KERNFORSCHUNGSZENTRUM

KARLSRUHE

Juni 1971 KFK 1428

Institut für Neutronenphysik und Reaktortechnik

Stabilität und Konvergenz nichtlinearer Differenzenoperatoren bei Anfangswertaufgaben

W. Kinnebrock

GESELLSCHAFT FÜR KERNFORSCHUNG M. B. H.

KARLSRUHE

GESELLSCHAFT FÜR KERNFORSCHUNG M.B.H.
KARLSRUHE

KERNFORSCHUNGSZENTRUM KARLSRUHE

Juni 1971 KFK 1428

Institut für Neutronenphysik und Reaktortechnik

Stabilität und Konvergenz nichtlinearer Differenzenoperatoren bei Anfangswertaufgaben

W. Kinnebrock

Gesellschaft für Kernforschung mbH., Karlsruhe

Zusammenfassung

Der Begriff der Stabilität bei der numerischen Lösung von Anfangswertaufgaben wird für nichtlineare Probleme so definiert, daß Stabilität und Konsistenz Konvergenz gewährleisten. Im Teil II werden einige Stabilitätskriterien bewiesen. Die Konvergenzaussagen aus Teil I werden an numerischen Beispielen nachgeprüft.

Abstract

Stability is defined for difference operators solving numerically nonlinear initial-value-problems so that stability and consistence imply convergence. In part II some criteria for stability are proved. Two numerical examples are treated.

Inhalt

Bei der numerischen Lösung von Anfangswertproblemen mit Hilfe des Differenzenverfahrens ersetzt man den kontinuierlichen Differentialoperator durch einen Differenzenoperator,der bei gegebenen diskreten Anfangswerten die sukzessive Berechnung einer auf endlich vielen Gitterpunkten definierten Näherungslösung gestattet. Bei Verkleinern der Maschenweiten des Gitters soll diese Näherung in die exakte kontinuierliche Lösung übergehen (Konvergenz).Zur Realisierung dieser Konvergenzforderung ist es notwendig,daß zwischen Differenzen- und Differentialoperator eine formale Verbindung besteht im Sinne der Konsistenz.Wie jedoch Courant,Friedrichs und Lewy ([1]) als erste 1928 entdeckten,reicht die Konsistenzeigenschaft nicht aus,um Konvergenz zu gewährleisten.Da nämlich jeder Lösungswert der Näherung mit einem Verfahrensfehler behaftet ist und gleichzeitig als Startwert für die Berechnung neuer Werte an benachbarten Gitterpunkten dient,wird er dort neue Fehler verursachen.Bei Verkleinern der Schrittweiten wird nun zwar der lokale Verfahrensfehler beliebig klein,jedoch kann die Fortpflanzung auf weitere Gitterpunkte so explosiv sein,daß das Wachstum des globalen Fehlers nicht durch die Reduktion des lokalen Fehlers aufgehoben wird.In diesem Fall sprechen wir von Instabilität. Die meisten Autoren nennen einen Differenzenoperator stabil, wenn bei Verkleinern der Schrittweiten die Fehler,die

durch Einbringen einer lokalen Störung entstehen, beschränkt bleiben. Dieser Arbeit liegt ein Stabilitätsbegriff zugrunde, der z.B. - zumindest in ähnlicher Form - verwendet wird bei Strang [4], Forsythe-Wasow [19], Spijker [17] und Watt [23]: Wenn h die maximale Schrittweite bedeutet, so heißt ein Differenzenoperator r-stabil, wenn für $h \to 0$ ein durch eine lokale Störung verursachter Fehler höchstens wie h^{-r} wächst.

Während bei linearen Differenzenoperatoren Stabilität eine Eigenschaft des Operators ist, hängt sie bei nichtlinearen Problemen sowohl vom Operator als auch von der Lösung des kontinuierlichen Problems ab. Daher ist es günstig, Stabilität und Konvergenz stets bezüglich fester singulärer Anfangswerte zu betrachten.

Der wesentliche Inhalt dieser Arbeit ist der Beweis folgenden Sachverhaltes: Ist eine nichtlineare Differenzengleichung von der Ordnung p und ist die an der Stelle der kontinuierlichen Lösung linearisierte Differenzengleichung r-stabil, sind zudem die für die sukzessive Rechnung benötigten Anfangswerte so beschaffen, daß sie für $h \to 0$ wie h^m in die exakten übergehen, so ist die ursprüngliche Differenzengleichung konvergent mit dem Konvergenzgrad $q = \min(p-1, m) - r$, falls die Ungleichungen $p \geq 2(r+1)$ und $m \geq 2r + 1$ erfüllt sind. Sind die Ungleichungen nicht erfüllt, gilt Konvergenz, falls die linearisierte Gleichung eine spezielle - hier als "streng r-stabil" bezeichnete - Eigenschaft erfüllt.

Im Kap.II werden vier verschiedene Stabilitätskriterien bewiesen, deren Anwendung in Kap.III an speziellen Beispielen erprobt wird.

Die Überlegungen werden der einfacheren Formulierung wegen nur für Differentialgleichungen durchgeführt und sind leicht auf Systeme zu verallgemeinern. Auch die Variablenanzahl läßt sich leicht von zwei auf n erhöhen. Ohne Schwierigkeiten könnte man die hier betrachtete Klasse echter Anfangswertaufgaben erweitern auf Anfangsrandwertaufgaben. Konvergenzen werden im Sinne der Maximumnorm betrachtet.

Die meisten Arbeiten über die numerische Lösung nichtlinearer Anfangswertaufgaben behandeln spezielle Differentialgleichungstypen (z.B. quasilinear, halblinear etc.). Die Ergebnisse dieser Arbeit sind auf alle Typen nichtlinearer Anfangswertaufgaben anwendbar, sofern eine eindeutige kontinuierliche Lösung existiert. Aussagen dieser allgemeinen Art findet man bei Ansorge [15] [24], Spijker [17], Stetter [12], [16] und Strang [9]. Der Stabilitätsbegriff von Spijker ist nicht identisch mit dem hier vorliegenden, für lineare Probleme jedoch decken sich beide Begriffe. Die hier benutzte Definition der Stabilität ist ähnlich der von Strang, jedoch - wie auch der entsprechende Konvergenzsatz - allgemeiner.

I) Allgemeine Theorie

1) Problemstellung

Sei G die Menge der reellen Zahlen x mit $0 \leqq x \leqq X$ und I die Menge der Zahlen t mit $0 \leqq t \leqq T$. Auf einem Teilgebiet B von $G \times I$, das die Punkte $\{(x,0);\ x \in G\}$ enthält, sei die reellwertige Funktion u(x,t) definiert und genüge der nicht notwendig linearen Gleichung

$$R[u(x,t)] = 0 \tag{1}$$

und den Anfangsbedingungen

$$\frac{\partial^j}{\partial t^j} u(x,0) = \varphi_j(x) \qquad (j=0, 1, \ldots k) \tag{2}$$

mit $\varphi_j(x) \in C_G^0$

Wir setzen voraus, daß das Anfangswertproblem (1),(2) genau eine Lösung auf B besitzt.

2) Differenzengleichungen

Zur numerischen Lösung von (1),(2) überziehen wir $G \times I$ mit einem achsenparallelen Gitter mit den Maschenweiten Δx und $h = \Delta t$, wobei

$$\Delta x = f(h) \to 0 \qquad (h \to 0)$$

Den Gitterpunkten $(j \cdot \Delta x, n \cdot h)$ auf B ordnen wir die Werte u_j^n zu, wobei diese Gitterwerte berechnet werden durch die Gleichungen:

$$(3) \qquad u_j^{n+1} = \psi_{j,n}(u_{j+l_1}^n, \ldots, u_{j+l_\mu}^n, u_{j+l_1}^{n-1}, \ldots, u_{j+l_\mu}^{n-1}, \ldots$$

$$\ldots, u_{j+l_1}^{n-k}, \ldots, u_{j+l_\nu}^{n-k}, h)$$

Durch Vorgabe von Anfangswerten

$$(4) \qquad u_j^m = \omega_m(j \cdot \Delta x) \qquad (m = 0, 1, \ldots, k); \ (j \cdot \Delta x \in G)$$

seien alle u_j^n auf B eindeutig bestimmt. Die Funktionen $\psi_{j,n}$ seien definiert für $0 < h \leq h_0$, $0 < n \cdot h < T$, $j \cdot \Delta x \in G$, sie seien stetig in h.

Die Differenzengleichungen (3) lassen sich in einer einfacheren Symbolik darstellen. Dazu sei B_n die Menge der Punkte (x,t) mit $x \in G$ und $t = n \cdot h$.

Es sei

$$G_n = B_n \cap B \qquad \text{und} \qquad G_n^* = \left\{ x \,/\, (x, n\cdot h) \in G_n \right\}$$

B sei so beschaffen, daß $G_{n+1}^* \subseteq G_n^*$. Wir fassen die Gitterwerte u_j^n, deren Gitterpunkte $(j\cdot\Delta x, n\cdot h)$ auf G_n liegen, zu einem Vektor $u^n = u^n(h)$ zusammen, und (3) ist darstellbar durch eine Zuordnung der Gestalt

$$(5) \qquad u^{n+1} = \tilde{\psi}_n(u^n, u^{n-1}, \ \ldots\ , u^{n-k}, h)$$

wobei $\tilde{\psi}_n$ eine Vektorfunktion ist. Die Vorgabe der Anfangswerte geschieht durch

$$(6) \qquad u^m = \tilde{\omega}_m \qquad\qquad (m = 0, 1, \ \ldots\ , k)$$

Abstände und Konvergenzen betrachten wir im folgenden stets im Sinne der Norm:

$$\left\| u^n \right\|_h = \sup_j \left| u_j^n \right|$$

wobei für jedes n ein Vektorraum mit anderer Dimension zugrunde liegen kann.

Die Gleichungen (5) können weiter vereinfacht werden, wenn man die Vektoren $u^n, u^{n-1}, \ \ldots\ , u^{n-k}$ zu einem Großvektor $v^n = v^n(h)$ zusammenfaßt. (vgl. z.B. Ansorge, [8]).

Es sei also

$$v^n = \begin{pmatrix} u^n \\ u^{n-1} \\ \vdots \\ u^{n-k} \end{pmatrix}$$

Dann gilt

$$v^{n+1} = \begin{pmatrix} \tilde{\psi}_n(u^n, u^{n-1}, \ldots, u^{n-k}, h) \\ u^n \\ u^{n-1} \\ \vdots \\ u^{n-k+1} \end{pmatrix} = \Phi_n(v^n, h)$$

Also

$$v^{n+1} = \Phi_n(v^n, h) \tag{7}$$

mit dem Anfangselement

$$v^k = \Omega_h = \begin{pmatrix} \tilde{\omega}_k \\ \tilde{\omega}_{k-1} \\ \vdots \\ \tilde{\omega}_o \end{pmatrix} \tag{8}$$

Als Norm sei zugrundegelegt

$$\| v^n \| = \max_{j=n-k}^{n} \| u^j \|_j$$

Um Vergleiche mit der exakten Lösung u(x,t) von (1),(2) durchführen zu können,führen wir die beiden folgenden Operatoren ein:

1) $$E_h \cdot u(x,n\cdot h) = \left(u(j\cdot \Delta x, n\ h) \right)_j$$

wobei der rechts stehende Ausdruck ein Vektor mit den Elementen $u(j\cdot\Delta x,t)$ sei für festes $t=n\cdot h$ und alle j mit $(j\cdot\Delta x,t)\in G_n$.

2) $$\hat{E}_h \cdot u(x,t) := \begin{pmatrix} E_h\ u(x,t) \\ E_h\ u(x,t-h) \\ \vdots \\ E_h\ u(x,t-kh) \end{pmatrix}$$

3) Konsistenz,Konvergenz

Wir definieren:

Definition 1: Die Differenzenoperatoren Φ_n aus (7) heißen zu R p-konsistent genau dann,wenn für jede auf B definierte Funktion v(x,t),die die Gleichung (1) erfüllt,gilt:

$$\hat{E}_h \cdot v(x,t+h) - \Phi_n(\hat{E}_h \cdot v(x,t),h) = O(h^p) \quad (h \to 0)$$

für $0 \leqq t \leqq T,\ 0 < h \leqq h_o\ ,\ 0 \leqq n\cdot h \leqq T$

Soll (3),(4) eine Näherungslösung für (1),(2) liefern, müssen die Anfangswerte $\omega_m(j\cdot\Delta x)$ aus (4) für $h\to 0$ in die echten Lösungswerte $u(j\Delta x, m\cdot h)$ übergehen, oder, in der Formulierung (7),(8), es muß $\Omega_h - \hat{E}_h\, u(x,kh) \to 0$ gelten. Das führt auf die

Definition 2: Der Anfangsvektor Ω_h aus (8) heißt bezüglich der Lösung $u(x,t)$ von (1),(2) s-zulässig genau dann, wenn

$$\hat{E}_h \cdot u(x,kh) - \Omega_h = O(h^s) \qquad (h\to 0)$$

für $0 < h \leqq h_o$

Die Anfangswerte $\omega_m(j\cdot\Delta x)$ werden meist durch weitere Differenzengleichungen bestimmt, deren Konsistenz mit den Anfangsbedingungen (2) i.A. schon die Zulässigkeit im Sinne von Definition 2 nach sich zieht. Der Begriff "zulässig" findet sich z.B. bei Ansorge ([8]).

Die meisten Autoren sprechen von Konvergenz, wenn die Lösung der Differenzengleichung für $h\to 0$ gegen die des Anfangswertproblems konvergiert. Vorteilhafter ist es, wenn man bei der Formulierung des Konvergenzbegriffes eventuelle Rundungsfehler brücksichtigt und fordert, daß die durch Rundungen gestörte Lösung gegen die des kontinuierlichen Problems konvergiert. Dabei wird man allerdings voraussetzen müssen, daß für $h\to 0$ die

Störungen ebenfalls beliebig klein werden, etwa wie h^q für ein $q > 0$. Diese Voraussetzung ist zwar in der Praxis nie erfüllt, jedoch wird man annehmen können, daß eine gewisse numerische Stabilität gewährleistet wird. Konvergenzbegriffe ähnlicher Art finden sich z.B. bei Ansorge ([8] , [24]) und Stetter ([6]).

Es gelte also die

Definition 3: Die Differenzenoperatoren ϕ_n mit den Störungen $\delta_n(h)$ und dem Anfangselement Ω_h heißen bezüglich $u(x,t)$ q-konvergent genau dann, wenn aus

$$v^{n+1} = \phi_n(v^n,h) + \delta_n(h)$$
$$v^k = \Omega_h$$

folgt:

$$v^n - \hat{E}_h\, u(x,n\cdot h) = O(h^q) \qquad (h \to 0)$$

für $0 < n\cdot h \leq T$, wenn $u(x,t)$ die Lösung von (1),(2) bedeutet.

4) Stabilität

Wir benutzen im folgenden die in Definition 4 festgelegten Sprechweisen:

Definition 4: 1)Die Funktionen $\psi_{j,n}$ aus (3) heißen "in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt" genau dann, wenn gilt:

Es existiert ein $\varepsilon > 0$ und ein $C > 0$, so daß für alle w(x,t) auf B mit

$$\sup_{(x,t)\in B} \left| w(x,t) - u(x,t) \right| \leqq \varepsilon$$

das Folgende erfüllt ist:

a) $\psi_{j,n}$ für $w^m_{j+l_i} = w((j+l_i)\Delta x, m\cdot h)$ definiert und dortselbst zweimal differenzierbar.

b) $$\left| \frac{\partial^2 \psi_{j,n}(w^n_{j+l_1}, w^n_{j+l_2}, \ldots\ldots, w^{n-k}_{j+l_r}, h)}{\partial w^{m_1}_{j+l_{i_1}} \cdot \partial w^{m_2}_{j+l_{i_2}}} \right| \leqq C$$

für $0 < h \leqq h_0, 0 < n\cdot h \leqq T, j\cdot\Delta x \in G^*_n$

2)Die Operatoren ϕ_n aus (7) heißen nach w(x,t) k mal stetig-differenzierbar genau dann, wenn die durch (3) gegebenen $\psi_{j,n}$ für alle j,n nach den Argumenten

$$w^m_{j+l_i} = w((j+l_i)\Delta x, m\cdot h)$$

k mal stetig-differenzierbar sind.

3)Die Operatoren ϕ_n aus (7) heißen "in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt"genau dann,wenn die Funktionen $\psi_{j,n}$ bei u(x,t) in der zweiten Ableitung gleichmäßig beschränkt sind.

Es sei bemerkt,daß die Operatoren ϕ_n in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt sind,wenn die Funktionen $\psi_{j,n}$ in der Umgebung der Werte $u((j+l_i)\Delta x, m\cdot h)$ zweimal stetig-differenzierbar sind sowie die zweiten Ableitungen für $0 \leqq h \leqq h_o$ definiert und stetig in h sind. Für spätere Zwecke benötigen wir den

Satz 1: Vor.:Die Operatoren ϕ_n sind in der zweiten Ableitung bei v(x,t) gleichmäßig beschränkt.

Beh.:Es existiert ein $\varepsilon > 0$,so daß für alle $\sigma = \sigma(h)$ aus dem Definitionsbereich von ϕ_n mit $\|\sigma\| \leqq \varepsilon$ gilt:

$$(9)\qquad \phi_n(\hat{E}_h \cdot v(x,n\cdot h) + \sigma, h) - \phi_n(\hat{E}_h \cdot v(x,n\cdot h), h)$$

$$= \phi'_n(\hat{E}_h \cdot v(x,n\cdot h), h)\cdot\sigma + \psi_n(\hat{E}_h \cdot v(x,n\cdot h), h, \sigma)\cdot\sigma$$

wobei ϕ'_n , ψ_n von v bzw v und σ sowie von h abhängige lineare Operatoren sind.Es ist zudem:

(10) $$\left\| \psi_n(\hat{E}_h \cdot v(x, n\cdot h), h, \sigma) \right\| \leqq M \|\sigma\|$$

Obiges gilt für alle n,h mit $0 \leqq n\cdot h \leqq T$ und für ein $M > 0$.

Beweis:

Wegen der Voraussetzung sind die Operatoren ϕ_n und damit die Funktionen $\psi_{j,n}$ aus (3) in der Umgebung von $v(x_i, t_n)$ zweimal differenzierbar, d.h.: Es existiert ein $\varepsilon > 0$, so daß $\psi_{j,n}$ nach $(v(x_i, t_n) + \sigma_{i,n})$ zweimal differenzierbar ist, falls $|\sigma_{i,m}| \leqq \varepsilon$. Sei nun $\sigma = \sigma(h)$ ein Vektor des Definitionsbereiches von ϕ_n mit den Komponenten $\sigma_{i,m} = \sigma_{i,m}(h)$ und $|\sigma_{i,m}| \leqq \varepsilon$. Dann gilt die Taylorentwicklung: ($v^m_{j+l_i} := v((j+l_i)\Delta x, m\cdot h)$)

(11) $$\psi_{j,n}(v^n_{j+l_1} + \sigma_{1,n}, \ \ldots\ldots \ , v^{n-k}_{j+l_r} + \sigma_{r,n-k}, h)$$

$$- \ \psi_{j,n}(v^n_{j+l_1}, \ \ldots\ldots \ , v^{n-k}_{j+l_r}, h)$$

$$= \sum_{i,m} c^m_{j+l_i} \cdot \sigma_{i,m} + \sum_{i,m} \left\{ \frac{1}{2} \sum_{\bar{i},\bar{m}} d^{m,\bar{m}}_{i,\bar{i}} \ \sigma_{\bar{i},\bar{m}} \right\} \cdot \sigma_{i,m}$$

mit

(12) $$c^m_{j+l_i} = \frac{\partial \psi_{j,n}(v^n_{j+l_1}, \ \ldots \ , v^{n-k}_{j+l_r}, h)}{\partial v^m_{j+l_i}}$$

$$(13)\qquad d_{i,\bar{i}}^{m,\bar{m}} = \frac{\partial^2 \Psi_{j,n}(v_{j+l_1}^{n} + \varphi\cdot\sigma_{1,n},\ \ldots\ ,v_{j+l_r}^{n-k} + \varphi\cdot\sigma_{r,n-k},h)}{\partial v_{j+l_i}^{m}\cdot\partial v_{j+l_{\bar{i}}}^{\bar{m}}}$$

$$(0 \leqq \varphi \leqq 1;\ 0 < h \leqq h_0)$$

Wegen der gleichmäßigen Beschränktheit in der zweiten Ableitung ist

$$(14)\qquad \left| d_{i,\bar{i}}^{m,\bar{m}} \right| \leqq C \qquad (C > 0)$$

für alle σ mit $\| \sigma \| \leqq \varepsilon$

Es sei nun σ^m ein Teilvektor von σ, bestehend aus den Komponenten $\sigma_{i,m}$, variiert über alle i. Dann ist σ^m erklärt für m = n-k, n-k+1, ... ,n, und (11) läßt sich in der Formulierung von (5) so schreiben:

$$(15)\qquad \tilde{\Psi}_n(E_h\cdot v(x,n\cdot h) + \sigma^n,\ \ldots\ ,E_h\cdot v(x,(n-k)h) + \sigma^{n-k},h)$$

$$- \tilde{\Psi}_n(E_h\, v(x,n\cdot h),\ \ldots\ ,E_h\, v(x,(n-k)h),h)$$

$$= \sum_{m=n-k}^{n} A_m\cdot\sigma^m \qquad + \sum_{m=n-k}^{n} B_m\cdot\sigma^m$$

Die Summanden sind Vektoren mit den Komponenten:

$$\left(A_m\cdot\sigma^m\right)_j = \sum_{i=1}^{r} c_{j+l_i}^{m}\cdot\sigma_{i,m}$$

$$(16)\qquad \left(B_m \cdot \tilde{\sigma}^m\right)_j = \sum_{i=1}^{r} \left\{ \frac{1}{2} \sum_{\bar{i},\bar{m}} d_{i,\bar{i}}^{m,\bar{m}}\, \sigma_{\bar{i},\bar{m}} \right\} \tilde{\sigma}_{i,m}$$

Schreibt man schließlich (15) in der Formulierung (7), so folgt die zu beweisende Relation (9) mit

$$\Phi_n'(\hat{E}_h v, h) \cdot \sigma = \begin{pmatrix} \sum_m A_m \cdot \sigma^m \\ \sigma^n \\ \sigma^{n-1} \\ \vdots \\ \sigma^{n-k+1} \end{pmatrix}$$

und

$$\Psi_n(\hat{E}_h v, h, \sigma) \cdot \sigma = \begin{pmatrix} \sum_m B_m \sigma^m \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Es bleibt noch der Beweis von (10). Sei x ein beliebiges Element aus dem Definitionsbereich von Ψ_n. Wenn x^m und $x_{i,m}$ die gleiche Bedeutung haben wie die entsprechenden Symbole σ^m und $\sigma_{i,m}$ von σ, so gilt:

$$\psi_n(\hat{E}_h v,h,\sigma)\,x = \begin{pmatrix} \sum_m B_m x^m \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

mit

$$\left(B_m x^m\right)_j = \sum_{i=1}^{r}\left\{\frac{1}{2}\sum_{\bar{i},\bar{m}} d_{i,\bar{i}}^{m,\bar{m}}\,\sigma_{\bar{i},\bar{m}}\right\}\cdot x_{i,m}$$

Wegen (14) ist

$$\left|\left(B_m x^m\right)_j\right| \leq r\cdot\frac{1}{2}\cdot k\cdot r\cdot C\,\|\sigma\|\sup_i |x_{i,m}|$$

und daher

$$\|\psi_n(\hat{E}_h v,h,\sigma)\,x\| \leq M\,\|\sigma\|\,\|x\|$$

woraus die Behauptung folgt.

Es sollen jetzt die Eigenschaften formuliert werden, die zusammen mit der Konsistenz die Konvergenz gewährleisten. Es gelte also die

Definition 5: 1) Die Operatoren ϕ_n aus (7) heißen bezüglich der Lösung u(x,t) von (1),(2) r-stabil genau dann, wenn sie in der zweiten Ableitung bei

$u(x,t)$ gleichmäßig beschränkt sind, und wenn für die in (9) gegebenen Operatoren $\phi'_n(\hat{E}_h \cdot u(x,t),h)$ gilt:

$$\left\| \prod_{j=0}^{1} \phi'_{n-j}(\hat{E}_h u(x,(n-j)h),h) \right\| \leqq C \cdot h^{-r}$$

für: $C > 0$
$r \geqq 0$
$0 < h \leqq h_o$
$k \cdot h \leqq (n-1)h \leqq n \cdot h \leqq T-h$

2) Die Operatoren ϕ_n heißen bezüglich der Lösung $u(x,t)$ von (1),(2) streng r-stabil genau dann, wenn sie in der zweiten Ableitung bei $u(x,t)$ gleichmäßig beschränkt sind, und wenn für die in (9) gegebenen Operatoren $\phi'_n(\hat{E}_h u(x,t),h)$ gilt:

$$\prod_{j=1}^{l} \left\| \phi'_{i_j}(\hat{E}_h u(x,i_j h),h) \right\| \leqq C \cdot h^{-r}$$

für: $C > 0$
$r \geqq 0$
$0 < h \leqq h_o$
$k \leqq i_1 < i_2 < \quad \ldots \quad < i_l \leqq \frac{T-h}{h}$

Der hier erklärte Stabilitätsbegriff geht für lineare Operatoren in bekannte Stabilitätsbegriffe über (Für r = 0 z.B. folgt die Lax-Richtmyer-Stabilität). Ist nämlich Φ_n linear, so gilt $\Phi_n = \Phi_n'$ und die Forderung für r-Stabilität lautet:

$$\left\| \prod_{j=0}^{l} \Phi_{n-j} \right\| \leq C \cdot h^{-r}$$

Ist Φ_j zudem von t bzw j unabhängig, folgt

$$\left\| \Phi^n \right\| \leq C \cdot h^{-r}$$

Für r=0 spricht man oft von "starker Stabilität" im Gegensatz zur "schwachen"Stabilität" oder "linearen Instabilität" bei $r > 0$.

5) Konvergenzsätze

Wie bereits erwähnt (vgl.Def.3), liegt für die numerische Berechnung des Problems (1),(2) der folgende Algorithmus vor:

$$\begin{aligned} v^{n+1} &= \Phi_n(v^n, h) + \delta_n \\ v^k &= \Omega_h \\ \delta_n &= O(h^p) \qquad (h \to 0) \end{aligned}$$

Die δ_n sind Rundungsfehler. Ω_h sei als Anfangsvektor

m-zulässig, d.h.:

$$\Omega_h - \hat{E}_h u(x,kh) = O(h^m) \qquad (h \to 0)$$

Dabei läßt sich $O(h^m)$ als Summe von Verfahrensfehler der für die Berechnung der Anfangswerte verwendeten Differenzengleichungen und der Rundungsfehler interpretieren, wenn man voraussetzt, daß letztere mit der Ordnung h^m gegen 0 konvergieren (für $h \to 0$). Da die δ_n ebenfalls Rundungen bedeuten, ist jede eventuelle Konvergenz des obigen Algorithmus in einem gewissen Sinne numerisch stabil.

Für die Konvergenz gilt nun der

Satz 2: Vor: 1) Die Operatoren Φ_n sind

a) zu R p-konsistent

b) bezüglich der Lösung u(x,t) von (1),(2) r-stabil

2) Ω_h ist m-zulässig zu u(x,t)

3) Es ist $p \geq 2(r+1)$ und $m \geq 2r + 1$.

Beh: Die Operatoren Φ_n mit den Störungen $\delta_n = O(h^p)$ und dem Anfangselement Ω_h sind bezüglich u(x,t) q-konvergent mit $q = \min(p-1, m) - r$.

Beweis:

Es sei

$$e^n = v^n - \hat{E}_h u(x,nh)$$

wobei v^n die (im obigen Algorithmus definierte) Näherungslösung bedeutet. Dann ist wegen der Konsistenz:

$$e^{n+1} = \phi_n(v^n,h) - \phi_n(\hat{E}_h u(x,nh),h) + \delta_n - O(h^p)$$

Daraus folgt:

$$(17) \qquad e^{n+1} = \phi_n(\hat{E}_h u(x,nh)+e^n,h) - \phi_n(\hat{E}_h u(x,nh),h) + \tau_n$$

mit

$$\| \tau_n \| \leqq c_1 \cdot h^p$$

Zu zeigen ist:

$$(18) \qquad \| e^j \| \leqq c_2 \cdot h^q \qquad \text{(für alle } j\text{)}$$

Wir beweisen dies durch Induktion. Für $j=k$ ist

$$e^k = \Omega_h - \hat{E}_h u(x,kh) = O(h^m)$$

also $\| e^k \| \leqq c_3 \cdot h^m \leqq c_3 \cdot h^q$ (denn $q \leqq m, h \leqq h_0$)

Die Ungleichung (18) gelte für j=k,k+1, ... ,n.Ohne Beschränkung der Allgemeinheit können wir $C_2 \geqq C_3$ voraussetzen. Dann läßt sich h so klein wählen,daß für (17) die Relation (9) von Satz 1 anwendbar ist,so daß gilt:

$$e^{n+1} = (\phi'_n + \psi_n)\, e^n + \tau_n$$

mit

$$\phi'_n = \phi'_n(\hat{E}_h u(x,nh),h)$$

$$\psi_n = \psi_n(\hat{E}_h u(x,nh),h,e^n)$$

$$\| \psi_n \| \leqq C_4 \cdot \| e^n \| \leqq C_5 \cdot h^q$$

Setzt man e^n rekursiv ein,folgt

$$(19) \qquad e^{n+1} = \prod_{j=0}^{n-k} (\phi'_{n-j} + \psi_{n-j})\, e^k + \sum_{m=k+1}^{n} \prod_{j=0}^{n-m} (\phi'_{n-j} + \psi_{n-j})\, \tau_{m-1} + \tau_n$$

Multipliziert man das Produkt $\prod_{j=0}^{m_o} (\phi'_{n-j} + \psi_{n-j})$ aus, erhält man eine Summe mit 2^{m_o+1} Summanden. $\binom{m_o+1}{j}$ von diesen enthalten genau j mal den Operator ψ und höchstens

(j+1) Operatorenprodukte der Art $\phi'_l \cdot \phi'_{l-1} \cdots \phi'_{l-s}$ dazwischen, d.h.: jeder Summand ist (wegen der r-Stabilität) abschätzbar durch:

$$\|\psi_{n_1}\| \cdot \|\psi_{n_2}\| \cdots \|\psi_{n_j}\| \cdot (C_6 h^{-r})^{j+1}$$

$$\leqq (C_5 \cdot h^q)^j \cdot (C_6 \cdot h^{-r})^{j+1}$$

Daraus folgt:

$$\left\| \prod_{j=0}^{m_o} (\phi'_{n-j} + \psi_{n-j}) \right\| \leqq \sum_{j=0}^{m_o+1} \binom{m_o+1}{j} \cdot (C_5 h^q)^j \cdot (C_6 h^{-r})^{j+1}$$

$$= C_6 \cdot h^{-r} \sum_{j=0}^{m_o+1} \binom{m_o+1}{j} (C_5 C_6)^j h^{(q-r)j}$$

$$= C_6 \cdot h^{-r} \left(1 + C_5 \cdot C_6 \cdot h^{q-r}\right)^{m_o+1}$$

Wegen der Vor.3) ist $q-r = \min(p-1,m) - 2r \geqq 1$, so daß $\left(1+ C_5 \cdot C_6 \cdot h^{q-r}\right)^{m_o+1} < \exp(C_5 \cdot C_6 \cdot T) = C_7$. Aus (19) folgt daher:

$$\left\|e^{n+1}\right\| \leqq C_6 \cdot C_7 \cdot h^{-r} \left\|e^k\right\| + n \cdot C_6 \cdot C_7 \cdot h^{-r} \cdot C_1 \cdot h^p$$

$$\leqq C_8 \cdot h^{m-r} + C_9 h^{p-r-1}$$

$$\leqq C_{10} h^{\min(p-1,m)-r} = C_{10} h^q \qquad (h \leqq h_o)$$

Wir haben gezeigt: Wenn $\|e^j\| \leq C_2 \cdot h^q$ für $j=k,k+1, \ldots ,n$, dann ist $\|e^{n+1}\| \leq C_{10} \cdot h^q$. Die Konstante C_{10} ist in keiner Weise von C_2 abhängig, so daß für alle m mit $0 \leq mh \leq T$ gilt:

$$\|e^m\| \leq \max(C_2, C_{10}) \cdot h^q$$

woraus die Behauptung folgt.

Dieser Satz wurde in ähnlicher Form von Strang in [9] bewiesen, jedoch nur für r=0 und in der L_2-Norm. Die zugrundeliegenden Differentialgleichungen waren quasilineare hyperbolische Systeme 1. Ordnung. Auch die Voraussetzungen waren verschieden von denen des Satzes 2. (vgl. auch [20], Seite 127).

Die Ungleichungen $p \geq 2(r+1)$ und $m \geq 2r + 1$ schränken die Klasse der anwendbaren Probleme stark ein, jedoch kann man annehmen, daß stabile Differenzengleichungen für Differentialgleichungen erster Ordnung i.A. 0-stabil sind und $p \geq 2, m \geq 1$ stets erfüllt ist, so daß auf Probleme dieser Art Satz 2 anwendbar ist. Auch für Probleme zweiter Ordnung (deren Differenzenoperatoren meist 1-stabil sind), gelten oft obige Ungleichungen. Für die Anfangswertprobleme, für die sie verletzt sind, gilt der folgende Satz 3:

Satz 3: Vor.: 1) Die Operatoren Φ_n sind

a) zu R p-konsistent

b) bezüglich der Lösung u(x,t) von (1),(2) streng r-stabil

2) Ω_h ist m-zulässig zu u(x,t)

3) $q = \min(p-1,m) - r \geq 1$

Beh.: Die Operatoren Φ_n mit Störungen $\delta_n = O(h^p)$ und dem Anfangselement Ω_h sind bezüglich u(x,t) q-konvergent.

Beweis:

Der erste Teil des Beweises verläuft wörtlich wie im Beweis zu Satz 2. Wie dort erhält man die Relation (19):

$$(19) \qquad e^{n+1} = \prod_{j=0}^{n-k} (\Phi'_{n-j} + \Psi_{n-j})\, e^k + \sum_{m=k+1}^{n} \prod_{j=0}^{n-m} (\Phi'_{n-j} + \Psi_{n-j})\, \tau^{m-1} + \tau^n$$

mit: $\|\tau^j\| \leq c_1 \cdot h^p$

$\|e^k\| \leq c_3 \cdot h^m \leq c_3 h^q$

$\|\Psi_m\| \leq c_5 \cdot h^q$

Ausmultiplizieren des Produktes $\prod_{j=0}^{m} (\Phi'_{n-j} + \Psi_{n-j})$

ergibt eine Summe von 2^{m+1} Summanden, deren jeder abschätzbar ist durch

$$(C_5 \cdot h^q)^j \left\| \Phi'_{n-i_1} \right\| \cdot \left\| \Phi'_{n-i_2} \right\| \cdot \quad \cdots\cdots \quad \cdot \left\| \Phi'_{n-i_{m+1-j}} \right\|$$

falls genau j mal der Operator Ψ in diesem Summanden vorkommt. Wegen der strengen Stabilität folgt die Abschätzung:

$$(C_5\, h^q)^j \cdot C_6 \cdot h^{-r}$$

Das ergibt:

$$\left\| \prod_{j=0}^{m} (\Phi'_{n-j} + \Psi_{n-j}) \right\| \leqq \sum_{j=0}^{m+1} \binom{m+1}{j} (C_5\, h^q)^j \cdot C_6 \cdot h^{-r}$$

$$= C_6\, h^{-r}\, (1 + C_5 h^q)^{m+1} \;<\; C_6 \cdot h^{-r} \cdot \exp(C_5 T) = C_7\, h^{-r}$$

Aus (19) folgt daher:

$$\left\| e^{n+1} \right\| \leqq C_7 h^{-r} \cdot \left\| e^k \right\| + \; n \cdot C_7 \cdot h^{-r} \cdot C_1 \cdot h^p$$

$$\leqq C_8 \cdot h^{m-r} + C_9 \cdot h^{p-r-1} \leqq C_{10}\, h^q \quad \text{mit } q = \min(p-1, m)-r$$

Mit der gleichen Schlußfolgerung wie im Beweis zu Satz 2 folgt daraus die Behauptung.

II. Stabilitätskriterien

Während es bei linearen Differenzengleichungen eine Reihe von brauchbaren Stabilitätskriterien gibt, existieren bei nichtlinearen Gleichungen kaum leistungsfähige Kriterien. Die Schwierigkeit rührt nicht nur von der Nichtlinearität, sondern auch daher, daß die Lösung bekannt sein muß. Auch die in diesem Kapitel bewiesenen Stabilitätskriterien lassen sich nur für einfache Gleichungstypen anwenden. Zuvor formulieren wir noch den (trivialen)

Satz 4: Vor.: Die Operatoren Φ_n sind bezüglich u(x,t) streng r-stabil

Beh.: Die Operatoren Φ_n sind bezüglich u(x,t) r-stabil.

Ein erstes Stabilitätskriterium ist der

Satz 5: Vor.: 1) Die Operatoren Φ_n sind in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt.

2) f(t,h) sei eine reellwertige Funktion für $0 \leqq t \leqq T, 0 < h \leqq h_0$, stetig-differenzierbar nach t für alle h. Es sei $f(t,h) > 0$ für $0 < t < T$ und $0 < h \leqq h_0$.

3) Für $k \cdot h \leqq t_1 < t_2 \leqq T - h$ ist

$$\int_{t_1}^{t_2} \frac{\frac{\partial}{\partial t} f(t,h)}{f(t,h)} \, dt \leqq C - r \cdot \ln(h)$$

mit $r \geqq 0$

4) $$\left\| \phi_n'(\hat{E}_h u(x,t),h) \right\| \leqq \frac{f(t+h,h)}{f(t,h)} + C_o h^{r+1}$$

$(C_o \geqq 0, t=nh)$

Beh.: 1) Die Operatoren ϕ_n sind bezüglich u(x,t) r-stabil.

2) Ist zudem f(t,h) monoton steigend in t, so sind die Operatoren ϕ_n bezüglich u(x,t) streng r-stabil.

Beweis:

1) Sei $\phi_n = \phi_n(\hat{E}_h u(x,t),h)$ und $f_j = f(jh,h)$. Dann ist

$$(20) \quad \left\| \prod_{j=0}^{m} \phi_{m-j}' \right\| \leqq \prod_{j=0}^{m} \left(\frac{f_{m-j+1}}{f_{m-j}} + C_o h^{r+1} \right)$$

Sei $A_i = \frac{f_{i+1}}{f_i}$. Es gilt die Abschätzung:

$$(21)\quad \left| \prod_{i=n_1}^{n_2} A_i \right| = \left| \frac{f_{n_2+1}}{f_{n_1}} \right| = \exp\left\{ \ln(f_{n_2+1}) - \ln(f_{n_1}) \right\}$$

$$= \exp\left\{ \int_{n_1 h}^{n_2 h+h} \frac{\frac{\partial}{\partial t} f(t,h)}{f(t,h)}\, dt \right\} \leqq \exp\left(C - r \ln(h)\right) = C_1 h^{-r}$$

Ausmultiplizieren des Produktes der rechten Seite von (20) führt auf eine Summe, deren Summanden Produkte aus $C_o h^{r+1}$ und A_i sind. $\binom{m+1}{j}$ dieser Summanden enthalten als Faktoren genau j mal $C_o h^{r+1}$ und höchstens (j+1) Produkte der Form $A_i A_{i+1} A_{i+2} \cdots$ dazwischen, d.h.: sie sind wegen (21) abschätzbar durch

$$(C_o h^{r+1})^j \, (C_1 h^{-r})^{j+1} = (C_o h \cdot C_1)^j \cdot C_1 h^{-r}$$

Daraus folgt:

$$\left\| \prod_{j=0}^{m} \Phi'_{m-j} \right\| \leqq \left\{ \sum_{j=0}^{m+1} \binom{m+1}{j} \cdot C_o{}^j \cdot h^j \cdot C_1{}^j \right\} C_1 h^{-r}$$

$$= C_1 h^{-r} \, (1 + C_o C_1 h)^{m+1} \quad < \quad C_1 h^{-r} \exp(C_o T \cdot C_1) = C_3 h^{-r}$$

woraus die Behauptung 1) folgt.

2) Es gilt:

$$(22)\qquad \prod_{j=0}^{m} \| \Phi_{i_j} \| \leqq \prod_{j=0}^{m} \left\{ \frac{f_{i_j+1}}{f_{i_j}} + C_o h^{r+1} \right\}$$

$$\leqq \prod_{j=i_o}^{i_m} \left\{ \frac{f_{j+1}}{f_j} + C_o h^{r+1} \right\}$$

Die letztere Abschätzung gilt wegen der Monotonie von $f(t,h)$, die

$$\frac{f_{i+1}}{f_i} \geqq 1$$

gewährleistet, so daß in das erste Produkt von (22) beliebig viele Faktoren $\frac{f_{i+1}}{f_i}$ eingefügt werden können, ohne daß sich die Ungleichung ändert. Der restliche Beweis verläuft dann wie in Teil 1).

Im folgenden wird die Anwendungsmöglichkeit von Satz 5 an zwei Beispielen demonstriert:

1) Sei $f(t,h) = \text{const.} = c$ mit $c > 0$.

Dann ist

$$\int_{t_1}^{t_2} \frac{\frac{\partial}{\partial t} f(t,h)}{f(t,h)} \, dt = 0$$

und

$$\frac{f(t+h,h)}{f(t,h)} = 1$$

Die Voraussetzungen von Satz 5 für f(t,h) sind also für r=0 erfüllt,und man erhält die Aussage:

Wenn die Operatoren Φ_n in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt sind und wenn

$$\| \Phi_n'(\hat{E}_h u(x,t),h) \| \leqq 1 + C\cdot h \qquad (C \geqq 0) \tag{23}$$

dann sind die Operatoren für u(x,t) O-stabil und streng O-stabil.Gleichung (23) ist die bekannte von Neumann - Bedingung. (vgl.[2], [20])

2) Sei $f(t,h) = t^r$ mit $r \geqq 0$.Dann ist

$$\int_{t_1}^{t_2} \frac{\frac{\partial}{\partial t} f(t,h)}{f(t,h)}\, dt = r(\ln(t_2) - \ln(t_1))$$

$$\leqq r \ln(T) - r \ln(kh) \leqq C_1 - r \ln(h)$$

Letzteres gilt wegen $t_1 \geqq k\,h$.

Wegen $\frac{f(t+h,h)}{f(t,h)} = (1 + \frac{h}{t})^r$ folgt:

Wenn die Operatoren ϕ_n in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt sind und wenn

$$(24) \qquad \left\| \phi_n'(\hat{E}_h u(x,t),h) \right\| \leqq (1 + \frac{h}{t})^r + C\, h^{r+1}$$

dann sind die Operatoren ϕ_n für u(x,t) r-stabil und streng r-stabil.

Für r=0 folgt wieder die Neumann-Bedingung.

Die folgenden beiden Sätze gelten für Operatoren, die als Summe zweier Differenzenoperatoren darstellbar sind. Ähnliche Stabilitätskriterien wurden z.B. von Ansorge in [18] und Spijker in [17] für die von ihnen betrachteten Stabilitätsbegriffe bewiesen.

Satz 6: Vor.: 1) Die Operatoren P_n sind in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt.

2) $\left\| P_n'(\hat{E}_h u(x,t),h) \right\| \leqq M$ für $0 \leqq t \leqq T, 0 < h \leqq h_o$

3) Die Operatoren T_n sind r-stabil bei u(x,t)

4) $1 \geqq r + 1$

Beh.: Die Operatoren

$$Q_n = T_n + h^1 \cdot P_n$$

sind r-stabil bei u(x,t).

Beweis:

Für die Ableitung $Q_n' = T_n' + h^1 \cdot P_n'$ gilt:

$$\left\| Q'_n\, Q'_{n-1} \cdots Q'_{n-j_o} \right\| = \left\| \prod_{j=0}^{j_o} (T'_{n-j} + h^{l} P'_{n-j}) \right\|$$

Ausmultiplizieren des Produktes führt auf eine Summe mit 2^{j_o+1} Summanden. Jeder Summand enthält j mal $h^{l} \cdot P'_i$ und (j_o+1) Gruppen $T'_m\, T'_{m-1} \cdots T'_{m-k}$ dazwischen für ein j mit $0 \leqq j \leqq j_o+1$.

Daher gilt die Abschätzung: (es ist $\left\| P'_i \right\| \leqq M$)

$$\left\| Q'_n\, Q'_{n-1} \cdots Q'_{n-j_o} \right\| \leqq \sum_j \binom{j_o+1}{j} (h^{l} M)^j \cdot (Ch^{-r})^{j+1}$$

$$= \sum_j \binom{j_o+1}{j} (h^{l-r} M \cdot C)^j \, C \cdot h^{-r} = C \cdot h^{-r} (1 + h^{l-r} MC)^{j_o+1}$$

$$< \exp(M \cdot C \cdot T) \cdot C \cdot h^{-r}$$

Die letzte Abschätzung gilt wegen $l - r \geq 1$.

Für streng stabile Operatoren gilt der

Satz 7: Vor.: 1) Die Operatoren P_n sind in der zweiten Ableitung bei u(x,t) gleichmäßig beschränkt.

2) $\left\| P'_n(\hat{E}_h u(x,t), h) \right\| \leqq M$ für $0 \leqq t \leqq T$ und $0 < h \leqq h_o$

3) Die Operatoren T_n sind streng r-stabil bei u(x,t)

Beh.: Die Operatoren $Q_n = T_n + h\,P_n$ sind streng r-stabil bei $u(x,t)$.

Im Gegensatz zu Satz 6 fehlt hier die Voraussetzung $q \geqq r+1$.

Beweis.: Es ist

$$\prod_{j=1}^{j_o} \| Q'_{n_j} \| \leqq \prod_{j=1}^{j_o} \left(\| T'_{n_j} \| + h \cdot \| P'_{n_j} \| \right)$$

Ausmultiplizieren des rechten Produktes führt auf eine Summe, deren Summanden abschätzbar sind durch

$$M^j \cdot h^j \cdot C \cdot h^{-r}$$

falls $\| h \cdot P'_m \|$ genau j mal in dem Summanden als Faktor auftaucht. Daher gilt:

$$\prod_j \| Q'_{n_j} \| \leqq \sum_{j=0}^{j_o} \binom{j_o}{j} M^j h^j C h^{-r} \leqq (1 + M \cdot h)^{j_o} \cdot C \cdot h^{-r}$$

$$< \exp(M \cdot T) \cdot C \cdot h^{-r}$$

woraus die Behauptung folgt.

Es sei angemerkt, daß Satz 6 und Satz 7 unter Umständen auch ohne Kenntnis der speziellen Lösung des kontinuierlichen

Problems zur Stabilitätsuntersuchung angewandt werden können.So ist z.B.die Differenzengleichung

$$u_j^{n+1} = u_j^n + h\, f(u_j^n) ,$$

die $u_t = f(u)$ approximiert,nach Satz 7 streng 0-stabil, falls $|f'_u(u)| \leqq M$.

Die Ableitungen der Differenzenoperatoren, $\phi'_n(\hat{E}_h u,h)$, sind Matrizen $\tilde{T}(u(x,t_n),x,t_n,h) = T(x,t_n,h)$ mit $x=j\cdot\Delta x, t_n = n\cdot h$. Wenn nun für die Matrizen

$$T_n = T(x,t_n,h)$$

gilt:

$$(25) \qquad \| T_n\, T_{n-1} \cdot \ldots \cdot T_{n-l} \| \leqq C_o$$

für $0 < nh < T, 0 \leqq l \leqq n$,dann ist der mit T assoziierte Differenzenoperator ϕ_n,für den $\phi'_n = T$, 0-stabil.
Der folgende Satz formuliert Voraussetzungen für T(x,t,h), die für das Produkt $(T_n\, T_{n-1} \cdots T_{n-l})$ die Abschätzung (25) gewährleisten,d.h. der Satz stellt ein Stabilitätskriterium für 0-Stabilität dar.

Satz 8. Vor.: 1) $T(x,t,h)$ sei eine quadratische, in x und h stetige und in t stetig-differenzierbare Matrix für $x \in G$, $0 \leqq t \leqq T$, $0 \leqq h \leqq h_0$.

2) Die Eigenwerte $\lambda_i(x,t,h)$ von T erfüllen die Bedingungen:

2.1) $|\lambda_i(x,t,h)| \leqq 1 + C \cdot h \qquad (C > 0)$

$(x \in G, 0 \leqq t \leqq T, 0 \leqq h \leqq h_0)$

2.2) $\lambda_i(x,t,h)$ stetig-differenzierbar in t

2.3) Die Vielfachheiten der Eigenwerte $\lambda_i(x,t,h)$ sind 1 für $x \in G, 0 \leqq t \leqq T, 0 \leqq h \leqq h_0$.

Beh: Für $T_n = T(x,nh,h)$ mit $x \in G, 0 \leqq h \leqq h_0, 0 \leqq nh \leqq T$ gilt die Abschätzung:

$$\| T_n \, T_{n-1} \cdots T_{n-l} \| \leqq C_0$$

für $0 \leqq l \leqq n$, $C_0 > 0$

Zum Beweis benötigen wir das

Lemma 1: Vor.: $B(x,t)$ sei eine $n \times n$- Matrix für $0 \leqq t \leqq T$ und $x \in G$. $B(x,t)$ sei stetig in x und stetig-differenzierbar in t. Der Rang von $B(x,t)$ sei $(n-1)$ für alle x,t.

Beh.: Es existiert ein Vektor $b(x,t)$ mit

1) $B(x,t)\cdot b(x,t) = 0$

2) $b(x,t) \neq 0$

3) $b(x,t)$ stetig in x und stetig-differenzierbar in t

für $x \in G$ und $0 \leq t \leq T$.

Beweis:

Sei $B(x,t) = \left(b_{ij}(x,t)\right)_{n,n}$ und $c_{ij}(x,t)$ das algebraische Komplement zu $b_{ij}(x,t)$. Dann ist

$$\sum_j b_{ij}(x,t)\, c_{lj}(x,t) = \begin{cases} \det B(x,t) = 0 & (l = i) \\ 0 & (l \neq i) \end{cases}$$

Die Vektoren

$$b_j = b_j(x,t) = \begin{pmatrix} \vdots \\ c_{j,k-1} \\ c_{j,k} \\ c_{j,k+1} \\ \vdots \end{pmatrix}$$

erfüllen dann die Behauptung 1) und 3).

Alle b_j sind Elemente eines eindimensionalen Unterraumes, so daß, falls $b_j \neq 0$,

$$c = \frac{b_j}{\|b_j\|}$$

bis auf das Vorzeichen unabhängig von j ist. Da nicht alle b_j

gleichzeitig verschwinden, erfüllt c (bei Vorgabe einer Orientierung) die Behauptung.

Es folgt der Beweis von Satz 8:

Es sei im folgenden stets $\lambda_i = \lambda_i(x,t,h)$, $T_n = T(x,nh,h)$. Für jeden Eigenwert λ_i gibt es einen Vektor $b_i = b_i(x,t,h)$ mit

$$(T - \lambda_i I)\cdot b_i = 0 \tag{26}$$

Wegen Lemma 1 sind die b_i stetig-differenzierbar in t und stetig in x und h wählbar. Sei $B = B(x,t,h)$ eine Matrix mit den linear unabhängigen Spaltenvektoren b_i. Dann ist

$$\| B(x,t,h) \| \leqq M_o \qquad (x \in G, 0 \leqq t \leqq T, 0 \leqq h \leqq h_o) \tag{27}$$

Zudem ist B überall differenzierbar nach t. Wenn $B = (b_{ij})$, $B^{-1} = (c_{ij})$, dann ist

$$c_{ij} = \frac{\text{Algebraisches Komplement von } b_{ji}}{\det B}$$

Daher ist auch B^{-1} stetig - differenzierbar in t und

stetig in x,h,und es ist

$$(28)\quad \left\| B^{-1}(x,t+h,h) - B^{-1}(x,t,h) \right\| \leqq M_1 h$$

Es gilt nun: $(B_n := B(x,nh,h))$

$$(29)\quad T_n \cdot T_{n-1} \cdot \ldots \cdot T_{n-l} =$$

$$B_{n+1}B_{n+1}^{-1}T_nB_nB_n^{-1}T_{n-1}B_{n-1}B_{n-1}^{-1}T_{n-2} \cdots\cdots T_{n-l}B_{n-l}B_{n-l}^{-1} =$$

$$B_{n+1}(J_n + R_n)(J_{n-1} + R_{n-1}) \cdots\cdots (J_{n-l} + R_{n-l})\, B_{n-l}^{-1}$$

mit

$$J_k = B_k^{-1}T_kB_k$$

und

$$R_k = (B_{k+1}^{-1} - B_k^{-1})\, T_k\, B_k$$

Wegen (27) und (28) ist

$$(30)\quad \left\| R_k \right\| \leqq M_2 h$$

J_k ist die Jordansche Normalform zu T_k,d.h.

$$J_k = \begin{pmatrix} \lambda_1(x,kh,h) & & & \\ & \lambda_2(x,kh,h) & & 0 \\ & & \ddots & \\ 0 & & & \lambda_N(x,kh,h) \end{pmatrix}$$

Daraus folgt mit Vor.2.1:

$$(31) \qquad \| J_k \| \leqq 1 + C\,h$$

Mit (29),(30),(31) folgt

$$\| T_n \cdot T_{n-1} \cdot \ldots \cdot T_{n-l} \| \leqq \| B_{n+1} \| \cdot \| B_{n-l}^{-1} \| \, (1 + C\,h + M_2 h)^{l+1}$$

$$\leqq \| B_{n+1} \| \, \| B_{n-l}^{-1} \| \, \exp((C + M_2)T)$$

Wegen der Stetigkeit von $B(x,t,h)$ und $B^{-1}(x,t,h)$ folgt schließlich:

$$\| T_n \, T_{n-1} \cdot \ldots \cdot T_{n-l} \| \leqq M_3$$

$$(M_3 > 0,\ 0 \leqq l \leqq n,\ 0 \leqq nh \leqq T)$$

III. Beispiele

1) Stabilitätsuntersuchungen

Nach Def.5 sind Operatoren ϕ_n r-stabil, wenn neben der Differenzierbarkeit das Produkt der Frechet-Ableitungen an der Stelle der kontinuierlichen Lösung mit Ch^{-r} beschränkt ist. Das heißt, daß die Lösungen von

$$v^{n+1} = \phi_n'(\hat{E}_h u(x,nh),h)\cdot v^n$$

mit Ch^{-r} beschränkt bleiben müssen. D.h.: Stabilität des linearisierten Problems impliziert Stabilität des nichtlinearen Problems, falls der nichtlineare Differenzenoperator bei u(x,t) in der zweiten Ableitung gleichmäßig beschränkt ist.

Mit Hilfe der in Kap.II bewiesenen Stabilitätskriterien sollen hier einige einfache Differenzengleichungen auf Stabilität untersucht werden:

1.1) $$\frac{\partial^p}{\partial t^p} u(x,t) = 0$$

werde approximiert durch den zugehörigen Differenzenquotienten

(32) $$\sum_{j=0}^{p} \binom{p}{j} \cdot (-1)^j \, u_l^{n+1-j} = 0 = D\cdot u_l^n$$

Die allgemeine Lösung ist

$$u_j^n = \sum_{q=0}^{p-1} c_q(j\Delta x)\, n^q$$

Daher ist (32) (p-1)-stabil.

1.2) $$\frac{\partial^p u}{\partial t^p} = f(u)$$

Es sei $\left|f'_u(u(x,t))\right| \leqq M$ für alle x,t.Wir ersetzen den linken Teil durch den Differenzenoperator (32)

(33) $$D\, u_j^n = h^p\, f(u_j^n)$$

Nach Satz 6 ist (33) (p-1)-stabil.Für p = 1 ist (33) nach Satz 7 streng 0-stabil.

1.3 $$\frac{\partial u}{\partial t} = f(u, \frac{\partial}{\partial x}u)$$

a) Wir lösen die Differentialgleichung mit

(34) $$u_j^{n+1} = u_j^n + h\, f(u_j^n, \frac{u_{j+1}^n - u_j^n}{\Delta x}) \qquad (h=\Delta x)$$

f sei zweimal stetig-differenzierbar.Die linearisierte Gleichung lautet

$$(35) \qquad v_j^{n+1} = v_j^n + h\, f'_u\, v_j^n + h\, f'_{u_x} \cdot \frac{v_{j+1}^n - v_j^n}{\Delta x}$$

Sei v^n ein Vektor mit den Komponenten v_j^n für alle j. Dann lassen sich die Gleichungen (35) für alle j zu dem System zusammenfassen:

$$v^{n+1} = B(h)\, v^n$$

mit

$$B(h) = \begin{pmatrix} (1+hf'_u-f'_{u_x})_1 & (f'_{u_x})_1 & & \\ & \ddots & \ddots & 0 \\ & (1+hf'_u-f'_{u_x})_j & (f'_{u_x})_j & \\ 0 & & \ddots & \ddots \end{pmatrix}$$

B(h) entspricht dem Operator ϕ und ist entscheidend für Stabilität.Die Eigenwerte von B(h) sind

$$\mu_j = (1 + h\, f'_u - f'_{u_x})_j$$

Für $0 \leqq f'_{u_x} \leqq 2$ gilt $|\mu_j| \leqq 1 + C\,h$

Sind die Eigenwerte zudem einfach, ist Satz 8 anwendbar und es gilt die Aussage:

Für $0 \leqq f'_{u_x}(u,u_x) \leqq 2$ ist (34) 0-stabil.

b) Ganz analog läßt sich beweisen: Die zu 1.3) konsistente Differenzengleichung

$$u_j^{n+1} = u_j^n + h \cdot f(u_j^n, \frac{u_j^n - u_{j-1}^n}{\Delta x})$$

ist 0-stabil, falls $-2 \leqq f'_{u_x}(u,u_x) \leqq 0$.

2) Numerische Beispiele

2.1) Wir betrachten

$$y'' = y^2$$
$$y(0) = 6$$
$$y'(0) = -2\,y(0)$$

Nach Beispiel 1.2) ist die Differenzengleichung

$$y^{n+1} = 2\,y^n - y^{n-1} + h^2(y^n)^2$$

1-stabil.Die Anfangswerte erhalten wir durch

$$y^0 = 6$$
$$y^1 = y(0) + h\,y'(0) + \frac{h^2}{2}\,y''(0)$$
$$= 6 - 12\,h + 18\,h^2$$

Nunmehr ist Satz 2 anwendbar für $p = 4$, $r = 1$, $m = 3$. Man erhält q-Konvergenz für $q = 2$, d.h.:

$$\left| y^n - y(nh) \right| \leqq C \cdot h^2 \tag{39}$$

In Figur 1 ist die maximale Abweichung der Näherungslösung von der exakten Lösung in Abhängigkeit von der Schrittweite h eingetragen.(Die exakte Lösung ist $y(x) = 6\,(x + 1)^{-2}$).Man sieht,daß die Fehlerkurve unter der in rot eingetragenen Kurve $\mathcal{E} = 40\,h^2$ liegt, was die Ungleichung (39) bestätigt.

2.2) Für lineare Differenzenoperatoren ist Satz 2 ohne die Voraussetzung 3),d.h. ohne die einschränkenden Ungleichungen

$$p \geqq 2(r+1)$$
$$m \geqq 2r + 1$$

gültig.Dies ist leicht nachprüfbar,wenn man im Beweis zu Satz 2 $\psi_n \equiv 0$ setzt.(Bei linearen Operatoren ist $\psi_n \equiv 0$).Das ermöglicht die folgenden,am Beispiel

der Wellengleichung durchgeführten Überlegungen:

Gegeben sei

$$\frac{\partial^2 u(x,t)}{\partial t^2} = \frac{\partial^2 u(x,t)}{\partial x^2}$$

mit den Anfangsbedingungen

$$u(x,0) = \frac{1}{\pi} \cdot \sin(\pi \cdot x)$$

$$\frac{\partial}{\partial t} u(x,0) = \cos(\pi \cdot x) \qquad (0 \leqq x \leqq 1{,}5)$$

Die Differentialgleichung wurde in dem durch die Charakteristiken bestimmten Gebiet gelöst mit Hilfe der Differenzengleichung

$$(40) \qquad u_j^{n+1} = 2\, u_j^n - u_j^{n-1} + (u_{j+1}^n - 2\, u_j^n + u_{j-1}^n)$$

(40) ist 4-konsistent und 1-stabil.

Im ersten Fall wurden die Anfangswerte berechnet durch

$$(41) \qquad \begin{aligned} u_j^0 &= \frac{1}{\pi} \sin(\pi \cdot x_j) \\ u_j^1 &= u_j^0 + h \cos(\pi \cdot x_j) \end{aligned}$$

Die Gleichungen (41) liefern 2-zulässige Anfangswerte, so daß für die maximale Abweichung der Näherungslösung von der exakten Lösung

$$u(x,t) = \frac{1}{\pi} \sin(\pi \cdot (x+t))$$

nach Satz 2 gilt:

$$(42) \qquad \left| u_j^n - u(x_j,t_n) \right| \leqq C \cdot h$$

Es sei $\varepsilon(h)$ der maximale Fehler. In Figur 2 sind die Werte $\varepsilon(h)$ sowie die Kurve $0{,}37\ h$ (in rot) aufgetragen. Man sieht die Bestätigung der Ungleichung (42).
Es gilt:

$$\left| \varepsilon(h) - 0{,}37\ h \right| \leqq 1{,}2\ 10^{-3}$$

für $0 < h \leqq 0{,}05$.

Im zweiten Fall wurden die Anfangswerte berechnet durch

$$u_j^0 = \frac{1}{\pi} \sin(\pi \cdot x_j)$$

$$u_j^1 = u_j^0 + h\ u_t(x_j,0) + \frac{h^2}{2} u_{tt}(x_j,0)$$

mit

$$u_t(x_j,0) = \cos(\pi \cdot x_j)$$

$$u_{tt}(x_j,0) = u_{xx}(x_j,0) = \frac{\partial^2}{\partial x^2} \left(\frac{1}{\pi} \sin(\pi x_j) \right.$$

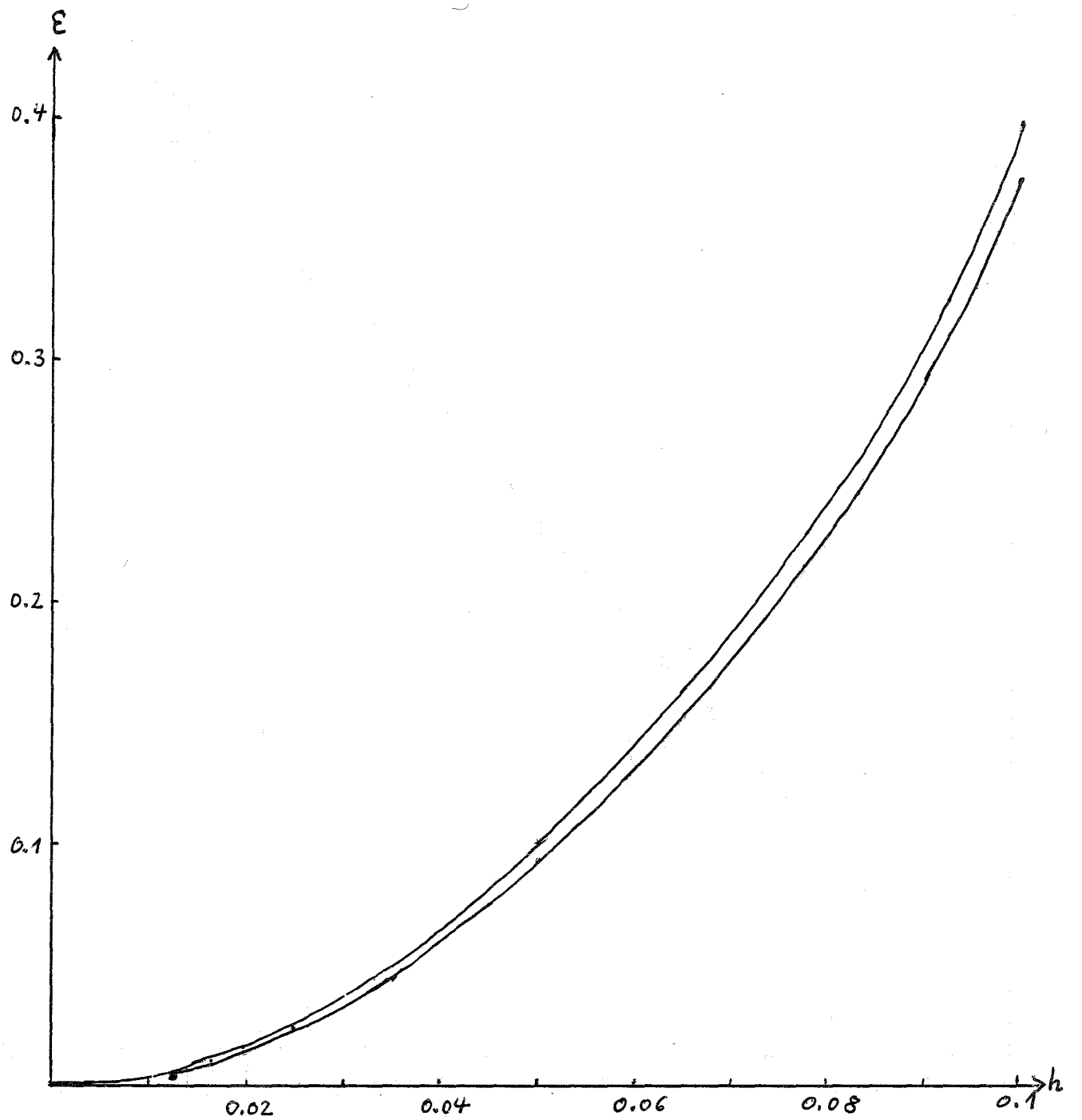
Die Anfangswerte sind jetzt 3-zulässig und für den maximalen Fehler $\varepsilon(h)$ gilt nach Satz 2:

$$\left| \varepsilon(h) \right| = \max_{j,n} \left| u_j^n - u(x_j, t_n) \right| \leqq C \cdot h^2$$

In Figur 3 sind $\varepsilon(h)$ und $0{,}63\ h^2$ für $0 < h \leqq 0{,}05$ eingetragen. Beide Kurven stimmen fast überein.
Es gilt:

$$\left| \varepsilon(h) - 0{,}63\ h^2 \right| \leqq 1{,}4 \cdot 10^{-5}$$

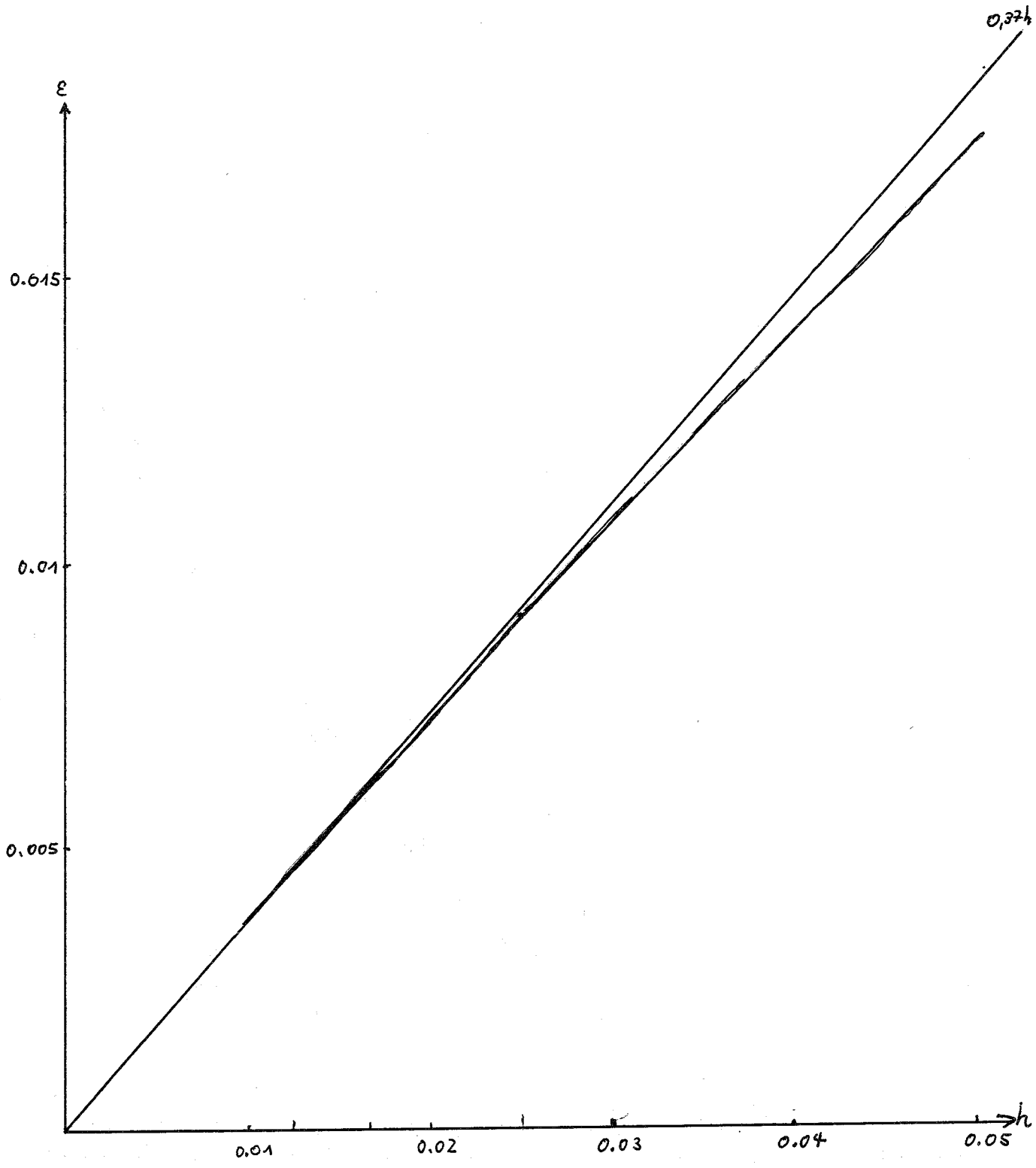
Auch hier wird daher durch die Rechnung die vorausgesagte Fehlerabschätzung bestätigt.



Figur 1

——— $\varepsilon = 40\ h^2$

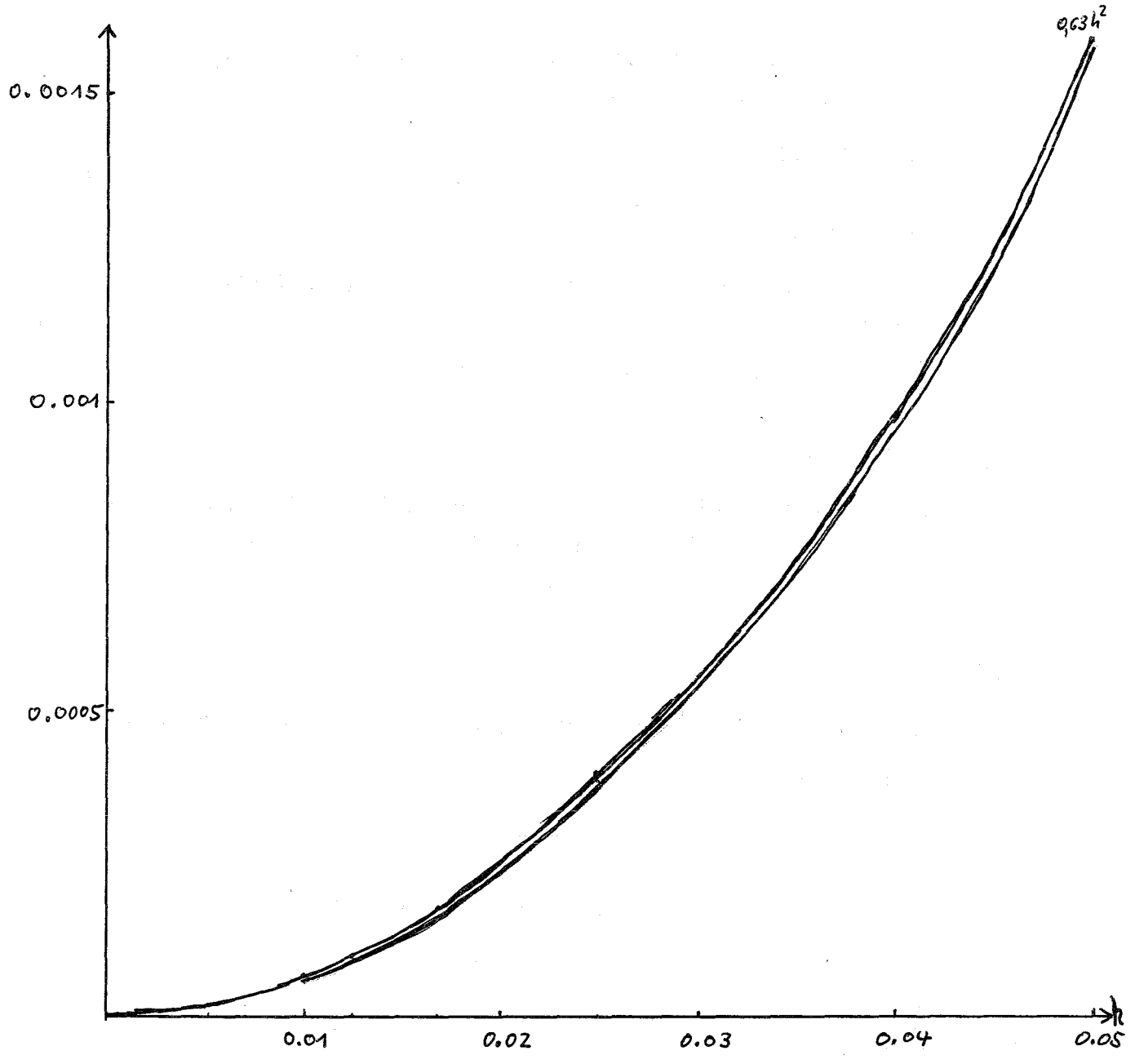
——— ε = max. Fehler
$= \mathrm{Max} | y'' - y(nh) |$



Figur 2

—— $\varepsilon = 0{,}37\ h$

—— $\varepsilon \neq \varepsilon(h)$



Figur 3

Literatur

[1] Courant,Friedrichs,Lewy

Über die partiellen Differentialgleichungen der mathematischen Physik

Math.Ann.100, 32-74, (1928)

[2] Von Neumann,Richtmyer

A method for the numerical calculation of hydrodynamic shocks

J.of appl.phys.21, 232-237 (1950)

[3] Dahlquist

Convergence and stability in the numerical integration of ordinary differential equations

Math.Scand 4,33-53 (1956)

[4] Strang

Difference methods for mixed boundary value problems

Duke Math.J.27, p.221 (1960)

[5] Ryabenkii,Fillipow

Über die Stabilität von Differenzengleichungen

Deutscher Verlag der Wissensch.,Berlin,1960

[6] Stetter

On the convergence of characteristic finite difference methods of high accuracy for quasi-linear hyperbolic equations

Num.Math.3,321-344, (1961)

[7] Henrici

Discrete variable methods in ordinary differential equations

John Wiley,New York,1962

[8] Ansorge

Zur Struktur gewisser Konvergenzkriterien bei der numerischen Lösung von Anfangswertaufgaben

Num.Math.6, 224-234 (1964)

[9] Strang

Accurate partial difference methods.
II.Non-linear problems

Num.Math.6, 37-46 (1964)

[10] Richtmyer,Morton

Stability studies for difference equations.

Rept.NYO-1480-5 Courant Inst.of Math.Sci. (1964)

[11] Gary

Finite difference schemes for hyperbolic systems.

Math.Comp.18, 1-18 (1964)

[12] Stetter

A study of strong and weak stability in discretization algorithms

SIAM Num.An.2, 265-280 (1965)

[13] Stetter

Asymptotic expansions for the error of discretization algorithms for non-linear functional equations

Num.Math.7,18-31 (1965)

[14] Ansorge

Der Äquivalenzsatz von Lax für halblineare Probleme

ZAMM 46, T35 (1966)

[15] Ansorge

Zur Frage der Verallgemeinerung des Äquivalenzsatzes von P.D.Lax.

ISNM 9, 13-23 (1966)

[16] Stetter

Stability of nonlinear discretization algorithms.

Num.sol.of part.diff.equ. - New York, 111-123 (1966)

[17] Spijker

Convergence and stability of step-by-step methods for the numerical solution of initial-value-problems.

Num.Math.8, 161 (1966)

[18] Ansorge

Konvergenz von Mehrschrittverfahren zur Lösung halblinearer Anfangswertaufgaben.

Num.Math.10, 209-219 (1967)

[19] Forsythe-Wasow

Finite difference methods for partial differential equations

Wiley & sons, New York, 1967

[20] Richtmyer, Morton

Difference methods for initial value problems.

Wiley & sons, New York, 1967

[21] Gourlay,Morris

A multistep formulation of the optimized Lax-Wendroff method for nonlinear hyperbolic systems in two space variables.

Math.Comp.22, 715-719 (1967)

[22] Strang

On the construction and comparison of difference equations.

SIAM Num.An.5, 506-517 (1968)

[23] Watt

Convergence and stability of discretization methods for functional equations.

Comp.J.11, 77-82 (1968)

[24] Ansorge

Problemorientierte Hierarchie von Konvergenzbegriffen bei der numerischen Lösung von Anfangswertaufgaben.

Math.Z.112,13-22 (1969)

[25] Rubin

Time dependent techniques for the solution of viscous heat conducting,chemically reacting,radiating discontinous flows.

Conf.on Num.sol.of diff.equ. Springer,234-243 (1969)

[26] Dahl

Approximation of nonlinear operators.

Conf.on Num.sol.of diff.equ. Springer, 148-153 (1969)

[27] Ansorge

Zur Existenz verallgemeinerter Lösungen nichtlinearer Anfangswertaufgaben.

ISNM 12, 13-22 (1969)

[28] Ansorge

Konvergenz von Differenzenverfahren für quasilineare Anfangswertaufgaben.

Num.Math. 13, 217-225 (1969)

[29] Rubin, Preiser

Three dimensional second-order accurate difference schemes for discontinous hydrodynamic flows.

Math.Comp. 24, 57-63 (1970)