Bemerkungen zur Stabilität von Differenzenverfahren für Anfangswertprobleme

(On the Stability of Difference Schemes for Initial Value Problems)

W. Höhn

IPP 6/121 October 1973

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

GARCHING BEI MÜNCHEN

MAX-PLANCK-INSTITUT FÜR PLASMAPHYSIK

GARCHING BEI MÜNCHEN

Bemerkungen zur Stabilität von Differenzenverfahren für Anfangswertprobleme

(On the Stability of Difference Schemes for Initial Value Problems)

W. Höhn

IPP 6/121 October 1973

Die nachstehende Arbeit wurde im Rahmen des Vertrages zwischen dem Max-Planck-Institut für Plasmaphysik und der Europäischen Atomgemeinschaft über die Zusammenarbeit auf dem Gebiete der Plasmaphysik durchgeführt.

IPP 6/121 W. Höhn On the Stability of Difference Schemes for Initial Value Problems

October, 1973 (in German)

ABSTRACT:

To establish stability of difference schemes for quasilinear initial value problems, criteria derived from the amplification matrix are widely used, but despite the papers of Strang [4] and Kreiss [1] they have not yet been extended to implicit schemes or to those with unsteady coefficients or unsteady solutions of the initial value problem involved.

This paper attempts to do so in the case of one space and one time variable. Difference operators are regarded as large matrices with non-zero elements near the main diagonal only. Their spectral norms have to be approximated by means of inequalities for the eigenvalues of, in general, similar matrices. For tridiagonal matrices such inequalities are proven. The method includes generalized eigenvalues needed for implicit difference schemes.

EINLEITUNG

Das Stabilitätskriterium aus der Amplifikationsmatrix für Differenzenverfahren zur numerischen Lösung von quasilinearen Anfangswertproblemen ist trotz der Arbeiten von Strang [4] und Kreiss [1] noch nicht auf implizite Verfahren, oder auf solche mit unstetigen Koeffizienten oder mit unstetigen Lösungen der zugehörigen Anfangswertaufgabe übertragen, obwohl praktische Rechnungen seine Gültigkeit mit gewissen Einschränkungen auch für diese Fälle nahelegen.

In dieser Arbeit wird ein Ansatz untersucht, der geeignet erscheint, im Falle einer Raum- und einer Zeitvariablen vollständige Klarheit über diese Fragen zu bringen. Differenzenoperatoren werden darin als Bandmatrizen betrachtet. Zur Abschätzung deren Spektralnormen werden scharfe Schranken für Eigenwerte möglicherweise anderer Bandmatrizen benötigt, die im Falle von Tridiagonalmatrizen auch angegeben werden. Die Methode umfaßt die Abschätzung verallgemeinerter Eigenwerte, die für Stabilitätsbetrachtungen bei impliziten Differenzenverfahren benötigt werden.

Die Arbeit wurde aufgrund meines Stellenwechsels vom IPP zur TH Darmstadt vorläufig abgebrochen.

Für die freundliche Unterstützung während meiner gesamten Arbeitszeit am IPP danke ich besonders Herrn Dr. Karl Graf Finck von Finckenstein und Herrn Dr. Karl-Ulrich von Hagenow.

ELEMENTARE KONVERGENZTHEORIE

Es seien $G \subset \mathbb{R}^k$, $k \in \mathbb{N}$ ein Gebiet und

$u(t,\cdot) : G \to \mathbb{C}$ für $t \in [0,T] =: I$, $T > 0$

eine Schar von Abbildungen. Es gelte $u(t) := u(t,\cdot)$,

$u(t)$, $\frac{d}{dt} u(t) \in L_2(G)$ für $t \in I$

mit der gewöhnlichen L_2-Norm:

$$\| u \|_{L_2(G)} := \left(\int_G |u|^2 dx \right)^{\frac{1}{2}}$$

$u(t)$ sei Lösung des Anfangswertproblems

$$\frac{d}{dt} u(t) = A(t)\, u(t) \qquad \text{für } t \in I \qquad (1)$$

$$u(0) = u_0 \in L_2(G)$$

und für $u(t)$ erklärte Operatoren $A(t)$

Für $h, t \geq 0$, $t \leq T - h$ sei definiert:

$$E(t, u(t), h) := u(t+h)$$

Es sei

$$G_h := \{ x \in \mathbb{R}^k, \; x_\nu = h \cdot R_\nu n_\nu, n_\nu \in \mathbb{Z}, \nu = 1, \dots, n \} \cap G$$

mit $0 < R_\nu, h$ für $\nu = 1, \dots, n$

ein Rechteckgitter und $L_2(G_h)$ der zu $L_2(G)$ analoge Banachraum auf dem Gitter G_h mit der Norm:

$$\| u_h \| := \left(\prod_{\nu=1}^{k} h R_\nu \sum_{x \in G_h} |u_h(x)|^2 \right)^{\frac{1}{2}} \quad \text{für } u_h \in L_2(G_h)$$

Es seien $u(t) \in L_2(G_h)$ für $0 \leq t \leq T$

Weiter seien die Differenzenoperatoren

$$\mathcal{C}(n,\cdot,h) : L_2(G_h) \to L_2(G_h) \quad \text{für} \quad 0 \leq n\cdot h \leq T-h$$

definiert, dann ist

$$u^{n+1} = \mathcal{C}(n, u^n, h) \qquad \text{mit vorgegebenem } u^0 \in L_2(G_h) \tag{2}$$

ein Differenzenschema; seine Lösung soll $u(n\cdot h)$ approximieren.

Dazu gelte für den Anfangsfehler

$$\| u^0 - u_0 \| \leq a\cdot h^\rho \qquad a, \rho > 0$$

mit einem ρ, das der Konsistenzordnung des Schemas entsprechen soll:

Definition: Das Schema (2) heißt konsistent von der Ordnung ρ $:\Leftrightarrow$

$$\| u((n+1)\cdot h) - \mathcal{C}(n, u(n\cdot h), h) \| \leq K\cdot h^{\rho+1} \tag{3}$$

für $0 \leq n\cdot h \leq T-h, K \geq 0$

Das Schema (2) heißt stark stabil:

$$\| \mathcal{C}(n, u(n\cdot h)+f, h) - \mathcal{C}(n, u(n\cdot h), h) \| \leq (1+S\cdot h)\cdot \|f\| \tag{4}$$

für alle $f \in L_2(G_h)$ mit $\| f \| \leq \varepsilon > 0$

und $0 \leq n\cdot h \leq T-h$, $0 < S$

Satz:

Das Schema (2) habe höchstens einen Anfangsfehler $a\cdot h^\rho$, sei konsistent von der Ordnung ρ und stark stabil $\Longrightarrow$

$$\| u(n\cdot h) - u^n \| \leq \mathrm{const}(K, S, T)\cdot h^\rho \quad \text{für} \quad 0 \leq n\cdot h \leq T$$

und genügend kleines h

Die Lösungen des Differenzenschemas liegen also dann in der Norm $\|\cdot\|$ sehr nahe bei der des Anfangswertproblems und konvergieren gegen sie für $h \to 0$.

Beweis der stärkeren Aussage

$$\| u(m\cdot h) - u^m \| \leq k\cdot h^{\rho+1} \sum_{\nu=0}^{m-1} (1+sh)^{\nu} + a\cdot h^{\rho}(1+sh)^m$$

$$0 \leq m\cdot h \leq T$$

für genügend kleines h

durch Induktion über m.

Für m = o ist die Aussage offenbar richtig.

Induktionsschluß: $\quad ((m+1)\cdot h \leq T)$

$$\| u((m+1)\cdot h) - u^{m+1} \| \leq \| E(m\cdot h, u(m\cdot h), h) - \mathcal{E}(m, u(m\cdot h), h) \|$$

$$+ \| \mathcal{E}(m, u(m\cdot h), h) - \mathcal{E}(m, u^m, h) \| \leq$$

$$\leq \left[k\cdot h^{\rho+1} \sum_{\nu=0}^{m-1} (1+sh)^{\nu} + a\cdot h^{\rho}\cdot(1+sh)^m \right] \cdot$$

$$\cdot (1+sh) + k\cdot h^{\rho+1} =$$

nach Induktionsschluß, Stabilität (4) und Konsistenz (3).

(man hat h nur so klein zu wählen, daß:

$$\| u(m\cdot h) - u^m \| \leq \varepsilon \quad)$$

$$= k\cdot h^{\rho+1} \cdot \sum_{\nu=0}^{m} (1+sh)^{\nu} + a\cdot h^{\rho}(1+sh)^m =$$

womit die Behauptung bewiesen ist.

$$= K \cdot h^{p+1} \frac{(1+sh)^{m+1}-1}{s \cdot h} + a \cdot h^{p} \cdot (1+sh)^{m} \leq$$

$$\leq h^{p} \left(\frac{K}{s} (e^{sT}-1) + a \cdot e^{sT} \right)$$

woraus der Satz folgt.

Die Stabilitätsbedingung (4) läßt sich vereinfachen, wenn $\mathcal{C}(n,\cdot,h)$ für $0 \leq n \cdot h \leq T$ in einer ε-Umgebung von $u(n \cdot h)$ Frechet-differenzierbar ist. Nach dem Mittelwertsatz gilt dann (vereinfachte Schreibweise $t = n \cdot h$)

$$\| \mathcal{C}(u(t)+f) - \mathcal{C}(u(t)) \| \leq \|f\| \cdot \sup_{r \in [0,1]} \| \mathcal{C}'(u(t)+r \cdot f) \|$$

(„'" bedeutet Frechet-Ableitung nach $u(t)$).

Man hat also zu zeigen:

$$\| \mathcal{C}'(u(t)+f) \| \leq 1 + O(h) \text{ für } \|f\| \leq \varepsilon$$

in einer geeigneten mit $\|\cdot\|$ verträglichen Operatornorm, die hier genauso geschrieben wurde.

DIFFERENZENOPERATOREN ALS BANDMATRIZEN

Bei fast allen quasilinearen parabolischen oder hyperbolischen Differentialgleichungssystemen ist die Konsistenz des approximierenden Differenzenschemas unproblematisch. Lokale Stabilität, d.h. lokal an einer Stelle im Gitter, bereitet ebenfalls im Prinzip keine Schwierigkeiten. Ihre Übertragung auf globale Stabilität wie in [1] oder der direkte Beweis für globale Stabilität ist erheblich komplizierter. Im folgenden wird für eine Raum- und eine Zeitvariable eine direkte Methode betrachtet. Seien $G \subset \mathbb{R}$ ein endliches Intervall, $G_h := \{ \ell \cdot R \cdot h, \ell \in \mathbb{Z} \} \cap G$ und

$$u^m := \left(u_0^{m\,T}, u_1^{m\,T}, \dots, u_n^{m\,T} \right)^T \in \mathbb{C}^{(n+1)\cdot k}, \quad u_\nu^m \in \mathbb{C}^k$$

für $0 \leq \nu \leq n$

dabei ist: $R > 0, \; k \in \mathbb{N}$

„T" bedeutet Transposition

$n :=$ Ordnung von $G_h - 1$

Das Differenzenschema sei

$$\sum_{\mu=-\tau}^{\tau} B_\mu T^\mu u_\ell^{m+1} = \sum_{\nu=-\delta}^{\delta} A_\nu T^\delta u_\ell^m \tag{5}$$

$$\tau, \delta \in \mathbb{N} \cup \{0\}$$

mit vom Ort $\ell \cdot R \cdot h$, von der Zeit abhängigen Koeffizienten $B_\mu \in \mathbb{C}^{k^2}$ und zusätzlich von u^m abhängigen $A_\nu \in \mathbb{C}^{k^2}$, sowie den Translationsoperatoren $T^\mu : \; T^\mu u_\rho^m := u_{\rho+\mu}^m$ im definierten Indexbereich. Das Schema (5) läßt sich bei geeignet vorhandenen Randbeziehungen schreiben als

$$B u^{m+1} = A u^m$$

mit quadratischen Bandmatrizen aus $2\tau+1$ bzw. $2\delta+1$ Diagonalen $B, A \in \mathbb{C}^{[(n+1)\cdot k]^2}$. B wird als invertierbar, A als nach u^m Frechet-differenzierbar vorausgesetzt. Die Frechet-Ableitung (dargestellt durch „ ' ") von $\mathcal{C}(m, u^m, h)$ nach u^m ist dann

$$B^{-1} A' \quad .$$

Die Norm von $L_2(G_h)$ ist bis auf einen konstanten Faktor dasselbe wie die Euklidische Vektornorm in $\mathbb{C}^{(n+1)\cdot k}$, und damit verträglich ist die Spektralnorm in $\mathbb{C}^{[(n+1)\cdot k]^2}$. Man hat daher den betragsgrößten Eigenwert von

$$B^{-1} A' (B^{-1} A')^+$$

(„+" bedeutet hermitesch konjugiert)

zu bestimmen. Gesucht sind daher alle λ mit

$$B^{-1} A' (B^{-1} A')^+ v = \lambda v \qquad \text{für ein} \qquad v \in \mathbb{C}^{(n+1)k}$$

oder aber mit

$$A'(A')^+ u = \lambda B v = \lambda B B^+ u \qquad \text{für } u = (B^{-1})^+ v$$

Die gesuchte Norm der Stabilitätsbedingung erhält man daher aus den Beträgen der allgemeinen Eigenwerte des Bandmatrizenpaares $A'(A')^+$, BB^+. Da Bandmatrizen schwach besetzt sind, darf man hoffen, scharfe Abschätzungen für ihre Eigenwerte zu finden.

EIGENWERTE VON BANDMATRIZEN

Zur Bestimmung der Eigenwerte λ werden alle möglichen Lösungen u der Gleichung

$$\mathcal{M}\, u = 0 \tag{6}$$

mit einer komplexen Bandmatrix $\mathcal{M}$ untersucht, die λ als Parameter enthält, z.B. $\mathcal{M} = A'(A')^{+} - \lambda B B^{+}$

Seien etwa

$$\mathcal{M} = \begin{pmatrix} a_{00} & a_{01} & \cdots & a_{0k} & & & & & 0 \\ a_{1,-1} & a_{10} & \cdots & & a_{1k} & & & & \\ \vdots & & \ddots & & & \ddots & & & \\ a_{\varkappa,-\varkappa} & \cdots & & a_{\varkappa 0} & \cdots & & a_{\varkappa k} & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & a_{n-k,-\varkappa} & & & a_{n-k,0} & & & a_{n-k,k} \\ & 0 & & \ddots & & & \ddots & & a_{n-k+1,k-1} \\ & & & & \ddots & & & \ddots & \vdots \\ & & & & & a_{n,-\varkappa} & & \cdots & a_{n0} \end{pmatrix}; \quad u = \begin{pmatrix} u^0 \\ \vdots \\ u^i \\ \vdots \\ u^n \end{pmatrix}$$

$\varkappa, k \in \mathbb{N}$

wobei in $\mathcal{M}$ der erste Index die Zeile der zweite die Nebendiagonale festlegt. Es wird $a_{\mu,-\varkappa},\ a_{\tau k} \neq 0$ für alle sinnvollen Indizes vorausgesetzt; das ist vernünftig, weil sonst das Differenzenschema, von dem $\mathcal{M}$ herstammt, an einer Stelle eine niedrigere Ordnung haben müßte.

Stellt man sich u am Anfang um die $\varkappa$ Komponenten

$(u^{-\varkappa}, u^{-\varkappa+1}, \ldots, u^{-1}) =: w_1^T$ und am Ende um

$(u^{n+1}, \ldots, u^{n+k}) =: w_2^T$ verlängert vor und erweitert

$\mathcal{M}$ um die beiden Dreiecksmatrizen

$$W_1 := \begin{pmatrix} a_{0,-\varkappa} & a_{0,-\varkappa+1} & \cdots & a_{0,-1} \\ & a_{1,-\varkappa} & & a_{1,-2} \\ & 0 & \ddots & \vdots \\ & & & a_{\varkappa-1,-\varkappa} \end{pmatrix} ; \quad W_2 := \begin{pmatrix} a_{n-k+1,k} & & 0 \\ a_{n-k+2,k-1} & a_{n-k+2,k} & \\ \vdots & & \ddots \\ a_{n1} & \cdots & a_{n,k} \end{pmatrix}$$

deren Elemente beliebig komplex gewählt aber in der Hauptdiagonale nicht verschwinden sollen, dann läßt sich unter der Voraussetzung

$$w_1 = 0 \quad , \quad w_2 = 0$$

die Relation (6) schreiben als

$$\underbrace{\left(\begin{array}{c|c|c} W_1 & & 0 \\ \cline{1-1} & M & \\ 0 & & \\ \cline{3-3} & & W_2 \end{array}\right)}_{=:M^*} \underbrace{\begin{pmatrix} w_1 \\ \hline u \\ \hline w_2 \end{pmatrix}}_{=:u^*} = 0 \qquad (7)$$

worin W_1 in der linken oberen und W_2 in der rechten unteren Ecke von M^* stehen sollen. Für die Komponenten von u^* bedeutet (7) das Erfülltsein der linearen Differenzengleichung

$$\sum_{\nu=-\varkappa}^{k} a_{m\nu}\, u^{m+\nu} = 0 \qquad m = 0, 1, \cdots, n \,, \qquad (8)$$

deren allgemeine Lösung v^m sich aus ihren $\varkappa + k$ linear unabhängigen Eigenlösungen $u^m_{-\varkappa}, u^m_{-\varkappa+1}, \cdots, u^m_{-1}, u^m_{1}, \cdots, u^m_{k}$, $m = -\varkappa, -\varkappa+1, \cdots, n$ linear kombinieren läßt (s.z.B. [3]).

Hier geht wieder $a_{\mu,-\varkappa}$, $a_{\tau,k} \neq 0$ ein. Unter den $\mathfrak{v}$ ist u^* genau die Lösung mit

$$W_1 \cdot \begin{pmatrix} v^{-\varkappa} \\ v^{-\varkappa+1} \\ \vdots \\ v^{-1} \end{pmatrix} = 0 \quad ; \quad W_2 \cdot \begin{pmatrix} v^{n+1} \\ v^{n+2} \\ \vdots \\ v^{n+k} \end{pmatrix} = 0 \qquad (9)$$

D.h.: jede Lösung von (6), also jeder Eigenvektor, ist eine Lösung von (8) mit den Randbedingungen (9). Umgekehrt folgt aus (8), (9) für v das Bestehen von (6), wenn man die Ränder wegläßt.

Da in den Hauptdiagonalen von W_1 und W_2 alle Elemente nicht verschwinden, ist (9) äquivalent zu:

$$\begin{pmatrix} v^{-\varkappa} \\ v^{-\varkappa+1} \\ \vdots \\ v^{-1} \end{pmatrix} = 0 \quad ; \quad \begin{pmatrix} v^{n+1} \\ v^{n+2} \\ \vdots \\ v^{n+k} \end{pmatrix} = 0 \qquad (10)$$

v besitzt eine Darstellung:

$$v = \sum_{\substack{\tau=-\varkappa \\ \tau\neq 0}}^{k} A_\tau u_\tau$$ mit $A_\tau \in \mathbb{C}$, die nicht alle verschwinden.

Sei $U \in \mathbb{C}^{(\varkappa+k)^2}$ die Matrix:

$$U = \left(\begin{array}{cccc|ccc} u_{-\varkappa}^{-\varkappa} & u_{-\varkappa+1}^{-\varkappa} & \cdots & u_{-1}^{-\varkappa} & u_{1}^{-\varkappa} & \cdots & u_{k}^{-\varkappa} \\ u_{-\varkappa}^{-\varkappa+1} & & & \vdots & \vdots & & \vdots \\ \vdots & & & & & & \\ u_{-\varkappa}^{-1} & \cdots & & u_{-1}^{-1} & u_{1}^{-1} & & u_{k}^{-1} \\ \hline u_{-\varkappa}^{n+1} & \cdots & & u_{-1}^{n+1} & u_{1}^{n+1} & \cdots & u_{k}^{n+1} \\ \vdots & & & \vdots & \vdots & & \vdots \\ u_{-\varkappa}^{n+k} & \cdots & & u_{-1}^{n+k} & u_{1}^{n+k} & \cdots & u_{k}^{n+k} \end{array}\right)$$

Dann gilt: $(10) \Longleftrightarrow$

es gibt Koeffizienten A_τ, die nicht alle verschwinden, mit: $U \begin{pmatrix} A_{-\varkappa} \\ A_{-\varkappa+1} \\ \vdots \\ A_{-1} \\ A_1 \\ \vdots \\ A_k \end{pmatrix} = 0$

$\Longleftrightarrow$

$$\det U = 0 \tag{11}$$

Diese Aussage läßt sich fast wörtlich genauso übertragen für Blockmatrizen und Matrizendifferenzengleichungen. U ist dann eine Blockmatrix.

Mit Hilfe von (11) hat man die Frage nach Eigenwerten zurückgeführt auf das Verhalten der Lösungen der linearen Differenzengleichung (8). Mit dieser Methode waren im Rahmen dieser Arbeit Eigenwertabschätzungen für Bandmatrizen mit beliebiger Breite geplant, gelungen ist das aber bisher nur für Blocktridiagonalmatrizen.

EIGENWERTE VON BLOCKTRIDIAGONALMATRIZEN

Es sei A die Blocktridiagonalmatrix

$$A := \begin{pmatrix} b_0 & c_0 & & & & \\ a_1 & b_1 & c_1 & & 0 & \\ & a_2 & b_2 & c_2 & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & a_{n-1} & b_{n-1} & c_{n-1} \\ & & & & a_n & b_n \end{pmatrix}$$

mit $a_\nu, b_\nu, c_\nu \in \mathbb{C}^{m^2}$ und $\det a_\nu, \det c_\nu \neq 0$ für alle vernünftigen Indizes;

sei $d_\nu := b_\nu - \lambda$.

Für $(A-\lambda)u = 0$, $\left(u = (u^{0T}, u^{1T}, \dots, u^{nT})^T, u^\nu \in \mathbb{C}^m\right)$ ist nach dem vorangegangenen Abschnitt notwendig und hinreichend, daß für ein Grundsystem $u_1^\nu, u_2^\nu \in \mathbb{C}^{m^2}$ der Lösungen von (8) (als Matrixgleichung) die Relation

$$\det U = \det \begin{pmatrix} u_1^{-1} & u_2^{-1} \\ u_1^{n+1} & u_2^{n+1} \end{pmatrix} = 0$$

erfüllt ist.

Die beiden Grundlösungen können durch die Wahl von Anfangswerten $u_1^{-1}, u_1^0; u_2^{-1}, u_2^0$ festgelegt werden. Hier wird gewählt:

$$\begin{pmatrix} u_1^{-1} & u_2^{-1} \\ u_1^0 & u_2^0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Wenn hier und im folgenden Skalare $s \in \mathbb{C}$ mit Matrizen gleichwertig behandelt werden, so wird darunter $s \times$ Einheitsmatrix verstanden.

Dann gilt: $(11) \iff$

$$\det u_2^{n+1} = 0 \tag{12}$$

$$u_2^1 = -c_0^{-1} d_0$$

und

Unter $\|\cdot\|$ wird im folgenden Satz die Spektralnorm der Matrizen verstanden.

Satz:

Es seien $\det a_\nu, \det c_\nu \neq 0$ für alle sinnvollen Indizes.

λ sei Eigenwert von A $\Longrightarrow$

für ein ν gilt: $\det(b_\nu - \lambda) = 0$ oder

$$\text{für ein } \mu \text{ gilt: } \quad \| a_\mu (b_{\mu-1} - \lambda)^{-1} \| \cdot \| c_{\mu-1} (b_\mu - \lambda)^{-1} \| > \frac{1}{4} \tag{13}$$

Beweis durch Widerspruch.

Angenommen es gilt $\det d_\nu \neq 0$ für alle ν und

$$\| a_\mu d_{\mu-1}^{-1} \| \cdot \| c_{\mu-1} d_\mu^{-1} \| \leq \frac{1}{4} \quad \text{für alle } \mu \tag{14}$$

Einen Widerspruch stellt nach dem Vorangegangenen schon

$\det u_2^{n+1} \neq 0$ dar,

es wird aber darüber hinaus gezeigt, daß unter der Voraussetzung (14) für die Lösung von

$$a_\ell u^{\ell-1} + d_\ell u^\ell + c_\ell u^{\ell+1} = 0 \tag{15}$$

mit den Anfangswerten $u^0 = 1,\ u^1 = -c_0^{-1} d_0$ sogar

$\det u^\nu = 0$ für $0 \leq \nu \leq n+1$ gilt.

Sei $W := \left\{ \| z - \frac{1}{1-r^2} \| \leq \frac{r}{1-r^2} \ ;\ z \in \mathbb{C}^{m^2} \right\}$ für $0 < r < 1$

Behauptung: Es gilt $\det u^{\nu+1} \neq 0$ und

$- d_\nu u^\nu (u^{\nu+1})^{-1} c_\nu^{-1} \in W$ für $0 \leq \nu \leq n$

Beweis durch vollständige Induktion über ν.

I) Für $\nu = 0$ ist $u^1 = -c_0^{-1} d_0 \Rightarrow \det u^1 \neq 0$, $u^0 = 1$

sowie $-d_0(-c_0^{-1} d_0)^{-1} = 1 \in W$, denn

$$1 - \frac{1}{1-r^2} = \frac{r^2}{1-r^2} \leq \frac{r}{1-r^2}$$

II) Sei die Behauptung für $\nu = n-1$ richtig

a) Determinante

Es ist $\quad - c_n u^{n+1} = d_n u^n + a_n u^{n-1}$

$$X := c_n u^{n+1} (u^n)^{-1} d_n^{-1} = 1 + a_n u^{n-1} (u^n)^{-1} d_n^{-1}$$

$$= 1 - a_n d_{n-1}^{-1} \underbrace{\left[-d_{n-1} u^{n-1} (u^n)^{-1} c_{n-1}^{-1} \right]}_{=: \, y \, \in W} c_{n-1} d_n^{-1} \quad (16)$$

y ist nach Induktionsvoraussetzung in W.

Sei $h \in \mathbb{C}^m \quad \Longrightarrow$

$$| h^+ x h | = | h^+ (1 - a_n d_{n-1}^{-1} y c_{n-1} d_n^{-1}) h | \geq$$

$$\geq h^+ h - \|h\| \cdot \| a_n d_{n-1}^{-1} \| \cdot \| y \| \cdot \| c_{n-1} d_n^{-1} \| \cdot \| h \| \geq$$

$$\geq h^+ h \left(1 - \frac{1}{4} y\right)$$

Sei im folgenden $r = \frac{1}{2}$. Für beliebiges $0 < r < 1$ würde man in (13) nur die schlechtere Abschätzung mit $r(1-r)$ statt $\frac{1}{4}$ erhalten. Das Polynom $r(1-r)$ wird für $r = \frac{1}{2}$ maximal mit dem Funktionswert $\frac{1}{4}$.

$$y \in W \Rightarrow \|y\| \leq \frac{r+1}{1-r^2} = \frac{1}{1-r} = 2$$

$\Longrightarrow$

$$|h^+ x h| \geq \frac{1}{2} h^+ h$$ für beliebiges $h \in \mathbb{C}^m$ mit $\|h\| = 1$

$\Longrightarrow \det x \neq 0 \quad \Longrightarrow \det u^{n+1} \neq 0$

b) zu zeigen ist: $\quad x^{-1} = -d_n u^n (u^{n+1})^{-1} c_n^{-1} \in W$

$$(16) \overset{a)}{\Longrightarrow} \quad x^{-1} = \left(1 - a_n d_{n-1}^{-1} y c_n d_n^{-1}\right)^{-1}$$

Behauptung: Sei $z \in W$, es gelte (14) $\Longrightarrow$

$$f(z) := \left(1 - a_n d_{n-1}^{-1} z c_n d_n^{-1}\right)^{-1} \in W$$

Beweis: $\quad z \in W \Rightarrow \det z \neq 0$

denn sonst hätte $-\left(z - \frac{1}{1-r^2}\right)$ den Eigenwert $\frac{1}{1-r^2}$, der sicher kleiner oder gleich der Norm ist; aber

$$\frac{1}{1-r^2} > \frac{r}{1-r^2}$$ und das ist ein Widerspruch

Daher ist

$$\varphi(z) := z^{-1}$$ für $z \in W$ wohldefiniert.

Es ist $\quad \varphi^{-1} = \varphi|_W$

Sei $\quad a := a_n d_{n-1}^{-1} \quad , \quad b := c_n d_n^{-1}$

sowie

$$\psi(z) := 1 - a z b \quad \text{für} \quad z \in \mathbb{C}^{m^2}$$

Für die Behauptung ist zu zeigen:

$$f(W) \subset W$$

$$\Longleftrightarrow \quad \varphi \circ \psi(W) \subset W$$

$$\Longleftrightarrow \quad \psi(W) \subset \varphi(W) \tag{17}$$

Berechnung von $\varphi(W)$

$$z \in W \Longleftrightarrow \left(z - \frac{1}{1-r^2}\right)^+ \left(z - \frac{1}{1-r^2}\right) \leq \frac{r^2}{(1-r^2)^2}$$

Die Ordnungsrelation bezieht sich auf die quadratische Form mit beliebigem Vektor als Argument.

$$\varphi(z) \in W \Longleftrightarrow z \in \varphi(W) \Longleftrightarrow$$

$$\left(z^{-1} - \frac{1}{1-r^2}\right)^+ \left(z^{-1} - \frac{1}{1-r^2}\right) - \frac{r^2}{(1-r^2)^2} \leq 0$$

Diese Relation wird von links mit z^+ und von rechts mit z multipliziert; das bedeutet für die quadratischen Formen lediglich Transformation.

$$\Longleftrightarrow$$

$$1 - \frac{1}{1-r^2}\left(z^+ + z\right) + z^+ z \underbrace{\left[\frac{1}{(1-r^2)^2} - \frac{r^2}{(1-r^2)^2}\right]}_{= \frac{1}{1-r^2}} \leq 0$$

$$\Longleftrightarrow \quad 1 - (z^+ + z) + z^+ z \leq r^2$$

$$\Longleftrightarrow \quad (z^+ - 1)(z - 1) \leq r^2$$

$$\Longleftrightarrow \quad \| z - 1 \| \quad \leq \quad r$$

$$\Longrightarrow \quad \varphi(W) = \{ z \in \mathbb{C}^{m^2} ; \| z - 1 \| \leq r \}$$

Um (17) zu erfüllen, muß für $z \in W$ gelten:

$$\| \psi(z) - 1 \| = \| 1 + a z b - 1 \| \overset{!}{\leq} r = \frac{1}{2}$$

$$\Longleftrightarrow \quad \| a z b \| \leq r \quad .$$

$$z \in W \quad \Longrightarrow \quad \| z \| \leq 2 \qquad \text{o.O.}$$

$$\| a z b \| \leq \| a \| \cdot \| b \| \cdot \| z \| \leq \frac{1}{4} \cdot 2 = \frac{1}{2}$$

und damit sind der Induktionsschluß und der Satz bewiesen.

Die dieser Herleitung zugrundeliegende Abschätzung des Wertebereichs eines Kettenbruchs stammt aus [2] , s. auch [5] .

Für gewöhnliche Tridiagonalmatrizen besagt (13), daß für einen Eigenwert λ von A ein μ existiert mit

$$| b_{\mu-1} - \lambda | \cdot | b_\mu - \lambda | \leq 4 \, | a_\mu \, c_{\mu-1} | \tag{18}$$

Das obige Verfahren beinhaltet den Fall verallgemeinerter Eigenwerte. Sei etwa $B \in \mathbb{C}^{(n+1)^2}$ hermitisch und positiv definit:

$$B := \begin{pmatrix} g_0 & h_0 & & & & \\ \bar{h}_0 & g_1 & h_1 & & \bigcirc & \\ & \bar{h}_1 & g_2 & h_2 & & \\ & & \ddots & \ddots & \ddots & \\ & \bigcirc & & \bar{h}_{n-2} & g_{n-1} & h_{n-1} \\ & & & & \bar{h}_{n-1} & g_n \end{pmatrix} \in \mathbb{C}^{(n+1)^2}$$

Dann gibt es für λ mit $\det(A - \lambda B) = 0$ ein μ, so daß:

$$|b_{\mu-1} - \lambda g_{\mu-1}| \cdot |b_\mu - \lambda g_\mu| \leq 4 \cdot |a_\mu - \lambda \bar{h}_{\mu-1}| \cdot |c_{\mu-1} - \lambda h_{\mu-1}|$$

Die Methode mit Hilfe der Amplifikationsmatrix oder andere direkte Methoden liefern für die symmetrische Tridiagonalmatrix

$$\begin{pmatrix} b & a & & & & & \\ a & b & a & & & & \\ & a & b & a & & & \\ & & a & b & a & \bigcirc & \\ & & & \ddots & \ddots & \ddots & \\ & \bigcirc & & & a & b & a \\ & & & & & a & b \end{pmatrix}$$

die Eigenwertabschätzung

$$|b - \lambda| \leq 2|a|$$

in Übereinstimmung mit (18).

LITERATURVERZEICHNIS

[1] Kreiss, H.O. (1964). On difference approximations of the dissipative type for hyperbolic difference equations. Comm. Pure Appl. Math., vol. 17, p. 335

[2] Lane, R.E. (1945). The value region problem for continued fractions. Duke Math. Jour., vol. 51, p. 246

[3] Miller, K.S. Linear difference equations. W.A. Benjamin, Inc., N.Y. 1968

[4] Strang, W.G. (1964). Accurate partial difference methods II: nonlinear problems. Numer. Math., vol. 6, p. 37

[5] Wall, H.S. Analytic theory of continued fractions. Chelsea Publishing Comp., Bronx, N.Y. 1967

INHALTSVERZEICHNIS