Das Lineare Komplementaritätsproblem mit Intervalleinträgen

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik der
Universität Karlsruhe
genehmigte

DISSERTATION

von

Dipl.-Math. Uwe Schäfer
aus Baden-Baden/Sandweier

Tag der mündlichen Prüfung:	29. November 1999
Referent:	Prof. Dr. Götz Alefeld
Korreferent:	Prof. Dr. Peter Volkmann

Inhaltsverzeichnis

Bezeichnungen

$\emptyset$	leere Menge
$\mathbf{N}$	Menge der natürlichen Zahlen
$\mathbf{R}$	Menge der reellen Zahlen
$\|a\|$	Betrag einer reellen Zahl a
$\mathbf{C}$	Menge der komplexen Zahlen
$\|\lambda\|$	Betrag einer komplexen Zahl λ
$\mathbf{R}^n$	Menge der reellen n-dimensionalen Vektoren
$w^{\mathrm{T}}z$	Skalarprodukt der Vektoren w und z
$\mathbf{R}^n_{\geq 0}$	Menge der reellen n-dimensionalen Vektoren mit nichtnegativen Komponenten
$\mathbf{R}^{m\times n}$	Menge der reellen $m \times n$ Matrizen
I	Einheitsmatrix
A^{-1}	Inverse Matrix von A
A^{T}	Transponierte Matrix von A
$\mathbf{IR}$	Menge der reellen Intervalle
$\mathbf{IR}^n$	Menge der reellen n-dimensionalen Intervallvektoren
$\mathbf{IR}^{m\times n}$	Menge der reellen $m \times n$ Intervallmatrizen
$i = m(1)n$	für alle natürlichen Zahlen von $i = m$ bis $i = n$
i.a.	im allgemeinen

Vektoren und reelle Zahlen werden mit kleinen, Matrizen mit großen lateinischen Buchstaben bezeichnet. Intervalle, Intervallvektoren und Intervallmatrizen werden zusätzlich mit eckigen Klammern versehen, also z.B. $[A]$.
Die Komponenten von (Intervall-) Vektoren und Matrizen indizieren wir durch Tiefstellung, also z.B. x_i, a_{ij}, $[q_i]$.
Iterationsindizes werden hochgestellt: x^k.
Die Arbeit ist unterteilt in Kapitel und diese wiederum in Abschnitte und Unterabschnitte. Sätze, Lemmata und Beispiele werden jeweils innerhalb eines Kapitels getrennt voneinander durchnumeriert und mit einer nachgestellten Kennung versehen, die aus zwei Zahlen besteht. Die erste Zahl bezeichnet das Kapitel, die zweite Zahl die Nummer des Satzes bzw. des Lemmas bzw. des Beispiels in diesem Kapitel. So bedeuten Satz 2.1 Satz 1 in Kapitel 2 und Lemma 2.1 Lemma 1 in Kapitel 2.

Einleitung

In den Büchern [10] und [28], aber auch schon im letzten Kapitel des Buches [6] widmet man sich dem linearen Komplementaritätsproblem, welches wir im folgenden stets mit LCP abkürzen werden. Darunter versteht man folgende Problemstellung:

> Gegeben sind eine Matrix $M \in \mathbf{R}^{n \times n}$ und ein Vektor $q \in \mathbf{R}^n$, und gesucht sind Vektoren $w, z \in \mathbf{R}^n$ (oder gesucht ist ein Beweis zur Nichtexistenz von Vektoren $w, z \in \mathbf{R}^n$) mit

$$\left.\begin{array}{rcl} w - Mz & = & q, \\ w^{\mathrm{T}} z & = & 0, \\ w \geq 0 \quad \text{und} & & z \geq 0. \end{array}\right\} \tag{1}$$

Es ist also $w_i = 0$ oder $z_i = 0$, $i = 1(1)n$. Daher der Name komplementär.

Eine zu (1) äquivalente Formulierung lautet:

> Gegeben sind eine Matrix $M \in \mathbf{R}^{n \times n}$ und ein Vektor $q \in \mathbf{R}^n$, und gesucht ist ein Vektor $z \in \mathbf{R}^n$ (oder gesucht ist ein Beweis zur Nichtexistenz eines Vektors $z \in \mathbf{R}^n$) mit

$$\left.\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z & = & 0. \end{array}\right\} \tag{2}$$

Dabei bedeutet die Äquivalenz zwischen (1) und (2):

> Sind die Vektoren $w, z \in \mathbf{R}^n$ eine Lösung von (1), so ist der Vektor $z \in \mathbf{R}^n$ eine Lösung von (2).
> Ist der Vektor $z \in \mathbf{R}^n$ eine Lösung von (2), so sind die Vektoren $w, z \in \mathbf{R}^n$ mit
>
> $$w := q + Mz$$
>
> eine Lösung von (1).

(1) hat zwar doppelt soviele Unbekannte wie (2). Dafür hat aber (2) mit $(q + Mz)^{\mathrm{T}} z = 0$ ein nichtlineares Teilproblem, während (1) hauptsächlich linear ist.

Der Grund, für ein LCP zwei Formulierungen einzuführen, liegt aber nicht so sehr darin, die jeweiligen Vorteile auszunutzen. Er liegt vielmehr in der Natur der Sache.

Einige Anwendungen (z.B. lineare und quadratische Programmierung, Zweipersonenspiele, siehe [10], [28]) führen auf ein LCP der Art (1), andere Anwendungen (z.B. freie Randwertprobleme, konvexe Hüllenbestimmung von Punkten in der Ebene, siehe [10]) führen auf ein LCP der Art (2).

Weiteres Auftreten und weitere Anwendungen eines LCPs findet man im Artikel [16].

Die Frage nach der Lösbarkeit eines LCPs ist schon hinlänglich beantwortet, und bereits in den sechziger Jahren hat man Algorithmen entwickelt, die eine Lösung eines LCPs bestimmen. Die beiden Hauptalgorithmen gehen dabei auf R. W. Cottle und G. B. Dantzig [9] und auf C. E. Lemke [23] zurück.

Anfang der siebziger Jahre hat C. W. Cryer [13] das S.O.R.-Verfahren auf das LCP übertragen und gezeigt, daß es konvergiert, falls M eine symmetrische positiv definite Matrix ist.

Viele weitere Verfahren wurden entwickelt (siehe etwa [10] und [28]). In diesen Arbeiten wird folgende Problematik aber nicht berücksichtigt:

> Welche Aussagen kann man noch treffen, wenn man die dem LCP zugrunde liegende Matrix M und/oder den dem LCP zugrunde liegenden Vektor q explizit gar nicht kennt ?
> Was kann man noch aussagen, wenn man lediglich untere und obere Schranken für M und/oder q kennt ?

Eine analoge Fragestellung hat schon in vielen anderen Bereichen der Mathematik zur Übertragung von klassischen Algorithmen auf Algorithmen, die auf der Intervallrechnung basieren, geführt. Man denke z.B. an den Intervall-Gauß-Algorithmus (siehe z.B. [2]).

Somit ist der Punkt, an dem diese Arbeit ein- bzw. ansetzt, folgende Problematik:

Gegeben sind eine quadratische $n \times n$ Intervallmatrix $[M]$, d.h. eine Matrix,

deren Elemente reelle Intervalle sind, und ein n-dimensionaler Intervallvektor $[q]$, d.h. ein Vektor, dessen Elemente reelle Intervalle sind. Dann betrachten wir die Menge von linearen Komplementaritätsproblemen:

$$\left.\begin{array}{rcl} w - Mz & = & q, \\ w^{\mathrm{T}} z & = & 0, \\ w \geq 0 \quad \text{und} & & z \geq 0, \end{array}\right\} \quad M \in [M],\, q \in [q], \tag{3}$$

bzw.

$$\left.\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z & = & 0, \end{array}\right\} \quad M \in [M],\, q \in [q]. \tag{4}$$

Es entstehen nun folgende Fragen:

- Kann man die Gesamtheit aller Lösungen von (3) bzw. von (4) bestimmen ?
- Läßt sich die Gesamtheit aller Lösungen von (3) bzw. von (4) durch einen Algorithmus intervallmäßig einschließen ?
- Wie schnell ist ein solcher Algorithmus ?
- Gibt es Anwendungen dazu ?
- Können aufgrund der Intervallrechnung Aussagen getroffen werden, die durch die klassische LCP - Betrachtung nicht abgedeckt werden ?

Das Ziel dieser Arbeit ist es, all diese Fragen positiv zu beantworten. Dazu gehen wir wie folgt vor:

In Kapitel 1 werden wir die wesentlichen Hilfsmittel aus der Intervallrechnung bereitstellen und einige Aussagen aus der Numerischen Linearen Algebra zurechtlegen.

In Kapitel 2 stellen wir sozusagen als Vorläufer für das LCP mit Intervalleinträgen all die Argumente bereit, die schon bei linearen Intervallgleichungssystemen zum Ziel geführt haben. Auch der oben schon erwähnte Intervall-Gauß-Algorithmus wird tiefgehend behandelt, da er im dritten Kapitel entscheidend mitwirkt.

In Kapitel 3 wenden wir uns den Problemen (3) und (4) zu. Die Lösungsmengen L_A und L_B, die die Gesamtheit aller Lösungen von (3) und von (4) beschreiben, werden eingeführt, und einige Lösungsmengen werden wir geometrisch darstellen. Danach wird eine Klasse von Intervallmatrizen vorgestellt, für die wir einen Algorithmus angeben können, der die Lösungsmengen L_A und L_B intervallmäßig einschließt.

In einem weiteren Abschnitt wird das LCP als Nullstellenproblem formuliert, und durch Vorgehensweisen zur Nullstelleneinschließung werden wir dann Lösungen eines LCPs mit Intervalleinträgen intervallmäßig einschließen können.

Anhand eines gewöhnlichen freien Randwertproblems zeigen wir dann in Kapitel 4 eine Anwendung. Die Ideen aus Kapitel 3 liefern dabei zwei Algorithmen, die Aussagen liefern, die ein klassischer Algorithmus für ein LCP nicht machen kann.

Kapitel 1

Grundbegriffe aus der Intervallrechnung und der Numerischen Linearen Algebra

Wir wollen in den ersten beiden Abschnitten kurz die wesentlichen Definitionen und Eigenschaften der Intervallrechnung zusammenfassen. Dabei orientieren wir uns an [3]. Im dritten Abschnitt sind dann einige Resultate aus der Numerischen Linearen Algebra zusammengestellt.

1.1 Reelle Intervallrechnung und Eigenschaften

In der Menge $\mathbf{R}$ der reellen Zahlen betrachten wir abgeschlossene, beschränkte Intervalle

$$[a] := [\underline{a}, \overline{a}] = \{x \in \mathbf{R} : \underline{a} \leq x \leq \overline{a}\}.$$

Die Gesamtheit dieser Intervalle bezeichnen wir mit $\mathbf{IR}$. Reelle Zahlen a sind spezielle Elemente von $\mathbf{IR}$ mit $[a] := [a, a]$. Wir schreiben dafür auch einfach a.
Bezeichnet $*$ eine der vier Verknüpfungen $+, -, \times, /$ für reelle Zahlen, so definiert man für zwei Elemente $[a]$ und $[b]$ in $\mathbf{IR}$ die entsprechenden Operationen durch

$$[a] * [b] := \{a * b : a \in [a], b \in [b]\}.$$

Bei der Division ist dabei $0 \notin [b]$ vorauszusetzen. Da die Funktion $f(a,b) = a * b$, $a \in [a], b \in [b], * \in \{+,-,\times,/\}$ stetig ist, ist $[a] * [b]$ wieder ein Element von $\mathbf{IR}$. Üblicherweise wird das $\times$-Zeichen durch einen Punkt ersetzt und dieser gewöhnlich weggelassen. Eine elementare Diskussion ergibt die folgenden Rechenregeln (siehe etwa [2]):

$$\begin{aligned}
[a] + [b] &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}], \\
[a] - [b] &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}], \\
[a] \times [b] &= [\min\{\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}, \max\{\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}\}], \\
[a]/[b] &= [\underline{a}\overline{a}] \times [\tfrac{1}{\overline{b}}, \tfrac{1}{\underline{b}}].
\end{aligned}$$

Eine grundlegende Eigenschaft ist die sogenannte Inklusionsmonotonie, d.h.:

$$[a] \subseteq [\tilde{a}],\ [b] \subseteq [\tilde{b}] \Rightarrow [a] * [b] \subseteq [\tilde{a}] * [\tilde{b}], \quad * \in \{+,-,\times,/\}.$$

Neben diesen vier elementaren Verknüpfungen definieren wir für Funktionen $r(\cdot)$, die in $\mathbf{R}$ (oder auf einer Teilmenge von $\mathbf{R}$) definiert sind,

$$r([a]) := \{r(a) : a \in [a]\},$$

wobei $[a]$ im Definitionsbereich von r liegt. Z.B. ist

$$abs([a]) := [\min\{|a| : a \in [a]\}, \max\{|a| : a \in [a]\}]$$

die intervallmäßige Betragsfunktion. Weitere wichtige Definitionen sind der Durchmesser $d([a])$ mit

$$d([a]) := \overline{a} - \underline{a},$$

der Mittelpunkt $\mathrm{mid}([a])$ mit

$$\mathrm{mid}([a]) := \frac{\overline{a} + \underline{a}}{2},$$

der Betrag $|[a]|$ mit

$$|[a]| := \max\{|a| : a \in [a]\}$$

und der Ostrowski-Operator $\langle [a] \rangle$ mit

$$\langle [a] \rangle := \min\{|a| : a \in [a]\}.$$

Es gilt $|[a]| = \max\{|\underline{a}|, |\overline{a}|\}$ und

$$\langle [a] \rangle = \begin{cases} 0 & \text{falls} \quad 0 \in [a], \\ \min\{|\underline{a}|, |\overline{a}|\} & \text{sonst.} \end{cases}$$

Speziell hat man

$$abs([a]) = [\langle [a] \rangle, |[a]|].$$

Nachfolgend findet man einige Rechenregeln für Durchmesser und Betrag:

$$\begin{aligned} d(a) &= 0 \quad \text{für } a \in \mathbf{R}, \\ d([a] \pm [b]) &= d([a]) + d([b]), \\ d(a[b]) &= |a| d([b]), \; a \in \mathbf{R}, \\ d([a] \cap [b]) &\leq d([a]) \,, d([b]), \\ d([a] \pm b) &= d([a]), \; b \in \mathbf{R}. \end{aligned}$$

Der Abstand zweier Intervalle $[a]$ und $[b]$ wird definiert als

$$\mathrm{q}([a], [b]) := \max\{|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|\}.$$

Es gelten

$$\begin{aligned} \mathrm{q}([a] + [b], [a] + [c]) &= \mathrm{q}([b], [c]), \\ \mathrm{q}(a[b], a[c]) &= |a| \mathrm{q}([b], [c]), \quad a \in \mathbf{R}, \\ \mathrm{q}([a][b], [a][c]) &\leq |[a]| \mathrm{q}([b], [c]). \end{aligned}$$

Die Beweise findet man in Einführungsbüchern zur Intervallrechnung (z.B. [2], [29]). Eine weitere Eigenschaft wollen wir aber speziell festhalten als

Lemma 1.1 *Es seien* $[x], [y] \in \mathbf{IR}$. *Dann gilt*

$$\mathrm{q}(abs([x]), abs([y])) \leq \mathrm{q}([x], [y]).$$

Einen Beweis findet man in [29].

1.2 Matrizenintervallrechnung

Mit einer Intervallmatrix $[A]$ bezeichnen wir eine Matrix, deren Elemente Intervalle $[a_{ij}] \in \mathbf{IR}$ sind. Wir schreiben dafür $[A] = ([a_{ij}])$. Hat eine Intervallmatrix m Zeilen und n Spalten, so schreiben wir $[A] \in \mathbf{IR}^{m \times n}$. Desweiteren wollen wir $\mathbf{IR}^{n \times 1}$ mit $\mathbf{IR}^n$ identifizieren.

Reelle Matrizen A sind spezielle Elemente von $\mathbf{IR}$ mit $[a_{ij}] := [a_{ij}, a_{ij}]$. Wir schreiben dafür auch einfach A und nennen A eine Punktmatrix. Entsprechend ist ein Intervallvektor bzw. ein Punktvektor $[a] = ([a_i])$ definiert. Verknüpfungen sind wie für reelle Matrizen und/oder Vektoren definiert:

$$\begin{aligned} [A] \pm [B] &= ([a_{ij}] \pm [b_{ij}]), \\ [A][B] &= \left(\sum_{k=1}^{n} [a_{ik}][b_{kj}] \right), \\ [A][x] &= \left(\sum_{k=1}^{n} [a_{ik}][x_k] \right), \end{aligned}$$

vorausgesetzt, die Dimensionen sind so, daß man die Verknüpfungen wie im Reellen durchführen kann.

Der Durchschnitt zweier Intervallmatrizen und die Teilmengenbeziehung zweier Intervallmatrizen sind über die Elemente bzw. Komponenten definiert:

$$\begin{aligned} [A] \cap [B] &= ([a_{ij}] \cap [b_{ij}]), \\ [A] \subseteq [B] &\Leftrightarrow [a_{ij}] \subseteq [b_{ij}]. \end{aligned}$$

Dabei wird $[A] \cap [B] = \emptyset$ gesetzt, falls i, j existieren mit $[a_{ij}] \cap [b_{ij}] = \emptyset$.

Es gilt i.a. nicht das Distributivgesetz, sondern die als Subdistributivität bezeichnete Eigenschaft

$$[A]([B] + [C]) \subseteq [A][B] + [A][C].$$

Für die Multiplikation von Intervallmatrizen (und/oder Intervallvektoren) gilt i.a. nicht das Assoziativgesetz. Ein Gegenbeispiel findet man z.B. in [2]. Es gilt jedoch das folgende

Lemma 1.2 *Sei e^i der i-te Einheitsvektor. Dann gilt für $[A], [B] \in \mathbf{IR}^{n \times n}$*

$$[A]([B]e^i) = ([A][B])e^i.$$

Einen Beweis findet man in [3].

Durchmesser, Mittelpunkt, Betrag und Abstand bei Intervallmatrizen und Intervallvektoren sind ebenfalls über die Elemente bzw. Komponenten definiert:

$$\begin{aligned} d([A]) &:= (d([a_{ij}])), \\ \mathrm{mid}([A]) &:= (\mathrm{mid}([a_{ij}])), \\ |[A]| &:= (|[a_{ij}]|), \\ \mathrm{q}([A],[B]) &:= (\mathrm{q}([a_{ij}],[b_{ij}])). \end{aligned}$$

Durchmesser, Betrag und Abstand von Intervallmatrizen sind also reelle Matrizen. Die entsprechenden Begriffe für Intervallvektoren sind reelle Vektoren. Ungleichungen zwischen reellen Matrizen bzw. reellen Vektoren sind elementweise bzw. komponentenweise zu verstehen. Für zwei reelle n-dimensionale Vektoren $x = (x_i)$ und $y = (y_i)$ gilt also

$$x \leq y \Leftrightarrow x_i \leq y_i, \quad i = 1(1)n.$$

Wie im eindimensionalen Fall gelten folgende Regeln:

$$\mathrm{q}([A]+[C],[B]+[C]) = \mathrm{q}([A],[B]), \tag{1.1}$$

$$\mathrm{q}([A][B],[A][C]) \leq |[A]|\mathrm{q}([B],[C]) \tag{1.2}$$

und

$$d(A[B]) = |A|d([B]), \tag{1.3}$$

$$d([A] \pm C) = d([A]), \tag{1.4}$$

$$d([A] \cap [B]) \leq d([A])\,, d([B]). \tag{1.5}$$

Die Beweise findet man in [2].

Als nächstes definieren wir den Ostrowski-Operator für quadratische Intervallmatrizen. Sei dazu $[A] \in \mathbf{IR}^{n\times n}$, dann setzen wir $\left\langle [A] \right\rangle = (c_{ij})$ mit

$$c_{ij} := \begin{cases} -|[a_{ij}]| & \text{für} \quad i \neq j, \\ \left\langle [a_{ij}] \right\rangle & \text{für} \quad i = j. \end{cases}$$

$\left\langle [A] \right\rangle$ wird auch oft Vergleichsmatrix von $[A]$ genannt. Sie wird bei den sogenannten H-Matrizen wichtig werden.

1.3 Hilfsmittel aus der Numerischen Linearen Algebra

Es sei $\mathbf{C}$ die Menge der komplexen Zahlen, und es sei A eine komplexe $n \times n$ Matrix. Wir schreiben dafür

$$A \in \mathbf{C}^{n \times n}$$

und meinen damit $a_{ij} \in \mathbf{C}$ für $i, j = 1(1)n$. Wir nennen $\lambda \in \mathbf{C}$ einen Eigenwert von A, falls es einen vom Nullvektor verschiedenen Vektor $x \in \mathbf{C}^n$ gibt mit

$$Ax = \lambda x.$$

x wird ein Eigenvektor von A genannt. Die Menge

$$\sigma(A) := \Big\{\lambda \in \mathbf{C} : \lambda \text{ ist ein Eigenwert von } A\Big\}$$

nennt man das Spektrum von A. Der betragsgrößte Eigenwert der Matrix A heißt Spektralradius und wird mit $\rho(A)$ bezeichnet. Es gilt der folgende

Satz 1.1 *Es sei $A \in \mathbf{C}^{n \times n}$. Dann gelten foldgende Aussagen:*

1. *Es gilt:*

$$\rho(A) < 1 \quad \Leftrightarrow \quad \begin{array}{l} \textit{Die Folge von Matrizen} \\ A, A^2, A^3, \ldots \\ \textit{konvergiert gegen die Nullmatrix.} \end{array}$$

2. *Aus $\rho(A) < 1$ folgt, daß $I - A$ nichtsingulär ist mit*

$$(I - A)^{-1} = \sum_{i=0}^{\infty} A^i.$$

Dabei bezeichnet I die $n \times n$ Einheitsmatrix und

$$A^i := \underbrace{A \cdots A}_{i \; mal}$$

mit $A^0 := I$.

Den Beweis zu Teil 1 findet man in [38] auf Seite 13 und den zweiten Teil findet man in [38] auf Seite 82.

Kapitel 2

Lineare Intervallgleichungssysteme

Lineare Intervallgleichungssysteme entstehen z.B. dann, wenn man ein lineares Gleichungssystem

$$\tilde{A}x = \tilde{b}, \quad \tilde{A} \in \mathbf{R}^{m\times n}, \tilde{b} \in \mathbf{R}^m, \tag{2.1}$$

zu lösen hat, wobei man aber für $\tilde{A}$ und $\tilde{b}$ lediglich untere und obere Schranken kennt.

Gegeben sind dann eine Intervallmatrix $[A] \in \mathbf{IR}^{m\times n}$ und ein Intervallvektor $[b] \in \mathbf{IR}^m$, und wir betrachten eine Menge von linearen Gleichungssystemen

$$Ax = b, \quad A \in [A], b \in [b]. \tag{2.2}$$

Die Gleichungen (2.2) werden als ein lineares Intervallgleichungssystem bezeichnet. Was uns jetzt interessiert, ist das Auffinden eines Intervallvektors $[x] \in \mathbf{IR}^n$, der die Menge

$$\Sigma := \Sigma([A],[b]) := \{x \in \mathbf{R}^n : Ax = b,\ A \in [A],\ b \in [b]\}$$

einschließt, falls Σ beschränkt ist. Denn dann können wir schließen:

Ist $\tilde{x}$ eine Lösung von (2.1), so gilt

$$\tilde{x} \in \Sigma([A],[b]) \subseteq [x].$$

Zur Warnung sei bemerkt, daß ein Intervallvektor $[x]$, der der Gleichung $[A][x] = [b]$ genügt, $\Sigma([A],[b])$ i.a. nicht einschließt. Dies zeigt das einfache

Beispiel 2.1 Gegeben sei die lineare Gleichung

$$\frac{4}{3}x = \frac{9}{7}. \tag{2.3}$$

Mit

$$\frac{4}{3} \in [1, \frac{16}{10}] =: [A] \in \mathbf{IR}^{1\times 1} \quad \text{und} \quad \frac{9}{7} \in [1,2] =: [b] \in \mathbf{IR}^1$$

erfüllt dann zwar $[x] = [1, \frac{5}{4}]$ die Gleichung $[A][x] = [b]$, aber die Lösung $\tilde{x} = \frac{27}{28}$ von (2.3) liegt nicht in $[x]$.

Eine Aussage, ob ein reeller Vektor $x \in \mathbf{R}^n$ zur Lösungsmenge Σ gehört, macht der folgende

Satz 2.1 *Es seien* $[A] \in \mathbf{IR}^{m\times n}$ *und* $[b] \in \mathbf{IR}^m$. *Dann sind folgende Aussagen äquivalent:*

a) $x \in \Sigma([A],[b])$.

b) $|mid([A])x - mid([b])| \leq \frac{1}{2}d([A])|x| + \frac{1}{2}d([b])$.

c) $[A]x \cap [b] \neq \emptyset$.

Die Äquivalenz *a)* $\Leftrightarrow$ *b)* ist der Satz von Oettli-Prager [31], und die Äquivalenz *a)* $\Leftrightarrow$ *c)* geht zurück auf den Satz von Beeck [5].

Einen Beweis insbesondere für den Fall $m \neq n$ findet man in [29] auf den Seiten 95/96. Dort wird die Äquivalenz *a)* $\Leftrightarrow$ *c)* als Satz formuliert, die Äquivalenz *a)* $\Leftrightarrow$ *b)* allerdings lediglich als Korollar.

Das läßt vermuten, daß der Satz von Beeck einen höheren Stellenwert hat als der Satz von Oettli-Prager. Tatsächlich wird auch der folgende Satz mit der Äquivalenz *a)* $\Leftrightarrow$ *c)* von Satz 2.1 bewiesen.

Satz 2.2 *Es sei* $[A]$ *eine reguläre Intervallmatrix, d.h.* $[A] \in \mathbf{IR}^{n\times n}$ *und jedes* $A \in [A]$ *sei invertierbar. Dann gelten folgende Aussagen:*

1. *Die Menge* $\Sigma([A],[b])$ *ist kompakt, zusammenhängend, aber i.a. nicht konvex.*

2. *Die Menge* $\Sigma([A],[b])$ *ist die Vereinigung von endlich vielen konvexen Polytopen.*

3. *Ist O ein festgewählter Orthant des $\mathbf{R}^n$ und die Menge $\Sigma([A],[b]) \cap O$ nichtleer, so ist $\Sigma([A],[b]) \cap O$ konvex, kompakt, zusammenhängend und ein Polytop.*

Den Beweis findet man in [4].

Eine Aussage, wann genau $\Sigma([A],[b])$ konvex ist, findet man in [33].

Eine Art, wie man eine intervallmäßige Einschließung von $\Sigma([A],[b])$ erhält, behandelt der nächste Abschnitt.

2.1 Der Intervall-Gauß-Algorithmus I.G.A.

Unter dem Intervall-Gauß-Algorithmus (I.G.A.) versteht man die direkte Übertragung des in der Linearen Algebra wohlbekannten Gauß-Algorithmus auf die Intervallrechnung. Es seien also $[A] \in \mathbf{IR}^{n \times n}$, $[b] \in \mathbf{IR}^n$.

Ist $0 \in [a_{11}]$, so sucht man eine Komponente $[a_{ij}]$ mit $0 \notin [a_{ij}]$. Diesen Vorgang nennt man wie im Reellen Pivotsuche.

Falls kein solches Element in der Intervallmatrix vorhanden ist, so ist der I.G.A. nicht durchführbar. Falls doch, so vertauscht man die erste Zeile mit der i -ten Zeile und die erste Spalte mit der j-ten Spalte. $[a_{ij}]$ nennt man das Pivotelement.

Das Pivotelement ist i.a. nicht eindeutig. Das Auswahlkriterium zur Bestimmung des Pivotelements nennt man Strategie der Pivotsuche.

Gilt also nach eventueller Vertauschung von Spalten und/oder Zeilen $0 \notin [a_{11}]$, dann versteht man unter dem ersten Intervall-Gauß-Schritt:

$$\begin{aligned}
[a'_{1j}] &:= [a_{1j}], & j &= 1(1)n, \\
[a'_{ij}] &:= [a_{ij}] - [a_{i1}][a_{1j}]/[a_{11}], & i,j &= 2(1)n, \\
[b'_i] &:= [b_i] - [a_{i1}][b_1]/[a_{11}], & i &= 2(1)n, \\
[a'_{i1}] &:= 0, & i &= 2(1)n.
\end{aligned}$$

Besitzt die Matrix $([a'_{ij}])$, $2 \leq i,j \leq n$, mindestens einen Koeffizienten, der die 0 nicht enthält, so ist nach eventueller Vertauschung von Spalten und/oder Zeilen ein weiterer Intervall-Gauß-Schritt anwendbar.

Der I.G.A. ist genau dann durchführbar, wenn $n-1$ Intervall-Gauß-Schritte

durchführbar sind und $0 \notin [a'_{nn}]$ gilt. Insbesondere folgt, daß die Durchführbarkeit nicht von der rechten Seite $[b]$ abhängt.

Im folgenden wollen wir keine Pivotsuche zulassen. Dann kann man den I.G.A. mit $[a_{ij}^{(1)}] := [a_{ij}]$, $1 \leq i,j \leq n$, und $[b_i^{(1)}] := [b_i]$, $1 \leq i \leq n$, folgendermaßen implementieren:

for $k := 1$ **to** $n-1$ **do**
begin
  for $i := k+1$ **to** n **do**
  begin
    for $j := k+1$ **to** n **do**
    $$[a_{ij}^{(k+1)}] := [a_{ij}^{(k)}] - [a_{ik}^{(k)}] \frac{[a_{kj}^{(k)}]}{[a_{kk}^{(k)}]};$$
    $$[b_i^{(k+1)}] := [b_i^{(k)}] - [a_{ik}^{(k)}] \frac{[b_k^{(k)}]}{[a_{kk}^{(k)}]};$$
  end;
  for $l := 1$ **to** k **do**
  begin
    for $j := l$ **to** n **do**
    $[a_{lj}^{(k+1)}] := [a_{lj}^{(k)}]$;
    $[b_l^{(k+1)}] := [b_l^{(k)}]$;
  end;
end;
$$[x_n] := \frac{[b_n^{(n)}]}{[a_{nn}^{(n)}]};$$
for $i := n-1$ **downto** 1 **do**
$$[x_i] := \left([b_i^{(n)}] - \sum_{j=i+1}^{n} [a_{ij}^{(n)}][x_j] \right) / [a_{ii}^{(n)}].$$

Dabei bezeichnen wir dann $IGA([A],[b]) := [x]$.

Aus dem Reellen wissen wir, daß durch den Gaußschen Algorithmus angewandt auf A eine Dreieckszerlegung generiert wird, d.h. man erhält eine linke untere Dreiecksmatrix L und eine rechte obere Dreiecksmatrix R mit $A = LR$. Um das reelle lineare Gleichungssystem $Ax = b$ zu lösen, löst

man dann zunächst $Ly = b$ (Vorwärtssubstitution) und danach $Rx = y$ (Rückwärtssubstitution).

Der Vorteil der Dreieckszerlegung ist offensichtlich. Hat man einmal die Dreieckszerlegung durchgeführt und ist z.B. öfters ein lineares Intervallgleichungssystem mit derselben Intervallmatrix aber mit verschiedenen rechten Seiten zu lösen (man denke z.B. an Iterationsverfahren), so kann man sich auf die Ausführung der Vorwärts- und Rückwärtssubstitution beschränken.

Ist der I.G.A. durchführbar für $[A] \in \mathbf{IR}^{n\times n}$, so führt die Übertragung der Dreieckszerlegung auf Intervallmatrizen auf eine linke untere Intervalldreiecksmatrix $[L]$ und auf eine rechte obere Intervalldreiecksmatrix $[R]$ mit $[A] \subseteq [L][R]$ (siehe etwa [29]):

$$[L] = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ [l_{21}] & 1 & \cdots & & 0 \\ \cdot & [l_{32}] & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot\ \cdot & \cdot \\ [l_{n1}] & [l_{n2}] & \cdots & [l_{nn-1}] & 1 \end{pmatrix}, \quad [R] = \begin{pmatrix} [r_{11}] & [r_{12}] & \cdots & [r_{1n}] \\ 0 & [r_{21}] & \cdots & [r_{2n}] \\ \cdot & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & \cdots & 0 & [r_{nn}] \end{pmatrix}.$$

Man erhält den I.G.A. mit Dreieckszerlegung:

{Dreieckszerlegung}

for $k := 1$ **to** $n-1$ **do**
begin
  for $i := k+1$ **to** n **do**
  begin
    $[l_{ik}] := [a_{ik}^{(k)}]/[a_{kk}^{(k)}]$;
    for $j := k+1$ **to** n **do**
    $[a_{ij}^{(k+1)}] := [a_{ij}^{(k)}] - [l_{ik}][a_{kj}^{(k)}]$;
  end;
end;

{ Vorwärtssubstitution }

for $i := 1$ **to** n **do**
$[y_i] := [b_i] - \sum_{j=1}^{i-1} [l_{ij}][y_j]$;

{ Rückwärtssubstitution }

for $i := n$ **downto** 1 **do**

$$[x_i] := \left([y_i] - \sum_{j=i+1}^{n} [a_{ij}^{(n)}][x_j] \right) / [a_{ii}^{(n)}].$$

Durch Induktion kann man zeigen, daß $[y_i] = [b_i^{(i)}] \equiv [b_i^{(n)}]$ für alle $i = 1(1)n$ gilt. Der I.G.A. und der I.G.A. mit Dreieckszerlegung liefern also das gleiche Ergebnis.

Definiert man $[C^{(k)}]$, $[T^{(k)}]$ und $[D^{(k)}]$ durch

$$[c_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j, \\ -[a_{ik}^{(k)}]/[a_{kk}^{(k)}] & \text{falls } j = k < i, \\ 0 & \text{sonst}, \end{cases}$$

$$[t_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j, \\ -[a_{kj}^{(n)}] & \text{falls } i = k < j, \\ 0 & \text{sonst}, \end{cases}$$

$$[d_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j \neq k, \\ 1/[a_{kk}^{(n)}] & \text{falls } i = j = k, \\ 0 & \text{sonst}, \end{cases}$$

so gilt

$$[x] = [D^{(1)}]([T^{(1)}] \cdots ([D^{(n-1)}]([T^{(n-1)}]([D^{(n)}]([C^{(n-1)}](\cdots [C^{(1)}][b]) \cdots).$$

Diese Darstellung wurde von H. Schwandt [35] angegeben. Dabei dürfen aufgrund der nicht gegebenen Assoziativität der Intervallmatrizenmultiplikation die Klammern nicht weggelassen werden.

Es gilt der grundlegende

Satz 2.3 *Es seien* $[A] \in \mathbf{IR}^{n \times n}$ *und* $[b] \in \mathbf{IR}^n$. *Ist der I.G.A. durchführbar für* $[A]$, *so gilt*

$$\Sigma([A], [b]) \subseteq IGA([A], [b]).$$

Den Beweis findet man in [2].

Die Intervall-Gauß-Inverse von $[A]$ bezeichnen wir mit $IGA([A])$, und sie ist definiert durch

$$IGA([A]) := [D^{(1)}]([T^{(1)}] \cdots ([D^{(n-1)}]([T^{(n-1)}]([D^{(n)}]([C^{(n-1)}](\cdots [C^{(1)}]) \cdots).$$

Es gilt das

Lemma 2.1 *Es sei $[A] \in \mathbf{IR}^{n \times n}$, und es sei e^i der i-te Einheitsvektor. Dann gilt*

$$IGA([A], e^i) = IGA([A])e^i,\ 1 \leq i \leq n.$$

Die Behauptung folgt aufgrund der obigen Darstellung für $IGA([A], e^i)$ sofort aus Lemma 1.2.

Um die Intervall-Gauß-Inverse $IGA([A])$ zu erhalten, braucht man also die Matrizen $[C^{(k)}]$, $[T^{(k)}]$ und $[D^{(k)}]$ nicht explizit auszurechnen. Man erhält $IGA([A])$, indem man mit Hilfe des Intervall-Gauß-Algorithmus formal die Intervallmatrix $[A]$ invertiert.

2.2 Durchführbarkeit des Intervall-Gauß-Algorithmus

Aufgrund von Satz 2.3 ist es wichtig, Kriterien zu finden, die zu einer Intervallmatrix Auskunft über die Durchführbarkeit des I.G.A. geben. Zum jetzigen Zeitpunkt ist es noch nicht gelungen, ein notwendiges und hinreichendes Kriterium dafür zu finden. Man kann aber Klassen von Intervallmatrizen angeben, für die der I.G.A. durchführbar ist.

Aus der Linearen Algebra weiß man, daß der Gauß-Algorithmus (eventuell mit Pivotsuche) für $A \in \mathbf{R}^{n \times n}$ genau dann durchführbar ist, wenn $\det A \neq 0$ gilt. Für $[A] \in \mathbf{IR}^{n \times n}$ ist die Sachlage anders. Aufgrund der Inklusionsmonotonie gilt für jeden Intervall-Gauß-Schritt:

$$[C] \subseteq [A] \Rightarrow [C^{(i)}] \subseteq [A^{(i)}] \tag{2.4}$$

für $i = 1(1)n$. Ist der I.G.A. also für $[A]$ durchführbar, so ist der I.G.A. auch durchführbar für jede Intervallmatrix $[C] \subseteq [A]$. (Läßt man Pivotsuche zu,

dann gilt die Bemerkung i.a. nur bei der gleichen Strategie der Pivotsuche.)

Man bekommt mit der Definition

$$\det[A] := \{\det A : A \in [A]\}$$

sofort als notwendige Bedingung für die Durchführbarkeit des I.G.A.:

$$0 \notin \det[A], \tag{2.5}$$

denn wäre $0 \in \det[A]$, so gäbe es ein $A \in [A]$, für welches der reelle Gauß-Algorithmus bei jeder Strategie der Pivotsuche nicht durchführbar wäre.

Wegen (2.4) ist dann der I.G.A. nicht durchführbar für $[A]$.

In [32] wurde allerdings gezeigt, daß die Bedingung (2.5) nicht hinreichend ist für die Durchführbarkeit des I.G.A. Als Gegenbeispiel genügte die Matrix

$$\begin{pmatrix} 1 & [0,\frac{2}{3}] & [0,\frac{2}{3}] \\ [0,\frac{2}{3}] & 1 & [0,\frac{2}{3}] \\ [0,\frac{2}{3}] & [0,\frac{2}{3}] & 1 \end{pmatrix}.$$

Wir geben nun einige Klassen von Intervallmatrizen an, für die der I.G.A. durchführbar ist.

2.2.1 Intervalltridiagonalmatrizen

Für Intervalltridiagonalmatrizen ist die Frage nach der Durchführbarkeit des I.G.A. schon vollständig beantwortet. Es gilt nämlich der folgende

Satz 2.4 *Es sei $[A]$ eine $n \times n$ Intervalltridiagonalmatrix. Dann ist der I.G.A. genau dann ohne Pivotsuche durchführbar, wenn für*

$$[A_k] := \begin{pmatrix} [a_1] & [b_1] & \dots & 0 & 0 \\ [c_1] & [a_2] & [b_2] & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & [c_{k-2}] & [a_{k-1}] & [b_{k-1}] \\ 0 & \dots & \dots & [c_{k-1}] & [a_k] \end{pmatrix}$$

gilt:

$$0 \notin \det[A_k], k = 1(1)n.$$

Den Beweis findet man in [32].
Bemerkung: In [25] findet man für Satz 2.4 eine andere Formulierung:

Satz 2.4' *Der I.G.A. ohne Pivotsuche ist für eine Intervalltridiagonalmatrix* $[A]$ *genau dann durchführbar, wenn der reelle Gauß-Algorithmus ohne Pivotsuche für jedes* $A \in [A]$ *durchführbar ist.*

2.2.2 M- und H-Matrizen

Wir beginnen mit einigen

Definitionen.

1. Die Menge $Z^{n\times n}$ ist definiert als die Menge aller Matrizen $A = (a_{ij}) \in \mathbf{R}^{n\times n}$, für die gilt
$$a_{ij} \leq 0 \quad \text{für } i \neq j.$$

2. Die Matrix $A \in \mathbf{R}^{n\times n}$ heißt M-Matrix, falls
$$A \in Z^{n\times n}$$
gilt und außerdem die Inverse A^{-1} existiert mit
$$A^{-1} \geq 0.$$

3. Eine Intervallmatrix $[A]$ heißt M-Matrix, wenn jede reelle Punktmatrix $A \in [A]$ eine M-Matrix ist.

4. Eine Intervallmatrix $[A]$ heißt H-Matrix, wenn $\langle [A] \rangle$ eine M-Matrix ist.

Auch hier ist die Frage nach der Durchführbarkeit des I.G.A. bereits beantwortet. Es gilt der folgende

Satz 2.5 *Es sei* $[A] \in \mathbf{IR}^{n\times n}$ *eine H-Matrix. Dann ist der I.G.A. ohne Pivotsuche durchführbar.*

Einen Beweis findet man in [2]. Als wesentliches Hilfsmittel wird dabei folgendes Lemma benutzt.

Lemma 2.2 *(Fan'sches Lemma) Es sei* $A \in Z^{n\times n}$*. Dann sind folgende Aussagen äquivalent:*

1. *A^{-1} existiert und $A^{-1} \geq 0$.*

2. *Es existiert ein Vektor $u \in \mathbf{R}^n$, $u > 0$, mit $Au > 0$.*

Einen Beweis findet man in [15]. Aus dem Fan'schen Lemma ergeben sich zwei einfache

Folgerungen:

1. *Die Diagonalelemente einer M-Matrix sind positiv.*

2. *Ist A eine M-Matrix und $B \in Z^{n\times n}$ mit $A \leq B$, dann ist auch B eine M-Matrix, und es gilt $B^{-1} \leq A^{-1}$.*

Lemma 2.3 *Es sei $[A] \in \mathbf{IR}^{n\times n}$. Dann gelten folgende Aussagen:*

1. *Ist $[A]$ eine M-Matrix, so ist $[A]$ insbesondere eine H-Matrix.*

2. *Ist $\left\langle [A] \right\rangle$ streng diagonal dominant, d.h.*

$$\left\langle [a_{ii}] \right\rangle > \sum_{\substack{j=1,\\ j\neq i}}^{n} |[a_{ij}]|, \quad i = 1(1)n,$$

dann ist $[A]$ eine H-Matrix.

Die Beweise findet man z.B. in [25], wo noch mehr Klassen von Matrizen als H-Matrizen erkannt werden, die uns aber hier nicht weiter interessieren.

Wir wollen im Hinblick auf das nächste Lemma Satz 2.5 noch etwas weiter fassen.

Satz 2.5' *Es sei $[A] \in \mathbf{IR}^{n\times n}$ eine H-Matrix. Läßt man keine Pivotsuche zu, so sind die Punktmatrizen*

$$\left(\left\langle [A] \right\rangle_{ij}^{(k)} \right)_{k\leq i,j\leq n}, \quad k = 2(1)n,$$

allesamt M-Matrizen und es gilt

$$\left(\left\langle [A] \right\rangle_{ij}^{(k)} \right)_{k\leq i,j\leq n} \leq \left(\left\langle [A^{(k)}] \right\rangle_{ij} \right)_{k\leq i,j\leq n}, \quad k = 2(1)n. \tag{2.6}$$

Alle

$$\left([a_{ij}^{(k)}]\right)_{k\le i,j\le n}, \quad k = 2(1)n,$$

sind also wiederum H-Matrizen.

Beweis: Ist $A^{(1)} := A \in \mathbf{R}^{n\times n}$ eine M-Matrix, so ist nach einem Gauß-Schritt die Matrix

$$\left(a_{ij}^{(2)}\right)_{2\le i,j\le n} \in \mathbf{R}^{(n-1)\times(n-1)}$$

wieder eine M-Matrix. Den Beweis findet man in [39].

Den Beweis für die Beziehung (2.6) findet man in Proposition 6 in [30]. Der Beweis an sich folgt dann durch Induktion. □

Zum Schluß dieses Abschnitts kommen wir zu einem Lemma, welches im nächsten Kapitel bei der Beweisführung eine große Rolle spielen wird. Die erste Aussage des Lemmas ist dabei ein Spezialfall von Lemma 2c in [24] und wird auch in [18] als Lemma 1 angegeben. Die zweite Aussage ist neu.

Lemma 2.4 *Es sei* $[A] \in \mathbf{IR}^{n\times n}$ *eine H-Matrix. Dann gelten für* $[x]$, $[y] \in \mathbf{IR}^n$ *folgende Aussagen:*

1. *Es gilt*

$$\mathrm{q}\Big(IGA([A],[x]),\, IGA([A],[y])\Big) \le \Big\langle [A]\Big\rangle^{-1} \mathrm{q}([x],[y]).$$

2. *Ist* $\underline{A} = \overline{A} =: A$, *so gilt*

$$d\Big(IGA(A,[x])\Big) \le \Big\langle A\Big\rangle^{-1} d([x]).$$

Beweis: Zu 1.: Aufgrund von Satz 2.5 ist der I.G.A. durchführbar für jedes $[b] \in \mathbf{IR}^n$ und es gilt mit der Darstellung von H. Schwandt [35]

$$\begin{aligned} &IGA([A],[b]) = \\ &[D^{(1)}]([T^{(1)}]\cdots([D^{(n-1)}]([T^{(n-1)}]([D^{(n)}]([C^{(n-1)}](\cdots[C^{(1)}][b])\cdots) \end{aligned}$$

mit

$$[c_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j, \\ -[a_{ik}^{(k)}]/[a_{kk}^{(k)}] & \text{falls } j = k < i, \\ 0 & \text{sonst}, \end{cases}$$

$$[t_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j, \\ -[a_{kj}^{(n)}] & \text{falls } i = k < j, \\ 0 & \text{sonst}, \end{cases}$$

$$[d_{ij}^{(k)}] := \begin{cases} 1 & \text{falls } i = j \neq k, \\ 1/[a_{kk}^{(n)}] & \text{falls } i = j = k, \\ 0 & \text{sonst}. \end{cases}$$

Wir erhalten damit

$$\begin{aligned} &\mathrm{q}\Big(IGA([A],[x])\,,\, IGA([A],[y])\Big) = \\ &\mathrm{q}\Big([D^{(1)}]([T^{(1)}]\cdots([D^{(n-1)}]([T^{(n-1)}]([D^{(n)}]([C^{(n-1)}](\cdots[C^{(1)}][x])\cdots)\,, \\ &\quad [D^{(1)}]([T^{(1)}]\cdots([D^{(n-1)}]([T^{(n-1)}]([D^{(n)}]([C^{(n-1)}](\cdots[C^{(1)}][y])\cdots)\Big). \end{aligned}$$

Mit (1.2) erhalten wir dann sukzessiv

$$\begin{aligned} &\mathrm{q}\Big(IGA([A],[x])\,,\, IGA([A],[y])\Big) \leq \\ &|[D^{(1)}]|\cdot|[T^{(1)}]|\cdots|[D^{(n-1)}]|\cdot|[T^{(n-1)}]|\cdot|[D^{(n)}]|\cdot|[C^{(n-1)}]|\cdots|[C^{(1)}]| \\ &\times \mathrm{q}([x],[y]). \end{aligned}$$

Wir zeigen nun

$$\left.\begin{aligned} &|[D^{(1)}]|\cdot|[T^{(1)}]|\cdots|[D^{(n-1)}]|\cdot|[T^{(n-1)}]|\cdot|[D^{(n)}]|\cdot|[C^{(n-1)}]|\cdots|[C^{(1)}]| \\ &\leq \Big\langle [A]\Big\rangle^{-1}. \end{aligned}\right\} \tag{2.7}$$

Dann ist der erste Teil des Lemmas vollständig bewiesen.

Es seien dazu $i \in \{1, ..., n\}$ und e^i der i-te Einheitsvektor. Dann gilt aufgrund der Formeln des Gauß-Algorithmus

$$\Big\langle [A]\Big\rangle^{-1} e^i = \hat{D}^{(1)} \cdot \hat{T}^{(1)} \cdots \hat{D}^{(n-1)} \hat{T}^{(n-1)} \cdot \hat{D}^{(n)} \cdot \hat{C}^{(n-1)} \cdots \hat{C}^{(1)} e^i$$

mit

$$\Big\langle [A]\Big\rangle^{(k)} =: (\hat{a}_{ij}^{(k)}), \quad k = 1(1)n,$$

und

$$\hat{c}_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j, \\ -\hat{a}_{ik}^{(k)}/\hat{a}_{kk}^{(k)} & \text{falls } j = k < i, \\ 0 & \text{sonst}, \end{cases}$$

$$\hat{t}_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j, \\ -\hat{a}_{kj}^{(n)} & \text{falls } i = k < j, \\ 0 & \text{sonst}, \end{cases}$$

$$\hat{d}_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j \neq k, \\ 1/\hat{a}_{kk}^{(n)} & \text{falls } i = j = k, \\ 0 & \text{sonst}. \end{cases}$$

Nun gilt aufgrund von Satz 2.5'

$$\left| \frac{1}{[a_{kk}^{(n)}]} \right| = \frac{1}{\left\langle [a_{kk}^{(n)}] \right\rangle} \leq \frac{1}{\hat{a}_{kk}^{(n)}},$$

sowie

$$| - [a_{kj}^{(n)}]| = |[a_{kj}^{(n)}]| \leq |\hat{a}_{kj}^{(n)}| = -\hat{a}_{kj}^{(n)}$$

und

$$\left| -\frac{[a_{ik}^{(k)}]}{[a_{kk}^{(k)}]} \right| = \frac{|[a_{ik}^{(k)}]|}{\left\langle [a_{kk}^{(k)}] \right\rangle} \leq \frac{|\hat{a}_{ik}^{(k)}|}{\hat{a}_{kk}^{(k)}} = \frac{-\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}}.$$

Es folgt also $|[C^{(k)}]| \leq \hat{C}^{(k)}$, $|[T^{(k)}]| \leq \hat{T}^{(k)}$ für $k = 1(1)n-1$ und $|[D^{(k)}]| \leq \hat{D}^{(k)}$ für $k = 1(1)n$, woraus dann (2.7) folgt.

Zu 2.: Auch hier ist aufgrund von Satz 2.5 der I.G.A. durchführbar für jedes $[b] \in \mathbf{IR}^n$ und es gilt mit der Darstellung von H. Schwandt [35]

$$IGA(A,[b]) = D^{(1)}(T^{(1)} \cdots (D^{(n-1)}(T^{(n-1)}(D^{(n)}(C^{(n-1)}(\cdots C^{(1)}[b]) \cdots)$$

mit

$$c_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j, \\ -a_{ik}^{(k)}/a_{kk}^{(k)} & \text{falls } j = k < i, \\ 0 & \text{sonst}, \end{cases}$$

$$t_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j, \\ -a_{kj}^{(n)} & \text{falls } i = k < j, \\ 0 & \text{sonst}, \end{cases}$$

$$d_{ij}^{(k)} := \begin{cases} 1 & \text{falls } i = j \neq k, \\ 1/a_{kk}^{(n)} & \text{falls } i = j = k, \\ 0 & \text{sonst.} \end{cases}$$

Wir erhalten damit

$$d\Big(IGA(A,[x])\Big) = d\Big(D^{(1)}(T^{(1)} \cdots (D^{(n-1)}(T^{(n-1)}(D^{(n)}(C^{(n-1)}(\cdots C^{(1)}[x])\cdots)\Big).$$

Mit (1.3) erhalten wir dann sukzessiv

$$d\Big(IGA(A,[x])\Big) =$$
$$|D^{(1)}| \cdot |T^{(1)}| \cdots |D^{(n-1)}| \cdot |T^{(n-1)}| \cdot |D^{(n)}| \cdot |C^{(n-1)}| \cdots |C^{(1)}| \cdot d([x]).$$

Setzt man nun

$$\begin{aligned} [D^{(k)}] &:= D^{(k)}, \quad k = 1(1)n, \\ [C^{(k)}] &:= C^{(k)}, \quad k = 1(1)n-1, \\ [T^{(k)}] &:= T^{(k)}, \quad k = 1(1)n-1, \end{aligned}$$

so erhält man mit (2.7) die Behauptung. □

2.2.3 Intervallpfeilmatrizen

Wir wollen nun eine neue Klasse von Intervallmatrizen vorstellen, für die der I.G.A. durchführbar ist.

Satz 2.6 *Es sei* $[A]$ *eine* $n \times n$ *Intervallpfeilmatrix:*

$$[A] = \begin{pmatrix} [a_1] & 0 & \dots & 0 & [b_1] \\ 0 & [a_2] & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & [a_{n-1}] & [b_{n-1}] \\ [c_1] & \dots & \dots & [c_{n-1}] & [a_n] \end{pmatrix}.$$

Dann ist der I.G.A. ohne Pivotsuche genau dann durchführbar, wenn gilt:

$$0 \notin [a_i], i = 1(1)n-1, \tag{2.8}$$

und

$$0 \notin \det[A]. \tag{2.9}$$

Beweis: (2.8) ist gleichbedeutend damit, daß die ersten $n-1$ Intervall-Gauß-Schritte durchführbar sind, und wir erhalten

$$[A^{(n)}] = \begin{pmatrix} [a_1] & 0 & \dots & 0 & [b_1] \\ 0 & [a_2] & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & [a_{n-1}] & [b_{n-1}] \\ 0 & \dots & \dots & 0 & [\tilde{a}_n] \end{pmatrix}$$

mit $[\tilde{a}_n] = [a_n] - \sum_{i=1}^{n-1} \frac{[b_i][c_i]}{[a_i]}$. Das heißt also, daß in dem Ausdruck für $[\tilde{a}_n]$ jedes Intervall $([a_1], .., [a_n], [b_1], .., [b_{n-1}], [c_1], .., [c_{n-1}])$ genau einmal vorkommt. Nach [27] ist $[\tilde{a}_n]$ gleich dem Wertebereich der Funktion:

$$f : [a_1] \times .. \times [a_n] \times [b_1] \times .. \times [b_{n-1}] \times [c_1] \times .. \times [c_{n-1}] \to \mathbf{R},$$

$$(a_1, .., a_n, b_1, .., b_{n-1}, c_1, .., c_{n-1}) \mapsto a_n - \sum_{i=1}^{n-1} \frac{b_i c_i}{a_i}.$$

Nehmen wir an, der I.G.A. sei nicht durchführbar. Das wäre gleichbedeutend mit $0 \in [\tilde{a}_n]$. Somit wäre

$$0 \in W(f; [a_1] \times .. \times [a_n] \times [b_1] \times .. \times [b_{n-1}] \times [c_1] \times .. \times [c_{n-1}]).$$

Das würde aber heißen, daß es in $[A]$ eine reelle Punktmatrix A gibt, die sich nach Anwendung des reellen Gauß-Algorithmus als singulär herausstellen würde, was nach Voraussetzung ausgeschlossen ist. □

Beispiel 2.2 Wir betrachten folgende 3×3 Intervallpfeilmatrix

$$[A] = \begin{pmatrix} [1,2] & 0 & 2 \\ 0 & 2 & 3 \\ [0,1] & 1 & [-4,1] \end{pmatrix}.$$

Nach zwei Intervall-Gauß-Schritten erhält man

$$[\tilde{a}_3] = [-4,1] - \left(\frac{2}{[1,2]}[0,1] + \frac{3}{2} \right) = [-\frac{15}{2}, -\frac{1}{2}].$$

Es ist also $0 \notin [\tilde{a}_3]$, und der I.G.A. ist somit durchführbar. Betrachten wir die zugehörige Vergleichsmatrix

$$\left\langle [A] \right\rangle = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 2 & -3 \\ -1 & -1 & 0 \end{pmatrix},$$

dann bekommt man mit

$$\left\langle [A] \right\rangle^{-1} = \frac{1}{7} \begin{pmatrix} 3 & -2 & -4 \\ -3 & 2 & -3 \\ -2 & -1 & -2 \end{pmatrix} \not\geq 0,$$

daß $\left\langle [A] \right\rangle$ keine M-Matrix und $[A]$ somit keine H-Matrix ist.

Kapitel 3

Das lineare Komplementaritätsproblem mit Intervalleinträgen

Wir wollen am Anfang dieses Kapitels die wichtigsten Resultate für lineare Komplementaritätsprobleme ohne Intervalleinträge zusammenfassen. Für weitere Ergebnisse verweisen wir auf die Bücher [10] und [28].

3.1 Das lineare Komplementaritätsproblem

Ein lineares Komplementaritätsproblem (LCP) kann man auf zwei Arten formulieren:

Mathematische Formulierung A: Gegeben sind eine Matrix $M \in \mathbf{R}^{n\times n}$ und ein Vektor $q \in \mathbf{R}^n$, und gesucht sind Vektoren $w, z \in \mathbf{R}^n$ (oder gesucht ist ein Beweis zur Nichtexistenz von Vektoren $w, z \in \mathbf{R}^n$) mit

$$\left.\begin{array}{rcl} w - Mz & = & q, \\ w^{\mathrm{T}} z & = & 0, \\ w \geq 0 \quad \text{und} & & z \geq 0. \end{array}\right\} \tag{3.1}$$

Es ist also $w_i = 0$ oder $z_i = 0$, $i = 1(1)n$. Daher der Name komplementär.

Mathematische Formulierung B: Gegeben sind eine Matrix $M \in \mathbf{R}^{n\times n}$

und ein Vektor $q \in \mathbf{R}^n$, und gesucht ist ein Vektor $z \in \mathbf{R}^n$ (oder gesucht ist ein Beweis zur Nichtexistenz eines Vektors $z \in \mathbf{R}^n$) mit

$$\left.\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z & = & 0. \end{array}\right\} \tag{3.2}$$

Lemma 3.1 *Gegeben seien eine Matrix $M \in \mathbf{R}^{n \times n}$ und ein Vektor $q \in \mathbf{R}^n$. Dann gelten folgende Aussagen:*

1. *Lösen die Vektoren $w, z \in \mathbf{R}^n$ das Problem (3.1), so löst der Vektor z das Problem (3.2).*
2. *Löst der Vektor $z \in \mathbf{R}^n$ das Problem (3.2), so lösen die Vektoren w, z mit $w := q + Mz$ das Problem (3.1).*

Beweis: Zu 1.: Seien die Vektoren $w, z \in \mathbf{R}^n$ eine Lösung von (3.1). Dann gilt:

$$\begin{array}{rcl} w - Mz & = & q, \\ w^{\mathrm{T}} z & = & 0, \\ w \geq 0 \quad \text{und} & & z \geq 0. \end{array}$$

Insbesondere gilt dann:

$$\begin{array}{rcl} q + Mz = w & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z = w^{\mathrm{T}} z & = & 0. \end{array}$$

Somit löst z das Problem (3.2).

Zu 2.: Sei $z \in \mathbf{R}^n$ eine Lösung von (3.2). Dann gilt:

$$\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z & = & 0. \end{array}$$

Es gilt dann mit $w := q + Mz$:

$$\begin{aligned} w - Mz &= q, \\ w^{\mathrm{T}} z = (q + Mz)^{\mathrm{T}} z &= 0, \\ w = q + Mz \geq 0 \quad \text{und} \quad z &\geq 0. \end{aligned}$$

Somit lösen die Vektoren w, z das Problem (3.1). □

Ob ein LCP lösbar oder sogar eindeutig lösbar ist, hängt davon ab, zu welcher Klasse von Matrizen die Matrix M gehört. Daher beginnen wir mit einigen

Definitionen.

1. Eine Matrix $M \in \mathbf{R}^{n \times n}$ heißt positiv definit, falls für alle $x \in \mathbf{R}^n - \{0\}$ gilt:
$$x^{\mathrm{T}} M x > 0.$$

2. Es sei $J \subset \{1, ..., n\}$. Dann bezeichnen wir mit $M(J)$ die Matrix, die aus M durch Streichen der j-ten Zeile und der j-ten Spalte hervorgeht für alle $j \in J$.

3. Unter einem Hauptminor verstehen wir $\det M(J)$ mit $J \subset \{1, ..., n\}$. Den führenden Hauptminoren entsprechen die Indexmengen $J = \{i, ..., n\}, i = 2(1)n$ und $J = \emptyset$.

4. Eine Matrix $M \in \mathbf{R}^{n \times n}$ heißt P-Matrix, wenn alle Hauptminoren positiv sind.

Beispiel 3.1 Die Matrix

$$M = \begin{pmatrix} 2 & -3 \\ 1 & -1 \end{pmatrix}$$

hat die führenden Hauptminoren $\det 2 = 2$ und $\det M = 1$. Es gilt jedoch $\det M(\{1\}) = -1$. M ist also keine P-Matrix.

Satz 3.1 *Gegeben sei $M \in \mathbf{R}^{n \times n}$. Dann gelten folgende Aussagen:*

1. *Ist M eine P-Matrix, so besitzt für jedes $q \in \mathbf{R}^n$ das zu M und q gehörende LCP genau eine Lösung.*

2. *Besitzt für jedes $q \in \mathbf{R}^n$ das zu M und q gehörende LCP genau eine Lösung, so ist M eine P-Matrix.*

Einen Beweis findet man in [10], Theorem 3.3.7.

Leider findet man in [10] auch die Aussage:

> At present, there is no efficient test to determine whether an arbitrary matrix is a P-matrix.

Daher ist es wichtig zu wissen, ob handhabbarere und bekanntere[1] Klassen von Matrizen enthalten sind in der Klasse der P-Matrizen. Wichtig ist in diesem Zusammenhang der folgende

Satz 3.2 *Es sei $M \in \mathbf{R}^{n\times n}$. Dann gelten folgende Aussagen:*

1. *Ist M eine H-Matrix mit positiven Diagonalelementen, so ist M eine P-Matrix.*

2. *Ist M positiv definit, so ist M eine P-Matrix.*

Den Beweis zu Teil 1 findet man in [10], Theorem 3.3.15. Es wird dort sogar mehr gezeigt. Zunächst wird gezeigt, daß sogenannte *diagonally stable* Matrizen P-Matrizen sind, und in Theorem 3.3.15 wird dann gezeigt, daß H-Matrizen mit positiven Diagonalelementen *diagonally stable* sind. Auf die Definition von *diagonally stable* wollen wir hier nicht eingehen.

Der Beweis von Teil 2 ist eine sehr einfache Folgerung aus

Lemma 3.2 *Es sei $M \in \mathbf{R}^{n\times n}$. Dann ist M genau dann eine P-Matrix, wenn für jedes $x \in \mathbf{R}^n$ gilt:*

$$[x_i(Mx)_i \leq 0 \, \textit{für alle} \; i\,] \Rightarrow [x = 0].$$

[1] Es soll jetzt aber nicht der Eindruck entstehen, daß P-Matrizen lediglich eingeführt wurden, um theoretisch die Frage nach der eindeutigen Lösbarkeit eines LCPs zu beantworten. Tatsächlich findet man schon in [19] den interessanten Satz, daß eine Abbildung $f : \Omega \to \mathbf{R}^n$, wobei $\Omega = \{x \in \mathbf{R}^n : p \leq x \leq q\}$ ist, injektiv ist, falls die Jacobimatrix für jedes $x \in \Omega$ eine P-Matrix ist.

Den Beweis zu Lemma 3.2 findet man in [19].

Die Umkehrung von Teil 2 von Satz 3.2 ist falsch, wie das abschließende Beispiel zeigen wird.

Beispiel 3.2 Die Matrix

$$M = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$$

ist offensichtlich eine P-Matrix. Wählt man aber $x = (1,1)^{\mathrm{T}}$, so erhält man $x^{\mathrm{T}}Mx = -1 < 0$. Das Beispiel stammt aus [10].

3.2 Die Lösungsmengen L_A und L_B

In diesem Abschnitt sind eine Intervallmatrix $[M] \in \mathbf{IR}^{n\times n}$ und ein Intervallvektor $[q] \in \mathbf{IR}^n$ gegeben, und wir betrachten die Mengen von linearen Komplementaritätsproblemen:

$$\left.\begin{array}{rcl} w - Mz & = & q, \\ w^{\mathrm{T}}z & = & 0, \\ w \geq 0 \quad \text{und} & & z \geq 0, \end{array}\right\} \quad M \in [M],\, q \in [q], \tag{3.3}$$

bzw.

$$\left.\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}}z & = & 0, \end{array}\right\} \quad M \in [M],\, q \in [q]. \tag{3.4}$$

Wir definieren die Lösungsmengen

$$\left.\begin{array}{rcl} L_A & := & \left\{ \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbf{R}^{2n} : \text{ Es gibt ein } M \in [M] \text{ und ein } q \in [q] \text{ mit} \right. \\ & & \left. w \geq 0, z \geq 0, w^{\mathrm{T}}z = 0,\, w - Mz = q \right\} \end{array}\right\} \tag{3.5}$$

und

$$\left.\begin{array}{rcl} L_B & := & \left\{ z \in \mathbf{R}^n : \text{ Es gibt ein } M \in [M] \text{ und ein } q \in [q] \text{ mit} \right. \\ & & \left. z \geq 0,\, q + Mz \geq 0,\, (q + Mz)^{\mathrm{T}}z = 0 \right\}. \end{array}\right\} \tag{3.6}$$

Der Index A bzw. B soll dabei an die zugrunde liegende mathematische Formulierung A bzw. B erinnern.

Lemma 3.3 *Es seien* $[M] \in \mathbf{IR}^{n\times n}$ *und* $[q] \in \mathbf{IR}^n$. L_A *und* L_B *seien wie in (3.5) bzw. (3.6) definiert. Dann gilt:*

$$z \in L_B \Leftrightarrow \textit{Es gibt ein } w \in \mathbf{R}^n \textit{ mit } \begin{pmatrix} w \\ z \end{pmatrix} \in L_A.$$

Beweis: a) Es sei $z \in L_B$. Nach (3.6) gibt es dann ein $M \in [M]$ und ein $q \in [q]$ mit $z \geq 0$, $q + Mz \geq 0$, $(q + Mz)^{\mathrm{T}} z = 0$. Setzen wir nun

$$w := q + Mz,$$

so gilt:

$$\left.\begin{array}{r} \text{Es existieren } M \in [M],\, q \in [q] \\ \text{mit } w - Mz = q \\ \text{und } w \geq 0, z \geq 0, w^{\mathrm{T}} z = 0 \end{array}\right\} \Rightarrow \begin{pmatrix} w \\ z \end{pmatrix} \in L_A.$$

b) Es sei $(w, z)^{\mathrm{T}} \in L_A$. Nach (3.5) gibt es dann ein $M \in [M]$ und ein $q \in [q]$ mit $w \geq 0, z \geq 0, w^{\mathrm{T}} z = 0$, $w - Mz = q$. Also gilt:

$$\left.\begin{array}{r} \text{Es existieren } M \in [M],\, q \in [q] \\ \text{mit } q + Mz = w \geq 0 \\ (q + Mz)^{\mathrm{T}} z = w^{\mathrm{T}} z = 0 \\ \text{und } z \geq 0 \end{array}\right\} \Rightarrow z \in L_B.$$

□

Um die Gestalt von L_A beschreiben zu können, werden wir denselben Weg wählen wie bei den linearen Intervallgleichungssystemen. Dabei werden wir explizit benutzen, daß Satz 2.1 auch für nichtquadratische Intervallmatrizen gilt.

Satz 3.3 *Es seien* $[M] \in \mathbf{IR}^{n\times n}$, I *die* $n \times n$ *Einheitsmatrix,* $[q] \in \mathbf{IR}^n$ *und* L_A *gemäß (3.5) definiert. Dann sind folgende Aussagen äquivalent:*

a) $\begin{pmatrix} w \\ z \end{pmatrix} \in L_A.$

b) $\left| \left(I \vdots - mid([M]) \right) \begin{pmatrix} w \\ z \end{pmatrix} - mid([q]) \right| \leq \frac{1}{2} d([M]) |z| + \frac{1}{2} d([q])$

und $w^{\mathrm{T}} z = 0$, $w \geq 0$, $z \geq 0$.

c) $(I \,\vdots\, -[M]) \begin{pmatrix} w \\ z \end{pmatrix} \cap [q] \neq \emptyset$ *und* $w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0.$

Dabei setzen wir für $[B] \in \mathbf{IR}^{n \times n}$

$$\big(I \,\vdots\, -[B]\big)) := \begin{pmatrix} 1 & 0 & \cdots & 0 & -[b_{11}] & \cdots & -[b_{1n}] \\ 0 & 1 & \ddots & \vdots & -[b_{21}] & \cdots & -[b_{2n}] \\ \vdots & \ddots & \ddots & 0 & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -[b_{n1}] & \cdots & -[b_{nn}] \end{pmatrix} \in \mathbf{IR}^{n \times 2n}.$$

Beweis: Wir setzen $[A] := (I \,\vdots\, -[M]) \in \mathbf{IR}^{n \times 2n}$ und $x := \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbf{R}^{2n}$. Wir beweisen zuerst die Äquivalenz $a) \Leftrightarrow b)$. Es gilt:

$$\begin{aligned} x = \begin{pmatrix} w \\ z \end{pmatrix} \in L_A \quad &\Leftrightarrow \quad \begin{pmatrix} w \\ z \end{pmatrix} = x \in \Sigma([A],[q]) \\ &\qquad \text{und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0 \\ &\Leftrightarrow \quad |\mathrm{mid}([A])x - \mathrm{mid}([q])| \leq \tfrac{1}{2} d([A])|x| + \tfrac{1}{2} d([q]) \\ &\qquad \text{mit } x = \begin{pmatrix} w \\ z \end{pmatrix} \text{ und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0 \\ &\Leftrightarrow \quad \left| \big(I \,\vdots\, -\mathrm{mid}([M])\big) \begin{pmatrix} w \\ z \end{pmatrix} - \mathrm{mid}([q]) \right| \leq \\ &\qquad \tfrac{1}{2}\big(O \,\vdots\, d([M])\big) \left| \begin{pmatrix} w \\ z \end{pmatrix} \right| + \tfrac{1}{2} d([q]) = \tfrac{1}{2} d([M])|z| + \tfrac{1}{2} d([q]) \\ &\qquad \text{und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0. \end{aligned}$$

Dabei galt der zweite Äquivalenzpfeil aufgrund der Äquivalenz $a) \Leftrightarrow b)$ von Satz 2.1. Nun beweisen wir die Äquivalenz $a) \Leftrightarrow c)$. Es gilt:

$$\begin{aligned} x = \begin{pmatrix} w \\ z \end{pmatrix} \in L_A \quad &\Leftrightarrow \quad \begin{pmatrix} w \\ z \end{pmatrix} = x \in \Sigma([A],[q]) \\ &\qquad \text{und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0 \\ &\Leftrightarrow \quad [A]x \cap [q] \neq \emptyset \\ &\qquad \text{mit } x = \begin{pmatrix} w \\ z \end{pmatrix} \text{ und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0 \\ &\Leftrightarrow \quad (I \vdots -[M]) \begin{pmatrix} w \\ z \end{pmatrix} \cap [q] \neq \emptyset \\ &\qquad \text{und } w^{\mathrm{T}} z = 0,\ w \geq 0,\ z \geq 0. \end{aligned}$$

Dabei galt der zweite Äquivalenzpfeil aufgrund der Äquivalenz $a) \Leftrightarrow c)$ von Satz 2.1. □

Wir wollen jetzt die Äquivalenz $a) \Leftrightarrow c)$ aus Satz 3.3 anwenden, um geometrische Aussagen über L_A zu gewinnen. Dazu bemerken wir, daß für Intervalle $[a], [b] \in \mathbf{IR}$ gilt:

$$[a] \cap [b] \neq \emptyset \Leftrightarrow \underline{a} \leq \overline{b} \text{ und } \underline{b} \leq \overline{a}.$$

Es sei $\begin{pmatrix} w \\ z \end{pmatrix} \in L_A$. Dann gilt mit der Äquivalenz $a) \Leftrightarrow c)$ aus Satz 3.3

$$\underline{w_i - \sum_{j=1}^{n} [m_{ij}] z_j \leq \overline{q}_i}, \quad i \in \{1, ..., n\},$$

und

$$\overline{\underline{q}_i \leq w_i - \sum_{j=1}^{n} [m_{ij}] z_j}, \quad i \in \{1, ..., n\}.$$

Wegen $z_j \geq 0$ folgt

$$w_i - \sum_{j=1}^{n} \overline{m}_{ij} z_j \leq \overline{q}_i, \quad i \in \{1, ..., n\} \tag{3.7}$$

und

$$\underline{q}_i \leq w_i - \sum_{j=1}^{n} \underline{m}_{ij} z_j, \quad i \in \{1, ..., n\}. \tag{3.8}$$

Wegen der Komplementarität hat man danach 2^n Fälle zu untersuchen, die alle einen Beitrag zu L_A liefern können.

Beispiel 3.3 a)

$$[M] = \begin{pmatrix} [\frac{3}{4}, 1] & [-\frac{1}{8}, 0] \\ [-\frac{1}{8}, 0] & [\frac{3}{4}, 1] \end{pmatrix}; \; [q] = \begin{pmatrix} [-1, -\frac{1}{10}] \\ [-3, -\frac{1}{2}] \end{pmatrix}.$$

Mit (3.7) und (3.8) bekommt man vier Ungleichungen.

$$\left.\begin{array}{ll} \text{(I)} \; w_1 - z_1 \leq -\frac{1}{10}\,; & \text{(II)} \; -1 \leq w_1 - (\frac{3}{4}\, z_1 - \frac{1}{8}\, z_2)\,; \\ \text{(III)} \; w_2 - z_2 \leq -\frac{1}{2}\,; & \text{(IV)} \; -3 \leq w_2 - (-\frac{1}{8}\, z_1 + \frac{3}{4}\, z_2)\,. \end{array}\right\} \tag{3.9}$$

Aufgrund der Komplementarität müssen nun vier Fälle untersucht werden:

$$1.\, w_1 = 0,\, w_2 = 0.$$
$$2.\, w_1 = 0,\, z_2 = 0.$$
$$3.\, z_1 = 0,\, w_2 = 0.$$
$$4.\, z_1 = 0,\, z_2 = 0.$$

Die Fälle mit $z_1 = 0$ bzw. $z_2 = 0$ führen eingesetzt in (3.9)(I) bzw. (3.9)(III) zu einem Widerspruch, da $w_1 \geq 0$ bzw. $w_2 \geq 0$ vorausgesetzt ist. Es muß also $z_1 > 0$ und $z_2 > 0$ gelten, was aufgrund der Komplementarität $w_1 = 0$ und $w_2 = 0$ impliziert. Eingesetzt in (3.9) ergibt sich:

$$\begin{array}{ll} \text{(I)} \; \frac{1}{10} \leq z_1\,; & \text{(II)} \; 6z_1 - 8 \leq z_2\,; \\ \text{(III)} \; \frac{1}{2} \leq z_2\,; & \text{(IV)} \; z_2 \leq \frac{1}{6} z_1 + 4\,. \end{array}$$

Es gilt also

$$L_A = \left\{ \begin{pmatrix} 0 \\ 0 \\ z_1 \\ z_2 \end{pmatrix} : z_1, z_2 \in L_1 \text{ (siehe Abbildung 3.1)} \right\}.$$

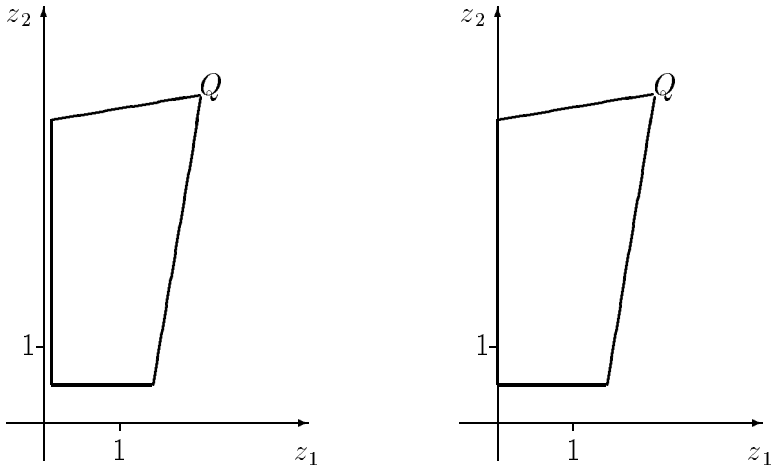

Abbildung 3.1: Das Polygon links bzw. rechts beschreibt L_1 bzw. L_2.

b) Wir ändern nun Beispiel a) an einer Stelle ab. Es soll jetzt

$$[q] = \begin{pmatrix} [-1,1] \\ [-3,-\frac{1}{2}] \end{pmatrix}$$

sein. Das hat zur Folge, daß sich bei (3.9) lediglich (I) verändert zu

$$\text{(I')}\ w_1 - z_1 \leq 1.$$

Wie in a) müssen vier Fälle untersucht werden:

$$1.\ w_1 = 0,\ w_2 = 0.$$
$$2.\ w_1 = 0,\ z_2 = 0.$$
$$3.\ z_1 = 0,\ w_2 = 0.$$
$$4.\ z_1 = 0,\ z_2 = 0.$$

Die Fälle 2. und 4. beinhalten $z_2 = 0$. Dann bekommt man aber in (3.9)(III) einen Widerspruch, da $w_2 \geq 0$ vorausgesetzt ist. Es muß also $z_2 > 0$ gelten,

was dann $w_2 = 0$ impliziert.

Wir betrachten nun Fall 1.: $w_1 = 0$, $w_2 = 0$. Man erhält:

$$\begin{aligned} &\text{(I')} \ -1 \le z_1\,; \qquad \text{(II)} \ 6z_1 - 8 \le z_2\,; \\ &\text{(III)} \ \tfrac{1}{2} \le z_2\,; \qquad \text{(IV)} \ z_2 \le \tfrac{1}{6}z_1 + 4\,. \end{aligned}$$

Der Fall 3.: $z_1 = 0$, $w_2 = 0$, liefert:

$$\begin{aligned} &\text{(I')} \ w_1 \le 1\,; \qquad \text{(II)} \ -8w_1 - 8 \le z_2\,; \\ &\text{(III)} \ \tfrac{1}{2} \le z_2\,; \qquad \text{(IV)} \ z_2 \le 4\,. \end{aligned}$$

Man bekommt also

$$L_A = \left\{ \begin{pmatrix} 0 \\ 0 \\ z_1 \\ z_2 \end{pmatrix} : z_1, z_2 \in L_2 \text{ (siehe Abbildung 3.1)} \right\} \cup \begin{pmatrix} [0,1] \\ 0 \\ 0 \\ [\frac{1}{2}, 4] \end{pmatrix}.$$

Wir wollen für spätere Zwecke noch anmerken, daß der Punkt Q in Abbildung 3.1 die Koordinaten

$$Q = \left(\frac{72}{35}, 4 + \frac{12}{35}\right) \approx (2.0571428, 4.3428571)$$

besitzt.

Bevor wir eine allgemeine Aussage über L_A machen, benötigen wir noch die

Definition: Eine Intervallmatrix $[M]$ heißt P-Matrix, wenn jede Punktmatrix $M \in [M]$ eine P-Matrix ist.

Satz 3.4 *Ist $[M] \in \mathbf{IR}^{n\times n}$ eine P-Matrix, so ist L_A die endliche Vereinigung von konvexen, kompakten und zusammenhängenden Polytopen.*

Beweis: Es sei $\alpha \subseteq \{1, ..., n\}$. Dann definieren wir die Intervallkomplementärmatrix $C_{[M]}(\alpha) \in \mathbf{IR}^{n\times n}$ mit

$$\left(C_{[M]}(\alpha)\right)_{.i} := \begin{cases} -[M]_{.i}\,, & \text{falls } i \in \alpha, \\ I_{.i}\,, & \text{falls } i \notin \alpha, \end{cases} \qquad i = 1(1)n.$$

Mit $I_{.i}$ bezeichnen wir dabei die i-te Spalte der $n \times n$ Einheitsmatrix, und mit $[M]_{.i}$ bezeichnen wir die i-te Spalte der Intervallmatrix $[M]$.

Da $[M]$ eine P-Matrix ist, ist für jedes $M \in [M]$ die Matrix $C_M(\alpha) \in C_{[M]}(\alpha)$ eine P-Matrix und somit insbesondere invertierbar. (Dies folgt unmittelbar aus der Definition einer P-Matrix.)

Daher ist $C_{[M]}(\alpha)$ regulär, und nach Teil 3 von Satz 2.2 ist die Lösungsmenge

$$\Sigma(C_{[M]}(\alpha), [q]) \cap \mathbf{R}^n_{\geq 0},$$

falls sie nichtleer ist, kompakt, konvex, zusammenhängend und ein Polytop.

Wir erweitern nun diese n-dimensionale Menge komplementär auf die Dimension $2n$. Dazu definieren wir eine durch $\alpha \subseteq \{1, ..., n\}$ bestimmte Projektion von $\mathbf{R}^{2n}$ auf $\mathbf{R}^n$:

$$\left(p_\alpha \begin{pmatrix} w \\ z \end{pmatrix} \right)_i := \begin{cases} w_i, & \text{falls } i \notin \alpha, \\ z_i, & \text{falls } i \in \alpha, \end{cases} \qquad i = 1(1)n,$$

für $w, z \in \mathbf{R}^n$. Damit setzen wir

$$\begin{aligned} \Sigma_\alpha \quad := \quad & \left\{ \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbf{R}^{2n} : p_\alpha \begin{pmatrix} w \\ z \end{pmatrix} \in \Sigma(C_{[M]}(\alpha), [q]) \cap \mathbf{R}^n_{\geq 0}\,, \right. \\ & w, z \in \mathbf{R}^n_{\geq 0}, \text{ und für } i = 1(1)n \text{ gilt:} \\ & \left. w_i = 0 \text{ falls } i \in \alpha,\ z_i = 0 \text{ falls } i \notin \alpha \right\}. \end{aligned}$$

Dann ist

$$L_A = \bigcup_{\alpha \subseteq \{1,\ldots,n\}} \Sigma_\alpha.$$

Da es bekanntlich 2^n Möglichkeiten gibt, ein $\alpha \subseteq \{1, ..., n\}$ auszuwählen, gibt es genau 2^n Intervallkomplementärmatrizen und damit höchstens 2^n Mengen Σ_α. □

Anhand von Beispiel 3.3b) sieht man, daß L_A selbst i.a. nicht konvex ist. Dazu wählen wir

$$w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ z_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix},\ w_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\ z_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Es gilt

$$\begin{pmatrix} w_1 \\ z_1 \end{pmatrix}, \begin{pmatrix} w_2 \\ z_2 \end{pmatrix} \in L_A,$$

aber es ist

$$\frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix} =: \begin{pmatrix} w \\ z \end{pmatrix} \notin L_A,$$

denn es gilt $w^{\mathrm{T}} z = \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot \frac{3}{2} = \frac{1}{4} \neq 0$.

Wir wollen weiter zeigen, daß auch L_B nicht konvex sein muß. Dazu betrachten wir

Beispiel 3.4

$$[M] = \begin{pmatrix} [\frac{1}{8}, 1] & [-\frac{1}{4}, -\frac{1}{5}] \\ [-\frac{1}{4}, -\frac{1}{10}] & 1 \end{pmatrix}; [q] = \begin{pmatrix} [-3, -1] \\ [1, 2] \end{pmatrix}.$$

Zunächst betrachten wir wieder L_A mit Hilfe von (3.7) und (3.8). Wir erhalten somit wieder vier Ungleichungen:

$$\begin{aligned} &\text{(I)}\ w_1 - (z_1 - \tfrac{1}{5} z_2) \leq -1\,; \qquad \text{(II)}\ -3 \leq w_1 - (\tfrac{1}{8} z_1 - \tfrac{1}{4} z_2)\,; \\ &\text{(III)}\ w_2 - (-\tfrac{1}{10} z_1 + z_2) \leq 2\,; \qquad \text{(IV)}\ 1 \leq w_2 - (-\tfrac{1}{4} z_1 + z_2)\,; \end{aligned}$$

und untersuchen vier Fälle:

$$\begin{aligned} &1.\ w_1 = 0,\ w_2 = 0. \\ &2.\ w_1 = 0,\ z_2 = 0. \\ &3.\ z_1 = 0,\ w_2 = 0. \\ &4.\ z_1 = 0,\ z_2 = 0. \end{aligned}$$

Die Fälle 3. und 4. liefern wegen (I) keinen Beitrag zu L_A. In Fall 1. bekommt man:

$$\begin{aligned} &\text{(I)}\ z_2 \leq 5 z_1 - 5\,; \qquad \text{(II)}\ \tfrac{1}{2} z_1 - 12 \leq z_2\,; \\ &\text{(III)}\ \tfrac{1}{10} z_1 - 2 \leq z_2\,; \qquad \text{(IV)}\ z_2 \leq \tfrac{1}{4} z_1 - 1\ . \end{aligned}$$

In Fall 2. bekommt man:

$$\text{(I) } 1 \leq z_1\,; \qquad \text{(II) } z_1 \leq 24\,;$$
$$\text{(III) } w_2 \leq -\tfrac{1}{10}\,z_1 + 2\,; \qquad \text{(IV) } -\tfrac{1}{4}\,z_1 + 1 \leq w_2\,.$$

Man erhält somit

$$L_A = \left\{ \begin{pmatrix} 0 \\ 0 \\ z_1 \\ z_2 \end{pmatrix} : z_1, z_2 \in L_3 \text{ (siehe Abbildung 3.2)} \right\} \cup \left\{ \begin{pmatrix} 0 \\ w_2 \\ z_1 \\ 0 \end{pmatrix} : w_2, z_1 \in L_4 \text{ (siehe Abbildung 3.2)} \right\}.$$

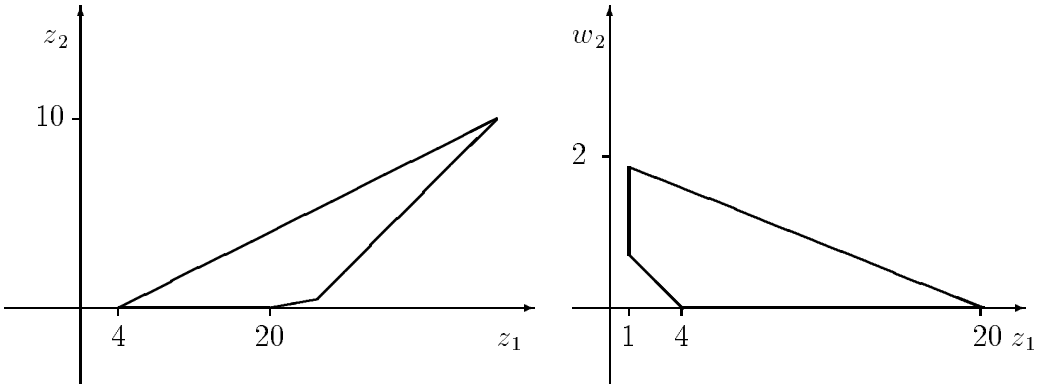

Abbildung 3.2: Das Polygon links bzw. rechts beschreibt L_3 bzw. L_4

Mit Lemma 3.3 bekommt man dann

$$L_B = L_3 \cup \begin{pmatrix} [1,20] \\ 0 \end{pmatrix}.$$

Diese Menge ist nicht konvex.

3.2.1 Intervallmäßige Einschließung von L_A bzw. L_B

In diesem Abschnitt werden wir in Satz 3.7 und Satz 3.8 zeigen, wie man die Lösungsmengen L_A (3.5) und L_B (3.6) intervallmäßig einschließen kann, falls $[M]$ eine H-Matrix mit $\underline{m_{ii}} > 0, i = 1(1)n$, ist.

Wir wissen ja bereits, daß unter dieser Voraussetzung für $[M]$ (siehe Satz 3.1 und Satz 3.2) jedes LCP mit $M \in [M], q \in \mathbf{R}^n$ eine (sogar eindeutige) Lösung besitzt. Daher sind die Mengen L_A und L_B nichtleer. Außerdem wissen wir aufgrund von Satz 3.2 und Satz 3.4, daß L_A und (wegen Lemma 3.3 dann auch) L_B kompakt sind. Es hat also auch Sinn, sie intervallmäßig einschließen zu wollen.

Dazu betrachten wir Funktionen $f : \mathbf{R}^n \to \mathbf{R}^n$, die außer von den Variablen $x_1, ..., x_n$ noch von Parametern $m_{11}, ..., m_{nn}, q_1, ..., q_n$ abhängen können, die unabhängig voneinander in Intervallen variieren:

$$\left.\begin{array}{c} f(x) = f(x_1, ..., x_n; m_{11}, ..., m_{nn}, q_1, ..., q_n) \quad \text{mit} \\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n,\ m_{ij} \in [m_{ij}],\ q_i \in [q_i],\ 1 \le i, j \le n. \end{array}\right\} \tag{3.10}$$

Definition. $f : \mathbf{R}^n \to \mathbf{R}^n$ habe die Bauart wie in (3.10).

1. $f(x)$ besitzt eine <u>intervallmäßige Auswertung</u>, wenn für alle $[x] = \begin{pmatrix} [x_1] \\ \vdots \\ [x_n] \end{pmatrix} \in \mathbf{IR}^n$ der Ausdruck

   $$f([x]) := f([x_1], .., [x_n]; [m_{11}], .., [m_{nn}], [q_1], .., [q_n])$$

   definiert ist.

2. $f(x)$ wird eine <u>P-Kontraktion</u> genannt, falls $f(x)$ eine intervallmäßige Auswertung besitzt und eine nichtnegative Matrix P existiert mit $\rho(P) < 1$, so daß

   $$\mathrm{q}\Big(f([x]), f([y])\Big) \le P \cdot \mathrm{q}\Big([x], [y]\Big)$$

   gilt für alle $[x], [y] \in \mathbf{IR}^n$. P nennen wir Kontraktionsmatrix.

Bevor wir zu den angekündigten Sätzen 3.7 und 3.8 kommen, zitieren wir noch zwei Sätze, die wir bei der Beweisführung benötigen.

Satz 3.5 *Es sei $f : \mathbf{R}^n \to \mathbf{R}^n$ eine P-Kontraktion. Dann gilt:*

1. *Es existiert genau ein Intervallvektor $[x^*] \in \mathbf{IR}^n$ mit*

$$f([x^*]) = [x^*],$$

und für jeden beliebigen Anfangsintervallvektor $[x^0] \in \mathbf{IR}^n$ konvergiert die Iteration

$$[x^{k+1}] := f([x^k]),\ k = 0, 1, 2...,$$

gegen $[x^*]$.

2. *Es gilt*

$$\{x \in \mathbf{R}^n : x = f(x),\ q \in [q],\ M \in [M]\} \subseteq [x^*].$$

Den Beweis zu Teil 1 findet man in [2], Seite 134/Theorem 4 und den Beweis zu Teil 2 findet man in [2], Seite 136/Corollary 6.

Satz 3.6 *Es seien $R, H, N \in \mathbf{R}^{n\times n}$ mit $R = H - N$. Weiter seien R und H invertierbar und $H^{-1}N \geq 0$. Dann gilt:*

$$R^{-1}N \geq 0 \Leftrightarrow \rho(H^{-1}N) < 1.$$

Einen Beweis findet man in [6], Theorem 7.5.2.

Damit kommen wir zum Hauptsatz dieses Abschnitts. Die wesentliche Idee stammt dabei aus [28], Abschnitt 9.2. Dort betrachtet man das Fixpunktproblem

$$f(x) = x,\ x \in \mathbf{R}^n,$$

mit

$$f(x) = \Big(I + M\Big)^{-1}\Big((I - M)|x| - q\Big)$$

unter der Voraussetzung, $M \in \mathbf{R}^{n\times n}$ sei symmetrisch und positiv definit. Es wird gezeigt, daß genau ein Fixvektor x^* existiert und die Vektoren

$$w := |x^*| - x^*, \quad z := |x^*| + x^*$$

die eindeutige Lösung des zu $M \in \mathbf{R}^{n\times n}$ und $q \in \mathbf{R}^n$ gehörenden LCPs bilden.

In Satz 3.7 und Satz 3.8 werden wir diesen Sachverhalt auf die Intervallrechnung übertragen. Dabei werden wir insbesondere zeigen, daß Theorem 9.1 aus [28] auch zum Ziel führt, wenn M eine H-Matrix mit positiven Diagonalelementen ist.

Satz 3.7 *Es seien* $[q] \in \mathbf{IR}^n$ *und* $[M] \in \mathbf{IR}^{n\times n}$ *gegeben. Dabei möge* $[M]$ *folgende Voraussetzungen erfüllen:*

(V) $\quad [M]$ *ist eine H-Matrix mit* $0 < \underline{m}_{ii}$, $i = 1(1)n$.

$(V1)$ $\quad$ *Es sind* $\overline{m}_{ii} \leq 1$ *für* $i = 1(1)n$.

Behauptung 1:

Die Funktion $f : \mathbf{R}^n \to \mathbf{R}^n$, *definiert durch*

$$\begin{aligned} f(x) &= f(x_1, .., x_n; m_{11}, .., m_{nn}, q_1, .., q_n) \\ &= \Big(I + M\Big)^{-1}\Big((I - M)|x| - q\Big),\ M \in [M],\ q \in [q], \end{aligned}$$

besitzt die intervallmäßige Auswertung

$$f([x]) = IGA\Big(I + [M]\,,\ (I - [M])abs([x]) - [q]\Big) \qquad (3.11)$$

und ist eine P-Kontraktion.

Dabei bezeichnet I *die* $n \times n$ *Einheitsmatrix, und es ist (siehe Abschnitt 1.1)*

$$abs([x]) = [\min\{|x| : x \in [x]\}, \max\{|x| : x \in [x]\}].$$

Gemäß Teil 1 von Satz 3.5 existiert dann genau ein Intervallvektor $[x^*] \in \mathbf{IR}^n$ *mit*

$$f([x^*]) = [x^*],$$

und für jeden beliebigen Anfangsintervallvektor $[x^0] \in \mathbf{IR}^n$ *konvergiert die Iteration*

$$[x^{k+1}] := f([x^k]),\ k = 0, 1, 2...,$$

gegen $[x^*]$. *Mit* $[x^*]$ *kann man dann eine intervallmäßige Einschließung von* L_A *und eine intervallmäßige Einschließung von* L_B *erzielen.*

Behauptung 2:

Es gilt:

a)

$$L_A \subseteq [SA]$$

mit

$$[SA] := \begin{pmatrix} abs([x^*]) - [x^*] \\ abs([x^*]) + [x^*] \end{pmatrix} \cap \mathbf{R}^{2n}_{\geq 0}.$$

b)

$$L_B \subseteq [SB]$$

mit

$$[SB] := (abs([x^*]) + [x^*]) \cap \mathbf{R}^{n}_{\geq 0}.$$

Beweis zu Behauptung 1: Wir setzen zunächst $[A] := I + [M]$. Dann ist wegen der Voraussetzung (V)

$$\Big\langle [M] \Big\rangle \leq \Big\langle I + [M] \Big\rangle = \Big\langle [A] \Big\rangle.$$

Gemäß den Folgerungen aus dem Lemma 2.2 ist $\Big\langle [A] \Big\rangle$ eine M-Matrix und $[A]$ somit eine H-Matrix. Es gilt also

$$\Big\langle I + [M] \Big\rangle^{-1} = \Big\langle [A] \Big\rangle^{-1} \geq 0, \tag{3.12}$$

und nach Satz 2.5 ist der I.G.A. durchführbar für $[A]$. (3.11) ist somit definiert.

Wir zeigen nun, daß

$$f([x]) = IGA\Big([A]\,,\,(I - [M])abs([x]) - [q]\Big)$$

eine P-Kontraktion ist. Dazu seien $[x]$, $[y] \in \mathbf{IR}^n$. Die Intervallmatrix $[A] = I + [M]$ ist eine H-Matrix. Daher erhält man mit Teil 1 von Lemma 2.4

$$\mathrm{q}\Big(f([x]), f([y])\Big) \leq \Big\langle [A] \Big\rangle^{-1} \mathrm{q}\Big((I - [M])abs([x]) - [q], (I - [M])abs([y]) - [q]\Big).$$

Mit (1.1), (1.2) und Lemma 1.1 erhält man dann

$$\begin{aligned} \mathrm{q}\Big(f([x]),f([y])\Big) &\leq \Big\langle [A]\Big\rangle^{-1}\mathrm{q}\Big((I-[M])abs([x]),(I-[M])abs([y])\Big) \\ &\leq \Big\langle [A]\Big\rangle^{-1}|I-[M]|\mathrm{q}\Big(abs([x]),abs([y])\Big) \\ &\leq \Big\langle [A]\Big\rangle^{-1}|I-[M]|\mathrm{q}\Big([x],[y]\Big). \end{aligned}$$

Man erhält insgesamt

$$\mathrm{q}\Big(f([x]),f([y])\Big) \leq P\cdot \mathrm{q}\Big([x],[y]\Big)$$

mit

$$P = \Big\langle I+[M]\Big\rangle^{-1}|I-[M]| \geq 0.$$

Wir setzen nun

$$H := \Big\langle I+[M]\Big\rangle,\ N := |I-[M]| \text{ und } R := H-N.$$

Es gilt

$$\Big\langle I+[M]\Big\rangle_{ij} = \begin{cases} \Big\langle 1+[m_{ii}]\Big\rangle, & i=j, \\ -|[m_{ij}]|, & i\neq j, \end{cases}$$

und

$$|I-[M]|_{ij} = \begin{cases} |1-[m_{ii}]|, & i=j, \\ |[m_{ij}]|, & i\neq j. \end{cases}$$

Man erhält mit den Voraussetzungen (V) und $(V1)$

$$R_{ij} = \Big(\Big\langle I+[M]\Big\rangle - |I-[M]|\Big)_{ij} = \begin{cases} 2\underline{m}_{ii}, & i=j, \\ -2|[m_{ij}]|, & i\neq j. \end{cases}$$

Somit gilt

$$R = 2\Big\langle [M]\Big\rangle.$$

$[M]$ war als H-Matrix vorausgesetzt. Man erhält daher

$$R^{-1} = \frac{1}{2}\Big\langle [M]\Big\rangle^{-1} \geq 0. \tag{3.13}$$

Es ist $N = |I - [M]| \geq 0$ und $H^{-1} = \left\langle I + [M] \right\rangle^{-1} \geq 0$ wegen (3.12). Mit (3.13) gilt dann

$$H^{-1}N \geq 0 \quad \text{und} \quad R^{-1}N \geq 0.$$

Mit Satz 3.6 bekommt man schließlich

$$\rho(P) = \rho(H^{-1}N) < 1.$$

Beweis zu Behauptung 2: Zu a): Sei $\begin{pmatrix} w \\ z \end{pmatrix} \in L_A$. Es gibt dann ein $M \in [M]$ und ein $q \in [q]$ mit

$$\begin{aligned} w - Mz &= q, \\ w^{\mathrm{T}} z &= 0, \\ w \geq 0 \quad &\text{und} \quad z \geq 0. \end{aligned}$$

Also ist

$$-2w + 2Mz = -2q. \tag{3.14}$$

Nun gilt

$$\left.\begin{aligned} -2w &= (z-w) - |z-w|, \\ 2z &= (z-w) + |z-w|, \end{aligned}\right\} \tag{3.15}$$

denn aufgrund der Komplementarität gilt

$$\begin{aligned} \{(z-w) - |z-w|\}_i &= \begin{cases} 0 & \text{für } z_i \geq 0,\ w_i = 0\,, \\ -2w_i & \text{für } z_i = 0,\ w_i \geq 0\,, \end{cases} \\ \{(z-w) + |z-w|\}_i &= \begin{cases} 2z_i & \text{für } z_i \geq 0,\ w_i = 0\,, \\ 0 & \text{für } z_i = 0,\ w_i \geq 0\,. \end{cases} \end{aligned}$$

Setzt man (3.15) in (3.14) ein, so resultiert

$$(z-w) - |z-w| + M((z-w) + |z-w|) = -2q$$

bzw.

$$(I+M)(z-w) = (I-M)(|z-w|) - 2q.$$

Schon im Beweis zu Behauptung 1 ist gezeigt worden, daß der I.G.A. für $I + [M]$ durchführbar ist. Daher existiert die Inverse von $I + M$. Wir erhalten

$$\frac{z-w}{2} = (I+M)^{-1}(I-M)\left|\frac{z-w}{2}\right| - (I+M)^{-1}q = f\left(\frac{z-w}{2}\right).$$

Mit Teil 2 von Satz 3.5 folgt nun

$$\frac{z-w}{2} \in [x^*]$$

und wegen (3.15) dann

$$\begin{aligned} w &\in abs([x^*]) - [x^*], \\ z &\in abs([x^*]) + [x^*]. \end{aligned}$$

Da $w \geq 0$ und $z \geq 0$ gilt, folgt der a)-Teil von Behauptung 2.

Zu b): Sei $z \in L_B$. Es gibt dann ein $M \in [M]$ und ein $q \in [q]$ mit

$$\begin{aligned} q + Mz &\geq 0, \\ (q + Mz)^{\mathrm{T}} z &= 0, \\ z &\geq 0. \end{aligned}$$

Wir setzen

$$w := q + Mz.$$

Dann kann man wie im Beweis zu a) zu Ende schließen. □

Bemerkung: Ist der I.G.A. durchführbar für $I + [M]$, so ist als intervallmäßige Auswertung von $f(x)$ auch

$$f([x]) = IGA\Big(I + [M]\Big) \cdot \Big((I - [M])abs([x]) - [q]\Big)$$

wählbar. Ist allerdings $[M]$ eine Intervalltridiagonalmatrix, so kann

$$IGA\Big(I + [M]\Big)$$

vollbesetzt sein, wohingegen man bei der Berechnung von $f([x])$ gemäß (3.11) die Tridiagonalgestalt ausnützen kann.

Wir wollen uns jetzt noch von der Voraussetzung $(V1)$ in Satz 3.7 befreien.

Satz 3.8 *Es seien* $[q] \in \mathbf{IR}^n$ *und* $[M] \in \mathbf{IR}^{n \times n}$ *gegeben. Dabei möge* $[M]$ *folgende Voraussetzung erfüllen:*

(V) $[M]$ *ist eine H-Matrix mit* $0 < \underline{m_{ii}}$, $i = 1(1)n$.

Mit

$$m := \max\left\{1, \overline{m}_{11}, .., \overline{m}_{nn}\right\}$$

seien dann

$$[\tilde{M}] := \frac{1}{m} \cdot [M] \quad \textit{und} \quad [\tilde{q}] := \frac{1}{m} \cdot [q]$$

gesetzt.

Behauptung: Es gilt:

a) Ist

$$[SA] = \begin{pmatrix} [SA1] \\ [SA2] \end{pmatrix}, \quad [SA1], [SA2] \in \mathbf{IR}^n,$$

eine intervallmäßige Einschließung von

$$\begin{aligned} \tilde{L}_A \;=\; \Bigg\{ \begin{pmatrix} w \\ z \end{pmatrix} \in \mathbf{R}^{2n} :\; & \textit{Es gibt ein } \tilde{M} \in [\tilde{M}] \;\; \textit{und ein } \tilde{q} \in [\tilde{q}] \;\; \textit{mit} \\ & w - \tilde{M}z = \tilde{q},\, w \geq 0, z \geq 0, w^{\mathrm{T}} z = 0 \quad \Bigg\}, \end{aligned}$$

so gilt

$$L_A \subseteq \begin{pmatrix} m \cdot [SA1] \\ [SA2] \end{pmatrix}.$$

b) Ist $[SB] \in \mathbf{IR}^n$ *eine intervallmäßige Einschließung von*

$$\begin{aligned} \tilde{L}_B \;=\; \Big\{ z \in \mathbf{R}^n :\; & \textit{Es gibt ein } \tilde{M} \in [\tilde{M}] \;\; \textit{und ein } \tilde{q} \in [\tilde{q}] \;\; \textit{mit} \\ & \tilde{q} + \tilde{M}z \geq 0,\, z \geq 0, (\tilde{q} + \tilde{M}z)^{\mathrm{T}} z = 0 \Big\}, \end{aligned}$$

so gilt $L_B \subseteq [SB]$.

Beweis: Zu a): Sei $\begin{pmatrix} w \\ z \end{pmatrix} \in L_A$. Es gibt dann ein $M \in [M]$ und ein $q \in [q]$ mit

$$\begin{aligned} w - Mz &= q, \\ w^{\mathrm{T}} z &= 0, \\ w \geq 0 \quad \text{und} \quad z &\geq 0. \end{aligned}$$

Hieraus folgt

$$\begin{aligned} \frac{1}{m} \cdot w - \frac{1}{m} \cdot Mz &= \frac{1}{m} \cdot q, \\ \left(\frac{1}{m} \cdot w\right)^{\mathrm{T}} z &= 0, \\ \frac{1}{m} \cdot w \geq 0 \quad \text{und} \quad z &\geq 0. \end{aligned}$$

Da

$$\frac{1}{m} \cdot M \in [\tilde{M}] \quad \text{und} \quad \frac{1}{m} \cdot q \in [\tilde{q}]$$

gilt, bekommen wir mit

$$\begin{pmatrix} \frac{1}{m} \cdot w \\ z \end{pmatrix} \in \tilde{L}_A \subseteq \begin{pmatrix} [SA1] \\ [SA2] \end{pmatrix}$$

dann letztlich

$$\begin{pmatrix} w \\ z \end{pmatrix} \in \begin{pmatrix} m \cdot [SA1] \\ [SA2] \end{pmatrix}.$$

Zu b): Sei $z \in L_B$. Es gibt dann ein $M \in [M]$ und ein $q \in [q]$ mit

$$\begin{aligned} q + Mz &\geq 0, \\ (q + Mz)^{\mathrm{T}} z &= 0, \\ z &\geq 0. \end{aligned}$$

Hieraus folgt

$$\begin{aligned} \frac{1}{m} \cdot q + \frac{1}{m} \cdot Mz &\geq 0, \\ \left(\frac{1}{m} \cdot q + \frac{1}{m} \cdot Mz\right)^{\mathrm{T}} z &= 0, \\ z &\geq 0. \end{aligned}$$

Da

$$\frac{1}{m} \cdot M \in [\tilde{M}] \quad \text{und} \quad \frac{1}{m} \cdot q \in [\tilde{q}]$$

gilt, bekommen wir

$$z \in \tilde{L}_B \subseteq [SB].$$

□

3.2.2 Praktische Umsetzung: Monotone Folgen

In diesem Abschnitt sei $[q] \in \mathbf{IR}^n$ beliebig und $[M] \in \mathbf{IR}^{n\times n}$ erfülle folgende Voraussetzung:

(V) $[M]$ ist eine H-Matrix mit $0 < \underline{m}_{ii}$, $i = 1(1)n$.

Nach Satz 3.7 und Satz 3.8 ist es möglich, L_A bzw. L_B intervallmäßig einzuschließen.

Glücklicherweise sind Satz 3.7 und Satz 3.8 nicht nur reine Existenzsätze, sondern auch Lieferanten für eine konkrete Vorgehensweise, wie man schließlich die Einschließung erhält.

Das Herz der Einschließung ist der Fixintervallvektor von

$$f([x]) = IGA\Big(I + [M]\,,\ (I - [M])abs([x]) - [q]\Big),$$

den man (theoretisch) über die Iteration

$$\begin{aligned} [x^0] &\in \mathbf{IR}^n \text{ beliebig,} \\ [x^{k+1}] &:= f([x^k]),\ k = 0, 1, 2, ..., \end{aligned}$$

erhält. Wünschenswert wäre jetzt die Gewißheit, daß die Intervalle monoton fallend gegen $[x^*]$ konvergieren, d.h.

$$[x^*] \subseteq \ldots \subseteq [x^1] \subseteq [x^0],$$

damit man sicher ist, daß die k-te Iterierte auf jeden Fall $[x^*]$ enthält und dadurch auf jeden Fall eine Einschließung von L_A bzw. L_B sicherstellt (Inklusionsmonotonie). Dazu benötigen wir den folgenden

Satz 3.9 *Es sei $f : \mathbf{R}^n \to \mathbf{R}^n$ eine Abbildung wie in (3.10) beschrieben, und es gebe ein $[y^0] \in \mathbf{IR}^n$ mit*

$$\{x \in \mathbf{R}^n : f(x) = x, M \in [M],\ q \in [q]\} \subseteq [y^0]. \tag{3.16}$$

Außerdem besitze $f(x)$ eine intervallmäßige Auswertung. Dann ist

$$[y^{k+1}] := f([y^k]) \cap [y^k],\ k = 0, 1, 2, ..., \tag{3.17}$$

eine monoton fallende Folge, und für $[y^*] := \cap_{k=0}^{\infty} [y^k]$ *gilt*

$$[y^*] = f([y^*]) \cap [y^*] \tag{3.18}$$

und

$$\{x \in \mathbf{R}^n : f(x) = x, M \in [M], q \in [q]\} \subseteq [y^k] \tag{3.19}$$

für alle $k \in \mathbf{N} \cup \{0\}$*, also insbesondere*

$$\{x \in \mathbf{R}^n : f(x) = x, M \in [M], q \in [q]\} \subseteq [y^*].$$

Beweis: Die ersten beiden Aussagen kann man in [2], Seite 133/Theorem 2 finden.

Für die letzte Aussage seien $M \in [M]$, $q \in [q]$ und

$$x = f(x). \tag{3.20}$$

Nach (3.16) gilt dann

$$x \in [y^0]$$

und wegen (3.20) ist

$$x \in f([y^0]).$$

Somit ist $x \in [y^0] \cap f([y^0]) = [y^1]$. Mit Induktion erhält man dann (3.19). □

Man erhält damit

Satz 3.10 *Es seien* $[q] \in \mathbf{IR}^n$ *und* $[M] \in \mathbf{IR}^{n \times n}$*.* $[M]$ *erfülle die Voraussetzung (V). Mit*

$$m := \max\left\{1, \overline{m}_{11}, .., \overline{m}_{nn}\right\}$$

seien dann

$$[\tilde{M}] := \frac{1}{m} \cdot [M] \quad \textit{und} \quad [\tilde{q}] := \frac{1}{m} \cdot [q]$$

gesetzt. Aufgrund von Satz 3.7 besitzt die Funktion

$$\begin{aligned} f(x) &= f(x_1, .., x_n; \tilde{m}_{11}, .., \tilde{m}_{nn}, \tilde{q}_1, .., \tilde{q}_n) \\ &= \left(I + \tilde{M}\right)^{-1} \left((I - \tilde{M})|x| - \tilde{q}\right),\ \tilde{M} \in [\tilde{M}],\ \tilde{q} \in [\tilde{q}] \end{aligned}$$

die intervallmäßige Auswertung

$$f([x]) = IGA\left(I + [\tilde{M}]\,,\ (I - [\tilde{M}])abs([x]) - [\tilde{q}]\right).$$

Desweiteren sei vorausgesetzt, daß es ein $[y^0] \in \mathbf{IR}^n$ *gibt mit*

$$\{x \in \mathbf{R}^n : f(x) = x, \tilde{M} \in [\tilde{M}],\, \tilde{q} \in [\tilde{q}]\} \subseteq [y^0]. \tag{3.21}$$

Definiert man nun $[y^{k+1}] := f([y^k]) \cap [y^k]$, *für* $k = 0, 1, 2, ...$, *so gilt für alle* $k \in \mathbf{N} \cup \{0\}$*:*

a)

$$L_A \subseteq [SA]$$

mit

$$[SA] := \begin{pmatrix} m \cdot \Big(abs([y^k]) - [y^k]\Big) \\ abs([y^k]) + [y^k] \end{pmatrix} \cap \mathbf{R}^{2n}_{\geq 0}\,.$$

b)

$$L_B \subseteq [SB]$$

mit

$$[SB] := (abs([y^k]) + [y^k]) \cap \mathbf{R}^n_{\geq 0}\,.$$

Beweis: 1. Fall: Es sei $m = 1$. Es gilt also $[\tilde{M}] = [M]$ und $[\tilde{q}] = [q]$. Die Intervallmatrix $[M]$ erfüllt also die Voraussetzungen (V) und $(V1)$ von Satz 3.7.

Zu a): Ist $\begin{pmatrix} w \\ z \end{pmatrix} \in L_A$, dann erhält man wie im Beweis zu Behauptung 2a) von Satz 3.7 ein $M \in [M]$ und ein $q \in [q]$ mit

$$\frac{z-w}{2} = -(I+M)^{-1}q + (I+M)^{-1}(I-M)\left|\frac{z-w}{2}\right|.$$

Mit Satz 3.9 angewandt auf

$$f(x) = \Big(I+M\Big)^{-1}\Big((I-M)|x| - q\Big),\; M \in [M],\; q \in [q]$$

erhält man dann

$$\frac{z-w}{2} \in [y^k],\;\; k = 0, 1, 2... \,.$$

Wie im Beweis zu Behauptung 2a) von Satz 3.7 folgt dann

$$\begin{aligned} w &\in abs([y^k]) - [y^k], \\ z &\in abs([y^k]) + [y^k], \end{aligned}$$

für jedes $k \in \mathbf{N} \cup \{0\}$. Wegen $w, z \geq 0$ folgt dann Behauptung a).

Zu b): Sei $z \in L_B$. Dann folgt wie im Beweis zu a)

$$z \in abs([y^k]) + [y^k]$$

für jedes $k \in \mathbf{N} \cup \{0\}$. Wegen $z \geq 0$ folgt dann Behauptung b).

2. Fall: Es sei $m > 1$. Dann erfüllt die Intervallmatrix $[\tilde{M}]$ die Voraussetzungen (V) und $(V1)$ von Satz 3.7. Die Behauptungen a) und b) folgen daher mit dem 1. Fall und Satz 3.8. □

Es bleibt jetzt nur noch zu zeigen, wie man ein $[y^0] \in \mathbf{IR}^n$ erhält, so daß (3.21) gilt. Dazu betrachten wir wieder die Iterationsvorschrift ohne Durchschnittsbildung:

$$\begin{aligned} [x^0] &\in \mathbf{IR}^n \text{ beliebig,} \\ [x^{k+1}] &:= f([x^k]),\ k = 0, 1, 2, \ldots . \end{aligned}$$

Lemma 3.4 *Gegeben seien eine Funktion $f : \mathbf{R} \to \mathbf{R}$ der Bauart (3.10) und ein Intervallvektor $[x^0] \in \mathbf{IR}^n$. Weiter sei f eine P-Kontraktion mit Kontraktionsmatrix P. Es seien dann definiert:*

$$\begin{aligned} [x^1] &:= f([x^0]), \\ v^1 &:= P(I - P)^{-1} \mathrm{q}([x^1], [x^0]), \end{aligned}$$

wobei I die $n \times n$ Einheitsmatrix bezeichnet.

Behauptung: Für

$$[y^0] := [\underline{x}^1 - v^1, \overline{x}^1 + v^1]$$

gilt

$$\{x \in \mathbf{R}^n : f(x) = x, M \in [M],\ q \in [q]\} \subseteq [y^0].$$

Beweis: Da f eine P-Kontraktion ist, gilt für $m \in \mathbf{N}$

$$\begin{aligned} \mathrm{q}([x^{m+1}], [x^m]) &= \mathrm{q}(f([x^m]), f([x^{m-1}])) \\ &\leq P \cdot \mathrm{q}([x^m], [x^{m-1}]) \\ &\cdots \\ &\leq P^m \cdot \mathrm{q}([x^1], [x^0]) =: P^m a. \end{aligned}$$

Man erhält für $l > k$ mit der Dreiecksungleichung, da $P \geq 0$ und wegen Teil 2 von Satz 1.1

$$\begin{aligned} \mathrm{q}([x^l],[x^k]) &\leq \mathrm{q}([x^l],[x^{l-1}]) + \ldots + \mathrm{q}([x^{k+1}],[x^k]) \\ &\leq P^{l-1}a + \ldots + P^k a \\ &= P^k(I + P + \ldots P^{l-k-1})a \\ &\leq P^k(\sum_{j=0}^{\infty} P^j)a \\ &= P^k(I-P)^{-1}a. \end{aligned}$$

Wegen $\lim_{l\to\infty}[x^l] = [x^*]$ gilt

$$\mathrm{q}([x^*],[x^k]) \leq P^k(I-P)^{-1}a =: v^k.$$

Setzt man $k = 1$, so erhält man

$$[x^*] \subseteq [\underline{x}^1 - v^1, \overline{x}^1 + v^1] = [y^0],$$

denn es gilt

$$\mathrm{q}([x^*],[x^1]) \leq v^1 \Leftrightarrow \begin{array}{l} |\overline{x}^* - \overline{x}^1| \leq v^1 \\ |\underline{x}^* - \underline{x}^1| \leq v^1 \end{array} \Rightarrow \begin{array}{l} \overline{x}^* \leq \overline{x}^1 + v^1 \\ \underline{x}^1 - v^1 \leq \underline{x}^* \end{array}.$$

Wegen Teil 2 von Satz 3.5 erhält man schließlich

$$\{x \in \mathbf{R}^n : f(x) = x, M \in [M],\, q \in [q]\} \subseteq [x^*] \subseteq [y^0]. \qquad \square$$

Bevor wir nun den Algorithmus angeben, geben wir noch an, wie die explizite Berechnung von P umgangen werden kann.

Lemma 3.5 *Es sei $[\tilde{M}] \in \mathbf{IR}^{n\times n}$ eine Intervallmatrix, die den Voraussetzungen (V) und $(V1)$ von Satz 3.7 genüge. Dann gilt für die Matrix*

$$P = \left\langle I + [\tilde{M}] \right\rangle^{-1} |I - [\tilde{M}]|$$

die Beziehung

$$P(I-P)^{-1} = \left\langle I + [\tilde{M}] \right\rangle^{-1} |I - [\tilde{M}]| \left(\left\langle [\tilde{M}] \right\rangle^{-1} + I \right)/2.$$

Beweis: Es ist

$$\begin{aligned}
(I-P)^{-1} &= \left(I - \left\langle I + [\tilde{M}]\right\rangle^{-1} |I - [\tilde{M}]|\right)^{-1} \\
&= \left(\left\langle I + [\tilde{M}]\right\rangle^{-1} (\left\langle I + [\tilde{M}]\right\rangle - |I - [\tilde{M}]|)\right)^{-1} \\
&= \left(\left\langle I + [\tilde{M}]\right\rangle^{-1} 2\left\langle [\tilde{M}]\right\rangle\right)^{-1} \\
&= \left\langle [\tilde{M}]\right\rangle^{-1} \left\langle I + [\tilde{M}]\right\rangle/2 \\
&= \left\langle [\tilde{M}]\right\rangle^{-1} \left(I + \left\langle [\tilde{M}]\right\rangle\right)/2 \\
&= \left(\left\langle [\tilde{M}]\right\rangle^{-1} + I\right)/2.
\end{aligned}$$

□

Algorithmus A

{ Gegeben seien eine Intervallmatrix $[M] \in \mathbf{IR}^{n\times n}$, die der Voraussetzung }
{ (V) genüge, und ein Intervallvektor $[q] \in \mathbf{IR}^n$. }

$m := \max\limits_{1\le i\le n} \{\overline{m}_{ii}\}$;

if $m > 1$ **then begin** $[M] := [M]/m$;

$[q] := [q]/m$;

end;

$[A] := I + [M]$;

$[x^0] := 0$;

$[x^1] := IGA([A], -[q])$;

$alpha :=$ q$([x^1], 0)$;

$[u] := IGA\left(\left\langle [M]\right\rangle, alpha\right)$;

$[v] := IGA\left(\left\langle I + [M]\right\rangle, |I - [M]|([u] + alpha)/2\right)$;

{ !!: $[v]$ ist ein Intervallvektor und kein Punktvektor wie in der Theorie }

$[y^0] := [\underline{x}^1 - \overline{v}, \overline{x}^1 + \overline{v}]$;

writeln('wieviele Iterationsschritte sollen höchstens getätigt werden');

readln$(stop)$;

$zaehler := 0;$

$[y^{k+1}] := [y^0];$

repeat

$zaehler := zaehler + 1;$

$[y^k] := [y^{k+1}];$

$[y^{k+1}] := IGA\Big([A],\, (I - [M]) \cdot abs([y^k]) - [q]\Big) \cap [y^k];$

until $(zaehler = stop)$ **or** $([y^k] = [y^{k+1}]);$

if $m > 1$ **then** $[SA][1..n] := m \cdot (abs([y^{k+1}]) - [y^{k+1}])$

else $[SA][1..n] := (abs([y^{k+1}]) - [y^{k+1}])$;

$[SA][n+1..2n] := (abs([y^{k+1}]) + [y^{k+1}]);$

for $i := 1$ **to** $2n$ **do if** $\underline{SA}_i < 0$ **then** $\underline{SA}_i := 0;$

{ bzw. }

$[S_B][1..n] := (abs([y^{k+1}]) + [y^{k+1}]);$

for $i := 1$ **to** n **do if** $\underline{SB}_i < 0$ **then** $\underline{SB}_i := 0;$

Beispiel 3.5 Wir betrachten Beispiel 3.3a). $[M]$ und $[q]$ erfüllen die Voraussetzung (V) wie man einfach sieht, und man erhält mit dem Algorithmus A nach zwei (!) Iterationsschritten

$$[SA] = \begin{pmatrix} [0 & , & 0.978571428571429] \\ [0 & , & 1.921428571428573] \\ [0.099999999999999 & , & 2.057142857142858] \\ [0.5 & , & 4.342857142857145] \end{pmatrix}.$$

Für Beispiel 3.3b) erhält man auch nach zwei Iterationsschritten

$$[SA] = \begin{pmatrix} [0 & , & 1.600000000000001] \\ [0 & , & 1.957142857142858] \\ [0 & , & 2.057142857142858] \\ [0.4285714285714285 & , & 4.342857142857145] \end{pmatrix}.$$

Für Beispiel 3.4, welches auch die Voraussetzung (V) erfüllt, erhält man nach 282 Iterationsschritten

$$[SA] = \begin{pmatrix} [0 & , & 21.62293314162479] \\ [0 & , & 5.983465132997859] \\ [0.754133716750539 & , & 44.00000000000012] \\ [0 & , & 10.00000000000004] \end{pmatrix}.$$

Wählt man $stop = 20$, so erhält man

$$[SA] = \begin{pmatrix} [0 & , & 21.88128448668272] \\ [0 & , & 6.052074062515644] \\ [0.754133716750539 & , & 44.51670269011598] \\ [0 & , & 10.13721785903560] \end{pmatrix}.$$

Es kommt also in der w-Komponente zu einer groben Überschätzung, was auch wegen $d([a] - [b]) = d([a]) + d([b])$ zu erwarten war.

Interessiert man sich allerdings lediglich für $[SB]$, siehe nächstes Kapitel, so braucht man sich um dieses Problem nicht zu kümmern.

Wir wollen zuvor noch den Spezialfall betrachten, welche Aussagen gelten, wenn ein LCP betrachtet wird mit einer Punktmatrix $M \in \mathbf{R}^{n \times n}$ und einem Punktvektor $q \in \mathbf{R}^n$, wobei M die Voraussetzung (V) erfüllt.

Dazu betrachten wir vorher noch das

Lemma 3.6 *Es sei* $[y] \in \mathbf{IR}$. *Dann gilt*

$$d(abs([y])) \leq d([y]).$$

Beweis: 1. Fall: $\underline{y} > 0$. Dann gilt:

$$abs([y]) = [y] \Rightarrow d(abs([y])) = d([y]).$$

2. Fall: $\overline{y} < 0$. Dann gilt:

$$abs([y]) = [-\overline{y}, -\underline{y}] \Rightarrow d(abs([y])) = -\underline{y} - (-\overline{y}) = d([y]).$$

3. Fall: $0 \in [y]$. Dann gilt:

$$abs([y]) = [0, \max\{-\underline{y}, \overline{y}\}] \Rightarrow d(abs([y])) = \max\{-\underline{y}, \overline{y}\} \leq \overline{y} - \underline{y} = d([y]).$$

Damit sind alle Fälle abgedeckt. □

Wir kommen damit zu

Satz 3.11 *Es seien $q \in \mathbf{R}^n$ und $M \in \mathbf{R}^{n\times n}$. Desweiteren sei M eine H-Matrix mit positiven Diagonalelementen. Dann gilt:*

Die in Algorithmus A (angewandt auf die Punktmatrix M und den Punktvektor q) berechnete monotone Iterationsfolge

$$\ldots \subseteq [y^k] \ldots \subseteq [y^2] \subseteq [y^1] \subseteq [y^0]$$

zieht sich zu einem Punktvektor zusammen:

$$\lim_{k\to\infty} d([y^k]) = 0. \tag{3.22}$$

Beweis: Wir setzen

$$m := \max\{1, m_{11}, .., m_{nn}\}$$

und damit $\tilde{M} := \frac{1}{m}M$ und $\tilde{q} := \frac{1}{m}q$. Für $A := I + \tilde{M}$ gilt mit (1.5)

$$\begin{aligned} 0 &\leq d([y^{k+1}]) = d\Big(IGA\Big(A\,,\,(I-\tilde{M})abs([y^k]) - \tilde{q}\Big) \cap [y^k]\Big) \\ &\leq d\Big(IGA\Big(A\,,\,(I-\tilde{M})abs([y^k]) - \tilde{q}\Big)\Big). \end{aligned}$$

Nun ist $\left\langle \tilde{M} \right\rangle \leq \left\langle I + \tilde{M} \right\rangle = \left\langle A \right\rangle$. Also ist $\left\langle A \right\rangle$ nach den Folgerungen aus Lemma 2.2 eine M-Matrix und A somit eine H-Matrix. Mit Teil 2 von Lemma 2.4 erhält man dann

$$0 \leq d([y^{k+1}]) \leq \left\langle I + \tilde{M} \right\rangle^{-1} d((I-\tilde{M})abs([y^k]) - \tilde{q}).$$

Mit (1.4), (1.3) und Lemma 3.6 folgt

$$\begin{aligned} 0 &\leq d([y^{k+1}]) \leq \left\langle I + \tilde{M} \right\rangle^{-1} d((I-\tilde{M})abs([y^k]) - \tilde{q}) \\ &= \left\langle I + \tilde{M} \right\rangle^{-1} d((I-\tilde{M})abs([y^k])) = \left\langle I + \tilde{M} \right\rangle^{-1} |I-\tilde{M}| d(abs([y^k])) \\ &\leq \left\langle I + \tilde{M} \right\rangle^{-1} |I-\tilde{M}| d([y^k]). \end{aligned}$$

Es resultiert also

$$0 \leq d([y^{k+1}]) \leq P \cdot d([y^k])$$

mit

$$P = \left\langle I + \tilde{M} \right\rangle^{-1} |I-\tilde{M}|.$$

Durch Induktion bekommt man dann

$$0 \leq d([y^{k+1}]) \leq P^{k+1} \cdot d([y^0]),\ k = 0, 1, 2, \ldots .$$

Wie im Beweis zu Satz 3.7 zeigt man $\rho(P) < 1$. Daraus folgt dann $\lim\limits_{k\to\infty} P^k = 0$ wegen Teil 1 von Satz 1.1. Somit gilt also (3.22). □

Satz 3.11 hat natürlich lediglich theoretisches Interesse, da in der Praxis, d.h. auf einer Rechenmaschine, Ergebnisse der Form

$$\frac{2}{3} \quad \text{oder} \quad \sqrt{2}$$

nicht exakt darstellbar sind. Daher wird Algorithmus A umgesetzt auf einer Rechenmaschine auch bei der Eingabe einer Punktmatrix und eines Punktvektors als Ergebnis einen Intervallvektor liefern.

Beispiel 3.6 Gegeben seien

$$M = \begin{pmatrix} 8 & 1 & 2 & 3 \\ 0 & 3 & 2 & 0 \\ 1 & 2 & 4 & 0 \\ -1 & -2 & 0 & 4 \end{pmatrix} \quad \text{und} \quad q = \begin{pmatrix} -1 \\ -2 \\ 3 \\ 4 \end{pmatrix}.$$

Die Matrix M ist streng diagonal dominant und daher wegen Teil 2 von Lemma 2.3 eine H-Matrix. Die Diagonalelemente sind positiv. Der Algorithmus A ist also anwendbar. Er liefert nach 133 Iterationsschritten:

$$[SA] = \begin{pmatrix} [0 & , & 0.0000000000000055] \\ [0 & , & 0.0000000000000076] \\ [4.374999999999992 & , & 4.37500000000008] \\ [2.624999999999993 & , & 2.62500000000007] \\ [0.041666666666665 & , & 0.041666666666668] \\ [0.666666666666665 & , & 0.666666666666668] \\ [0 & , & 0.0000000000000010] \\ [0 & , & 0.0000000000000008] \end{pmatrix}.$$

3.3 Das lineare Komplementaritätsproblem als Nullstellenproblem

Definition. Sind $x, y \in \mathbf{R}^n$, so definieren wir den Vektor $\min(x,y) \in \mathbf{R}^n$ durch

$$(\min(x_i, y_i))_i := \begin{cases} x_i, & \text{falls } x_i \leq y_i\,, \\ y_i, & \text{sonst}\,, \end{cases} \qquad i = 1(1)n.$$

Beginnen wollen wir diesen Abschnitt dann mit

Satz 3.12 *Es seien $q \in \mathbf{R}^n$ und $M \in \mathbf{R}^{n\times n}$ gegeben. Dann gilt für jedes $z \in \mathbf{R}^n$:*

$$\min(q+Mz, z) = 0 \Leftrightarrow \begin{cases} q+Mz & \geq & 0, \\ z & \geq & 0, \\ (q+Mz)^{\mathrm{T}} z & = & 0. \end{cases}$$

Der Beweis ist offensichtlich.

Satz 3.12 offenbart uns die Möglichkeit, eine Lösung eines LCPs zu bestimmen, indem wir eine Nullstelle einer Funktion bestimmen. Dabei ändert sich die Nullstelle nicht, wenn man q und M mit einer positiven Zahl durchdividiert. Dies zeigen wir in

Lemma 3.7 *Es seien $q \in \mathbf{R}^n$, $M \in \mathbf{R}^{n\times n}$ und $m > 0$. Dann gilt:*

$$\min(q+Mz, z) = 0 \quad \Leftrightarrow \quad \min\left(\frac{1}{m}q + \frac{1}{m}Mz, z\right) = 0.$$

Beweis: Mit Satz 3.12 ist

$$\min(q+Mz, z) = 0 \Leftrightarrow \begin{cases} q+Mz & \geq & 0, \\ z & \geq & 0, \\ (q+Mz)^{\mathrm{T}} z & = & 0. \end{cases}$$

Wegen $m > 0$ gilt dann

$$\left.\begin{array}{rcl} q+Mz & \geq & 0, \\ z & \geq & 0, \\ (q+Mz)^{\mathrm{T}} z & = & 0, \end{array}\right\} \Leftrightarrow \begin{cases} \frac{1}{m}q + \frac{1}{m}Mz & \geq & 0, \\ z & \geq & 0, \\ \left(\frac{1}{m}q + \frac{1}{m}Mz\right)^{\mathrm{T}} z & = & 0. \end{cases}$$

Die rechte Seite ist dann wieder mit Satz 3.12 äquivalent zu

$$\min\left(\frac{1}{m}q + \frac{1}{m}Mz, z\right) = 0.$$

□

Leider muß die Funktion $H(z) := \min(q+Mz, z)$, $z \in \mathbf{R}^n_{\geq 0}$, nicht differenzierbar sein, so daß viele klassische Verfahren zur Nullstellenbestimmung nicht angewandt werden können.

In [1] wurde gezeigt, wie man mit dem verallgemeinerten Krawczyk-Operator von einem Intervallvektor nachprüfen kann, ob er eine Nullstelle von $H(z)$ beinhaltet.

Das Ziel dieses Abschnitts ist es, diese Methode zu verallgemeinern für den Fall, daß die Ausgangsdaten Intervalle sind.

Es seien also $[q] \in \mathbf{IR}^n$ und $[M] \in \mathbf{IR}^{n\times n}$ gegeben. Wir betrachten das LCP mit Intervalleinträgen:

$$\left.\begin{array}{rcl} q + Mz & \geq & 0, \\ z & \geq & 0, \\ (q + Mz)^{\mathrm{T}} z & = & 0, \end{array}\right\} \quad q \in [q],\ M \in [M]. \tag{3.23}$$

Wir definieren nun im Hinblick auf Satz 3.12 die Funktion

$$H(z; q, M) := \min(q + Mz, z), \quad q \in [q],\ M \in [M],$$

und suchen dazu einen Intervallvektor $[z] \in \mathbf{IR}^n$, der die folgende Eigenschaft erfüllt:

$$\begin{array}{c} \text{Für jedes } q \in [q] \text{ und jedes } M \in [M] \\ \text{gibt es ein } z \in [z] \text{ mit } H(z; q, M) = 0. \end{array} \tag{3.24}$$

Ein $[z] \in \mathbf{IR}^n$ mit der Eigenschaft (3.24) wird die Menge

$$\begin{array}{rcl} L_B & = & \Big\{z \in \mathbf{R}^n : \text{ Es gibt ein } M \in [M] \text{ und ein } q \in [q] \text{ mit} \\ & & H(z; q, M) = \min(q + Mz, z) = 0\Big\} \end{array}$$

i.a. nicht einschließen.

Beispiel 3.7 Wir betrachten

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 0 \\ 1 & 3 & 0 \end{pmatrix}, \quad q = \begin{pmatrix} -6 \\ -5 \\ -4 \end{pmatrix}.$$

Ein $[z] \in \mathbf{IR}^3$ mit der Eigenschaft (3.24) wäre z.B.

$$[z] = z = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

denn es gilt

$$H(z; q, M) = \min \left(\begin{pmatrix} -6 \\ -5 \\ -4 \end{pmatrix} + M \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) = 0.$$

Allerdings gilt auch $H(z; q, M) = 0$ für

$$z = \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}.$$

Dieses z liegt aber nicht in $[z]$.

Beispiel 3.7 zeigt, daß das zu M und q gehörende LCP mindestens zwei Lösungen besitzt. Wenn wir voraussetzen, daß $[M]$ eine P-Matrix ist, so wird die in Beispiel 3.7 angedeutete Schwierigkeit nicht auftreten. Dies ist die Aussage des nächsten Lemmas.

Lemma 3.8 *Es sei $[q] \in \mathbf{IR}^n$ und es sei $[M] \in \mathbf{IR}^{n \times n}$ eine P-Matrix. Dann gelten folgende Aussagen:*

1. *Ein Intervallvektor $[z] \in \mathbf{IR}^n$ mit*

$$[z] \supseteq L_B \;=\; \Big\{ z \in \mathbf{R}^n : \text{ Es gibt ein } M \in [M] \text{ und ein } q \in [q] \text{ mit } H(z; q, M) = \min(q + Mz, z) = 0 \Big\}$$

   erfüllt die Eigenschaft (3.24).

2. *Für einen Intervallvektor* $[z] \in \mathbf{IR}^n$*, der die Eigenschaft (3.24) erfüllt, gilt*

$$[z] \supseteq L_B = \Big\{ z \in \mathbf{R}^n : \textit{ Es gibt ein } M \in [M] \textit{ und ein } q \in [q] \textit{ mit } H(z;q,M) = \min(q + Mz, z) = 0 \Big\}.$$

Beweis: Zu 1.: Es seien $q \in [q]$ und $M \in [M]$. Da $[M]$ eine P-Matrix ist, hat das zu q und M gehörende LCP genau eine Lösung z. Es gilt also

$$z \in L_B \subseteq [z]$$

und

$$H(z;q,M) = 0.$$

Zu 2.: Es sei $z \in L_B$. Dann gibt es ein $q \in [q]$ und ein $M \in [M]$ mit

$$H(z;q,M) = 0.$$

Da $[z]$ die Eigenschaft (3.24) erfüllt, gibt es für dasselbe $q \in [q]$ und dasselbe $M \in [M]$ ein $\tilde{z} \in [z]$ mit

$$H(\tilde{z};q,M) = 0.$$

M ist als P-Matrix vorausgesetzt. Das zu M und q gehörende LCP hat daher eine eindeutige Lösung. Es folgt somit

$$z = \tilde{z} \in [z].$$

□

Um die wesentlichen Ideen aus [1] umsetzen zu können, beginnen wir mit

Satz 3.13 *Gegeben seien* $[q] \in \mathbf{IR}^n$ *und* $[M] \in \mathbf{IR}^{n \times n}$*. Damit definieren wir* $H(z;[q],[M]) \in \mathbf{IR}^n$ *durch*

$$(H(z;[q],[M]))_i := \begin{cases} z_i & \textit{falls} \quad z_i < \underline{q_i} + \underline{([M]z)_i}, \\ [q_i] + ([M]z)_i & \textit{falls} \quad z_i > \overline{q}_i + \overline{([M]z)_i}, \\ [\underline{q_i} + \underline{([M]z)_i}, z_i] & \textit{falls} \quad z_i \in [q_i] + ([M]z)_i, \end{cases} \quad i = 1(1)n.$$

Es sei $[z] \in \mathbf{IR}^n$ *ein Intervallvektor, von dem wir zeigen wollen, daß er die Eigenschaft (3.24) erfüllt bzw. nicht erfüllen kann. Dazu seien* $x \in [z]$, $q \in [q]$

und $M \in [M]$ beliebig, aber fest gewählt. Wir nehmen an, wir kennen eine Intervallmatrix

$$G(x,[z],[q],[M]) = \begin{pmatrix} G_1(x,[z],[q],[M]) \\ \vdots \\ G_n(x,[z],[q],[M]) \end{pmatrix} \in \mathbf{IR}^{n\times n},$$

die folgendes erfüllt:

$$\left.\begin{array}{rcll} H(x;q,M) - H(y;q,M) & = & G(x,y,q,M)(x-y) & \\ & \in & G(x,[z],[q],[M])(x-y) & \text{für alle } y \in [z]. \end{array}\right\} \tag{3.25}$$

Damit definieren wir den Operator

$$N(x,[z],[q],[M]) := x - IGA\Big(G(x,[z],[q],[M]),\, H(x;[q],[M])\Big).$$

Behauptung 1: Ist der I.G.A. durchführbar für $G(x,[z],[q],[M])$ und gilt $N(x,[z],[q],[M]) \subseteq [z]$, dann gibt es für jedes $q \in [q]$ und für jedes $M \in [M]$ ein $z \in N(x,[z],[q],[M])$ mit $H(z;q,M) = 0$.

Behauptung 2: Ist der I.G.A. durchführbar für $G(x,[z],[q],[M])$ und gilt $N(x,[z],[q],[M]) \cap [z] = \emptyset$, dann gilt für jedes $q \in [q]$, für jedes $M \in [M]$ und für jedes $z \in [z]$: $H(z;q,M) \neq 0$.

Behauptung 3: Gibt es ein $z \in [z]$, ein $q \in [q]$ und ein $M \in [M]$ mit $H(z;q,M) = 0$, und ist der I.G.A. durchführbar für $G(x,[z],[q],[M])$, so gilt

$$z \in N(x,[z],[q],[M]).$$

Behauptung 4: Es sei $[z^0] \in \mathbf{IR}^n$. Gibt es ein $z \in [z^0]$, ein $q \in [q]$ und ein $M \in [M]$ mit $H(z;q,M) = 0$, so gilt $z \in [z^k] \in \mathbf{IR}^n$, $k \in \mathbf{N} \cup \{0\}$, mit

$$[z^{k+1}] := N(x^k,[z^k],[q],[M]) \cap [z^k],$$

vorausgesetzt der I.G.A. ist durchführbar für $G(x^k,[z^k],[q],[M])$. x^k ist dabei beliebig aus $[z^k]$ gewählt.

Beweis zu Behauptung 1: Da der I.G.A. durchführbar ist für $G(x,[z],[q],[M])$, existiert die Inverse von $G(x,y,q,M)$ für jedes $y \in [z]$, $q \in [q]$ und $M \in [M]$.

Es seien nun $q \in [q]$ und $M \in [M]$ beliebig, aber fest gewählt. Dann definieren wir $p(y) := y - \Big(G(x,y,q,M)\Big)^{-1} H(y;q,M)$, $y \in [z]$. Es gilt

$$\left.\begin{array}{rcl} p(y) & = & y - \Big(G(x,y,q,M)\Big)^{-1} H(y;q,M) \\ & = & y + \Big(G(x,y,q,M)\Big)^{-1} (H(x;q,M) - H(y;q,M)) \\ & & -\Big(G(x,y,q,M)\Big)^{-1} H(x;q,M) \\ & = & y + x - y - \Big(G(x,y,q,M)\Big)^{-1} H(x;q,M) \\ & \in & x - IGA\Big(G(x,[z],[q],[M]),\, H(x;[q],[M])\Big) \\ & = & N(x,[z],[q],[M]) \subseteq [z] \end{array}\right\} \tag{3.26}$$

für jedes $y \in [z]$. Die stetige Funktion $p(\cdot)$ bildet also die nichtleere, konvexe und kompakte Menge $[z]$ in sich ab. Nach dem Fixpunktsatz von Brouwer [26] gibt es daher ein $z \in [z]$ mit

$$z = p(z) = z - \Big(G(x,z,q,M)\Big)^{-1} H(z;q,M). \tag{3.27}$$

Somit ist $H(z;q,M) = 0$. Zuletzt gilt wegen (3.27) und (3.26)

$$z = p(z) \in N(x,[z],[q],[M]).$$

Beweis zu Behauptung 2: Wir nehmen an, es gäbe ein $q \in [q]$, ein $M \in [M]$ und ein $z \in [z]$ mit $H(z;q,M) = 0$. Dann gilt

$$\begin{array}{rcl} z & = & z - \Big(G(x,z,q,M)\Big)^{-1} H(z;q,M) \\ & = & z - \Big(G(x,z,q,M)\Big)^{-1} (\, H(z;q,M) + H(x;q,M) - H(x;q,M)\,) \\ & = & z - \Big(G(x,z,q,M)\Big)^{-1} H(x;q,M) - (z-x) \\ & = & x - \Big(G(x,z,q,M)\Big)^{-1} H(x;q,M) \\ & \in & x - IGA\Big(G(x,[z],[q],[M]),\, H(x;[q],[M])\Big) \\ & = & N(x,[z],[q],[M]). \end{array}$$

Daraus würde $z \in N(x,[z],[q],[M]) \cap [z]$ folgen, was nach Voraussetzung ausgeschlossen war.

Beweis zu Behauptung 3: Man erhält mit $H(z;q,M)=0$ und (3.25)

$$H(x;q,M)=H(x;q,M)-H(z;q,M)=G(x,z,q,M)(x-z).$$

Der I.G.A. ist durchführbar für $G(x,[z],[q],[M])$. Daher ist $G(x,z,q,M)$ invertierbar, und es gilt

$$\Big(G(x,z,q,M)\Big)^{-1}H(x;q,M)=x-z.$$

Es resultiert

$$\begin{aligned} z &= x-\Big(G(x,z,q,M)\Big)^{-1}H(x;q,M)\\ &\in x-IGA\Big(G(x,[z],[q],[M]),\,H(x;[q],[M])\Big)\\ &= N(x,[z],[q],[M]). \end{aligned}$$

Beweis zu Behauptung 4: Es ist $z\in[z^0]$, $H(z;q,M)=0$ und der I.G.A. ist durchführbar für $G(x^0,[z^0],[q],[M])$. Nach Behauptung 3 ist dann $z\in N(x^0,[z^0],[q],[M])$. Somit erhalten wir

$$z\in N(x^0,[z^0],[q],[M])\cap[z^0]=[z^1].$$

Die eigentliche Aussage erfolgt dann durch vollständige Induktion. □

Bemerkungen:

1. Die wesentliche Idee zum Beweis von Satz 3.13 findet man auch in [3].

2. In [1] wird der Operator $L(x,A,[z],q,M)=$

   $$x-A^{-1}H(x;q,M)+(I-A^{-1}G(x,[z],q,M))([z]-x)$$

   betrachtet. Dieser Operator würde für den allgemeinen Fall dann $L(x,A,[z],[q],[M])=$

   $$x-A^{-1}H(x;[q],[M])+(I-A^{-1}G(x,[z],[q],[M]))([z]-x)$$

   lauten. Es ist aber i.a. $d(H(x;[q],[M]))>0$, und es gilt ganz allgemein für $[a]$, $[b]\in\mathbf{IR}$:

   $$d([a]\pm[b])=d([a])+d([b]).$$

Daher wird sich der Operator $L(x, A, [z], [q], [M])$ aufblähen, und er wird für den Test

$$L(x, A, [z], [q], [M]) \subseteq [z]$$

nicht sehr geeignet sein. Wir haben uns daher in Satz 3.13 für den Operator

$$N(x, [z], [q], [M]) := x - IGA\Big(G(x, [z], [q], [M]),\, H(x; [q], [M])\Big)$$

entschieden.

3. Behauptung 4 von Satz 3.13 eröffnet uns sofort einen Algorithmus für den folgenden Fall:

   Gegeben seien $[M] \in \mathbf{IR}^{n\times n}$, $[q] \in \mathbf{IR}^n$, das dazugehörende LCP mit Intervalleinträgen (3.23) und ein $[z] \in \mathbf{IR}^n$.

   Gesucht ist ein bestimmtes

   $$\begin{aligned} z \in L_B \;=\; \Big\{ & z \in \mathbf{R}^n : \text{ Es gibt ein } M \in [M] \text{ und ein } q \in [q] \text{ mit} \\ & H(z; q, M) = \min(q + Mz, z) = 0 \Big\}, \end{aligned}$$

   von dem wir wissen, daß

   $$z \in [z] \tag{3.28}$$

   gilt. Dann konvergiert die Iteration

   $$\begin{aligned} [z^0] \;&:=\; [z]; \\ [z^{k+1}] \;&:=\; N(x^k, [z^k], [q], [M]) \cap [z^k],\; x^k \in [z^k],\; k = 0, 1, 2, ..., \end{aligned}$$

   gegen einen Intervallvektor $[z^*] \in \mathbf{IR}^n$ (siehe [2], Corollary 10.8), falls $N(x^k, [z^k], [q], [M])$ für $k = 0, 1, 2, ...$, definiert ist, und es gilt mit der Behauptung 4 von Satz 3.13

   $$z \in [z^*].$$

4. Ist für den gesuchten Vektor $z \in L_B$ aus 3. kein Intervallvektor $[z] \in \mathbf{IR}^n$ bekannt mit (3.28), so läßt sich ein solches $[z] \in \mathbf{IR}^n$ konstruieren, falls $[M]$ eine H-Matrix mit $0 < \underline{m_{ii}}$, $i = 1(1)n$, ist. Wir setzen gemäß Satz 3.10

   $$[z] := [SB] := (abs([y^0]) + [y^0]) \cap \mathbf{IR}^n_{\geq 0}\,,$$

wobei wir $[y^0]$ mit Lemma 3.4 bestimmen. Mit Satz 3.10 gilt dann

$$z \in L_B \subseteq [z].$$

Die zentrale Rolle in Satz 3.13 spielt die Intervallmatrix $G(x,[z],[q],[M])$, deren Bestimmung wir jetzt in Angriff nehmen.

Gegeben seien $[q] \in \mathbf{IR}^n$, $[M] \in \mathbf{IR}^{n\times n}$ und $[z] \in \mathbf{IR}^n$. (Der Fall $[q] = q \in \mathbf{R}^n$, $[M] = M \in \mathbf{R}^{n\times n}$ wurde bereits in [1] behandelt.) Für festes $q \in [q]$ und für festes $M \in [M]$ definieren wir die Funktion

$$H(z;q,M) := \min(q + Mz, z), \quad z \in [z].$$

Es sei $x \in [z]$ fest gewählt. Dann ist eine Intervallmatrix

$$G(x,[z],[q],[M]) = \begin{pmatrix} G_1(x,[z],[q],[M]) \\ \vdots \\ G_n(x,[z],[q],[M]) \end{pmatrix} \in \mathbf{IR}^{n\times n}$$

gesucht, die folgendes erfüllt:

$$\left.\begin{array}{rcll} H(x;q,M) - H(y;q,M) & = & G(x,y,q,M)(x-y) & \\ & \in & G(x,[z],[q],[M])(x-y) & \text{für alle } y \in [z]. \end{array}\right\} \tag{3.29}$$

Es bezeichne im folgenden

$$\begin{array}{cl} [m_i]^{\mathrm{T}} & \text{die } i\text{-te Zeile von } [M], \\ e_i & \text{den } i\text{-ten Einheitsvektor}, \\ \Big(H(\cdot;\cdot,\cdot)\Big)_i & \text{die } i\text{-te Komponente von } H(\cdot;\cdot,\cdot). \end{array}$$

$y \in [z]$ sei beliebig, und es sei $i \in \{1,...,n\}$. Um $\Big(H(x;q,M)\Big)_i$ bzw. $\Big(H(y;q,M)\Big)_i$ für festes $q \in [q]$ und festes $M \in [M]$ übersichtlicher ausrechnen zu können, betrachten wir die Ausdrücke

$$(m_i - e_i)^{\mathrm{T}} x + q_i \quad \text{bzw.} \quad (m_i - e_i)^{\mathrm{T}} y + q_i.$$

3.3.1 Bestimmung der Intervallmatrix $G(x,[z],[q],[M])$ für den Fall: $[M] = M \in \mathbf{R}^{n\times n}$

Es gelten alle Bezeichnungen und Vereinbarungen aus dem übergeordneten Abschnitt.

Der Spezialfall $[M] = M \in \mathbf{R}^{n\times n}$ ist deshalb so wichtig, weil er in den Anwendungen verstärkt auftritt (siehe Kapitel 4).

Die Konstruktion von $G(x,[z],[q],M)$ vollzieht sich zeilenweise, und zwar durch drei Fallunterscheidungen mit jeweils drei Unterfällen.

1. Fall: $(m_i - e_i)^{\mathrm{T}}x + [q_i] > 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + [q_i] \geq 0$.

Wir erhalten also für jedes feste $q_i \in [q_i]$:

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}}x - e_i^{\mathrm{T}}y = e_i^{\mathrm{T}}(x-y).$$

2. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + [q_i] \leq 0$.

Dann gilt für festes $q_i \in [q_i]$:

$$\begin{aligned}\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i &= e_i^{\mathrm{T}}x - m_i^{\mathrm{T}}y - q_i \\ &= \left(m_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{(m_i - e_i)^{\mathrm{T}}(x-y)}(e_i - m_i)^{\mathrm{T}}\right)(x-y).\end{aligned}$$

Nun beobachten wir, daß

$$0 \leq \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{\underbrace{(m_i - e_i)^{\mathrm{T}}x + q_i}_{>0} - \underbrace{((m_i - e_i)^{\mathrm{T}}y + q_i)}_{\leq 0}} \leq 1$$

gilt. Ist $y^i \in [z]$ ein Vektor, für welchen

$$(m_i - e_i)^{\mathrm{T}}y, \quad y \in [z],$$

minimal wird - wir wollen diesen Sachverhalt im folgenden mit

$$y^i = \arg\min_{y\in[z]}(m_i - e_i)^{\mathrm{T}}y$$

bezeichnen -, so ist

$$(m_i - e_i)^{\mathrm{T}} y + q_i \in [(m_i - e_i)^{\mathrm{T}} y^i + q_i, 0].$$

Daher können wir

$$m_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{(m_i - e_i)^{\mathrm{T}}(x - y)} (e_i - m_i)^{\mathrm{T}}$$

in

$$m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \underline{q_i}}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1 \right] (e_i - m_i)^{\mathrm{T}}$$

einschließen und erhalten

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \underline{q_i}}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1 \right] (e_i - m_i)^{\mathrm{T}} \right) (x - y).$$

3. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [q_i] \ni 0$.

Dann existiert ein $q_i^* \in [q_i]$ mit

$$[q_i] = [\underline{q_i}, q_i^*] \cup [q_i^*, \overline{q_i}]$$

und

$$\begin{aligned} (m_i - e_i)^{\mathrm{T}} y + [\underline{q_i}, q_i^*] &\leq 0, \\ (m_i - e_i)^{\mathrm{T}} y + [q_i^*, \overline{q_i}] &\geq 0. \end{aligned}$$

Ist $q_i \in [\underline{q_i}, q_i^*]$, so ergibt sich nach dem 2. Unterfall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \underline{q_i}}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1 \right] (e_i - m_i)^{\mathrm{T}} \right) (x - y).$$

Ist $q_i \in [q_i^*, \overline{q_i}]$, so ergibt sich nach dem 1. Unterfall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}}(x - y).$$

Insgesamt ergibt sich also (für $q_i \in [q_i]$)

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}}x + \underline{q_i}}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1\right](e_i - m_i)^{\mathrm{T}}\right)(x - y).$$

2. Fall: $(m_i - e_i)^{\mathrm{T}}x + [q_i] < 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + [q_i] \geq 0$.

Dann gilt für festes $q_i \in [q_i]$:

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}}x + q_i - e_i^{\mathrm{T}}y$$
$$= \left(e_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{(m_i - e_i)^{\mathrm{T}}(x - y)}(m_i - e_i)^{\mathrm{T}}\right)(x - y).$$

Nun beobachten wir, daß

$$0 \leq \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{\underbrace{(m_i - e_i)^{\mathrm{T}}x + q_i}_{<0} - \underbrace{((m_i - e_i)^{\mathrm{T}}y + q_i)}_{\geq 0}} \leq 1$$

gilt. Ist $z^i \in [z]$ ein Vektor, für welchen

$$(m_i - e_i)^{\mathrm{T}}y, \quad y \in [z],$$

maximal wird - wir wollen diesen Sachverhalt im folgenden mit

$$z^i = \arg\max_{y \in [z]} (m_i - e_i)^{\mathrm{T}}y$$

bezeichnen -, so ist

$$(m_i - e_i)^{\mathrm{T}}y + q_i \in [0, (m_i - e_i)^{\mathrm{T}}z^i + q_i].$$

Daher können wir

$$e_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{(m_i - e_i)^{\mathrm{T}}(x - y)}(m_i - e_i)^{\mathrm{T}}$$

in

$$e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \overline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1 \right] (m_i - e_i)^{\mathrm{T}}$$

einschließen und erhalten

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \overline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1 \right] (m_i - e_i)^{\mathrm{T}} \right) (x - y).$$

2. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [q_i] \leq 0$.

Dann gilt für jedes feste $q_i \in [q_i]$:

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}} x + q_i - m_i^{\mathrm{T}} y - q_i = m_i^{\mathrm{T}}(x - y).$$

3. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [q_i] \ni 0$.

Dann existiert ein $q_i^{**} \in [q_i]$ mit

$$[q_i] = [\underline{q_i}, q_i^{**}] \cup [q_i^{**}, \overline{q_i}]$$

und

$$\begin{aligned} (m_i - e_i)^{\mathrm{T}} y + [\underline{q_i}, q_i^{**}] &\leq 0, \\ (m_i - e_i)^{\mathrm{T}} y + [q_i^{**}, \overline{q_i}] &\geq 0. \end{aligned}$$

Ist $q_i \in [\underline{q_i}, q_i^{**}]$, so ergibt sich nach dem 2. Unterfall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}}(x - y).$$

Ist $q_i \in [q_i^*, \overline{q_i}]$, so ergibt sich nach dem 1. Unterfall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \overline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1 \right] (m_i - e_i)^{\mathrm{T}} \right) (x - y).$$

Insgesamt ergibt sich also (für $q_i \in [q_i]$)

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$
$$\left(e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \overline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1 \right] (m_i - e_i)^{\mathrm{T}} \right) (x - y).$$

3. Fall: $(m_i - e_i)^{\mathrm{T}} x + [q_i] \ni 0$.

Dann existiert ein $q_i^{***} \in [q_i]$ mit

$$(m_i - e_i)^{\mathrm{T}} x + [\underline{q}_i, q_i^{***}) \;<\; 0, \tag{3.30}$$

$$(m_i - e_i)^{\mathrm{T}} x + q_i^{***} \;=\; 0, \tag{3.31}$$

$$(m_i - e_i)^{\mathrm{T}} x + (q_i^{***}, \overline{q}_i] \;>\; 0. \tag{3.32}$$

Zu (3.30): $(m_i - e_i)^{\mathrm{T}} x + [\underline{q}_i, q_i^{***}) < 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [\underline{q}_i, q_i^{***}] \leq 0$.

Dann gilt gemäß dem 2. Unterfall im 2. Fall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}}(x-y).$$

Für den

2. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [\underline{q}_i, q_i^{***}] \geq 0$

bzw. für den

3. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + [\underline{q}_i, q_i^{***}] \ni 0$

erhält man gemäß dem 1. Unterfall im 2. Fall bzw. gemäß dem 3. Unterfall im 2. Fall

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in$$

$$\left(e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + q_i^{***}}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1\right](m_i - e_i)^{\mathrm{T}}\right)(x-y)$$

$$= \Big(e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}}\Big)(x-y).$$

Zu (3.31): $(m_i - e_i)^{\mathrm{T}} x + q_i^{***} = 0$.

Nun gilt entweder $(m_i - e_i)^{\mathrm{T}} y + q_i^{***} \geq 0$, was

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}}(x-y)$$

nach sich zieht, oder es gilt $(m_i - e_i)^{\mathrm{T}} y + q_i^{***} < 0$. Dann gilt

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}}(x-y).$$

Insgesamt kann man also

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in [\min\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}, \max\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}](x-y)$$

schließen.

Zu (3.32): $(m_i - e_i)^{\mathrm{T}} x + (q_i^{***}, \overline{q}_i] > 0$.

$$\text{1. Unterfall: } (m_i - e_i)^{\mathrm{T}} y + [q_i^{***}, \overline{q}_i] \geq 0.$$

Gemäß dem 1. Unterfall im 1. Fall gilt dann

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}}(x-y).$$

Für den

$$\text{2. Unterfall: } (m_i - e_i)^{\mathrm{T}} y + [q_i^{***}, \overline{q}_i] \leq 0$$

bzw. für den

$$\text{3. Unterfall: } (m_i - e_i)^{\mathrm{T}} y + [q_i^{***}, \overline{q}_i] \ni 0$$

erhält man gemäß dem 2. Unterfall im 1. Fall bzw. gemäß dem 3. Unterfall im 1. Fall

$$\begin{aligned}
&\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in \\
&\left(m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + q_i^{***}}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1\right](e_i - m_i)^{\mathrm{T}}\right)(x-y) \\
&= \Big(m_i^{\mathrm{T}} + [0,1](e_i - m_i)^{\mathrm{T}}\Big)(x-y).
\end{aligned}$$

Beachtet man nun

$$\left.\begin{aligned}
m_i^{\mathrm{T}} + [0,1](e_i - m_i)^{\mathrm{T}} &= e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}} \\
&= [\min\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}, \max\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}],
\end{aligned}\right\} \tag{3.33}$$

so kann man die Ergebnisse zu (3.30),(3.31),(3.32) zusammenfassen in

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in \Big(e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}}\Big)(x-y).$$

Um die Bestimmung von $G(x,[z],[q],M)$ letztendlich durchführen zu können, benötigen wir noch

$$y^i = \arg\min_{y \in [z]} (m_i - e_i)^{\mathrm{T}} y,$$

$$z^i = \arg\max_{y \in [z]} (m_i - e_i)^{\mathrm{T}} y.$$

Dies läßt sich aber sehr einfach bewerkstelligen. Es ist

$$y_j^i = \begin{cases} \underline{z}_j & \text{falls } (m_i - e_i)_j > 0, \\ \overline{z}_j & \text{sonst,} \end{cases} \qquad j = 1(1)n, \tag{3.34}$$

und

$$z_j^i = \begin{cases} \underline{z}_j & \text{falls } (m_i - e_i)_j \leq 0, \\ \overline{z}_j & \text{sonst,} \end{cases} \qquad j = 1(1)n. \tag{3.35}$$

Algorithmus zur Bestimmung von $G(x, [z], [q], M)$

Gegeben seien $[z], [q] \in \mathbf{IR}^n$, $M \in \mathbf{R}^{n \times n}$ und $x \in [z]$. Wir bestimmen $G(x, [z], [q], M)$ zeilenweise.

for $i := 1$ **to** n **do**
begin

1. Bestimme
$$y^i = \arg\min_{y \in [z]} (m_i - e_i)^{\mathrm{T}} y,$$
gemäß (3.34). Ist dann
$$(m_i - e_i)^{\mathrm{T}} y^i + [q_i] \geq 0,$$
so setze
$$G_i(x, [z], [q], M) := e_i^{\mathrm{T}}.$$
Ansonsten fahre fort mit

2. Bestimme
$$z^i = \arg\max_{y \in [z]} (m_i - e_i)^{\mathrm{T}} y,$$
gemäß (3.35). Ist dann
$$(m_i - e_i)^{\mathrm{T}} z^i + [q_i] \leq 0,$$
so setze
$$G_i(x, [z], [q], M) := m_i^{\mathrm{T}}.$$
Ansonsten fahre fort mit

3. (a) Ist

$$(m_i - e_i)^{\mathrm{T}} x + [q_i] > 0,$$

dann setze

$$G_i(x,[z],[q],M) := m_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \underline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - y^i)}, 1 \right] (e_i - m_i)^{\mathrm{T}}.$$

Ansonsten fahre fort mit

(b) Ist

$$(m_i - e_i)^{\mathrm{T}} x + [q_i] < 0,$$

dann setze

$$G_i(x,[z],[q],M) := e_i^{\mathrm{T}} + \left[\frac{(m_i - e_i)^{\mathrm{T}} x + \overline{q}_i}{(m_i - e_i)^{\mathrm{T}}(x - z^i)}, 1 \right] (m_i - e_i)^{\mathrm{T}}.$$

Ansonsten fahre fort mit

(c) Ist

$$(m_i - e_i)^{\mathrm{T}} x + [q_i] \ni 0,$$

dann setze

$$G_i(x,[z],[q],M) := e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}}.$$

end.

3.3.2 Allgemeiner Fall

Gegeben seien $[q], [z] \in \mathbf{IR}^n$ und $[M] \in \mathbf{IR}^{n \times n}$. Für festes $x \in [z]$ suchen wir eine Intervallmatrix $G(x,[z],[q],[M]) \in \mathbf{IR}^{n \times n}$, die der Bedingung (3.29) genügt. Wir bestimmen $G(x,[z],[q],[M])$ zeilenweise.

Dazu seien $q \in [q]$ und $M \in [M]$ beliebig, aber fest, und $y \in [z]$.

1. Fall: $(m_i - e_i)^{\mathrm{T}} x + q_i > 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + q_i > 0$.

Dann ist

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}} x - e_i^{\mathrm{T}} y = e_i^{\mathrm{T}}(x - y).$$

2. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + q_i \leq 0$.

Dann ist

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}} x - m_i^{\mathrm{T}} y - q_i$$
$$= \left(m_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{(m_i - e_i)^{\mathrm{T}}(x-y)} (e_i - m_i)^{\mathrm{T}}\right)(x-y).$$

Wegen

$$0 \leq \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{\underbrace{(m_i - e_i)^{\mathrm{T}} x + q_i}_{>0} - \underbrace{((m_i - e_i)^{\mathrm{T}} y + q_i)}_{\leq 0}} \leq 1$$

und (3.33) erhält man dann

$$\begin{aligned}\left(m_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{(m_i - e_i)^{\mathrm{T}}(x-y)} (e_i - m_i)^{\mathrm{T}}\right) &\in m_i^{\mathrm{T}} + [0,1](e_i - m_i)^{\mathrm{T}} \\ &= e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}} \\ &\in e_i^{\mathrm{T}} + [0,1]([m_i] - e_i)^{\mathrm{T}}.\end{aligned}$$

Also gilt

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in \Big(e_i^{\mathrm{T}} + [0,1]([m_i] - e_i)^{\mathrm{T}}\Big)(x-y).$$

2. Fall: $(m_i - e_i)^{\mathrm{T}} x + q_i < 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}} y + q_i > 0$.

Nun gilt

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}} x + q_i - e_i^{\mathrm{T}} y$$
$$= \left(e_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{(m_i - e_i)^{\mathrm{T}}(x-y)} (m_i - e_i)^{\mathrm{T}}\right)(x-y).$$

Wegen

$$0 \leq \frac{(m_i - e_i)^{\mathrm{T}} x + q_i}{\underbrace{(m_i - e_i)^{\mathrm{T}} x + q_i}_{<0} - \underbrace{((m_i - e_i)^{\mathrm{T}} y + q_i)}_{>0}} \leq 1$$

gilt

$$\begin{aligned}\left(e_i^{\mathrm{T}} + \frac{(m_i - e_i)^{\mathrm{T}}x + q_i}{(m_i - e_i)^{\mathrm{T}}(x-y)}(m_i - e_i)^{\mathrm{T}}\right) &\in e_i^{\mathrm{T}} + [0,1](m_i - e_i)^{\mathrm{T}} \\ &\in e_i^{\mathrm{T}} + [0,1]([m_i] - e_i)^{\mathrm{T}}.\end{aligned}$$

Somit bekommt man

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in \left(e_i^{\mathrm{T}} + [0,1]([m_i] - e_i)^{\mathrm{T}}\right)(x-y).$$

2. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + q_i \leq 0$.

Hier ergibt sich

$$\begin{aligned}\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i &= m_i^{\mathrm{T}}x + q_i - m_i^{\mathrm{T}}y - q_i \\ &= m_i^{\mathrm{T}}(x-y) \\ &\in [m_i]^{\mathrm{T}}(x-y).\end{aligned}$$

3. Fall: $(m_i - e_i)^{\mathrm{T}}x + q_i = 0$.

1. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + q_i > 0$.

Dann gilt

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = e_i^{\mathrm{T}}(x-y).$$

2. Unterfall: $(m_i - e_i)^{\mathrm{T}}y + q_i \leq 0$.

Hier gilt dann

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i = m_i^{\mathrm{T}}(x-y).$$

Man kann beide Unterfälle zusammenfassen zu

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in [\min\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}, \max\{e_i^{\mathrm{T}}, m_i^{\mathrm{T}}\}](x-y),$$

was dann mit Hilfe von (3.33) auf

$$\Big(H(x;q,M)\Big)_i - \Big(H(y;q,M)\Big)_i \in \left(e_i^{\mathrm{T}} + [0,1]([m]_i - e_i)^{\mathrm{T}}\right)(x-y)$$

führt.

Algorithmus zur Bestimmung von $G(x,[z],[q],[M])$

Gegeben seien $[z], [q] \in \mathbf{IR}^n$, $[M] \in \mathbf{IR}^{n\times n}$ und $x \in [z]$. Wir bestimmen $G(x,[z],[q],[M])$ zeilenweise.

for $i := 1$ **to** n **do**
begin
if $([m_i] - e_i)^{\mathrm{T}}[z] + [q_i] \geq 0$
then $G_i(x, [z], [q], [M]) := e_i^{\mathrm{T}}$
else if $([m_i] - e_i)^{\mathrm{T}}[z] + [q_i] \leq 0$
then $G_i(x, [z], [q], [M]) := [m_i]^{\mathrm{T}}$
else $G_i(x, [z], [q], [M]) := e_i^{\mathrm{T}} + [0, 1]([m_i] - e_i)^{\mathrm{T}}$
end.

Dieser Algorithmus ist übersichtlich und sehr einfach zu programmieren. Auch für den Fall $[M] = M \in \mathbf{R}^{n \times n}$ liefert er das Gewünschte. Allerdings gilt im Fall $[M] = M \in \mathbf{R}^{n \times n}$ für die Intervallmatrix $G(x, [z], [q], M)$ aus Abschnitt 3.3.1 stets

$$G(x, [z], [q], M) \subseteq G(x, [z], [q], [M]),$$

wie man sehr einfach in den Abschnitten 3.3.1 und 3.3.2 erkennen kann.

Beispiel 3.8 Gegeben seien

$$[M] = M = \begin{pmatrix} 1 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 \end{pmatrix}, \quad [q] = \begin{pmatrix} [1,2] \\ \frac{3}{4} \\ 0 \end{pmatrix}$$

und

$$[z] = \begin{pmatrix} [1,2] \\ [0,1] \\ 0 \end{pmatrix}.$$

Als $x \in [z]$ sei $\underline{z} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ gewählt. Der Algorithmus zur Bestimmung von $G(x, [z], [q], [M])$ aus diesem Abschnitt liefert für $i = 1$:

$$([m_1] - e_1)^{\mathrm{T}}[z] + [q_1] = [-\frac{1}{2}, 0] + [1, 2] \geq 0.$$

Man erhält

$$G_1(x,[z],[q],[M]) = e_1^{\mathrm{T}} = (1 \quad 0 \quad 0).$$

Für $i = 2$ bekommt man:

$$([m_2] - e_2)^{\mathrm{T}}[z] + [q_2] = [-1, -\frac{1}{2}] + \frac{3}{4} = [-\frac{1}{4}, \frac{1}{4}].$$

Es wird daher

$$G_2(x,[z],[q],[M]) = e_2^{\mathrm{T}} + [0,1]([m_2] - e_2)^{\mathrm{T}} = \left([-\frac{1}{2}, 0] \quad 1 \quad [-\frac{1}{2}, 0]\right)$$

gesetzt. Schließlich liefert der Algorithmus für $i = 3$:

$$G_3(x,[z],[q],[M]) = [m_3]^{\mathrm{T}} = \left(0 \quad -\frac{1}{2} \quad 1\right),$$

da

$$([m_3] - e_3)^{\mathrm{T}}[z] + [q_3] = [-\frac{1}{2}, 0] + 0 \leq 0$$

gilt. Insgesamt erhält man die Matrix

$$G(x,[z],[q],[M]) = \begin{pmatrix} 1 & 0 & 0 \\ [-\frac{1}{2}, 0] & 1 & [-\frac{1}{2}, 0] \\ 0 & -\frac{1}{2} & 1 \end{pmatrix}.$$

Der Algorithmus zur Bestimmung von $G(x,[z],[q],M)$ aus Abschnitt 3.3.1 berechnet für $i = 1$:

$$y^1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \text{(gemäß (3.34))}$$

und

$$(m_1 - e_1)^{\mathrm{T}} y^1 + [q_1] = -\frac{1}{2} + [1,2] \geq 0.$$

Man erhält

$$G_1(x,[z],[q],M) = e_1^{\mathrm{T}} = (1 \quad 0 \quad 0).$$

Für $i = 2$ bekommt man mit

$$y^2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad \text{(gemäß (3.34))}$$

die Aussage

$$(m_2 - e_2)^{\mathrm{T}} y^2 + [q_2] = -1 + \frac{3}{4} \not\geq 0,$$

und mit

$$z^2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{(gemäß (3.35))}$$

die Aussage

$$(m_2 - e_2)^{\mathrm{T}} z^2 + [q_2] = -\frac{1}{2} + \frac{3}{4} \not\leq 0.$$

Wegen

$$(m_2 - e_2)^{\mathrm{T}} x + [q_2] = -\frac{1}{2} + \frac{3}{4} > 0$$

setzt dann der Algorithmus

$$\begin{aligned} G_2(x, [z], [q], M) &:= m_2^{\mathrm{T}} + \left[\frac{(m_2 - e_2)^{\mathrm{T}} x + \underline{q_2}}{(m_2 - e_2)^{\mathrm{T}} (x - y^2)}, 1 \right] (e_2 - m_2)^{\mathrm{T}} \\ &= \left(-\tfrac{1}{2} \quad 1 \quad -\tfrac{1}{2} \right) + [\tfrac{1}{2}, 1] \left(\tfrac{1}{2} \quad 0 \quad \tfrac{1}{2} \right) \\ &= \left(\left[-\tfrac{1}{4}, 0 \right] \quad 1 \quad [-\tfrac{1}{4}, 0] \right). \end{aligned}$$

Für die dritte Zeile erhalten wir

$$G_3(x, [z], [q], M) = m_3^{\mathrm{T}} = \begin{pmatrix} 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

wegen

$$(m_3 - e_3)^{\mathrm{T}} y^3 + [q_3] = \begin{pmatrix} 0 & -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + 0 = -\frac{1}{2} \not\geq 0$$

und

$$(m_3 - e_3)^{\mathrm{T}} z^3 + [q_3] = \begin{pmatrix} 0 & -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 0 = 0 \leq 0.$$

Insgesamt erhält man die Matrix

$$G(x,[z],[q],M) = \begin{pmatrix} 1 & 0 & 0 \\ [-\frac{1}{4},0] & 1 & [-\frac{1}{4},0] \\ 0 & -\frac{1}{2} & 1 \end{pmatrix}.$$

Es gilt $2 \cdot d(G(x,[z],[q],M)) = d(G(x,[z],[q],[M]))$.

Im Hinblick auf die 3. Bemerkung zu Satz 3.13 interessiert uns jetzt die Frage, ob man unter gewissen Voraussetzungen an $[M]$ garantieren kann, daß der I.G.A. durchführbar ist für $G(x,[z],[q],[M])$. Dazu betrachten wir den folgenden

Satz 3.14 *Ist die Intervallmatrix $[M]$ eine H-Matrix mit $0 < \underline{m}_{ii}$, $\overline{m}_{ii} \leq 1$, $i = 1(1)n$, so ist die Intervallmatrix $G(x,[z],[q],[M])$ für jedes $[z],[q] \in \mathbf{IR}^n$ und für jedes $x \in [z]$ ebenfalls eine H-Matrix.*

Beweis: Aufgrund der Konstruktion von $G(x,[z],[q],[M])$ gilt:

$$G(x,[z],[q],[M]) \subseteq \begin{pmatrix} e_1^{\mathrm{T}} + [0,1]([m_1] - e_1)^{\mathrm{T}} \\ \vdots \\ e_n^{\mathrm{T}} + [0,1]([m_n] - e_n)^{\mathrm{T}} \end{pmatrix} =: [H].$$

Für $i,j \in \{1,...,n\}$, $i \neq j$, gilt dann

$$\begin{aligned} \left\langle [H] \right\rangle_{ii} &= \left\langle 1 + [0,1]([m_{ii}] - 1) \right\rangle \\ &= \left\langle 1 + [-1 + \underline{m}_{ii}, 0] \right\rangle \\ &= \left\langle [\underline{m}_{ii}, 1] \right\rangle = \underline{m}_{ii} \end{aligned}$$

und

$$\left\langle [H] \right\rangle_{ij} = -|[0,1] \cdot [m_{ij}]| = -|[m_{ij}]|.$$

Es ist also $\langle [H] \rangle = \langle [M] \rangle$. Daher ist wegen

$$\langle [H] \rangle \leq \langle G(x, [z], [q], [M]) \rangle$$

aufgrund der Folgerungen von Lemma 2.2 die Matrix

$$\langle G(x, [z], [q], [M]) \rangle$$

eine M-Matrix. $G(x, [z], [q], [M])$ ist daher eine H-Matrix. □

Aufgrund von Satz 2.5 ist der I.G.A. also durchführbar für $G(x, [z], [q], [M])$, falls $[M]$ die Voraussetzung von Satz 3.14 erfüllt, und die Iteration in der 3. Bemerkung zu Satz 3.13 kann stets durchgeführt werden.

Kapitel 4

Anwendung bei gewöhnlichen freien Randwertproblemen

4.1 Freie Randwertprobleme

Unter einer Lösung eines Randwertproblems versteht man eine Funktion, die auf einem vorgegebenen Gebiet einer Differentialgleichung genügt und auf dem Rand des Gebiets vorgegebene Bedingungen erfüllt.

Im Unterschied dazu ist bei einem freien Randwertproblem der Rand des Gebiets im voraus nicht bekannt und muß als Teil der Lösung mitbestimmt werden.

Man denke z.B. an einen schmelzenden Eiswürfel in einem Glas Wasser. Der Inhalt des Glases trennt sich in ein 'Wassergebiet' und in ein 'Eisgebiet'. Um die Temperatur im Wassergebiet als eine Funktion von Raum und Zeit zu bestimmen, muß man eine parabolische Differentialgleichung lösen (siehe etwa [7]). Der Rand des Wassergebiets bewegt[1] sich aber mit der Zeit, weil der Eiswürfel im Wasser schmilzt, und ist somit nicht im voraus bekannt.

Ein weiteres Beispiel für ein freies Randwertproblem ist das sogenannte Dammproblem. Wir wollen hier nicht näher darauf eingehen, sondern nur bemerken, daß bei dem Dammproblem der freie Rand den Damm in ein von Wasser getränktes und in ein trockenes Gebiet teilt und daß in dem vom

[1]Freie Randwertprobleme, bei denen sich der freie Rand mit der Zeit ändert/bewegt, nennt man im Englischen daher auch oft *moving boundary problems*. Siehe etwa [12].

Wasser getränkten Gebiet eine elliptische Differentialgleichung zu lösen ist. Für Details verweisen wir auf [12].

Als abschließendes Beispiel betrachten wir ein gewöhnliches freies Randwertproblem. In der (x, y)-Ebene sei ein Seil im Punkt $x = 0$, $y = y_0$ aufgehängt

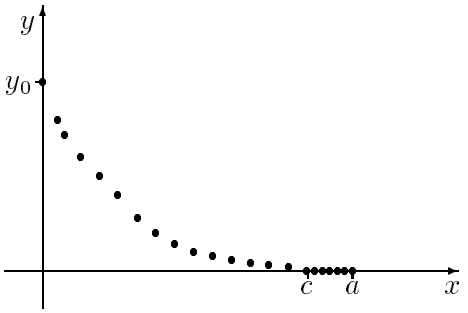

Abbildung 4.1: Freie Randwertaufgabe bei einem Seil

und liege ein Stück auf dem Boden, der durch $y = 0$ beschrieben sei. Siehe Abbildung 4.1.

Das Seil berühre den Boden mit horizontaler Tangente im Punkt $x = c$ und im Intervall $[0, c]$ genüge die Form des Seils, die durch $y(x)$ beschrieben sei, der Differentialgleichung

$$y''(x) = \sqrt{1 + (y'(x))^2}, \tag{4.1}$$

wobei die Stelle c im voraus nicht bekannt ist. Um dieses Problem zu lösen, muß man also den freien Rand c bestimmen und auf $[0, c]$ die gewöhnliche Differentialgleichung (4.1) lösen. Wir wollen dieses Problem mit 'Seilproblem' bezeichnen. Dieses Beispiel stammt aus [8].

4.2 Problemstellung, Ziel und Vorgehensweise

In dieser Arbeit betrachten wir ausschließlich gewöhnliche freie Randwertprobleme. Die dem freien Randwertproblem zugrunde liegende Differentialgleichung ist also eine gewöhnliche Differentialgleichung.

Es seien also eine auf $[0,\infty) \times \mathbf{R} \times \mathbf{R}$ definierte Funktion $f(x,s,t)$ und ein $y_0 > 0$ gegeben. Dann betrachten wir das folgende gewöhnliche freie Randwertproblem:

$$\left.\begin{aligned} \text{Finde } c \in \mathbf{R} \quad \text{und} \quad & y(\cdot) : [0,\infty) \to \mathbf{R} \text{ mit} \\ y''(x) = f(x, y(x), y'(x)) \quad & \text{für } x \in [0,c], \\ y(x) > 0 \quad & \text{für } x \in [0,c), \\ y(x) \equiv 0 \quad & \text{für } x \in [c,\infty), \\ y'(c) = 0, & \\ y(0) = y_0. & \end{aligned}\right\} \tag{4.2}$$

Wir beschäftigen uns in dieser Arbeit nicht mit der Frage nach der Existenz oder Eindeutigkeit einer Lösung von (4.2). Wir verweisen diesbezüglich auf [36] und [37]. Dort werden sogar Systeme von gewöhnlichen freien Randwertproblemen betrachtet, nachdem man mit Hilfe der Linienmethode (vgl. [34]) einem parabolischen freien Randwertproblem ein in t-Richtung diskretisiertes Problem zugeordnet hat, das aus einem System von gewöhnlichen freien Randwertproblemen besteht.

Wir gehen in dieser Arbeit davon aus, das Problem (4.2) habe eine Lösung $\tilde{c} \in \mathbf{R}$ und $\tilde{y}(\cdot) : [0,\infty) \to \mathbf{R}$. Ob $\tilde{c}$, $\tilde{y}(\cdot)$ dabei eindeutig bestimmt sind, spielt keine Rolle.

In den nächsten beiden Abschnitten werden wir zeigen, wie man unter gewissen Voraussetzungen an $f(x,s,t)$ ein Intervall $[C] \in \mathbf{IR}$ angeben kann mit

$$\tilde{c} \in [C],$$

und wir werden zeigen, wie man zu jedem $\tilde{x} \in [0,\infty)$ ein Intervall $[Y(\tilde{x})] \in \mathbf{IR}$ bestimmen kann mit

$$\tilde{y}(\tilde{x}) \in [Y(\tilde{x})],$$

falls $\tilde{c}$ und $\tilde{y}(\cdot)$ eine Lösung von (4.2) bilden.

Die Vorgehensweise ist dabei die folgende:

Unter gewissen Voraussetzungen an $f(x,s,t)$ kann man sofort ein $a \in \mathbf{R}$ erzielen mit

$$\tilde{c} \leq a,$$

und man kann zeigen, daß $\tilde{y}(\cdot)$ auf $[0, \tilde{c}]$ streng monoton fallend sein muß. Man hat also sofort

$$\tilde{c} \in [0, a] \quad \text{und} \quad \tilde{y}(\tilde{x}) \in [0, y_0] \text{ für jedes } \tilde{x} \in [0, \infty). \tag{4.3}$$

Um engere Einschließungen zu erzielen, werden wir n Stützstellen, $x_1, ..., x_n$, aus dem Intervall $(0, a)$ wählen und zeigen, daß der Vektor

$$\tilde{y} := \begin{pmatrix} \tilde{y}(x_1) \\ \vdots \\ \tilde{y}(x_n) \end{pmatrix} \in \mathbf{R}^n$$

in der Lösungsmenge L_B eines LCPs mit Intervalleinträgen liegt.

Wir werden dann zwei Algorithmen (Algorithmus B und Algorithmus C) vorstellen, die jeweils den Vektor $\tilde{y}$ in einem Intervallvektor, etwa $[L] \in \mathbf{IR}^n$, einschließen. Dabei gilt bei beiden Algorithmen $\underline{L}_i \geq 0$, $i = 1(1)n$.

Existiert dann ein $s \in \{1, ..., n\}$ mit

$$\underline{L}_s > 0, \quad \underline{L}_{s+1} = 0, \quad ([L_{n+1}] := [0, 0]),$$

so kann man wegen (4.2)

$$x_s < \tilde{c}$$

schließen. Wir erhalten dann

$$\tilde{c} \in [x_s, a] =: [C],$$

für jedes $\tilde{x} = x_i$, $i \in \{1, ..., n\}$,

$$\tilde{y}(\tilde{x}) \in [L_i] =: [Y(\tilde{x})]$$

und für jedes $\tilde{x} \in (x_i, x_{i+1})$, $i \in \{0, ..., n\}$, $x_0 := 0$, $x_{n+1} := a$,

$$\tilde{y}(\tilde{x}) \in [\underline{L}_{i+1}, \overline{L}_i] =: [Y(\tilde{x})],$$

da $\tilde{y}(\cdot)$ monoton fallend sein wird.

Gibt es außerdem ein ß $\in \{1, ..., n\}$ mit

$$[L_ß] = 0,$$

so wird

$$a := x_{\text{ß}}$$

gesetzt und die ganze Prozedur wird ab (4.3) wiederholt bis a nicht mehr verbessert werden kann.

Algorithmus B und Algorithmus C unterscheiden sich dabei wie folgt:

Algorithmus B wird mit Hilfe von Algorithmus A aus Abschnitt 3.2.2 die Lösungsmenge L_B in einen Intervallvektor $[SB] \in \mathbf{IR}^n$ einschließen und damit

$$\tilde{y} \in L_B \subseteq [SB]$$

erzielen. Algorithmus C wird die Iteration aus der 3. Bemerkung zu Satz 3.13 anwenden und dabei

$$[z^0] := \begin{pmatrix} [0, y_0] \\ \vdots \\ [0, y_0] \end{pmatrix} \ni \tilde{y}$$

benutzen. Man erhält somit $\tilde{y} \in [z^*]$.

4.3 Zusammenhang zwischen einer Lösung eines gewöhnlichen freien Randwertproblems und einem LCP mit Intervalleinträgen

Das Ziel dieses Abschnitts ist es, folgenden Sachverhalt zu beweisen:

Sind $\tilde{c} \in \mathbf{R}$ und $\tilde{y}(\cdot) : [0, \infty) \to \mathbf{R}$ eine Lösung des gewöhnlichen freien Randwertproblems (4.2) und sind $x_1, ..., x_n$ Stützstellen, die später noch genauer spezifiziert werden, dann liegt der Vektor

$$\tilde{y} := \begin{pmatrix} \tilde{y}(x_1) \\ \vdots \\ \tilde{y}(x_n) \end{pmatrix} \in \mathbf{R}^n$$

in der Lösungsmenge L_B eines LCPs mit Intervalleinträgen. Diese Aussage werden wir in Satz 4.1 beweisen. Dabei verbinden wir unter anderem die

Ideen aus [11] und [14] mit den Ideen aus [21].

In [11] und [14] wurde bereits gezeigt, wie die Diskretisierung eines freien Randwertproblems auf ein LCP führt. Da aber dort bei der zugrunde liegenden Taylorentwicklung das Restglied einfach weggelassen wird, bekommt man lediglich Näherungswerte für die exakte Lösung.

In [21] wurde gezeigt, wie man bei der Diskretisierung eines gewöhnlichen Randwertproblems (also mit festem, bekanntem Rand) die exakte Lösung an diskreten Stellen intervallmäßig einschließen kann, indem man bei der Taylorentwicklung das Restglied in einem Intervall einschließt und dann ein lineares Intervallgleichungssystem betrachtet.

Bevor wir aber jetzt stur die Taylorsche Formel auf $\tilde{y}(\cdot) : [0, \infty) \to \mathbf{R}$ anwenden, müssen wir auf eine typische Eigenschaft von Lösungen eines freien Randwertproblems hinweisen, aufgrund derer wir im Beweis zu Satz 4.1 eine umfangreiche Fallunterscheidung betrachten müssen. Wir wollen diese Eigenschaft vorweg an einem Beispiel erläutern.

Beispiel 4.1 Es seien $\kappa, v_0 > 0$. Dann betrachten wir das gewöhnliche freie Randwertproblem:

$$\left.\begin{array}{rll} \text{Finde } c \in \mathbf{R} & \text{und} & y(\cdot) : [0, \infty) \to \mathbf{R} \text{ mit} \\ y''(x) = \kappa & & \text{für } x \in [0, c], \\ y(x) > 0 & & \text{für } x \in [0, c), \\ y(x) \equiv 0 & & \text{für } x \in [c, \infty), \\ y'(c) = 0, & & \\ y(0) = v_0. & & \end{array}\right\} \tag{4.4}$$

Eine sehr einfache Rechnung zeigt, daß

$$c_p = \sqrt{\frac{2v_0}{\kappa}}$$

und

$$v(x) = \begin{cases} \dfrac{\kappa}{2}(x - c_p)^2, & x \in [0, c_p), \\ 0, & x \in [c_p, \infty), \end{cases}$$

die eindeutige Lösung von (4.4) bilden. Für $x \in [0,\infty) - \{c_p\}$ ist $v(x)$ beliebig oft differenzierbar. In $x = c_p$ ist $v(x)$ stetig differenzierbar, aber es gilt aufgrund von (4.4)

$$\lim_{\substack{h\to 0,\\ h>0}} \frac{v'(c_p-h)-v'(c_p)}{-h} = v''_-(c_p) = \kappa > 0 = v''_+(c_p) = \lim_{\substack{h\to 0,\\ h>0}} \frac{v'(c_p+h)-v'(c_p)}{h}.$$

An der Stelle $x = c_p$ besitzt $v''(x)$ demnach eine endliche Sprungstelle. Wir wollen diese Eigenschaft in einem Lemma präzisieren.

Lemma 4.1 *Gegeben seien eine auf* $[0,\infty)\times\mathbf{R}\times\mathbf{R}$ *stetig differenzierbare Funktion* $f(x,s,t)$ *und ein* $y_0 > 0$. *Es seien* $\tilde{c} \in \mathbf{R}$ *und* $\tilde{y}(\cdot) : [0,\infty) \to \mathbf{R}$ *eine Lösung des freien Randwertproblems (4.2), d.h. es wird vorausgesetzt, daß überhaupt eine Lösung von (4.2) existiert. Dann gelten folgende Aussagen:*

1. *Auf* $[\tilde{c},\infty)$ *ist* $\tilde{y}(x)$ *beliebig oft differenzierbar und es gilt*

$$\tilde{y}^{(n)}(x) \equiv 0 \quad \textit{für } x \in [\tilde{c},\infty) \quad \textit{und } n \in \mathbf{N}.$$

2. *Auf* $[0,\infty)$ *ist* $\tilde{y}(x)$ *stetig differenzierbar.*

3. *Auf* $[0,\tilde{c}]$ *ist* $\tilde{y}(x)$ *dreimal differenzierbar und es gilt* $\tilde{y}'''(x) =$

$$f_x(x,\tilde{y}(x),\tilde{y}'(x)) + f_s(x,\tilde{y}(x),\tilde{y}'(x))\cdot\tilde{y}'(x) + f_t(x,\tilde{y}(x),\tilde{y}'(x))\cdot\tilde{y}''(x),$$

$x \in [0,\tilde{c}]$.

4. *Die Funktion* $\tilde{y}''(x)$, $x \in [0,\infty)$, *hat an der Stelle* $x = \tilde{c}$ *genau dann eine Unstetigkeitsstelle, wenn*

$$f(\tilde{c},0,0) \neq 0$$

gilt. Hat $\tilde{y}''(x)$ *an der Stelle* $x = \tilde{c}$ *eine Unstetigkeitsstelle, so ist sie eine endliche Sprungstelle.*

Beweis: Zu 1.: Für $x \in [\tilde{c},\infty)$ ist $\tilde{y}(x) \equiv 0$. Eine konstante Funktion ist beliebig oft differenzierbar und die Ableitungen sind konstant null.

Zu 2.: $\tilde{y}(x)$ löst insbesondere das Problem

$$\begin{aligned} y''(x) &= f(x, y(x), y'(x)) \text{ für } x \in [0, \tilde{c}], \\ y'(\tilde{c}) &= 0, \\ y(\tilde{c}) &= 0. \end{aligned}$$

Daher ist $\tilde{y}(x)$ auf $[0, \tilde{c}]$ stetig differenzierbar, und es ist

$$\lim_{\substack{x \to \tilde{c}, \\ x < \tilde{c}}} \tilde{y}'(x) = \tilde{y}'(\tilde{c}) = 0.$$

Mit Teil 1 folgt dann die Behauptung.

Zu 3.: Es sei zunächst $x \in (0, \tilde{c})$. Dann ist wegen (4.2)

$$\frac{\tilde{y}''(x+h) - \tilde{y}''(x)}{h} = \frac{f(x+h, \tilde{y}(x+h), \tilde{y}'(x+h)) - f(x, \tilde{y}(x), \tilde{y}'(x))}{h}.$$

Da $f(x, s, t)$ als stetig differenzierbar vorausgesetzt ist, gilt

$$\lim_{h \to 0} \frac{f(x+h, \tilde{y}(x+h), \tilde{y}'(x+h)) - f(x, \tilde{y}(x), \tilde{y}'(x))}{h} =$$

$$f_x(x, \tilde{y}(x), \tilde{y}'(x)) + f_s(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}'(x) + f_t(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}''(x).$$

Somit folgt $\tilde{y}'''(x) =$

$$f_x(x, \tilde{y}(x), \tilde{y}'(x)) + f_s(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}'(x) + f_t(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}''(x).$$

Für $x = 0$ bzw. $x = \tilde{c}$ hat man die rechtsseitige bzw. die linksseitige dritte Ableitung zu betrachten. Man erhält völlig analog die entsprechenden Ergebnisse.

Zu 4.: Da $f(x, s, t)$, $\tilde{y}(x)$ und $\tilde{y}'(x)$ stetig sind, bekommt man mit (4.2)

$$\lim_{\substack{x \to \tilde{c}, \\ x < \tilde{c}}} \tilde{y}''(x) = \lim_{\substack{x \to \tilde{c}, \\ x < \tilde{c}}} f(x, \tilde{y}(x), \tilde{y}'(x)) = f(\tilde{c}, \tilde{y}(\tilde{c}), \tilde{y}'(\tilde{c})) = f(\tilde{c}, 0, 0).$$

Nach Teil 1 gilt

$$\lim_{\substack{x \to \tilde{c}, \\ x > \tilde{c}}} \tilde{y}''(x) = 0.$$

An der Stelle $x = \tilde{c}$ hat $\tilde{y}''(x)$ also genau dann eine Unstetigkeitsstelle, wenn

$$f(\tilde{c}, 0, 0) \neq 0$$

gilt. Da $f(x, s, t)$ auf $[0, \infty) \times \mathbf{R} \times \mathbf{R}$ definiert ist, ist der Wert $f(\tilde{c}, 0, 0)$ auf jeden Fall endlich. Daher ist, falls $f(\tilde{c}, 0, 0) \neq 0$ gilt, die Unstetigkeitsstelle an der Stelle $x = \tilde{c}$ eine endliche Sprungstelle. □

Wir kommen nun zu dem angekündigten

Satz 4.1 *Gegeben seien eine auf* $[0, \infty) \times \mathbf{R} \times \mathbf{R}$ *stetig differenzierbare Funktion* $f(x, s, t)$ *und ein* $y_0 > 0$. *Es seien* $\tilde{c} \in \mathbf{R}$ *und* $\tilde{y}(\cdot) : [0, \infty) \to \mathbf{R}$ *eine Lösung des freien Randwertproblems (4.2), d.h. es wird vorausgesetzt, daß überhaupt eine Lösung von (4.2) existiert. Außerdem seien bekannt:*

1. *Ein* $a \in \mathbf{R}$ *mit* $\tilde{c} \leq a$.
2. *Ein Intervall* $[F] = [\underline{F}, \overline{F}]$ *mit*

$$\{\eta \in \mathbf{R} : \eta = f(x, \tilde{y}(x), \tilde{y}'(x)) \text{ für } x \in [0, a]\} \subseteq [F].$$

3. *Ein* $F' \in \mathbf{R}$ *mit*

$$\left| \frac{\partial f(x, \tilde{y}(x), \tilde{y}'(x))}{\partial x} \right| \leq F' \text{ für } x \in [0, \tilde{c}].$$

Es sei weiter $n \in \mathbf{N}$ *beliebig gewählt. Aus dem Intervall* $[0, a]$ *seien dann folgendermaßen* $n + 2$ *Stützstellen bestimmt:*

$$\begin{aligned} x_0 &:= 0, \\ x_i &:= i \cdot h, \quad i = 1(1)n, \\ x_{n+1} &:= a, \end{aligned}$$

mit $h = a/(n + 1)$.

Behauptung: Ist $\underline{F} \geq 0$, *dann gilt für*

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_n \end{pmatrix} \in \mathbf{R}^n, \ \tilde{y}_i := \tilde{y}(x_i), \ 1 \leq i \leq n,$$

folgende Aussage:

$$\tilde{y} \in L_B = \Big\{ z \in \mathbf{R}^n : \textit{Es existiert ein } q \in [q] \textit{ mit} \\ (q + Mz)^{\mathrm{T}} z = 0,\ q + Mz \geq 0,\ z \geq 0 \Big\}$$

mit

$$M := \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathbf{R}^{n \times n}$$

und

$$[q] := \begin{pmatrix} h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{1}{2} h^3 \cdot [-F', F'] - y_0 \\ h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{1}{2} h^3 \cdot [-F', F'] \\ \vdots \\ h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{1}{2} h^3 \cdot [-F', F'] \end{pmatrix} \in \mathbf{IR}^n.$$

Beweis: Es sei zunächst $i \in \{2, ..., n\}$. Dann ist zu zeigen: Es existiert ein $q_i \in h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{1}{2} h^3 \cdot [-F', F']$, so daß

$$q_i + (M\tilde{y})_i \geq 0, \quad \tilde{y}_i = 0$$

oder

$$q_i + (M\tilde{y})_i = 0, \quad \tilde{y}_i \geq 0$$

gilt. Dabei gilt $\tilde{y}_i \geq 0$ immer wegen (4.2).

Wir setzen

$$f(x) := f(x, \tilde{y}(x), \tilde{y}'(x)), \quad x \in [0, a],$$

und $\tilde{y}'''(x) = f'(x) :=$

$$f_x(x, \tilde{y}(x), \tilde{y}'(x)) + f_s(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}'(x) + f_t(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}''(x)$$

für $x \in [0, \tilde{c}]$.

Da wir aufgrund von Lemma 4.1 davon ausgehen müssen, daß $\tilde{y}''(x)$ an der

Stelle $x = \tilde{c}$ eine endliche Sprungstelle besitzt, unterscheiden wir fünf Fälle:

$$\begin{array}{l} 1.\, x_i + h \leq \tilde{c}. \\ 2.\, \tilde{c} \leq x_i - h. \\ 3.\, x_i < \tilde{c} < x_i + h. \\ 4.\, x_i - h < \tilde{c} < x_i. \\ 5.\, x_i = \tilde{c}. \end{array}$$

Zu 1.: Es ist $x_i + h \leq \tilde{c}$.

Mit dem Satz von Taylor bekommt man

$$\tilde{y}(x_i + h) = \tilde{y}(x_i) + h\tilde{y}'(x_i) + \frac{h^2}{2}\tilde{y}''(x_i) + \frac{h^3}{6}\tilde{y}'''(\theta_{i1})$$

mit $x_i < \theta_{i1} < x_i + h \leq \tilde{c}$ und

$$\tilde{y}(x_i - h) = \tilde{y}(x_i) - h\tilde{y}'(x_i) + \frac{h^2}{2}\tilde{y}''(x_i) - \frac{h^3}{6}\tilde{y}'''(\theta_{i2})$$

mit $x_i - h < \theta_{i2} < x_i < \tilde{c}$. Man erhält durch Addition der beiden Gleichungen

$$\tilde{y}(x_i + h) + \tilde{y}(x_i - h) = 2\tilde{y}(x_i) + h^2\tilde{y}''(x_i) + \frac{h^3}{6}\tilde{y}'''(\theta_{i1}) - \frac{h^3}{6}\tilde{y}'''(\theta_{i2}).$$

Wegen $\theta_{i1}, \theta_{i2}, x_i \in (0, \tilde{c})$ gilt $\tilde{y}'''(\theta_{i1}) = f'(\theta_{i1})$, $\tilde{y}'''(\theta_{i2}) = f'(\theta_{i2})$ und $\tilde{y}''(x_i) = f(x_i)$. Daher ist

$$h^2 f(x_i) + \frac{h^3}{6}(f'(\theta_{i1}) - f'(\theta_{i2})) - \tilde{y}_{i-1} + 2\tilde{y}_i - \tilde{y}_{i+1} = 0.$$

Wählt man $q_i = 1 \cdot h^2 \cdot f(x_i) + h^3 \cdot (f'(\theta_{i1}) - f'(\theta_{i2}))/6$, so ist

$$q_i \in [\frac{1}{2}, 1] \cdot h^2 \cdot [F] + \frac{h^3}{6} \cdot [-F', F'] + \frac{h^3}{6} \cdot [-F', F'] \subseteq [q_i]$$

und $q_i + (M\tilde{y})_i = 0$.

Zu 2.: Es ist $\tilde{c} \leq x_i - h$.

In diesem Falle ist $\tilde{y}(x_i - h) = \tilde{y}(x_i) = \tilde{y}(x_i + h) = 0$. Wählt man $q_i = 1 \cdot h^2 f(x_i) \in [q_i]$, dann gilt $q_i + (M\tilde{y})_i = q_i = h^2 f(x_i) \geq 0$ und $\tilde{y}(x_i) = \tilde{y}_i = 0$.

Zu 3.: Es ist $x_i < \tilde{c} < x_i + h$.

Wir setzen

$$\begin{aligned} h_1 &:= \tilde{c} - x_i, \\ h_2 &:= x_i + h - \tilde{c}. \end{aligned}$$

Es gilt dann $h = h_1 + h_2$ und der Satz von Taylor liefert

$$\tilde{y}(x_i+h) = \tilde{y}(x_i+h_1+h_2) = \tilde{y}(\tilde{c}+h_2) = \tilde{y}(\tilde{c}) + h_2\tilde{y}'(\tilde{c}) + \frac{h_2^2}{2}\tilde{y}'_+(\tilde{c}) + \frac{h_2^3}{6}\tilde{y}'''(\theta_{i3})$$

mit $\tilde{c} < \theta_{i3} < x_i + h$. Es gilt also $\tilde{y}'''(\theta_{i3}) = 0$. Weiter resultiert aus dem Satz von Taylor

$$\tilde{y}(\tilde{c}) = \tilde{y}(x_i + h_1) = \tilde{y}(x_i) + h_1\tilde{y}'(x_i) + \frac{h_1^2}{2}\tilde{y}''(x_i) + \frac{h_1^3}{6}\tilde{y}'''(\theta_{i4})$$

mit $x_i < \theta_{i4} < x_i + h_1 = \tilde{c}$. Man erhält

$$\tilde{y}(x_i+h) = \tilde{y}(x_i) + h_1\tilde{y}'(x_i) + \frac{h_1^2}{2}\tilde{y}''(x_i) + \frac{h_1^3}{6}\tilde{y}'''(\theta_{i4}) + h_2\tilde{y}'(\tilde{c}) + \frac{h_2^2}{2}\tilde{y}'_+(\tilde{c}). \quad (4.5)$$

Mit denselben Überlegungen erzielt man

$$\left.\begin{aligned} \tilde{y}(x_i - h) &= \tilde{y}(x_i) - h_1\tilde{y}'(x_i) + \frac{h_1^2}{2}\tilde{y}''(x_i) - \frac{h_1^3}{6}\tilde{y}'''(\theta_{i6}) \\ &\quad -h_2\tilde{y}'(x_i - h_1) + \frac{h_2^2}{2}\tilde{y}''(x_i - h_1) - \frac{h_2^3}{6}\tilde{y}'''(\theta_{i5}) \end{aligned}\right\} \quad (4.6)$$

mit $x_i - h < \theta_{i5} < x_i - h_1 < \theta_{i6} < x_i$. Addiert man die Gleichungen (4.5) und (4.6), so ergibt sich

$$\begin{aligned} \tilde{y}(x_i + h) + \tilde{y}(x_i - h) &= 2\tilde{y}(x_i) + h_1^2\tilde{y}''(x_i) \\ &\quad + h_2\Big(\tilde{y}'(\tilde{c}) - \tilde{y}'(x_i - h_1)\Big) \\ &\quad + \frac{h_2^2}{2}\Big(\tilde{y}'_+(\tilde{c}) + \tilde{y}''(x_i - h_1)\Big) \\ &\quad + \frac{h_1^3}{6}\Big(\tilde{y}'''(\theta_{i4}) - \tilde{y}'''(\theta_{i6})\Big) \\ &\quad - \frac{h_2^3}{6}\tilde{y}'''(\theta_{i5}). \end{aligned}$$

Da $x_i < \tilde{c}$ gilt, folgt mit dem Mittelwertsatz der Differentialrechnung

$$\tilde{y}'(\tilde{c}) - \tilde{y}'(x_i - h_1) = \tilde{y}''(\theta_{i7})(\tilde{c} - (x_i - h_1))$$

mit $x_i - h_1 < \theta_{i7} < \tilde{c} = x_i + h_1$. Somit ist

$$\tilde{y}'(\tilde{c}) - \tilde{y}'(x_i - h_1) = \tilde{y}''(\theta_{i7})2h_1.$$

Wegen $\tilde{y}''_+(\tilde{c}) = 0$ erhält man insgesamt

$$0 = q_i - \tilde{y}_{i-1} + 2\tilde{y}_i - \tilde{y}_{i+1}$$

mit

$$q_i = h_1^2 f(x_i) + 2h_1 h_2 f(\theta_{i7}) + \frac{h_2^2}{2} f(x_i - h_1) + \frac{h_1^3}{6}\Big(f'(\theta_{i4}) - f'(\theta_{i6})\Big) - \frac{h_2^3}{6} f'(\theta_{i5}).$$

Es gilt also

$$q_i \in \Big(h_1^2 + 2h_1 h_2 + \frac{h_2^2}{2}\Big) \cdot [F] + \frac{h_1^3}{3}[-F', F'] + \frac{h_2^3}{6}[-F', F'].$$

Da

$$h_1^2 + 2h_1 h_2 + \frac{h_2^2}{2} \leq h_1^2 + 2h_1 h_2 + h_2^2 = (h_1 + h_2)^2 = h^2$$

und

$$h_1^2 + 2h_1 h_2 + \frac{h_2^2}{2} = h_1(h_1 + 2h_2) + \frac{h_2^2}{2} = h_1 h + h_2(h_1 + \frac{h_2}{2}) \geq h_1 h + \frac{1}{2} h_2 h \geq \frac{h^2}{2}$$

gilt, folgt schließlich $q_i \in [q_i]$ und $q_i + (M\tilde{y})_i = 0$.

Zu 4.: Es ist $x_i - h < \tilde{c} < x_i$.

Hier setzen wir $h_1 = \tilde{c} - (x_i - h)$ und $h_2 = x_i - \tilde{c}$, dann erhält man mit denselben Argumenten wie zu 3.

$$\begin{aligned} \tilde{y}(x_i - h) = \tilde{y}(x_i - h_2 - h_1) \quad &= \quad \tilde{y}(x_i) - h_2\tilde{y}'(x_i) + \frac{h_2^2}{2}\tilde{y}''(x_i) - \frac{h_2^3}{6}\tilde{y}'''(\theta_{i2}) \\ &\quad -h_1\tilde{y}'(\tilde{c}) + \frac{h_1^2}{2}\tilde{y}''_-(\tilde{c}) - \frac{h_1^3}{6}\tilde{y}'''(\theta_{i1}) \end{aligned}$$

mit $\tilde{c} - h_1 < \theta_{i1} < \tilde{c} = x_i - h_2 < \theta_{i2} < x_i$. Unter Berücksichtigung von $\tilde{y}(x) \equiv 0$ für $x \geq \tilde{c}$ resultiert schließlich

$$0 = \frac{h_1^2}{2} f(\tilde{c}) - \frac{h_1^3}{6} f'(\theta_{i1}) - \tilde{y}_{i-1} + 2\tilde{y}_i - \tilde{y}_{i+1}.$$

Wählt man also $q_i = h^2 f(\tilde{c}) + \frac{1}{2} h^3 F' \in [q_i]$, so gilt $q_i + (M\tilde{y})_i \geq 0$, und $\tilde{y}_i = \tilde{y}(x_i) = 0$.

Zu 5.: Es ist $\tilde{c} = x_i$.

Hier liefert der Satz von Taylor

$$\tilde{y}(\tilde{c} + h) = \tilde{y}(\tilde{c}) + h\tilde{y}'(\tilde{c}) + \frac{h^2}{2} \tilde{y}''_+(\tilde{c}) + \frac{h^3}{6} \tilde{y}'''(\theta_{i1})$$

mit $\tilde{c} < \theta_{i1} < \tilde{c} + h$ und

$$\tilde{y}(\tilde{c} - h) = \tilde{y}(\tilde{c}) - h\tilde{y}'(\tilde{c}) + \frac{h^2}{2} \tilde{y}''_-(\tilde{c}) - \frac{h^3}{6} \tilde{y}'''(\theta_{i2})$$

mit $\tilde{c} - h < \theta_{i2} < \tilde{c}$. Man erhält

$$\tilde{y}(\tilde{c} + h) - 2\tilde{y}(\tilde{c}) + \tilde{y}(\tilde{c} - h) = \frac{h^2}{2} \tilde{y}''_-(\tilde{c}) - \frac{h^3}{6} \tilde{y}'''(\theta_{i2}).$$

Wählt man also $q_i = h^2 f(\tilde{c})/2 - h^3 f'(\theta_{i2})/6$, so gilt $q_i + (M\tilde{y})_i = 0$ und $\tilde{y}_i = \tilde{y}(x_i) = \tilde{y}(\tilde{c}) = 0$.

Den Fall $i = 1$ beweist man mit denselben Ideen unter Berücksichtigung von $\tilde{y}(x_1 - h) = y_0$. □

4.4 Einschließung einer Lösung eines gewöhnlichen freien Randwertproblems

In Satz 4.1 haben wir bereits den Zusammenhang zwischen einer Lösung $\tilde{c} \in \mathbf{R}$ und $\tilde{y}(\cdot) : [0, \infty) \to \mathbf{R}$ eines gewöhnlichen freien Randwertproblems der Gestalt (4.2) und einem LCP mit Intervalleinträgen gezeigt.

In diesem Abschnitt wollen wir Voraussetzungen an $f(x, s, t)$ vorgeben, so daß

die Voraussetzungen von Satz 4.1 ausgerechnet werden können. Dies werden wir in Satz 4.2 tun. Außerdem werden wir zeigen, daß man den dem LCP mit Intervalleinträgen zugrunde liegenden Intervallvektor $[q]$ iterativ verbessern kann.

Wir wollen zunächst die iterative Verbesserung von $[q]$ erläutern. Es seien dazu die Voraussetzungen von Satz 4.1 erfüllt, und $x_0, ..., x_{n+1}$ seien die dort definierten Stützstellen. Das Ziel ist es, den Vektor

$$\tilde{y} = \begin{pmatrix} \tilde{y}(x_1) \\ \vdots \\ \tilde{y}(x_n) \end{pmatrix}$$

in einen Intervallvektor einzuschließen. Dazu geben wir zwei Möglichkeiten an:

1. Möglichkeit: Die Matrix M aus Satz 4.1 ist eine M-Matrix (siehe z.B. [29], Seite 105). Mit Teil 1 von Lemma 2.3 ist M eine H-Matrix mit positiven Diagonaleinträgen, und wir können Algorithmus A aus Abschnitt 3.2.2 anwenden, um L_B aus Satz 4.1 intervallmäßig einzuschließen.

Nach jeweils einer festen Anzahl von Iterationsschritten, etwa k, gilt nach (3.19)

$$\{x \in \mathbf{R}^n : f(x) = x,\, q \in [q]\} \subseteq [y^k], \tag{4.7}$$

und nach Satz 3.10 und Satz 4.1 folgt

$$\tilde{y} \in L_B \subseteq \Big(abs([y^k]) + [y^k]\Big) \cap \mathbf{R}^n_{\geq 0}. \tag{4.8}$$

Die Grundidee ist nun die folgende:

Existiert ein $s \in \{1, ..., n\}$ mit

$$\underline{y}^k_s > 0 \quad \text{und} \quad \underline{y}^k_{s+1} = 0, \quad ([y^k_{n+1}] := [0, 0]),$$

so gilt wegen (4.8)

$$\tilde{y}(x_s) > 0.$$

Da $\tilde{y}(x) > 0$ für $x \in [0, \tilde{c})$ gilt, muß

$$x_s < \tilde{c}$$

gelten. Für $i \in \{1, ..., s-1\}$ gilt nach dem ersten Fall aus dem Beweis zu Satz 4.1

$$q_i = h^2 f(x_i) + \frac{h^3}{6}(f'(\theta_{i1}) - f'(\theta_{i2}))$$

mit θ_{i1}, $\theta_{i2} \in (x_{i-1}, x_{i+1})$.

Für $i \in \{s, ..., n\}$ bekam man in den Fällen 2.-5. aus dem Beweis zu Satz 4.1 entweder

$$q_i = h_1^2 f(x_i) + 2h_1 h_2 f(\theta_{i7}) + \frac{h_2^2}{2} f(x_i - h_1) + \frac{h_1^3}{6}\Big(f'(\theta_{i4}) - f'(\theta_{i6})\Big) - \frac{h_2^3}{6} f'(\theta_{i5})$$

mit $x_i - h_1, \theta_{i4}\theta_{i5}, \theta_{i6}, \theta_{i7} \in (x_{i-1}, \tilde{c}), h_1 > 0, h_2 > 0, h = h_1 + h_2$, oder

$$q_i = h^2 \cdot f(x_i)$$

oder

$$q_i \geq \frac{h_1^2}{2} f(\tilde{c}) - \frac{h_1^3}{6} f'(\theta_{i1})$$

mit $\theta_{i1} \in (x_{i-1}, \tilde{c}), 0 \leq h_1 \leq h$ oder

$$q_i = \frac{h^2}{2} f(\tilde{c}) - \frac{h^3}{6} f'(\theta_{i2})$$

mit $\theta_{i2} \in (x_{i-1}, \tilde{c})$.

Wir nehmen nun an, wir kennen zusätzlich zu $a, F' \in \mathbf{R}$ und $[F] \in \mathbf{IR}$ aus Satz 4.1 auch noch $[F_i] \in \mathbf{IR}$ und $F_i' \in \mathbf{R}$, $i = 1(1)s$, mit:

1. $f(x_i, \tilde{y}(x_i), \tilde{y}'(x_i)) \in [F_i] \subseteq [F], \quad i = 1(1)s-1.$
2. $\left|\frac{\partial f(x, \tilde{y}(x), \tilde{y}'(x))}{\partial x}\right| \leq F_i' \leq F'$ für $x \in [x_{i-1}, x_{i+1}]$, $i = 1(1)s-1$.
3. $\{\eta \in \mathbf{R} : \eta = f(x, \tilde{y}(x), \tilde{y}'(x)) \text{ für } x \in [x_{s-1}, a]\} \subseteq [F_s] \subseteq [F]$.
4. $\left|\frac{\partial f(x, \tilde{y}(x), \tilde{y}'(x))}{\partial x}\right| \leq F_s' \leq F'$ für $x \in [x_{s-1}, \tilde{c}]$.

Dann gilt

$$\begin{aligned} \tilde{y} \in L_B = \; & \{z \in \mathbf{R}^n : \text{Es existiert ein } q \in [\tilde{q}] \text{ mit} \\ & (q + Mz)^{\mathrm{T}} z = 0,\ q + Mz \geq 0,\ z \geq 0\} \end{aligned}$$

mit

$$M := \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathbf{R}^{n \times n}$$

und $[\tilde{q}] \in \mathbf{IR}^n$ mit

$$[\tilde{q}_i] := \begin{cases} h^2[F_1] + \frac{h^3}{3}[-F_1', F_1'] - y_0 & \text{für } i = 1, \\ h^2[F_i] + \frac{h^3}{3}[-F_i', F_i'] & \text{für } i = 2(1)s-1, \\ h^2 \cdot [\frac{1}{2}, 1] \cdot [F_s] + \frac{1}{2}h^3 \cdot [-F_s', F_s'] & \text{für } i = s(1)n, \end{cases}$$

und man hat quasi die Lösungsmenge L_B eines neuen LCPs intervallmäßig einzuschließen, wobei man als grobe Ersteinschließung von

$$\{x \in \mathbf{R}^n : f(x) = x,\, q \in [\tilde{q}]\}$$

$[y^k]$ verwendet, denn es gilt aufgrund der Punkte 1.-4. und wegen (4.7)

$$\{x \in \mathbf{R}^n : f(x) = x,\, q \in [\tilde{q}]\} \subseteq \{x \in \mathbf{R}^n : f(x) = x,\, q \in [q]\} \in [y^k].$$

2. Möglichkeit: In Satz 4.1 wurde gezeigt, daß es für den Vektor $\tilde{y} \in \mathbf{R}^n$ ein $q \in [q]$ gibt mit

$$\begin{aligned} q + M\tilde{y} &\geq 0, \\ \tilde{y} &\geq 0, \\ (q + M\tilde{y})^{\mathrm{T}}\tilde{y} &= 0, \end{aligned}$$

wobei $[q] \in \mathbf{IR}^n$ mit

$$[q_i] := \begin{cases} h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{h^3}{2}[-F', F'] - y_0 & \text{für } i = 1, \\ h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + \frac{h^3}{2}[-F', F'] & \text{für } i = 2(1)n, \end{cases}$$

und

$$M := \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathbf{R}^{n \times n}$$

galt. Nach Satz 3.12 gilt dann

$$H(\tilde{y}; q, M) = \min(q + M\tilde{y}, \tilde{y}) = 0,$$

und mit Lemma 3.7 folgt

$$H\left(\tilde{y}; \frac{1}{2}q, \frac{1}{2}M\right) = 0.$$

In Satz 4.2 werden wir zeigen, daß die Funktion $\tilde{y}(\cdot)$ unter gewissen Voraussetzungen monoton fallend sein wird. Dann gilt wegen $\tilde{y}(0) = y_0$:

$$\tilde{y} \in \begin{pmatrix} [0, y_0] \\ \vdots \\ [0, y_0] \end{pmatrix} =: [z^0] \in \mathbf{IR}^n.$$

Wir starten daher gemäß Behauptung 4 von Satz 3.13 die Iteration

$$[z^{k+1}] := N\left(x^k, [z^k], \frac{1}{2}[q], \frac{1}{2}M\right) \cap [z^k], \quad x^k \in [z^k],\ k = 0, 1, 2, 3, \ldots . \quad (4.9)$$

Die Matrix $\frac{1}{2}M$ erfüllt die Voraussetzungen von Satz 3.14. Somit ist die Matrix $G(x^k, [z^k], \frac{1}{2}[q], \frac{1}{2}M)$ eine H-Matrix. Die Iteration (4.9) ist somit wegen Satz 2.5 und wegen

$$N\left(x^k, [z^k], \frac{1}{2}[q], \frac{1}{2}M\right) = x^k - IGA\left(G(x^k, [z^k], \frac{1}{2}[q], \frac{1}{2}M),\ H(x^k; \frac{1}{2}[q], \frac{1}{2}M)\right)$$

stets durchführbar. Außerdem kann man $G(x^k, [z^k], \frac{1}{2}[q], \frac{1}{2}M)$ gemäß dem Algorithmus aus Abschnitt 3.3.1 berechnen.

Nach einer festen Anzahl von Iterationsschritten, etwa k, der Iterationsvorschrift (4.9) gilt dann mit der Behauptung 4 von Satz 3.13

$$\tilde{y} \in [z^k].$$

Existiert nun ein $t \in \{1, \ldots, n\}$ mit

$$\underline{z}_t^k > 0 \quad \text{und} \quad \underline{z}_{t+1}^k = 0, \quad ([z_{n+1}^k] := [0, 0]),$$

so gilt völlig analog zur 1. Möglichkeit

$$x_t < \tilde{c}$$

und

$$H\left(\tilde{y}; \frac{1}{2}\tilde{q}, \frac{1}{2}M\right) = 0.$$

Dabei ist $\tilde{q} \in [\tilde{q}] \in \mathbf{IR}^n$ mit

$$[\tilde{q}_i] := \begin{cases} h^2[F_1] + \frac{h^3}{3}[-F_1', F_1'] - y_0 & \text{für } i = 1, \\ h^2[F_i] + \frac{h^3}{3}[-F_i', F_i'] & \text{für } i = 2(1)t-1, \\ h^2 \cdot [\frac{1}{2}, 1] \cdot [F_t] + \frac{1}{2}h^3 \cdot [-F_t', F_t'] & \text{für } i = t(1)n. \end{cases}$$

Bevor wir einige Voraussetzungen an $f(x,s,t)$ angeben, mit denen wir dann $a, F' \in \mathbf{R}$, $[F] \in \mathbf{IR}$, $[F_i] \in \mathbf{IR}$ und $F_i' \in \mathbf{R}$, $i = 1(1)s$ bzw. $i = 1(1)t$, ausrechnen können, ziehen wir noch ein Lemma vor, welches wir dazu benötigen werden.

Danach werden wir dann die 1. Möglichkeit in einen Algorithmus B und die 2. Möglichkeit in einen Algorithmus C umsetzen.

Lemma 4.2 *Es seien $\tilde{c} \in \mathbf{R}$ und $\tilde{y}(\cdot) : [0,\infty) \to \mathbf{R}$ eine Lösung von (4.2), wobei an die Funktion $f(x,s,t)$ vorausgesetzt sei:*

$$\text{Es gibt ein } \kappa > 0 \text{ mit } f(x, \tilde{y}(x), \tilde{y}'(x)) > \kappa, \quad \text{für } x \in (0, \tilde{c}). \tag{4.10}$$

Weiter seien

$$c_p = \sqrt{\frac{2v_0}{\kappa}}$$

und

$$v(x) = \begin{cases} \frac{\kappa}{2}(x - c_p)^2, & x \in [0, c_p), \\ 0, & x \in [c_p, \infty), \end{cases}$$

die Lösung des freien Randwertproblems (4.4) aus Beispiel 4.1.

Behauptung: Gilt $y_0 \le v_0$, so folgt

$$\tilde{y}(x) \le v(x) \quad \text{für } x \in [0,\infty)$$

und

$$\tilde{c} \le c_p.$$

Beweis: Wir zeigen zuerst $\tilde{c} \leq c_p$. Dazu nehmen wir an, es gelte

$$c_p < \tilde{c}. \tag{4.11}$$

Wir betrachten die Funktion

$$g(x) := \tilde{y}(x) - v(x), x \in [0, \tilde{c}].$$

$g(x)$ ist stetig und nimmt daher sein Maximum auf dem Kompaktum $[0, \tilde{c}]$ an. Wegen $y_0 \leq v_0$ gilt

$$g(0) = \tilde{y}(0) - v(0) = y_0 - v_0 \leq 0,$$

und wegen (4.11) gilt

$$\begin{array}{rcccccc} g(\tilde{c}) & = & \tilde{y}(\tilde{c}) - v(\tilde{c}) & = & 0 - 0 & = & 0, \\ g(c_p) & = & \tilde{y}(c_p) - v(c_p) & = & \tilde{y}(c_p) & > & 0. \end{array}$$

Es gibt somit ein $x_0 \in (0, \tilde{c})$ mit

$$g(x_0) = \max_{x \in [0,\tilde{c}]} g(x) > 0.$$

Da $g(x)$ differenzierbar ist, gilt $0 = g'(x_0)$ und es gibt ein $h_0 > 0$ mit

$$g'(x_0 - h) \geq 0 \quad \text{für jedes } h \in (0, h_0).$$

Daraus resultiert für $h \in (0, h_0)$

$$\frac{g'(x_0 - h) - g'(x_0)}{-h} = \frac{g'(x_0 - h)}{-h} \leq 0. \tag{4.12}$$

Es ist $x_0 \in (0, \tilde{c})$. Daher gilt

$$\tilde{y}''_-(x_0) = \tilde{y}''(x_0)$$

und

$$v''_-(x_0) \in \{0, \kappa\}. \tag{4.13}$$

Wir erhalten daher durch Grenzübergang in (4.12)

$$\tilde{y}''(x_0) - v''_-(x_0) = g''_-(x_0) \leq 0.$$

Es ist also

$$-\tilde{y}''(x_0) \geq -v_-''(x_0).$$

Mit (4.10) und (4.13) erhält man dann durch

$$0 = f(x_0, \tilde{y}(x_0), \tilde{y}'(x_0)) - \tilde{y}''(x_0) > \kappa - \tilde{y}''(x_0) \geq \kappa - v_-''(x_0) \geq 0$$

einen Widerspruch. Die Annahme (4.11) ist also falsch und es gilt

$$\tilde{c} \leq c_p.$$

Nun zeigen wir

$$\tilde{y}(x) \leq v(x) \quad \text{für } x \in [0, \infty).$$

Dabei folgt aus $\tilde{c} \leq c_p$ unmittelbar

$$\tilde{y}(x) \leq v(x) \quad \text{für } x \in [\tilde{c}, \infty). \tag{4.14}$$

Nehmen wir nun an, daß es ein $x_{00} \in (0, \tilde{c})$ gibt mit $\tilde{y}(x_{00}) > v(x_{00})$, dann können wir wieder mit $g(x) := \tilde{y}(x) - v(x)$, $x \in [0, \tilde{c}]$, einen Widerspruch erzeugen. Die Argumentation startet dabei mit

$$\begin{array}{rcccccl} g(x_{00}) & = & \tilde{y}(x_{00}) - v(x_{00}) & > & 0, & & \\ g(\tilde{c}) & = & \tilde{y}(\tilde{c}) - v(\tilde{c}) & = & -v(\tilde{c}) & \leq & 0, \\ g(0) & = & \tilde{y}(0) - v(0) & = & y_0 - v_0 & \leq & 0 \end{array}$$

und kann wie im Beweis zu $\tilde{c} \leq c_p$ zu Ende geführt werden. □

Damit kommen wir zu dem angekündigten

Satz 4.2 *Es seien $\tilde{c} \in \mathbf{R}$ und $\tilde{y}(\cdot) : [0, \infty) \to \mathbf{R}$ eine Lösung von (4.2), und die Funktion $f(x, s, t) : [0, \infty) \times \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ erfülle folgende Voraussetzungen:*

[V1] $f(x, s, t)$ ist stetig differenzierbar auf $[0, \infty) \times [0, y_0] \times \mathbf{R}$,
und f, f_x, f_s, f_t besitzen eine intervallmäßige Auswertung.

[V2] Es gibt ein $\kappa > 0$ mit $f(x, \tilde{y}(x), \tilde{y}'(x)) > \kappa$ für $x \in (0, \tilde{c})$.

[V3] Es gibt ein $K \geq 0$ mit $f(x, s, 0) \leq K$ für $(x, s) \in [0, \tilde{c}] \times [0, y_0]$.

[V4] Es gibt ein $L \geq 0$ mit $|f(x, s, t_1) - f(x, s, t_2)| \leq L|t_1 - t_2|$
für (x, s, t_1), $(x, s, t_2) \in [0, \tilde{c}] \times [0, y_0] \times \mathbf{R}$.

Dann gilt:

1. $\tilde{c} \leq \sqrt{\dfrac{2y_0}{\kappa}} =: a.$

2. $\tilde{y}(\cdot)$ *ist auf* $[0, \tilde{c}]$ *streng monoton fallend.*

3. $\{\eta \in \mathbf{R} : \eta = f(x, \tilde{y}(x), \tilde{y}'(x))$ *für* $x \in [0, a]\} \subseteq$

   $f\Big([0,a], [0,y_0], [-(Ly_0 + Ka), 0]\Big) =: [F].$

4. $\left|\frac{\partial f(x,\tilde{y}(x),\tilde{y}'(x))}{\partial x}\right| \leq \Big|f_x\Big([0,a], [0,y_0], [-(Ly_0 + Ka), 0]\Big) +$

   $f_s\Big([0,a], [0,y_0], [-(Ly_0 + Ka), 0]\Big) \cdot [-(Ly_0 + Ka), 0] +$

   $f_t\Big([0,a], [0,y_0], [-(Ly_0 + Ka), 0]\Big) \cdot [F]\Big| =: F'$ *für* $x \in [0, \check{c}].$

Weiter seien nun $x_i < \check{c}$, $i = 1(1)s$, *und* $\tilde{y}(x_i) \in [SB_i]$, $i = 1(1)n$. *Dann gilt:*

5. $f(x_i, \tilde{y}(x_i), \tilde{y}'(x_i)) \in$

   $f\Big(x_i, [SB_i], [-(L \cdot \overline{SB}_i + K(a - x_i)), 0]\Big) =: [F_i] \subseteq [F],\ i = 1(1)s - 1.$

6. $\{\eta \in \mathbf{R} : \eta = f(x, \tilde{y}(x), \tilde{y}'(x))$ *für* $x \in [x_{s-1}, a]\} \subseteq$

   $f\Big([x_{s-1}, a], [0, \overline{SB}_{s-1}], [-(L \cdot \overline{SB}_{s-1} + K(a - x_{s-1})), 0]\Big) =: [F_s] \subseteq [F].$

7. *Für* $x \in [x_{i-1}, x_{i+1}]$, $i = 1(1)s - 1$, *gilt:* $\left|\frac{\partial f(x,\tilde{y}(x),\tilde{y}'(x))}{\partial x}\right| \leq$

   $\Big|f_x\Big([x_{i-1}, x_{i+1}], [\underline{SB}_{i+1}, \overline{SB}_{i-1}], [-(L \cdot \overline{SB}_{i-1} + K(a - x_{i-1})), 0]\Big) +$

   $f_s\Big([x_{i-1}, x_{i+1}], [\underline{SB}_{i+1}, \overline{SB}_{i-1}], [-(L \cdot \overline{SB}_{i-1} + K(a - x_{i-1})), 0]\Big) \cdot$

   $[-(L \cdot \overline{SB}_{i-1} + K(a - x_{i-1})), 0] +$

   $f_t\Big([x_{i-1}, x_{i+1}], [\underline{SB}_{i+1}, \overline{SB}_{i-1}], [-(L \cdot \overline{SB}_{i-1} + K(a - x_{i-1})), 0]\Big) \cdot$

   $f\Big([x_{i-1}, x_{i+1}], [\underline{SB}_{i+1}, \overline{SB}_{i-1}], [-(L \cdot \overline{SB}_{i-1} + K(a - x_{i-1})), 0]\Big)\Big| =: F_i'$

   $\leq F'.$

8. Für $x \in [x_{s-1}, \tilde{c}]$ gilt: $\left|\frac{\partial f(x,\tilde{y}(x),\tilde{y}'(x))}{\partial x}\right| \leq$

$$\Big|f_x\Big([x_{s-1},a],[0,\overline{SB}_{s-1}],[-(L\cdot\overline{SB}_{s-1}+K(a-x_{s-1})),0]\Big)+$$

$$f_s\Big([x_{s-1},a],[0,\overline{SB}_{s-1}],[-(L\cdot\overline{SB}_{s-1}+K(a-x_{s-1})),0]\Big)\cdot$$

$$[-(L\cdot\overline{SB}_{s-1}+K(a-x_{s-1})),0]+$$

$$f_t\Big([x_{s-1},a],[0,\overline{SB}_{s-1}],[-(L\cdot\overline{SB}_{s-1}+K(a-x_{s-1})),0]\Big)\cdot[F_s]\Big|$$

$$=: F_s' \leq F'.$$

Beweis: Es bilden

$$c_p = \sqrt{\frac{2v_0}{\kappa}}$$

und

$$v(x) = \begin{cases} \frac{\kappa}{2}(x-c_p)^2, & x \in [0, c_p), \\ 0, & x \in [c_p, \infty), \end{cases}$$

die eindeutige Lösung der freien Randwertaufgabe (4.4), wie wir schon in Beispiel 4.1 gesehen haben. Wählt man $v_0 = y_0$, so erhält man mit [V2] und Lemma 4.2

$$\tilde{c} \leq c_p = \sqrt{\frac{2y_0}{\kappa}} = a.$$

Wiederum mit [V2] erhalten wir, daß

$$\tilde{y}''(x) > \kappa > 0 \quad \text{für } x \in (0, \tilde{c})$$

gilt. Somit ist $\tilde{y}'(x)$ auf $[0, \tilde{c}]$ streng monoton wachsend. Wegen $\tilde{y}'(\tilde{c}) = 0$ folgt dann $\tilde{y}'(x) < 0$ für $x \in [0, \tilde{c})$, d.h. $\tilde{y}(x)$ ist streng monoton fallend auf $[0, \tilde{c}]$.

Wir zeigen jetzt: Für jedes $i \in \{0, 1, ..., s\}$ gilt:

$$\tilde{y}'(x) \in [-(L\cdot\overline{SB}_i + K(a-x_i)), 0] \quad \text{für } x \in [x_i, \infty).$$

Wegen $\tilde{y}'(x) \equiv 0$ für $\tilde{c} \leq x$ ist es hinreichend,

$$-(L\cdot\overline{SB}_i + K(a-x_i)) \leq \tilde{y}'(x_i)$$

zu zeigen, wobei wir aus beweistechnischen Gründen $[SB_0] := [y_0, y_0]$ setzen. Mit [V4] bekommen wir für $x \in [x_i, \tilde{c}]$:

$$\begin{aligned} f(x, \tilde{y}(x), \tilde{y}'(x)) - f(x, \tilde{y}(x), 0) &\leq |f(x, \tilde{y}(x), \tilde{y}'(x)) - f(x, \tilde{y}(x), 0)| \\ &\leq L|\tilde{y}'(x) - 0| = -L\tilde{y}'(x). \end{aligned}$$

[V3] liefert dann

$$f(x, \tilde{y}(x), \tilde{y}'(x)) \leq -L\tilde{y}'(x) + f(x, \tilde{y}(x), 0) \leq -L\tilde{y}'(x) + K$$

für $x \in [x_i, \tilde{c}]$. Damit erhalten wir einerseits

$$\begin{aligned} \int_{x_i}^{\tilde{c}} \tilde{y}''(x)\, dx &= \int_{x_i}^{\tilde{c}} f(x, \tilde{y}(x), \tilde{y}'(x))\, dx \\ &\leq \int_{x_i}^{\tilde{c}} (-L\tilde{y}'(x) + K)\, dx \\ &= -L(\tilde{y}(\tilde{c}) - \tilde{y}(x_i)) + K(\tilde{c} - x_i) \\ &= L\tilde{y}(x_i) + K(\tilde{c} - x_i) \leq L \cdot \overline{SB}_i + K(a - x_i). \end{aligned}$$

Andererseits gilt

$$\int_{x_i}^{\tilde{c}} \tilde{y}''(x)\, dx = \tilde{y}'(\tilde{c}) - \tilde{y}'(x_i) = -\tilde{y}'(x_i).$$

Insgesamt gilt also $-\tilde{y}'(x_i) \leq L \cdot \overline{SB}_i + K(a - x_i)$ bzw.

$$-(L \cdot \overline{SB}_i + K(a - x_i)) \leq \tilde{y}'(x_i).$$

Mit

$$\frac{\partial f(x, \tilde{y}(x), \tilde{y}'(x))}{\partial x} =$$

$$f_x(x, \tilde{y}(x), \tilde{y}'(x)) + f_s(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}'(x) + f_t(x, \tilde{y}(x), \tilde{y}'(x)) \cdot \tilde{y}''(x)$$

für $x \in [0, \tilde{c}]$ und der Inklusionsmonotonie folgen dann die Aussagen 3. - 8. nacheinander. □

Algorithmus B

program AlgorithmusB;

{ Dieser Algorithmus ist eine modifizierte Version von Algorithmus A }
{ speziell zur Bearbeitung von freien Randwertproblemen der Art (4.2). }
{ Eingabe sind der linke Randwert y_0 }
{ und die Anzahl der Stützstellen n. }
{ Es wird vorausgesetzt: }
{ 1. Es existiert eine Lösung $\tilde{c}$ und $\tilde{y}(\cdot)$. }
{ 2. $f(x,s,t)$ erfüllt die Voraussetzungen aus Satz 4.2. }
{ 3. Für $[F]$ aus Satz 4.2 gilt $\underline{F} \geq 0$. }
{ Ausgabe sind ein Intervall $[C]$ mit $\tilde{c} \in [C]$ }
{ und ein n-dimensionaler Intervallvektor $[SB] \in \mathbf{IR}^n$ }
{ mit $\tilde{y}(x_i) \in [SB_i]$, $i = 1(1)n$. }
{ Die Stützstellen, $x_1, ..., x_n$, sind dabei wie in Satz 4.1 definiert. }

procedure algB($y_0 \in \mathbf{R}$; $n \in \mathbf{N}$;**var** $a \in \mathbf{R}$);

var $i, stop, zaehler, s \in \mathbf{N}$;
$h, randwert \in \mathbf{R}$;
$[C] \in \mathbf{IR}$;
$alpha \in \mathbf{R}^n$;
$[M] \in \mathbf{IR}^{n \times n}$;
$[yneu], [yalt], [ystart], [q], [x0], [x1], [u], [v], [SB] \in \mathbf{IR}^n$;

begin

$h := a/(n+1)$;

$[M] := 0$;

for $i := 1$ **to** n **do**

begin $[M_{ii}] := 1$;
if $(i > 1)$ **then** $[M_{i-1i}] := -\frac{1}{2}$;
if $(i < n)$ **then** $[M_{i+1i}] := -\frac{1}{2}$;
end;

{ $[M]$ ist bereits durch 2 dividiert! }

for $i := 1$ **to** n **do**

$[q_i] := h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + h^3 \cdot [-F', F']/2;$

$[q_1] := [q_1] - y_0;$

$[q] := [q]/2;$

$[x0] := 0;$

$[x1] := IGA(I + [M], -[q]);$

$alpha :=$ q$([x1], [x0]);$

$[u] := IGA(\langle [M] \rangle, alpha);$

$[v] := IGA\left(\langle I + [M] \rangle, (I - [M])([u] + alpha)/2\right);$

$[ystart] := [\underline{x1} - \overline{v}, \overline{x1} + \overline{v}];$

{ erste Grobeinschließung gemäß Algorithmus A }

writeln('wieviele Iterationsschritte sollen höchstens getätigt werden ?');

readln($stop$);

$zaehler :=$ 0;

$[yneu] := [ystart];$

repeat

$zaehler := zaehler + 1;$

$[yalt] := [yneu];$

$[yneu] := IGA\Big(I + [M], (I - [M]) \cdot abs([yalt]) - [q]\Big) \cap [yalt];$

{ jetzt kommt die iterative Verbesserung von $[q]$ }
{ 10 ist dabei ein default-Wert }

if $zaehler$ **mod** 10 = 0 **then**

begin

$s := 0;$

while $(s < n)$ **and** $(\underline{yneu}_{s+1} > 0)$ **do** $s := s + 1;$

if $s > 1$ **then**

begin

for $i := 1$ **to** $s - 1$ **do**

```
            [q_i] := h^2 · [F_i] + h^3 · [−F_i', F_i']/3;
            for i := s to n do
            [q_i] := h^2 · [1/2, 1] · [F_s] + h^3 · [−F_s', F_s']/2;
            [q_1] := [q_1] − y_0;
            [q] := [q]/2;
            end
            else writeln('noch keine Verbesserung von [q] möglich');
    end;
until (zaehler = stop) or ([yalt] = [yneu]);
[SB] := (abs([yneu]) + [yneu]);
for i := 1 to n do
if SB_i < 0 then SB_i := 0;
randwert := s · h;
s := n + 1;
while (s > 1) and (yneu_{s−1} = 0) do s := s − 1;
if s < n + 1 then a := s · h;
[C] := [randwert, a];
writeln('Der freie Rand liegt in');
write([C]);
writeln('An den ',n,' diskreten Stützstellen ergibt sich die Einschließung:');
write([SB]);
end;

var n ∈ N;
    y_0, a, sicherheit ∈ R;
begin
read(y_0);
read(n);
```

$a := \sqrt{(2 \cdot y_0)/\kappa}$;

repeat

$sicherheit := a$;

algB(y_0, n, a);

until $sicherheit = a$;

end.

Algorithmus C

program AlgorithmusC;

{ Dieser Algorithmus setzt die Behauptung 4 von Satz 3.13 um }
{ speziell zur Bearbeitung von freien Randwertproblemen der Art (4.2). }
{ Eingabe sind der linke Randwert y_0 }
{ und die Anzahl der Stützstellen n. }
{ Es wird vorausgesetzt: }
{ 1. Es existiert eine Lösung $\tilde{c}$ und $\tilde{y}(\cdot)$. }
{ 2. $f(x, s, t)$ erfüllt die Voraussetzungen aus Satz 4.2. }
{ 3. Für $[F]$ aus Satz 4.2 gilt $\underline{F} \geq 0$. }
{ Ausgabe sind ein Intervall $[C]$ mit $\tilde{c} \in [C]$ }
{ und ein n-dimensionaler Intervallvektor $[z^*] \in \mathbf{IR}^n$ }
{ mit $\tilde{y}(x_i) \in [z_i^*]$, $i = 1(1)n$. }
{ Die Stützstellen, $x_1, ..., x_n$, sind dabei wie in Satz 4.1 definiert. }

procedure algC$(y_0 \in \mathbf{R}; n \in \mathbf{N}; \mathbf{var}\ a \in \mathbf{R})$;

var $i, zaehler, t \in \mathbf{N}$;
$h, randwert \in \mathbf{R}$;
$x \in \mathbf{R}^n$;
$[C] \in \mathbf{IR}$;
$[G] \in \mathbf{IR}^{n \times n}$;
$M \in \mathbf{R}^{n \times n}$;
$[zneu], [zalt], [zstart], [q], [z^*], [H] \in \mathbf{IR}^n$;

begin

$h := a/(n+1)$;

$M := 0$;

for $i := 1$ **to** n **do**

begin $M_{ii} := 1$;
if $(i > 1)$ **then** $M_{i-1i} := -\frac{1}{2}$;
if $(i < n)$ **then** $M_{i+1i} := -\frac{1}{2}$;
end;

{ M ist bereits durch 2 dividiert! }

for $i := 1$ **to** n **do**

$[q_i] := h^2 \cdot [\frac{1}{2}, 1] \cdot [F] + h^3 \cdot [-F', F']/2$;

$[q_1] := [q_1] - y_0$;

$[q] := [q]/2$;

for $i := 1$ **to** n **do**

$[zstart_i] := [0, y_0]$;

$zaehler := 0$;

$[zneu] := [zstart]$;

repeat

$zaehler := zaehler + 1$;

$[zalt] := [zneu]$;

wähle x aus $[zalt]$;

$[G] := G(x, [zalt], [q], M)$; { Berechnung gemäß Abschnitt 3.3.1 }

$[H] := H(x; [q], M)$; { Berechnung wie in Satz 3.13 }

$[zneu] := \Big(x - IGA\big([G], [H]\big)\Big) \cap [zalt]$;

{ jetzt kommt die iterative Verbesserung von $[q]$ }
{ 10 ist dabei ein default-Wert }

if $zaehler$ **mod** $10 = 0$ **then**

begin

$t := 0$;

while $(t < n)$ **and** $(\underline{zneu}_{t+1} > 0)$ **do** $t := t + 1$;

if $t > 1$ **then**

```
            begin
            for i := 1 to t − 1 do
            [q_i] := h^2 · [F_i] + h^3 · [−F'_i, F'_i]/3;
            for i := t to n do
            [q_i] := h^2 · [1/2, 1] · [F_t] + h^3 · [−F'_t, F'_t]/2;
            [q_1] := [q_1] − y_0;
            [q] := [q]/2;
            end
            else writeln('noch keine Verbesserung von [q] möglich');
    end;
until [zneu] = [zalt];
[z*] := [zneu];
randwert := t · h;
t := n + 1;
while (t > 1) and (z*_{t−1} = 0) do t := t − 1;
if t < n + 1 then a := t · h;
[C] := [randwert, a];
writeln('Der freie Rand liegt in');
write([C]);
writeln('An den ',n,' diskreten Stützstellen ergibt sich die Einschließung:');
write([z*]);
end;

var n ∈ N;
    y_0, a, sicherheit ∈ R;
begin
read(y_0);
read(n);
```

$a := \sqrt{(2 \cdot y_0)/\kappa}$;

repeat

$sicherheit := a$;

algC(y_0, n, a);

until $sicherheit = a$;

end.

Bemerkungen:

1. $\kappa \in \mathbf{R}$, $[F] \in \mathbf{IR}$, $F' \in \mathbf{R}$, $[F_i] \in \mathbf{IR}$ und $F_i' \in \mathbf{R}$, $i = 1(1)s$ bzw. $i = 1(1)t$, sind mit Satz 4.2 bestimmbar.

2. Algorithmus B und Algorithmus C machen in dieser Form keine Aussage über Rundungsfehler, die bei der direkten Übertragung auf einen Computer entstehen können. Z.B. muß der Ausdruck $a/(n+1)$ auf einer Rechenmaschine nicht exakt darstellbar sein.

   Es ist also notwendig außer den Diskretisierungsfehler auch die etwaigen Rundungsfehler 'einzufangen'.

   Mit der Programmiersprache Pascal-XSC z.B. bekommt man diese Notwendigkeit (siehe etwa [20], [22]) in den Griff. Unter Benutzung von Pascal-XSC schlagen wir dann folgende spezielle Deklarationen und Zuweisungen vor:

   $h, a, randwert, c, y_0, sicherheit : interval$;
   $\vdots$
   $a := sqrt((2 * y_0)/\kappa)$;
   $h := a.sup/intval(n+1)$;
   $\vdots$
   $randwert := s * h$;
   $c := intval(randwert.inf, a.sup)$;
   $\vdots$

   Die restlichen Zuweisungen und Deklarationen in Algorithmus B und C sind dann 1:1 übertragbar, d.h. aus $[M] \in \mathbf{IR}^{n \times n}$; wird

M : imatrix[1..n,1..n];

usw.

Um eine Aussage über die Konvergenzgeschwindigkeit von Algorithmus B machen zu können, betrachten wir

Lemma 4.3 *Der Spektralradius der Matrix* $P = (I + \tilde{M})^{-1}(I - \tilde{M})$ *mit*

$$\tilde{M} := \begin{pmatrix} 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & \cdots & 0 & -\frac{1}{2} & 1 \end{pmatrix} \in \mathbf{R}^{n \times n}$$

lautet

$$\rho(P; n) = \frac{1 - 2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}{1 + 2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}.$$

Beweis: Das Spektrum der Matrix

$$M = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix} \in \mathbf{R}^{n \times n}$$

ist bekannt. Es lautet (siehe etwa [17], Seite 174)

$$\sigma(M) = \left\{ 4\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right),\ 4\sin^2\left(\frac{\pi}{2}\frac{2}{n+1}\right),\ ...,\ 4\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right) \right\}.$$

Damit hat (siehe auch [28], Seite 363) P das Spektrum

$$\sigma(P) = \left\{ \frac{1 - 2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right)}{1 + 2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right)} \ :\ i = 1(1)n \right\}.$$

Es ist nun der betragsgrößte Eigenwert von P zu bestimmen.

Es ist

$$\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right)} = \frac{2}{1+2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right)} - 1$$

streng monoton fallend in $i \in \{1, ..., n\}$. Daher gilt

$$\rho(P;n) = \max\left\{\left|\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}\right|, \left|\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)}\right|\right\}.$$

Mit

$$1-2\sin^2\left(\frac{\pi}{2}\frac{i}{n+1}\right) \geq 0 \Leftrightarrow \frac{\sqrt{2}}{2} \geq \sin\left(\frac{\pi}{2}\frac{i}{n+1}\right) \Leftrightarrow \frac{n+1}{2} \geq i$$

bekommt man

$$\left|\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}\right| = \frac{1-2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}$$

und

$$\left|\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)}\right| = \frac{2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)-1}{1+2\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right)} \leq \frac{2\cdot 1-1}{1+2\sin^2\left(\frac{\pi}{2}\frac{n+1}{2(n+1)}\right)} = \frac{1}{2}.$$

Nun ist für $n > 2$

$$3\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right) \leq 3\sin^2\left(\frac{\pi}{2}\frac{1}{3+1}\right) = 3\sin^2\left(\frac{\pi}{8}\right) = 0.439... \leq \frac{1}{2}.$$

Damit bekommt man dann für $n > 2$

$$\begin{aligned} 3\sin^2\left(\tfrac{\pi}{2}\tfrac{1}{n+1}\right) \leq \tfrac{1}{2} \quad &\Leftrightarrow \quad \tfrac{1}{2} + \sin^2\left(\tfrac{\pi}{2}\tfrac{1}{n+1}\right) \leq 1 - 2\sin^2\left(\tfrac{\pi}{2}\tfrac{1}{n+1}\right) \\ &\Leftrightarrow \quad \frac{1}{2} \leq \frac{1-2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{1}{n+1}\right)}. \end{aligned}$$

Für $n = 2$ erhält man

$$\left|\frac{1-2\sin^2\left(\frac{\pi}{2}\frac{2}{2+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{2}{2+1}\right)}\right| = \left|\frac{-\frac{1}{2}}{\frac{5}{2}}\right| = \frac{1}{5} \leq \frac{1}{3} = \frac{\frac{1}{2}}{\frac{3}{2}} = \frac{1-2\sin^2\left(\frac{\pi}{2}\frac{1}{2+1}\right)}{1+2\sin^2\left(\frac{\pi}{2}\frac{1}{2+1}\right)}.$$

Den betragsgrößten Eigenwert von P erhält man also in $\sigma(P)$ für $i = 1$. $\square$

Bemerkung: Für den Spektralradius aus Lemma 4.3 erkennt man, daß für $1 \leq n_1 < n_2$ gilt:

$$\rho(P; n_1) < \rho(P; n_2).$$

Für größeres n muß der Algorithmus B also nicht nur mehr Rechenoperationen pro Iterationsschritt durchführen, sondern die Konvergenzgeschwindigkeit an sich wird auch langsamer. Beispiel:

$$\begin{aligned} \rho(P; 50) &\approx 0.996213, \\ \rho(P; 150) &\approx 0.999567, \\ \rho(P; 300) &\approx 0.999891. \end{aligned}$$

Fazit: Mit dem Abbruchkriterium $[yneu] = [yalt]$ wird man in Algorithmus B keine schnellen Ergebnisse bekommen.

Die wesentliche Beobachtung ist vielmehr, daß der Benutzer selbst durch die Eingabe von '*stop*' die Iterationsschritte und somit die Rechenzeit begrenzen kann und trotzdem stets verifizierte Aussagen bekommt aufgrund von Satz 3.10 und Satz 4.1.

Über die Konvergenzgeschwindigkeit von Algorithmus C wollen wir keine theoretische Aussage machen. Die zeitlichen Unterschiede zwischen Algorithmus B und Algorithmus C werden wir im nächsten Abschnitt anhand von Beispielen demonstrieren.

4.5 Beispiele

In diesem Abschnitt werden wir den Algorithmus B und den Algorithmus C auf einige Beispiele anwenden. Dabei legen wir unser Augenmerk u.a. auf zwei Dinge:

- Welcher Algorithmus liefert die besseren Ergebnisse ?
- Welcher Algorithmus liefert die schnelleren Ergebnisse ?

Intuitiv haben wir schon in Algorithmus B zwei Abbruchkriterien eingebaut. Beide werden wir mit Algorithmus C vergleichen. Zu Algorithmus C sei bemerkt, daß wir in den Beispielen hier in der Zeile:

$$\text{wähle } x \text{ aus } [zalt];$$

stets $x := \underline{zalt}$ gewählt haben.

Beispiel 4.2 Als erstes Beispiel betrachten wir das Seilproblem, welches wir bereits in Abschnitt 4.1 vorgestellt haben. Die Stelle c (siehe Abbildung 4.1) und die Form des Seils $y(x)$ lösen das gewöhnliche freie Randwertproblem:

$$\left.\begin{aligned} \text{Finde } c \in \mathbf{R} \quad &\text{und} \quad y(\cdot) : [0,\infty) \to \mathbf{R} \text{ mit} \\ y''(x) = \sqrt{1 + (y'(x))^2} \quad &\text{für } x \in [0,c], \\ y(x) > 0 \quad &\text{für } x \in [0,c), \\ y(x) \equiv 0 \quad &\text{für } x \in [c,\infty), \\ y'(c) = 0, & \\ y(0) = y_0. & \end{aligned}\right\} \tag{4.15}$$

Wir wollen Algorithmus B und Algorithmus C anwenden und prüfen daher, ob $f(x,s,t) = \sqrt{1+t^2}$ die Voraussetzungen von Satz 4.2 und von Satz 4.1 erfüllt.

[V1] ist offensichtlich erfüllt. Es ist

$$y''(x) = f(x, y(x), y'(x)) = \sqrt{1 + (y'(x))^2} \geq 1 > \frac{1}{2}$$

auf $[0,c]$. Daher ist $y'(x)$, $x \in [0,c]$, streng monoton wachsend. Wegen $y'(c) = 0$ ist dann $y'(x) < 0$ in $(0,c)$. Daraus resultiert

$$f(x, y(x), y'(x)) = \sqrt{1 + (y'(x))^2} > 1 \quad \text{für } x \in (0,c).$$

Wir können demnach $\kappa = 1$ in [V2] wählen. Weiter ist

$$f(x,s,0) = \sqrt{1+0^2} = 1.$$

[V3] ist also mit $K = 1$ erfüllt. Zuletzt seien (x, s, t_1), $(x, s, t_2) \in [0, c] \times [0, y_0] \times \mathbf{R}$ und ohne Einschränkung sei $t_1 \neq t_2$. Dann gilt

$$\begin{aligned} |f(x,s,t_1) - f(x,s,t_2)| &= \left|\sqrt{1+t_1^2} - \sqrt{1+t_2^2}\right| = \left|\frac{t_1^2 - t_2^2}{\sqrt{1+t_1^2} + \sqrt{1+t_2^2}}\right| \\ &< \left|\frac{(t_1 - t_2)\cdot(t_1 + t_2)}{|t_1| + |t_2|}\right| \leq |t_1 - t_2|. \end{aligned}$$

[V4] ist somit erfüllt mit $L = 1$. Mit Satz 4.2 erhält man dann

$$a := \sqrt{2y_0} \tag{4.16}$$

und

$$[F] := f\Big([0,a],[0,y_0],[-(y_0+a),0]\Big) = \left[1, \sqrt{1+(y_0+a)^2}\right].$$

Es ist somit $\underline{F} = 1 \geq 0$. Damit können wir Algorithmus B und Algorithmus C anwenden. Dabei kommen wir an den Punkt, an dem wir uns entscheiden müssen, wie wir die Algorithmen anwenden wollen.

1. Möglichkeit: Wir entscheiden uns für schnelle Resultate und wählen daher eine kleine Anzahl von Stützstellen. Zudem können wir Algorithmus B durch das Abbruchkriterium $zaehler = stop$ oder durch das Abbruchkriterium $[yalt] = [yneu]$ beenden lassen.

In Abbildung 4.2 sieht man in der ersten Spalte das Ergebnis von Algorithmus B angewandt auf das Seilproblem mit folgenden Eingabewerten:

$$\begin{aligned} y_0 &= 0.1, \\ n &= 10, \\ stop &= 200. \end{aligned}$$

In der zweiten Spalte sieht man das Ergebnis von Algorithmus C bei den Eingabewerten:

$$\begin{aligned} y_0 &= 0.1, \\ n &= 10. \end{aligned}$$

Die dritte Spalte besitzt die exakten Werte, die man sehr einfach berechnen kann, denn es sind

$$c = \ln\left(y_0 + 1 + \sqrt{y_0^2 + 2y_0}\right)$$

	Einschließung $[SB_i]$	Einschließung $[z_i^*]$	$y(x_i)$
$i = 1$	$[0.08210179, 0.08276674]$	$[0.08207610, 0.08291841]$	0.08227326
$i = 2$	$[0.06605772, 0.06717242]$	$[0.06600934, 0.06747573]$	0.06633565
$i = 3$	$[0.05182709, 0.05321877]$	$[0.05176104, 0.05367363]$	0.05216082
$i = 4$	$[0.03937435, 0.04090745]$	$[0.03929666, 0.04151363]$	0.03972534
$i = 5$	$[0.02866892, 0.03023995]$	$[0.02858580, 0.03099719]$	0.02900866
$i = 6$	$[0.01968505, 0.02121768]$	$[0.01960226, 0.02212569]$	0.01999304
$i = 7$	$[0.01240158, 0.01384193]$	$[0.01232393, 0.01490040]$	0.01266360
$i = 8$	$[0.00680169, 0.00811392]$	$[0.00673282, 0.00932256]$	0.00700821
$i = 9$	$[0.00287264, 0.00403479]$	$[0.00281489, 0.00539333]$	0.00301753
$i = 10$	$[0.00060543, 0.00160561]$	$[0.00057062, 0.00228581]$	0.00068496
Zeit:	5 Sekunden	2 Sekunden	

Abbildung 4.2: Grobe aber schnelle Einschließung

und

$$y(x) = \begin{cases} \cosh(x-c) - 1, & x \in [0, c), \\ \\ 0, & x \in [c, \infty), \end{cases}$$

die exakte Lösung, wie man sehr leicht nachrechnet. Für $y_0 = 0.1$ ergibt sich

$$c \approx 0.443568254385.$$

Sowohl Algorithmus B als auch Algorithmus C offenbaren $y(x_{10}) > 0$. Der freie Rand liegt also rechts von $x_{10} = 10 \cdot a/(10+1)$. Mit (4.16) erhalten wir dann

$$c \in [0.4065578140908707, 0.4472135954999580].$$

Wenn wir die Ergebnisse von Algorithmus B mit den Ergebnissen von Algorithmus C vergleichen, stellen wir fest, daß Algorithmus B schärfere Einschließungen für $y(x_i)$ liefert als Algorithmus C.

Läßt man den Algorithmus B durch das Abbruchkriterium $[yalt] = [yneu]$ terminieren, so dauert der Algorithmus B 11 Sekunden, also mehr als fünf mal solange wie Algorithmus C.

Um diesen zeitlichen Unterschied noch drastischer zu machen, betrachten wir die

2. Möglichkeit: Wir wählen eine große Anzahl von Stützstellen und lassen solange iterieren bis $[yalt] = [yneu]$ in Algorithmus B gilt bzw. $[zalt] = [zneu]$ in Algorithmus C.

Algorithmus B terminierte bei den Eingabewerten:

$$\begin{aligned} y_0 &= 0.1, \\ n &= 300, \\ stop &= 500000, \end{aligned}$$

durch das Abbruchkriterium $[yneu] = [yalt]$ nach 265258 Iterationsschritten, und Algorithmus C hatte die Eingabewerte:

$$\begin{aligned} y_0 &= 0.1, \\ n &= 300. \end{aligned}$$

In Abbildung 4.3 sehen wir das Ergebnis. Da wir nicht alle 300 Komponenten niederschreiben wollten, haben wir uns bis auf die entscheidenden Komponenten auf jede 15. Komponente beschränkt.

Interessanterweise sind diesmal alle Einschließungen aus Algorithmus C besser als die aus Algorithmus B. Insbesondere haben wir

$$\underline{z}^*_{297} > \underline{SB}_{297} = 0$$

erhalten, was sich als sehr entscheidend herausstellt, denn mit

$$h := \frac{\sqrt{2 \cdot 0.1}}{300 + 1} \in [0.00148575945348823 2, 0.00148575945348823 3] =: [J]$$

resultiert aus Algorithmus C

$$c \in [t \cdot \underline{J}, a] = [297 \cdot \underline{J}, \sqrt{2 \cdot 0.1}] \subseteq [0.4412705576, 0.4472135955].$$

Mit Algorithmus B bekommt man lediglich

$$c \in [s \cdot \underline{J}, a] = [296 \cdot \underline{J}, \sqrt{2 \cdot 0.1}] \subseteq [0.4397847982, 0.4472135955].$$

Beim zeitlichen Vergleich zwischen Algorithmus B und Algorithmus C sprechen 38 Stunden gegenüber 7 Minuten für sich.

Wir wollen kurz resümieren:

i	Einschließung $[SB_i]$	Einschließung $[z_i^*]$	$y(x_i)$
1	[0.09931468, 0.09933677]	[0.09931468, 0.09933676]	0.09932035
15	[0.08998325, 0.09028325]	[0.08998329, 0.09028298]	0.09005943
30	[0.08052542, 0.08106304]	[0.08052550, 0.08106249]	0.08066034
45	[0.07161950, 0.07233938]	[0.07161962, 0.07233856]	0.07179794
60	[0.06325894, 0.06411228]	[0.06325910, 0.06411118]	0.06346794
75	[0.05543767, 0.05638175]	[0.05543788, 0.05638038]	0.05566618
90	[0.04815005, 0.04914780]	[0.04815028, 0.04914615]	0.04838876
105	[0.04139089, 0.04241044]	[0.04139117, 0.04240851]	0.04163209
120	[0.03515546, 0.03616967]	[0.03515577, 0.03616747]	0.03539280
135	[0.02943942, 0.03042551]	[0.02943977, 0.03042302]	0.02966779
150	[0.02423888, 0.02517796]	[0.02423927, 0.02517521]	0.02445422
165	[0.01955035, 0.02042703]	[0.01955077, 0.02042400]	0.01974950
180	[0.01537074, 0.01617272]	[0.01537120, 0.01616942]	0.01555130
195	[0.01169733, 0.01241505]	[0.01169783, 0.01241148]	0.01185752
210	[0.00852780, 0.00915402]	[0.00852834, 0.00915017]	0.00866634
225	[0.00586017, 0.00638964]	[0.00586074, 0.00638551]	0.00597616
240	[0.00369280, 0.00412191]	[0.00369341, 0.00411751]	0.00378566
255	[0.00202439, 0.00235084]	[0.00202503, 0.00234616]	0.00209374
270	[0.00085394, 0.00107643]	[0.00085462, 0.00107148]	0.00089957
285	[0.00018076, 0.00029870]	[0.00018148, 0.00029347]	0.00020255
295	[0.00000802, 0.00005614]	[0.00000876, 0.00005073]	0.00001388
296	[0.00000289, 0.00004402]	[0.00000363, 0.00003860]	0.00000715
297	[0, 0.00003307]	[0.00000071, 0.00002867]	0.00000263
298	[0, 0.00002310]	[0, 0.00001985]	0.00000032
299	[0, 0.00001430]	[0, 0.00001213]	0
300	[0, 0.00000660]	[0, 0.00000552]	0
Zeit:	38:44:07 h	7:32 min	

Abbildung 4.3: Vergleich von Algorithmus B mit Algorithmus C

Algorithmus C scheint immer schneller zum Ziel zu führen als Algorithmus B. Zudem lieferte Algorithmus C bei der Stützstellenzahl $n = 300$ auch die besseren Ergebnisse. Für $n = 10$ waren die Ergebnisse aus Algorithmus B besser. Dies inspirierte uns zu folgender

Idee: Zunächst verwenden wir Algorithmus B mit einer kleinen Stützstellenanzahl n (z.B. $n = 10$) und wenig Iterationsschritten (z.B. $stop = 200$). Für den gesuchten Intervallvektor $y \in \mathbf{R}^n$ liefert dann Algorithmus B einen Intervallvektor $[SB] \in \mathbf{IR}^n$ mit

$$y \in [SB].$$

Als nächstes wählen wir ein neues $N \in \mathbf{N}$, daß wesentlich größer ist als n und setzen $h := a/(N+1)$. Da $y(\cdot)$ monoton fallend ist, erhalten wir sehr einfach mit Hilfe von $[SB]$ einen Intervallvektor $[zstart] \in \mathbf{IR}^N$ mit

$$y(x_i) \in [zstart_i], \quad x_i := i \cdot h, \quad i = 1(1)N.$$

Damit beginnen wir dann die Iteration aus Algorithmus C. Mit dieser Idee war es möglich, mit $n = 10$, $stop = 200$ und $N = 300$ nach 23 Sekunden auf

$$c \in [0.4412705576, 0.4472135955]$$

zu schließen. Auf die Details wollen wir nicht eingehen.

In diesem Beispiel war es nie möglich, die obere Grenze a für den freien Rand iterativ zu verbessern. Das ist aber kein großer Nachteil, wie wir kurz an einer Anwendung des Seilproblems erläutern wollen.

Wir betrachten den Fall einer Stromleitung, die an der rechten Befestigung zu reißen droht. Nach einem eventuellen Abriß haben wir die Situation in Abbildung 4.1.

Die Verantwortlichen interessieren sich nun hauptsächlich für den freien Randwert c. Mit $y_0 = 0.1$ und $n = 10$ erhalten sie schon nach 200 Iterationsschritten (das sind auf einer sun-workstation 5 Sekunden) die Aussage, daß der freie Randwert c rechts von 0.4065578140908707 liegen würde. Befände sich also z.B. lediglich im Bereich $0 \leq x \leq 0.4$ stromleitendes Material, so bräuchte man mit keiner größeren Katastrophe rechnen...

Wir wollen noch darauf hinweisen, daß aufgrund von Trägheitsmomenten

und Reibungskräften die heruntergesackte Stromleitung nicht mehr unbedingt durch die gleiche Differentialgleichung beschrieben werden muß. Wie sich die Differentialgleichung allerdings ändert, darauf wollen wir hier nicht eingehen. Wir wollen aber bemerken, daß sich die Differentialgleichung in eine Differentialgleichung ändern könnte, die nicht mehr explizit gelöst werden kann, während Algorithmus B und Algorithmus C trotzdem verifizierte Aussagen liefern, falls die Voraussetzungen von Satz 4.1 und Satz 4.2 erfüllbar bleiben.

Beispiel 4.3 Wir betrachten das gewöhnliche freie Randwertproblem:

$$\left.\begin{aligned} \text{Finde } c \in \mathbf{R} \quad \text{und} \quad & y(\cdot) : [0,\infty) \to \mathbf{R} \text{ mit} \\ y''(x) = y + \arctan(xy') + \pi \quad & \text{für } x \in [0,c], \\ y(x) > 0 \quad & \text{für } x \in [0,c), \\ y(x) \equiv 0 \quad & \text{für } x \in [c,\infty), \\ y'(c) = 0, & \\ y(0) = \tfrac{1}{10}. & \end{aligned}\right\} \tag{4.17}$$

Mit Hilfe von Theorem 3.1 bzw. Corollary 3.3 aus [36] kann man zeigen, daß (4.17) genau eine Lösung besitzt. Es bezeichne $\tilde{c}$ und $\tilde{y}(\cdot) : [0,\infty) \to \mathbf{R}$ diese (eindeutige) Lösung.

$\tilde{c}$ und $\tilde{y}(\cdot)$ sind explizit nicht bestimmbar, so daß Algorithmus B und Algorithmus C umso wertvoller werden.

In diesem Beispiel wird sich sogar herausstellen, daß im Gegensatz zu Beispiel 4.2 hier die obere Grenze für $\tilde{c}$ iterativ verbessert werden kann. Wir wollen aber zunächst die Voraussetzungen von Satz 4.2 und Satz 4.1 ausrechnen.

[V1] ist offensichtlich erfüllt.

$$f(x,s,t) = s + \arctan(xt) + \pi$$

ist stetig differenzierbar auf dem $\mathbf{R}^3$. Weiter ist

$$f(x,\tilde{y}(x),\tilde{y}'(x)) = \tilde{y}(x) + \arctan(x\tilde{y}'(x)) + \pi > 0 - \frac{\pi}{2} + \pi = \frac{\pi}{2}.$$

Wir setzen somit $\kappa := \frac{\pi}{2}$ in [V2] und erhalten mit Lemma 4.2

$$\tilde{c} \leq a := \sqrt{\frac{2\frac{1}{10}}{\frac{\pi}{2}}} = 2\sqrt{\frac{1}{10 \cdot \pi}}.$$

Für $(x, s) \in [0, \tilde{c}] \times [0, \frac{1}{10}]$ gilt

$$f(x, s, 0) = s + \pi \leq \frac{1}{10} + \pi.$$

[V3] ist also mit

$$K = \frac{1}{10} + \pi$$

erfüllt. Zuletzt gilt für $(x, s, t_1), (x, s, t_2) \in [0, \tilde{c}] \times [0, \frac{1}{10}] \times \mathbf{R}$, $t_1 \neq t_2$,

$$\begin{aligned} |f(x, s, t_1) - f(x, s, t_2)| &= |f_t(x, s, \xi)||t_1 - t_2| \\ &= \left|\frac{x}{1 + (x\xi)^2}\right| |t_1 - t_2| \\ &\leq \tilde{c}\,|t_1 - t_2| \leq a\,|t_1 - t_2|. \end{aligned}$$

Dabei liegt ξ zwischen t_1 und t_2 (Mittelwertsatz).

[V4] ist somit erfüllt mit $L := a$.

Es ist

$$[F] := f\left([0, a], [0, \frac{1}{10}], [-(L \cdot \frac{1}{10} + K \cdot a), 0]\right).$$

Wir erhalten

$$\underline{F} \geq 0 - \frac{\pi}{2} + \pi = \frac{\pi}{2} > 0.$$

Somit können wir Algorithmus B und Algorithmus C anwenden. Algorithmus B terminierte bei den Eingabewerten:

$$n = 300, \quad stop = 500000,$$

durch das Abbruchkriterium $[yneu] = [yalt]$ nach 158439 Iterationsschritten, was etwas mehr als 17 Stunden dauerte.

i	Einschließung $[SB_i]$	Einschließung $[z_i^*]$
10	$[0.0907547020, 0.0908553282]$	$[0.0907547510, 0.0908551426]$
20	$[0.0819638001, 0.0821628800]$	$[0.0819638980, 0.0821625090]$
30	$[0.0736260658, 0.0739196490]$	$[0.0736262125, 0.0739190925]$
40	$[0.0657403340, 0.0661228306]$	$[0.0657405296, 0.0661220886]$
50	$[0.0583055024, 0.0587698218]$	$[0.0583057467, 0.0587688944]$
60	$[0.0513205311, 0.0518582199]$	$[0.0513208239, 0.0518571070]$
70	$[0.0447844421, 0.0453858214]$	$[0.0447847834, 0.0453845232]$
80	$[0.0386963195, 0.0393506212]$	$[0.0386967089, 0.0393491376]$
90	$[0.0330553085, 0.0337508109]$	$[0.0330557460, 0.0337491421]$
100	$[0.0278606157, 0.0285847780]$	$[0.0278611010, 0.0285829240]$
110	$[0.0231115088, 0.0238511045]$	$[0.0231120416, 0.0238490653]$
120	$[0.0188073161, 0.0195485656]$	$[0.0188078962, 0.0195463415]$
130	$[0.0149474266, 0.0156761291]$	$[0.0149480536, 0.0156737200]$
140	$[0.0115312897, 0.0122329535]$	$[0.0115319633, 0.0122303596]$
150	$[0.0085584152, 0.0092183880]$	$[0.0085591350, 0.0092156094]$
160	$[0.0060283732, 0.0066319704]$	$[0.0060291388, 0.0066290073]$
170	$[0.0039407980, 0.0044734235]$	$[0.0039416090, 0.0044702759]$
180	$[0.0022953880, 0.0027426534]$	$[0.0022962440, 0.0027393215]$
190	$[0.0010919021, 0.0014397529]$	$[0.0010928026, 0.0014362369]$
200	$[0.0003301593, 0.0005650014]$	$[0.0003311038, 0.0005613014]$
210	$[0.0000100387, 0.0001188639]$	$[0.0000110267, 0.0001149802]$
211	$[0.0000023130, 0.0000978474]$	$[0.0000033053, 0.0000939453]$
212	$[0, 0.0000789502]$	$[0, 0.0000750297]$
219	$[0, 0.0000601021]$	$[0, 0.0000019612]$
220	$[0, 0.0000040675]$	0
300	$[0, 0.0000017739]$	0
Zeit:	17:08:11 h	6:17 min

Abbildung 4.4: Vergleich von Algorithmus B mit Algorithmus C

In Abbildung 4.4 sehen wir das Ergebnis.

In der zweiten Spalte sehen wir das Ergebnis aus Algorithmus C nach dem ersten Aufruf der Prozedur algC für $n = 300$. Die Rechenzeit betrug bis dorthin (auf einer sun-workstation) etwa 6 Minuten. Wir haben uns bis auf die entscheidenden Komponenten auf jede 10. Komponente beschränkt.

Die Einschließungen von Algorithmus C sind besser als die Einschließungen von Algorithmus B. Jedoch können wir mit beiden Algorithmen wegen

$$h := \frac{1}{301} \cdot 2\sqrt{\frac{1}{10 \cdot \pi}} \in [0.001185464529005163, 0.001185464529005164] =: [J]$$

auf

$$\tilde{c} \in [211 \cdot \underline{J}, a] = \left[211 \cdot \underline{J}, 2\sqrt{\frac{1}{10 \cdot \pi}}\right] \subseteq [0.2501330156, 0.3568248233]$$

schließen. Die entscheidende Beobachtung in diesem Beispiel ist aber, daß nach dem ersten Aufruf von algC in Algorithmus C für $i = 220(1)300$

$$[z_i^*] = 0$$

gilt. Es wurde also erneut die Prozedur algC aufgerufen, diesmal mit

$$a = \frac{220}{301} \cdot 2\sqrt{\frac{1}{10 \cdot \pi}}.$$

Dabei konnte sich auch die untere Grenze für $\tilde{c}$ verbessern. Insgesamt wurde die Prozedur algC dreimal aufgerufen bis Algorithmus C dann schließlich

$$\tilde{c} \in [0.2505055169, 0.2564699340]$$

lieferte. Die Rechenzeit betrug 24 Minuten und 7 Sekunden.

Literaturverzeichnis

[1] G. Alefeld, X. Chen und F. Potra, *Numerical Validation of Solutions of Linear Complementarity Problems*, Numer. Math. 83 (1999), pp. 1-23.

[2] G. Alefeld und J. Herzberger, *Introduction to Interval Computations*, Academic Press (1983).

[3] G. Alefeld und G. Mayer, *Einschließungsverfahren.* In: Wissenschaftliches Rechnen: eine Einführung in das Scientific Computing. J. Herzberger (Editor). Akademie Verlag Berlin (1995), pp. 155-186.

[4] G. Alefeld und G. Mayer, *On the symmetric and unsymmetric Solution Set of Interval Systems*, SIAM J. Matrix Anal. Appl. 16 (1995), pp. 1223-1240.

[5] H. Beeck, *Über Struktur und Abschätzungen der Lösungsmenge von linearen Gleichungssystemen mit Intervallkoeffizienten*, Computing 10 (1972), pp. 231-244.

[6] A. Berman und R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* Academic Press (1979).

[7] J. R. Cannon, *Multiphase parabolic free Boundary Value Problems.* In: Moving Boundary Problems. D. G. Wilson, A. D. Solomon, P. T. Boggs (Editors). Academic Press (1978), pp. 3-24.

[8] L. Collatz, *Differentialgleichungen*, Teubner (1981).

[9] R. W. Cottle und G. B. Dantzig, *Complementary Pivot Theory of Mathematical Programming*, Linear Algebra and Appl. 1 (1968), pp. 103-125.

[10] R. W. Cottle, J. S. Pang und R. E. Stone, *The Linear Complementarity Problem*, Academic Press (1992).

[11] R. W. Cottle und R. S. Sacher, *On the Solution of large, structured Linear Complementarity Problems: the tridiagonal Case*, Appl. Math. Optim., Vol. 3, No. 4 (1977), pp. 321-340.

[12] J. Crank, *Free and Moving Boundary Problems*, Clarendon Press (1984).

[13] C. W. Cryer, *The Solution of a Quadratic Programming Problem using systematic Overrelaxation*, SIAM J. Control 9 (1971), pp. 385-392.

[14] C. W. Cryer und Y. Lin, *An alternating Direction implicit Algorithm for the Solution of Linear Complementarity Problems arising from free Boundary Problems*, Appl. Math. Optim. 13 (1985), pp. 1-17.

[15] K. Fan, *Topological Proof for certain Theorems on Matrices with non-negative Elements*, Monatsh. Math. 62 (1958), pp. 219-237.

[16] M. C. Ferris und J. S. Pang, *Engineering and Economic Applications of Complementarity Problems*, Siam Rev. Vol 39, No. 4 (1997), pp. 669-713.

[17] Graf Fink von Finkenstein, *Einführung in die Numerische Mathematik Band 1*, Carl Hanser Verlag (1977).

[18] A. Frommer und G. Mayer, *Parallel Interval Multisplitting*, Numer. Math. 56 (1989), pp. 255-267.

[19] D. Gale und H. Nikaido, *The Jacobian Matrix and global Univalence of Mappings*, Math. Annalen 159 (1965), pp. 81-93.

[20] R. Hammer, M. Hocks, U. Kulisch und D. Ratz, *Numerical Toolbox for Verified Computing I*, Springer Verlag (1991).

[21] E. Hansen, *On Solving Two-point Boundary-value Problems using Interval Arithmetic*. In: Topics in Interval Analysis. E. Hansen (Editor), Clarendon Press, Oxford (1969), pp. 74-90.

[22] R. Klatte, U. Kulisch, M. Neaga, D. Ratz und Ch. Ullrich, *Pascal-XSC-Sprachbeschreibung mit Beispielen*, Springer Verlag (1991).

[23] C. E. Lemke, *Bimatrix Equilibrium Points and Mathematical Programming*, Management Science 11 (1965), pp. 681-689.

[24] G. Mayer, *Enclosing the Solution Set of Linear Systems with inaccurate Data by iterative Methods based on incomplete LU-Decomposition*, Computing 35 (1985), pp. 189- 206.

[25] G. Mayer, *Old and new Aspects for the Interval Gaussian Algorithm.* In: Computer Arithmetic, Scientific Computation and Mathematical Modelling. E. Kaucher, S.M. Markov, G. Mayer (Editors), J.C. Baltzer AG, Basel (1991), pp. 329-349.

[26] J. Milnor, *Analytic Proofs of the 'Hairy Ball Theorem' and the Brouwer Fixed Point Theorem*, Amer. Math. Monthly 85, No. 7 (1978), pp. 521-524.

[27] R. Moore, *Intervallanalyse*, Oldenbourg Verlag (1969).

[28] K. G. Murty, *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann Verlag (1988).

[29] A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press (1990).

[30] A. Neumaier, *New Techniques for the Analysis of Linear Interval Equations*, Linear Algebra Appl. 58 (1984), pp. 273-325.

[31] W. Oettli und W. Prager, *Compatibility of approximate Solutions of Linear Equations with given Error Bounds for Coefficients and right-hand Sides*, Numer. Math. 6 (1964).

[32] K. Reichmann, *Abbruch beim Intervall-Gauß-Algorithmus*, Computing 22 (1979), pp. 355-361.

[33] J. Rohn, *On Nonconvexity of the Solution Set of a System of Linear Interval Equations*, BIT 30 (1989), pp. 161-165.

[34] E. Rothe, *Zweidimensionale parabolische Randwertaufgaben als Grenzwert eindimensionaler Randwertaufgaben*, Math. Ann. 102 (1930), pp. 650-670.

[35] H. Schwandt, *Schnelle fast global konvergente Verfahren für die Fünf-Punkt-Diskretisierung der Poissongleichung mit Dirichletschen Randbedingungen auf Rechteckgebieten*, Dissertation, Techn. Univ. Berlin (1981).

[36] R. C. Thompson und W. Walter, *Convergence of the Line Method Approximation for a parabolic free Boundary Problem*, Differential and Integral Equations 3 (1990), pp. 335-351.

[37] R. C. Thompson und W. Walter, *An Existence Theorem for a parabolic free Boundary Problem*, Differential and Integral Equations 5 (1992), pp. 43-54.

[38] R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall (1962).

[39] R. S. Varga und Do-Young Cai, *On the LU-Factorization of M-Matrices*, Numer. Math. 38 (1981), pp. 179-192.

Lebenslauf

Name:	Uwe Schäfer	
Geburtstag:	01.03.1969	
Geburtsort:	Baden-Baden	
Eltern:	Ignaz Schäfer (geb.: 30.07.1939) Rita Schäfer; geb. Fortenbacher (geb.: 04.08.1941)	
Geschwister:	Ulrike, Petra Lux; geb. Schäfer (geb.: 03.01.1963)	
	Jürgen Schäfer (geb.: 17.08.1966)	
Schulbildung:	1975 – 1979:	Grundschule in Baden-Baden Stadtteil Sandweier
	1979 – 1988:	Richard-Wagner Gymnasium in Baden-Baden Abiturprüfung am 3.5.1988
Grundwehrdienst:	1988 – 1989:	Fernmeldebataillon 230 in Dillingen/Donau
Studium:	1989 – 1994:	Diplom-Mathematik mit Nebenfach Informatik an der Universität Karlsruhe Diplomprüfung am 22.12.1994
Berufstätigkeit:	August 1992:	Programmierer bei Stierlen-Maquet/Rastatt
	1992 – 1994:	Wissenschaftliche Hilfskraft am Mathematischen Institut I bzw. II
	1995:	Promotionsstipendiat nach dem Landesgraduiertenförderungsgesetz in Baden-Württemberg
	seit Dez. 1995:	Wissenschaftlicher Angestellter am Institut für Angewandte Mathematik der Universität Karlsruhe