta_M \oplus \mathfrak{R}_M) \odot \mathfrak{x}_{\nu\,M} \oplus \mathfrak{b}_M\} \cap \mathfrak{x}_{\nu M}$$

Hat die zugrundegelegte Maschinenintervallarithmetik die Teilmengeneigenschaft, so kann man wie beim Gesamtschrittverfahren vielfach auf die Durchschnittsbildung verzichten.

In diesem Falle gilt nämlich auch für die Iterationsvorschrift

$$(IE)\quad \mathfrak{y}_{\nu+1\,M} := \mathfrak{A}_M \odot \mathfrak{y}_{\nu+1\,M} \oplus (\vartheta_M \oplus \mathfrak{R}_M) \odot \mathfrak{y}_{\nu M} \oplus \mathfrak{b}_M \quad :$$

Aus $\mathfrak{y}_{oM} \supset \mathfrak{y}_{1M}$ folgt $\mathfrak{y}_{\nu M} \supset \mathfrak{y}_{\nu+1\,M} \supset \mathfrak{x}^*$ für alle ν.

Man beweist das genauso, wie die entsprechende Aussage für das Gesamtschrittverfahren (siehe Korollar zu Satz 8.3)

Im Falle $\mathfrak{y}_{oM} := \mathfrak{x}_{oM} \supset \mathfrak{y}_{1M}$ ist also die Vorschrift zur Durchschnittsbildung im Verfahren (DE) wieder ohne Wirkung und wird deshalb besser weggelassen; d. h. man arbeitet mit der Vorschrift (IE).

Bemerkung 2: Wir wollen das oben angegebene Verfahren noch bei seiner Anwendung auf reelle lineare Gleichungssysteme $\mathfrak{x} = \mathfrak{A}\mathfrak{x} + \mathfrak{b}$ mit $\mathfrak{A} \in M_n(R)$ und $\mathfrak{b} \in V_n(R)$ interpretieren:
Hier ist $\overset{c}{\mathfrak{x}}_M$ die auf der Maschine durch reelle Rechnung mit üblicher Rundung berechnete Approximation für die exakte Lösung; für diese werden in $\mathfrak{x}_{oM}$ zunächst Fehlerschranken bestimmt und anschließend durch intervallmäßige Nachiteration verbessert.

9. Literaturverzeichnis

[1] APOSTOLATOS,N. und KULISCH,U.: Grundlagen einer Maschinenintervallarithmetik, Computing, Vol.2, Fasc.2, 1967, S.89-104.

[2] APOSTOLATOS,N. und KULISCH,U.: Approximation der erweiterten Intervallarithmetik durch die einfache Maschinenintervallarithmetik, Computing, Vol.2, Fasc.3, 1967, S.181-194.

[3] APOSTOLATOS,N. und KULISCH,U.: Grundzüge einer Intervallrechnung für Matrizen und einige Anwendungen. Elektronische Rechenanlagen, Band 10, 1968, Heft 2, S.73-83.

[4] APOSTOLATOS,N. und KULISCH,U.: Über die Konvergenz des Relaxationsverfahrens bei nicht-negativen und diagonaldominanten Matrizen, Computing, Vol.2, Fasc.1, 1967, S.17-24.

[5] COLLATZ,L.: Funktionalanalysis und numerische Mathematik, Springer-Verlag, 1964.

[6] DUNFORD-SCHWARTZ: Linear Operators, Part I, Interscience Publishers, Inc., New-York, 1958.

[7] HOUSEHOLDER,A.S.: The Theorie of Matrices in Numerical Analysis, Blaisdell Publishing Company, New-York - Toronto - London, 1964.

[8] KULISCH,U.: Grundzüge der Intervallrechnung (Auszug aus einer vom Verfasser im WS 67/68 an der Universität Karlsruhe gehaltenen Vorlesung). Erscheint in einem Taschenbuch des Bibliographischen Instituts.

[9] MOORE,R.E.: Intervall Analysis. Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1966.

[10] NICKEL,K.: Über die Notwendigkeit einer Fehlerschrankenarithmetik für Rechenautomaten. Numerische Mathematik, Band 9, Heft 1, 1966, S.69-79.

[11] NICKEL,K.: Anwendungen einer Fehlerschrankenarithmetik. Erscheint in "Numerische Mathematik, Differentialgleichungen, Approximationstheorie". Internationale Schriftenreihe zur Numerischen Mathematik. Birkhäuser-Verlag, Basel 1968.

[12] NICKEL,K.: Error-bounds and Computer-arithmetik. Vortrag JFIP Congress 1968, Edinbourgh.

[13] SCHÄFER,H.H.: Topological Vector Spaces. Macmillan Series in Advanced Mathematics and Theoretical Physics, 1966 .

[14] VARGA,R.S.: Matrix Iterative Analysis. Prentice-Hall, Inc.Englewood Cliffs , New Jersey 1962.

[15] WIPPERMANN,H.W.: Realisierung einer Intervallarithmetik in einem Algol-60-System. Elektronische Rechenanlagen, Band 9, 1967, Heft 5, S.224-233.

[16] WULICH,B.S.: Einführung in die Funktionalanalysis. Teubner-Verlag, Leipzig 1961.

[17] YOUNG,R.C.: The algebra of many-valued quantities. Mathematische Annalen, Band 104, 1931, S.261-290.

Lebens- und Bildungsgang

Otto Mayer

geboren	am 24. 6. 1939 in München
Eltern	Otto Mayer, Leutnant, gefallen 1944; Zenta Mayer, geborene Kremer
Religion	römisch-katholisch
Schulbildung	von 1945 bis 1949 Besuch der Volksschule, von 1949 bis 1958 des Humanistischen Gymnasiums in Freising
Reifeprüfung	im Juli 1958 am Dom-Gymnasium Freising
Studium	der Mathematik und Physik an der Technischen Hochschule München von Wintersemester 1958/59 bis Sommersemester 1964
Wissenschaftliche Prüfungen	Diplomprüfung für Mathematiker an der Technischen Hochschule München im Oktober 1963, Wissenschaftliche Prüfung (1. Staatsexamen) für das Lehramt an Höheren Schulen im November 1964
Studienreferendar	von Dezember 1964 bis August 1966
Lehramtsprüfung	Pädagogische Prüfung (2. Staatsexamen) für das Lehramt an Höheren Schulen im August 1966
Tätigkeit als Wissenschaftlicher Assistent	vom 1.11.1966 bis 31.7.1967 am Mathematischen Institut der Technischen Hochschule München, seit 1.8.1967 am Institut für Angewandte Mathematik der Universität Karlsruhe

Karlsruhe, den 1. Dezember 1968

Gahmig-Druck · 63 Gießen 1 · Klinikstr. 21 · Tel. 79 19 72