

---

Preprint No. M 02/12

**Zwei Anwendungen von  
Diskretisierungsverfahren für  
nichtlineare Operatorgleichungen**

Vogt, Werner

Oktober 2002

**Impressum:**

Hrsg.: Leiter des Instituts für Mathematik  
Weimarer Straße 25  
98693 Ilmenau

Tel.: +49 3677 69 3621

Fax: +49 3677 69 3270

<http://www.tu-ilmenau.de/ifm/>

ISSN xxxx-xxxx

ilmedia

# **Zwei Anwendungen von Diskretisierungsverfahren für nichtlineare Operatorgleichungen**

Werner Vogt  
Technische Universität Ilmenau  
Institut für Mathematik  
Postfach 100565  
98684 Ilmenau

Ilmenau, den 15.10.2002

**Zusammenfassung** Ein grundlegender numerischer Zugang zur Lösung nichtlinearer Operatorgleichungen in Banachräumen durch Diskretisierungsverfahren wird allgemein dargestellt. Konsistenz, Stabilität des diskreten Problems und Konvergenz der Näherungslösungen werden anhand von Differenzenverfahren für Zweipunkt-Randwertprobleme bei gewöhnlichen Differentialgleichungen sowie für quasilineare Systeme partieller Differentialgleichungen auf dem 2-Torus nachgewiesen.

**MSC 2000:** 65J15, 65L10, 65L12, 65L20, 65M06

**Keywords:** Discretization methods, Boundary value problems, Invariant torus

## 1 Diskretisierung nichtlinearer Operatorgleichungen

In diesem Beitrag betrachten wir allgemeine Aspekte von Diskretisierungsverfahren. Während die Lösung von Anfangswertproblemen

$$\frac{dx}{dt} = \dot{x} = f(t, x), \quad x(a) = x_0, \quad f : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^m, \quad (1)$$

beginnend mit dem gegebenen Anfangswert  $x_0$ , schrittweise mittels eines Einschritt- oder Mehrschrittverfahrens erfolgen kann, ist diese Vorgehensweise bei *Randwertproblemen* der allgemeinen Form

$$\dot{x} = f(t, x), \quad g(x(a), x(b)) = 0 \quad (2)$$

mit  $f : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  nicht möglich. Deshalb sind die Begriffe *Konsistenz*, *Stabilität* und *Konvergenz* nun so allgemein zu definieren, daß sie für beide Problemklassen zutreffen. Der von H.B.KELLER [3] und H.J.STETTER [9] entwickelte und bei ASCHER et al. [1] genutzte Zugang soll hier vorgestellt werden.

Das gegebene Problem wird dazu als Operatorgleichung in den Funktionenräumen  $B$  (Urbildraum) und  $B^0$  (Bildraum)

$$F(x) = 0, \quad F : B \rightarrow B^0 \quad (3)$$

betrachtet. Dabei sind der Urbildraum  $B$  und der Bildraum  $B^0$  vollständige normierte Räume (Banachräume). Das Tripel  $\mathcal{P} = (B, B^0, F)$  heißt *Ausgangsproblem*, für das ein  $x^* \in B$  gesucht wird, das dieser Operatorgleichung  $F(x^*) = 0$  genügt. Jedes derartige  $x^*$  heißt klassische Lösung. Die theoretische Frage, unter welchen weiteren Voraussetzungen eine klassische Lösung  $x^*$  existiert, ist ein zentraler Gegenstand der qualitativen analytischen Theorie und wird deshalb hier ausgeklammert.

**Beispiel 1** Das Anfangswertproblem

$$\dot{x} = t - x^2, \quad x(0) = 0, \quad 0 < t < 1$$

läßt sich in die Operatorform  $F(x) = 0$  mit dem Differentialoperator

$$Fx(t) \equiv \dot{x}(t) - t + x(t)^2 \quad (4)$$

überführen. Urbildraum sei der Banachraum (B-Raum)

$$B = \{x | x \in C^1(0, 1), x(0) = 0\}$$

der auf dem Intervall  $I = [0, 1]$  stetig differenzierbaren Funktionen, die die homogene Anfangsbedingung erfüllen und mit der  $C^1$ -Norm

$$\|x\|_B = \max_{0 \leq t \leq 1} |x(t)| + \max_{0 \leq t \leq 1} |\dot{x}(t)|.$$

versehen ist. Als Bildraum empfiehlt sich der B-Raum der stetigen Funktionen

$$B^0 = C^0(0, 1) \quad \text{mit der Norm} \quad \|y\|_{B^0} = \max_{0 \leq t \leq 1} |y(t)|. \quad \blacktriangleleft$$

Eine geeignete Wahl der Funktionenräume  $B$  und  $B^0$  ist wesentlich, wenn bestimmte Lösungseigenschaften erwünscht sind. So bedeutet die Abschätzung  $\|x\|_B \leq C$  mit einer Konstanten  $C$ , daß außer den Funktionswerten auch die Ableitungswerte der Lösung durch  $C$  beschränkt sind, wogegen dies bei  $\|x\|_{B^0} \leq C$  nur für die Funktionswerte gilt. Aus Vereinfachungsgründen betrachtet man dennoch oft den B-Raum

$$B = \{x | x \in C^0(0, 1), x(0) = 0\}$$

stetiger Funktionen als Urbildraum und schränkt den Definitionsbereich des Operators  $F$  auf  $D = C^1(0, 1) \cap B$  ein.

Alternativ dazu kann man mittels Integration der Differentialgleichung und Berücksichtigung der Anfangsbedingung zu einer äquivalenten Integralgleichung

$$x(t) = \int_0^t [s - x(s)^2] ds$$

übergehen. Da jede Lösung dieser Gleichung die Eigenschaft  $x(0) = 0$  besitzt, kann man nun den Integraloperator

$$Fx(t) \equiv x(t) - \int_0^t [s - x(s)^2] ds \quad (5)$$

und beide Räume mit  $B = B^0 = C^0(0, 1)$  wählen.

Um das Ausgangsproblem numerisch auf einem Computer darstellen und lösen zu können, muß man es „diskretisieren“. Dazu betrachtet man zwei Folgen endlichdimensionaler B-Räume,

- die Urbildräume  $\{B_n\}$ ,  $n \in N$ , mit  $\dim B_n < \infty$ ,
- die Bildräume  $\{B_n^0\}$ ,  $n \in N$ , mit  $\dim B_n^0 < \infty$

und definiert eine Folge von Operatoren  $\{F_n\}$ ,  $n \in N$ , mit  $F_n : B_n \rightarrow B_n^0$ . Dabei ist  $N \subset \mathbb{N}$  eine unendliche Teilmenge der natürlichen Zahlen. Die Folge  $\{\mathcal{P}_n\}$  der Tripel  $\mathcal{P}_n = (B_n, B_n^0, F_n)$  wird als *diskretes Problem* bezeichnet. Jede Elementfolge  $\{u_n^*\}$ ,  $n \in N$ , mit  $u_n^* \in B_n$  ist eine *Lösung des diskreten Problems*, falls

$$F_n(u_n^*) = 0, \quad F_n : B_n \rightarrow B_n^0, \quad n \in N$$

(6)

gilt.

**Beispiel 2** (vgl. Bsp. 1) Auf dem Grundintervall  $[0, 1]$  kann man mit der Indexmenge  $N = \mathbb{N}$  zu jedem  $n \in N$  ein Gitter der Schrittweite  $h = 1/n$  mit  $n + 1$  Knoten

$$I_n = \{t_j | t_j = jh, j = 0(1)n\}$$

definieren. Urbildraum sei der  $(n+1)$ -dimensionale Vektorraum  $B_n = \{u | u = (u_0, u_1, \dots, u_n)^T, u_0 = 0\}$  mit der diskreten  $C^1$ -Norm

$$\|u\|_{B_n} = \max_{j=0(1)n} |u_j| + \max_{j=1(1)n} |\partial u_j|, \quad \partial u_j = \frac{1}{h}(u_j - u_{j-1}).$$

Als Bildraum empfiehlt sich  $B_n^0 = \mathbb{R}^n$  mit der Norm  $\|v\|_{B_n^0} = \max_{j=1(1)n} |v_j|$ . Den diskreten Operator  $F_n$  definieren wir komponentenweise, indem wir die 1.Ableitung  $\dot{x}(t_{j-1})$  in (4) durch den Differenzenquotienten ersetzen

$$\{F_n u\}_j \equiv \frac{1}{h}(u_j - u_{j-1}) - t_{j-1} + u_{j-1}^2, \quad j = 1(1)n. \quad \blacktriangleleft \quad (7)$$

An diesem Beispiel erkennt man, daß die endlichdimensionalen Ersatzräume  $B_n$  und  $B_n^0$  keine Unterräume der Originalräume  $B$  und  $B^0$  sind. Hierin unterscheiden sich Diskretisierungsverfahren von der Klasse der Projektionsverfahren. Um die Beziehung zwischen Ausgangsproblem  $\mathcal{P}$  und diskretem Problem  $\{\mathcal{P}_n\}$ ,  $n \in N$ , herzustellen, müssen lineare Abbildungen zwischen den entsprechenden Räumen, sogenannte Restriktionsoperatoren, eingeführt werden.

**Definition 3 (Restriktion)** Die Folge  $\{p_n\}$ ,  $n \in N$ , linearer beschränkter Operatoren mit  $p_n : B \rightarrow B_n$  sei normkonsistent, d.h. für alle  $x \in B$  gilt

$$\lim_{n \rightarrow \infty} \|p_n x\|_{B_n} = \|x\|_B, \quad n \in N.$$

Dann heißt  $p_n$  Restriktionsoperator (Restriktion, Einschränkung) von  $B$  auf  $B_n$ . Analoges gelte für die Restriktion  $p_n^0 : B^0 \rightarrow B_n^0$  des Bildraumes.

Wegen der Normkonsistenz ist die Folge der Quotienten  $\|p_n x\|_{B_n} / \|x\|_B$  für jedes Element  $x \in B$ ,  $x \neq 0$ , beschränkt. Nach dem Satz von Banach–Steinhaus sind dann die Normen der Operatoren  $p_n$  und  $p_n^0$  sogar gleichmäßig beschränkt. Es existieren Konstanten  $P \geq 1$  und  $P^0 \geq 1$ , so daß

$$\|p_n\| \leq P, \quad \|p_n^0\| \leq P^0, \quad n \in N \quad (8)$$

mit den induzierten Operatornormen gilt.

**Beispiel 4** (vgl. Bsp. 1 und 2) Zu den Räumen  $B$  und  $B_n$  definieren wir nun die Operatoren  $p_n$  komponentenweise auf dem Gitter  $I_n$  durch

$$\{p_n x\}_j = x(t_j), \quad j = 0(1)n.$$

Das Bild der Funktion  $x$  ist eine *Gitterfunktion*. Offenbar bildet  $p_n : B \rightarrow B_n$  ab und ist linear. Die Beschränktheit von  $p_n$  folgt mit dem Mittelwertsatz aus

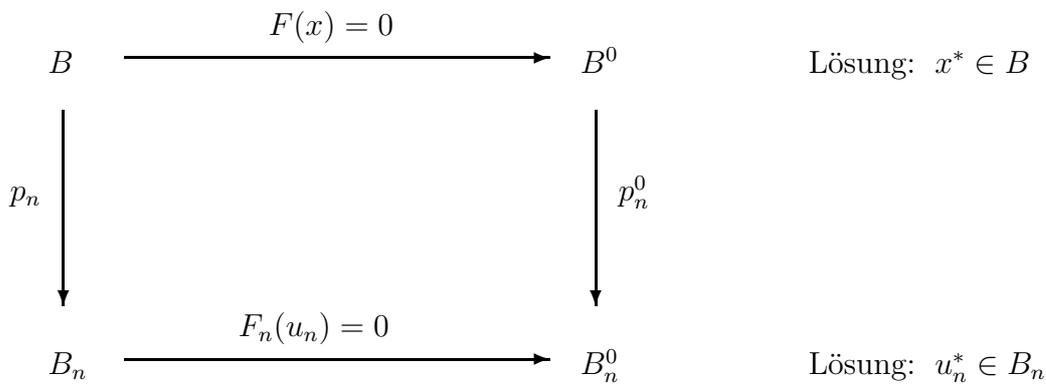
$$\begin{aligned} \|p_n x\|_{B_n} &= \max_{j=0(1)n} |x(jh)| + \max_{j=1(1)n} \left| \frac{1}{h}[x(jh) - x((j-1)h)] \right| \\ &\leq \max_{0 \leq t \leq 1} |x(t)| + \max_{0 \leq t \leq 1} |\dot{x}(t)| = \|x\|_B, \end{aligned}$$

womit sogar  $\|p_n\| \leq P = 1$  gilt. Die Normkonsistenz erhält man mit aufwendigeren Abschätzungen wegen der Stetigkeit von  $x(t)$  und  $\dot{x}(t)$ . Analog hierzu definiert man die Restriktion  $p_n^0$  zu den Bildräumen  $B^0$  und  $B_n^0$  der Beispiele 1 und 2 komponentenweise durch

$$\{p_n^0 y\}_j = y(t_{j-1}), \quad j = 1(1)n.$$

Auch hier zeigt man leicht, daß  $\|p_n^0\| \leq P^0 = 1$  gilt. ◀

Faßt man nun die 4 eingeführten Banachräume  $B, B^0, B_n, B_n^0$  und die 4 Operatoren  $F, F_n, p_n, p_n^0$  zusammen, so liefern sie ein *Diskretisierungsverfahren*. Den Zusammenhang zwischen Ausgangsproblem  $\mathcal{P}$  und diskretem Problem  $\{\mathcal{P}_n\}, n \in N$ , erhält man in anschaulicher Form durch das *Approximationsschema*



Dieses allgemeine Schema verdeutlicht zugleich das Ziel, für  $x \in B$  anstelle der vorgegebenen Gleichung  $F(x) = 0$  das diskrete Problem  $F_n(u_n) = 0$  zu lösen. Das gelingt offenbar, wenn der Operator  $F$  und die Diskretisierung  $p_n$  für  $n \rightarrow \infty$  miteinander vertauschbar sind.

## 2 Konsistenz und Stabilität

Wie man an Beispiel 1 – 2 sieht, ist die Wahl der Operatoren  $F, p_n, p_n^0$  und der entsprechenden B-Räume meistens evident. Die grundlegende Frage eines Diskretisierungsverfahrens lautet dann: Wie ist zu gegebenem Operator  $F$  der diskrete Operator  $F_n$  zu konstruieren, damit  $F_n$  das gegebene Problem mit  $F$  in einem gewissen Sinne „approximiert“? Dazu sollten die Bilder  $F_n(p_n x)$  und  $p_n^0 F(x)$  für alle  $x$  in einer Umgebung der Lösung  $x^*$  ebenfalls nahe beieinander liegen.

### Definition 5 (Konsistenz)

- (i) Das diskrete Problem  $\mathcal{P}_n = (B_n, B_n^0, F_n)$  heißt konsistent mit dem Ausgangsproblem  $\mathcal{P} = (B, B^0, F)$  im Punkt  $x \in B$ , falls

$$\lim_{n \rightarrow \infty} \|F_n(p_n x) - p_n^0 F(x)\|_{B_n^0} = 0, \quad n \in N \tag{9}$$

gilt. Ist  $\mathcal{P}_n$  konsistent mit  $\mathcal{P}$  für alle  $x \in D \subset B$ , so ist das Diskretisierungsverfahren auf  $D$  konsistent.

(ii)  $\mathcal{P}_n$  ist konsistent (in  $x \in B$ ) mit  $\mathcal{P}$  mit der Ordnung  $p \in \mathbb{N}$ , falls

$$\|F_n(p_n x) - p_n^0 F(x)\|_{B_n^0} = \mathcal{O}(n^{-p}), \quad n \in N \quad (10)$$

ist, d.h. falls Konstanten  $n_0 \in N$  und  $M > 0$  existieren, so daß für alle  $n \in N, n \geq n_0$ ,

$$\|F_n(p_n x) - p_n^0 F(x)\|_{B_n^0} \leq M n^{-p}$$

gilt.

Anschaulich bedeutet Konsistenz die asymptotische Vertauschbarkeit von  $F_n$  und  $p_n$  in obigem Approximationsschema. Um die Konsistenz eines Diskretisierungsverfahrens nachzuweisen und seine Konsistenzordnung zu bestimmen, entwickelt man in der Regel die Funktion  $x(t) \in B$  an geeigneten Argumentstellen  $t_j$  in eine *Taylor-Reihe* in Termen von  $n^{-1}$ . Voraussetzung ist dafür jedoch eine hinreichende Glattheit aller beteiligten Funktionen.

**Beispiel 6** Wir betrachten zur Veranschaulichung der Definition die skalaren Anfangswertprobleme 1.Ordnung

$$\dot{x} = f(t, x), \quad x(0) = 0, \quad 0 < t < 1 \quad (11)$$

mit einer als stetig differenzierbar vorausgesetzten Funktion  $f$ . Der Differentialoperator  $F$  lautet dann offenbar

$$Fx(t) \equiv \dot{x}(t) - f(t, x(t)). \quad (12)$$

Wir wenden nun beispielhaft das einfachste Diskretisierungsverfahren zur Lösung von Anfangswertproblemen an. Das *explizite Euler-Verfahren* (*Euler-Cauchy-Verfahren*) lautet

$$u_j = u_{j-1} + hf(t_{j-1}, u_{j-1}), \quad j = 1(1)n \quad (13)$$

mit  $h = 1/n$ ,  $t_j = jh$ . Definiert man den diskreten Operator  $F_n$  dieses Verfahrens durch Umstellung dieser Vorschrift und anschließende Division durch  $h$  zu

$$\{F_n u\}_j \equiv \frac{1}{h}(u_j - u_{j-1}) - f(t_{j-1}, u_{j-1}), \quad j = 1(1)n, \quad (14)$$

so kann man die Konsistenz des expliziten Euler-Verfahrens für alle 2-mal stetig differenzierbaren Funktionen  $x \in B$  leicht nachweisen. Wir definieren dazu die Abbildung  $G_n : B \rightarrow B_n^0$  mit

$$G_n(x) \equiv F_n(p_n x) - p_n^0 F(x), \quad x \in B$$

und erhalten für deren  $j$ -te Komponente mittels Taylor-Entwicklung

$$\begin{aligned} \{G_n x\}_j &= \frac{1}{h} [x(t_j) - x(t_{j-1})] - f(t_{j-1}, x(t_{j-1})) - \dot{x}(t_{j-1}) + f(t_{j-1}, x(t_{j-1})) \\ &= \frac{1}{h} \left[ x(t_{j-1}) + h\dot{x}(t_{j-1}) + \frac{h^2}{2}\ddot{x}(\xi_j) - x(t_{j-1}) \right] - \dot{x}(t_{j-1}) \\ &= \frac{h}{2}\ddot{x}(\xi_j), \quad t_{j-1} \leq \xi_{j-1} \leq t_j, \end{aligned} \quad (15)$$

woraus

$$\|G_n(x)\|_{B_n^0} \leq \frac{h}{2} \max_{0 \leq t \leq 1} |\ddot{x}(t)| = M \cdot n^{-1}$$

mit einer Konstanten  $M > 0$  folgt. Somit ist das Verfahren konsistent mit Konsistenzordnung  $p = 1$ . ◀

Eine besondere Rolle unter allen Elementen  $x \in B$  spielt die vorausgesetzte exakte Lösung  $x^* \in D \subset B$  des Ausgangsproblems  $\mathcal{P}$ .

**Definition 7 (Lokaler Diskretisierungsfehler)**

$x^* \in D \subset B$  sei Lösung des Ausgangsproblems  $F(x) = 0$ . Dann heißt das Element  $\tau \in B_n^0$  mit

$$\tau = F_n(p_n x^*) - p_n^0 F(x^*) = F_n(p_n x^*), \quad n \in N \quad (16)$$

lokaler Diskretisierungsfehler des Verfahrens  $\mathcal{P}_n$ .

**Beispiel 6** Hier erhält man als lokalen Diskretisierungsfehler

$$\tau_j = \frac{1}{h} [x^*(t_j) - x^*(t_{j-1})] - f(t_{j-1}, x^*(t_{j-1})), \quad j = 1(1)n.$$

Umstellung dieser Formel nach dem Wert  $x^*(t_j)$  ergibt die Darstellung

$$x^*(t_j) = x^*(t_{j-1}) + hf(t_{j-1}, x^*(t_{j-1})) + h\tau_j,$$

mit der eine anschauliche Interpretation dieses Fehlers für das explizite Euler-Verfahren möglich wird: Führt man nämlich mit dem Anfangswert  $\eta_{j-1} = x^*(t_{j-1})$  einen einzelnen Verfahrensschritt aus, so liefert

$$\eta_j = x^*(t_{j-1}) + hf(t_{j-1}, x^*(t_{j-1}))$$

genau den Näherungswert des Verfahrens mit der Fehlerdarstellung  $x^*(t_j) = \eta_j + h\tau_j$ . Der Wert  $\tau_j$  stellt folglich in diesem Verfahren den durch die Schrittweite  $h$  dividierten Fehler eines einzelnen Integrationsschrittes (local error per unit step) dar. ◀

Die Konsistenz eines Diskretisierungsverfahrens bedeutet wegen (9), daß der diskrete Operator  $F_n$  den gegebenen Operator  $F$  in der Umgebung der Lösung  $x^*$  approximiert. Falls diese Grundbedingung nicht erfüllt ist, so kann im allgemeinen nicht erwartet werden, daß die Näherungslösungen  $u^*$  – falls sie überhaupt existieren – die exakte Lösung  $x^*$  approximieren. In vielen Fällen ist die Konsistenz allein jedoch nicht hinreichend für die Konvergenz der Näherungslösung, d.h. für

$$\lim_{n \rightarrow \infty} \|u^* - p_n x^*\|_{B_n} = 0.$$

Kleine Störungen wie Rundungsfehler, Abschneidefehler etc. dürfen sich möglichst nicht verstärken und kumulieren. Das Verfahren muß „robust“ oder auch „stabil gegenüber Störungen“ sein. Ersetzt man die rechte Seite des diskreten Problems  $F_n(u) = 0$  durch kleine Änderungen  $\delta^1, \delta^2$ , so lauten nun die gestörten Probleme

$$F_n(u^1) = \delta^1, \quad F_n(u^2) = \delta^2$$

mit den vorerst als existent vorausgesetzten Lösungen  $u^1, u^2 \in B_n$ . Aus der Theorie der algebraischen Gleichungssysteme ist bekannt, daß die Abbildung  $F_n$  stabil ist, wenn kleine Störungen der rechten Seiten zu endlichen Störungen der Lösungen führen. Das ist offenbar garantiert, wenn eine Beziehung

$$\|u^1 - u^2\|_{B_n} \leq S \cdot \|\delta^1 - \delta^2\|_{B_n^0}$$

mit einer Konstanten  $S > 0$  nachgewiesen werden kann. Damit für  $n \rightarrow \infty$  diese Stabilitätsrelation nicht verlorengelht, fordert man die Unabhängigkeit der Konstanten  $S$  von  $n$  und gelangt so zum Begriff der diskreten Stabilität.

**Definition 8 (Diskrete Stabilität)**

$F_n$  sei stetig auf der Menge  $S(u^0, r) = \{u \in B_n \mid \|u - u^0\| \leq r\}$ . Falls für alle  $u^1, u^2 \in S(u^0, r)$

$$\|u^1 - u^2\|_{B_n} \leq S \cdot \|F_n(u^1) - F_n(u^2)\|_{B_n^0} \quad \forall n \in N \quad (17)$$

mit Konstanten  $S > 0, r > 0$  gilt, so ist der Operator  $F_n$  stabil auf  $u^0$  mit der Stabilitätsgrenze  $S$  und der Stabilitätsschwelle  $s = r/S$ .

Für Fréchet-differenzierbare (kurz: F-differenzierbare) Operatoren  $F_n$  kann die unhandliche Bedingung (17) durch eine Bedingung an die Inverse  $[F'_n(u^0)]^{-1}$  ersetzt werden, die mitunter einfacher zu verifizieren ist. Dazu nutzt man den

**Lemma 9**  $F_n$  sei für  $n \in N$  auf den Mengen  $S(u^0, R)$  F-differenzierbar und genüge den Voraussetzungen

- (i)  $F'_n(u^0)^{-1}$  existiert und  $\|F'_n(u^0)^{-1}\| \leq S$  mit  $S > 0$  unabhängig von  $n$ .
- (ii)  $F'_n(u)$  ist gleichmäßig Lipschitz-stetig (kurz: L-stetig) auf  $S(u^0, R)$ , d.h. es existiert eine von  $n$  unabhängige Konstante  $L \geq 0$  mit

$$\|F'_n(u) - F'_n(v)\| \leq L \cdot \|u - v\| \quad \forall u, v \in S(u^0, R).$$

Wählt man  $r = \min[\kappa/(SL), R]$ ,  $\kappa \in (0, 1)$ , so ist  $F_n$  stabil auf  $u^0$  mit Stabilitätsgrenze  $S_0 = S/(1 - SLr)$  und Stabilitätsschwelle  $s_0 = r(1 - SLr)/S$ .

BEWEIS: Seien  $u^1, u^2 \in S(u^0, r) \subset S(u^0, R)$ . Für  $t \in [0, 1]$  ist dann  $u := tu^1 + (1 - t)u^2 \in S(u^0, R)$ . Folglich existiert der beschränkte lineare Operator  $L : B_n \rightarrow B_n^0$  mit

$$L := \int_0^1 F'_n(tu^1 + (1 - t)u^2) dt.$$

Mit den Voraussetzungen (i) und (ii) schätzt man leicht ab

$$\|F'_n(u^0)^{-1}\| \cdot \|L - F'_n(u^0)\| \leq SLr =: q.$$

Da  $q \leq \kappa < 1$  ist, existiert nach dem Störungslemma der inverse Operator  $L^{-1}$  mit der Normabschätzung  $\|L^{-1}\| \leq S_0$  gemäß Definition von  $S_0$ . Mit dem Mittelwertsatz gilt desweiteren

$$F_n(u^1) - F_n(u^2) = L(u^1 - u^2),$$

woraus nach Anwendung von  $L^{-1}$  und Abschätzung

$$\|u^1 - u^2\| \leq S_0 \cdot \|F_n(u^1) - F_n(u^2)\|$$

folgt.  $F_n$  ist stetig auf  $S(u^0, r)$ ; die Stabilitätsschwelle ergibt sich zu  $s_0 := r/S_0 = r(1 - SLr)/S$ . Damit ist Definition 8 gültig.  $\square$

Der Nachweis der diskreten Stabilität eines Verfahrens ist im Gegensatz zum Konsistenzbeweis oft aufwendig und soll deshalb hier beispielhaft für das explizite Euler-Verfahren, angewandt auf das Anfangswertproblem (11)

$$\dot{x} = f(t, x), \quad x(0) = 0, \quad 0 < t < 1,$$

geführt werden.

**Beispiel 6** Die Funktion  $f$  der rechten Seite wird weiterhin als stetig differenzierbar vorausgesetzt (Hier wäre auch die L-Stetigkeit hinreichend). Wir weisen die Gültigkeit der Stabilitätsdefinition mit den B-Räumen  $B_n$  und  $B_n^0$  nach. Setzen wir dazu  $u^1, u^2 \in S(u^0, r) \subset B_n$ ,  $d := F_n(u^1) - F_n(u^2)$ ,  $e := u^1 - u^2$  an. Mit der Definition des Operators  $F_n$  erhalten wir nach Umstellung für die  $j$ -ten Komponenten der Gitterfunktionen  $u^1, u^2$

$$u_j^k = u_{j-1}^k + hf(t_{j-1}, u_{j-1}^k) + h\{F_n(u^k)\}_j, \quad k = 1, 2.$$

Subtraktion  $u^1 - u^2$  liefert mit dem Mittelwertsatz

$$\begin{aligned} e_j &= e_{j-1} + h[f(t_{j-1}, u_{j-1}^1) - f(t_{j-1}, u_{j-1}^2)] + hd_j \\ &= e_{j-1} + hf_x(t_{j-1}, \eta)(u_{j-1}^1 - u_{j-1}^2) + hd_j, \quad \eta \in S(u^0, r), \end{aligned}$$

woraus wegen der Stetigkeit von  $f_x(t, x)$  auf  $S(u^0, r)$  die Existenz einer Konstanten  $L > 0$  folgt, mit der

$$\begin{aligned} |e_j| &\leq |e_{j-1}| + hL|e_{j-1}| + h|d_j| \\ &\leq (1 + hL)|e_{j-1}| + h \cdot \max_{k=1(1)n} |d_k| \end{aligned}$$

gilt. Setzen wir rekursiv ein und beachten die Tatsache, daß wegen  $u^1 = u^2 = 0$  auch  $e^0 = 0$  gilt, so erhalten wir

$$\begin{aligned} |e_{j-1}| &\leq h \sum_{k=1}^j (1 + hL)^{j-k} \cdot \|d\|_{B_n^0} \\ &= h \frac{(1 + hL)^j - 1}{(1 + hL) - 1} \cdot \|d\|_{B_n^0} \\ &\leq \frac{1}{L} \cdot e^{j(hL)} \cdot \|d\|_{B_n^0}, \quad jh < 1 \\ &\leq \frac{1}{L} \cdot e^L \cdot \|d\|_{B_n^0}, \quad j = 1(1)n, \quad \text{also} \\ \max_{j=0(1)n} |e_j| &\leq \frac{1}{L} \cdot e^L \cdot \|d\|_{B_n^0}. \end{aligned}$$

Dabei wurde von der Eigenschaft der Exponentialfunktion  $1 + x \leq e^x$ ,  $\forall x \in \mathbb{R}$ , Gebrauch gemacht. Da in der Norm des B-Raumes  $B_n$  auch  $\max_j |\partial u_j|$  benötigt wird, notieren wir mit der Definition des Operators  $F_n$  unmittelbar

$$\partial u_j^k := \frac{1}{h}(u_j^k - u_{j-1}^k) = f(t_{j-1}, u_{j-1}^k) + \{F_n(u^k)\}_j, \quad k = 1, 2.$$

Subtraktion und Abschätzung liefert mit dem Mittelwertsatz

$$\begin{aligned}
 \partial e_j &= f(t_{j-1}, u_{j-1}^1) - f(t_{j-1}, u_{j-1}^2) + d_j \\
 |\partial e_j| &\leq L|e_{j-1}| + |d_j| \\
 &\leq L \cdot \max_{j=0(1)n} |e_j| + \max_{j=1(1)n} |d_j| \\
 &\leq L \cdot \left( \frac{1}{L} \cdot e^L \cdot \|d\|_{B_n^0} \right) + \|d\|_{B_n^0}, \quad \text{also} \\
 \max_{j=1(1)n} |\partial e_j| &\leq (e^L + 1) \cdot \|d\|_{B_n^0}.
 \end{aligned}$$

Zusammenfassung der beiden Abschätzungen ergibt für die  $B_n$ -Norm schließlich

$$\begin{aligned}
 \|e\|_{B_n} &= \max_{j=0(1)n} |e_j| + \max_{j=1(1)n} |\partial e_j| \\
 &\leq \left( e^L + \frac{1}{L} e^L + 1 \right) \cdot \|d\|_{B_n^0} \quad \text{bzw.} \\
 \|u^1 - u^2\|_{B_n} &\leq S \cdot \|F_n(u^1) - F_n(u^2)\|_{B_n^0},
 \end{aligned}$$

also ist das explizite Euler-Verfahren diskret stabil mit Stabilitätsgrenze  $S = e^L + \frac{1}{L}e^L + 1$  und beliebigem festen Wert  $r > 0$ . ◀

### 3 Existenz und Konvergenz diskreter Lösungen

Wir nehmen nun vereinfachend an, daß die diskrete Aufgabe  $\mathbb{P}_n$  eine Lösung  $u_n^*$  für  $n \in N$  besitzt. Um die Konvergenz dieser Näherungslösungen gegen eine exakte Lösung  $x^*$  zu beschreiben, führen wir die folgenden Begriffe ein.

#### Definition 10 (Diskrete Konvergenz)

$x^* \in D \subset B$  und  $u_n^* \in B_n$  seien Lösungen der Probleme  $\mathbb{P}$  bzw.  $\mathbb{P}_n$ .

(i) Die Größe  $e_n := u_n^* - p_n x^*$  in  $B_n$  nennt man globalen Diskretisierungsfehler des Verfahrens  $\mathcal{P}_n$ .

(ii) Das Verfahren  $\mathcal{P}_n = (B_n, B_n^0, F_n)$  konvergiert diskret gegen  $\mathcal{P} = (B, B^0, F)$ , falls

$$\lim_{n \rightarrow \infty} \|e_n\|_{B_n} = \lim_{n \rightarrow \infty} \|u_n^* - p_n x^*\|_{B_n} = 0, \quad n \in N \quad (18)$$

*gilt.*

(iii)  $\mathcal{P}_n$  konvergiert diskret gegen  $\mathcal{P}$  mit der Ordnung  $p \in \mathbb{N}$ , falls

$$\|u_n^* - p_n x^*\|_{B_n} = \mathcal{O}(n^{-p}), \quad n \in N \quad (19)$$

*ist.*

Beziehung (19) schreibt man dann häufig in der Form  $u_n^* = p_n x^* + \mathcal{O}(n^{-p})$  und sagt, daß  $u_n^*$  mit Ordnung  $p$  gegen  $x^*$  konvergiert.

Kann man die Existenz einer Näherungslösung  $u_n^*$  des Problems  $\mathcal{P}_n$  voraussetzen, so lassen sich hinreichende Bedingungen für deren diskrete Konvergenz angeben.

**Satz 11 (Konvergenz)**

Mit der Konstanten  $P$  aus (8) seien folgende Voraussetzungen erfüllt:

- (i)  $x^* \in D \subset B$  ist Lösung von  $F(x) = 0$ .
- (ii) Für  $n \in \mathbb{N}$  existiert ein  $R > 0$ , so daß die diskrete Gleichung  $F_n(u) = 0$  in der Kugelmenge  $S(p_n x^*, PR)$  eine Lösung  $u_n^*$  besitzt.
- (iii)  $F_n$  ist stabil auf  $p_n x^*$  in der Kugel  $S(p_n x^*, PR)$ .
- (iv)  $F_n$  ist konsistent mit  $F$  auf  $x^*$ .

Dann konvergiert  $u_n^*$  diskret gegen  $x^*$ .

BEWEIS: Nach Voraussetzung (iii) schätzt man mit den Lösungen  $u_n^*$  und  $x^*$  ab

$$\begin{aligned} \|u_n^* - p_n x^*\|_{B_n} &\leq S \cdot \|F_n(u_n^*) - F_n(p_n x^*)\|_{B_n^0} \\ &= S \cdot \|F_n(p_n x^*) - p_n^0 F(x^*)\|_{B_n^0}, \quad \text{d.h.} \\ \|e_n\|_{B_n} &\leq S \cdot \|\tau\|_{B_n^0}. \end{aligned} \quad (20)$$

Mit Voraussetzung (iv) folgt für  $n \rightarrow \infty$  unmittelbar die Konvergenz mit  $\|e_n\| \rightarrow 0$ .  $\square$

Die Aussage dieses grundlegenden Satzes der Numerischen Mathematik läßt sich in der einfachen Formel

**Konsistenz + Stabilität  $\implies$  Konvergenz**

zusammenfassen. Voraussetzung für ihre Gültigkeit ist jedoch, daß alle betrachteten Lösungen existieren. Kann man zusätzlich nachweisen, daß der Operator  $F_n$  eine bestimmte Konsistenzordnung  $p \in \mathbb{N}$  hat, so folgt aus Ungleichung (20) mit einer Konstanten  $M > 0$  unmittelbar

$$\|e_n\|_{B_n} \leq S \cdot \|\tau\| \leq S \cdot M \cdot n^{-p},$$

so daß das Verfahren auch dieselbe *Konvergenzordnung*  $p$  besitzt.

**Beispiel 6** Das betrachtete Anfangswertproblem besitze die eindeutige Lösung  $x^*(t)$ , und die rechte Seite  $f$  sei stetig differenzierbar. Da das explizite Euler-Verfahren unter diesen Voraussetzungen stabil und konsistent mit Ordnung 1 ist, liefert der Konvergenzsatz für die Näherungslösung  $u_n^*$  deren Konvergenz gegen  $x^*(t)$  mit Ordnung 1. Für die globalen Diskretisierungsfehler  $e_j = u_j^* - x^*(t_j)$  erhält man so die qualitative Aussage

$$\max_{j=0(1)n} |e_j| + \max_{j=1(1)n} \left| \frac{1}{h} (e_j - e_{j-1}) \right| \leq C \cdot h$$

mit einer Konstanten  $C > 0$ . ◀

Im Gegensatz zum expliziten Euler-Verfahren für Anfangswertprobleme

$$u_j = u_{j-1} + hf(t_{j-1}, u_{j-1}), \quad j = 1(1)n,$$

bei dem die diskrete Lösung  $u^*$  stets existiert, falls nur die rechte Seite  $f$  definiert ist, muß dies bei impliziten Verfahren nicht der Fall sein. Hier ist der Näherungswert  $u_j$  in jedem Schritt  $j$  als Lösung einer im allgemeinen nichtlinearen Gleichung zu ermitteln, z.B. beim *impliziten Euler-Verfahren*

$$u_j = u_{j-1} + hf(t_j, u_j), \quad j = 1(1)n.$$

Noch diffiziler stellt sich die Situation bei der Lösung von Randwertproblemen (2) dar. Hier muß die Frage beantwortet werden, unter welchen Voraussetzungen das diskrete Problem  $\mathbb{P}_n$  – zumindest für hinreichend großen Parameter  $n$  – eine Lösung  $u^*$  besitzt. Für das Ausgangsproblem  $\mathbb{P}$  ist in der Regel eine Lösung  $x^*$  vorauszusetzen, die in einem gewissen Sinne „isoliert“ sein muß, also geometrisch getrennt von eventuell weiteren Lösungen. Für Fréchet-differenzierbare Operatoren  $F$  mit regulärer Lösung läßt sich eine derartige Voraussetzung leicht formulieren.

### Definition 12 (Reguläre Lösung)

Eine Lösung  $x^* \in D \subset B$  der Gleichung  $f(x) = 0$  heißt regulär (isoliert), falls die Fréchet-Ableitung  $L = F'(x^*)$  existiert und einen beschränkten inversen Operator  $L^{-1}$  besitzt.

Eine reguläre Lösung  $x^*$  ist auch geometrisch isoliert, d.h. sie besitzt eine Umgebung, in der keine weitere Lösung existiert. Für derartige reguläre Lösungen liefert der folgende Satz neben der Konvergenzaussage auch eine lokale Existenz- und Eindeutigkeitsaussage.

### Satz 13 (Existenz, Eindeutigkeit und Konvergenz)

Mit der Konstanten  $P$  aus (8) seien folgende Voraussetzungen erfüllt:

- (i)  $F(x) = 0$  besitze eine reguläre Lösung  $x^*$ , die lokal eindeutig in der Kugelumgebung  $S(x^*, R) \subset D$  ist.
- (ii)  $F$  und  $F_n$  sind  $F$ -differenzierbar auf  $S(x^*, R) \subset D$  bzw.  $S(p_n x^*, PR) \subset B_n$ .
- (iii)  $F'_n(p_n x^*)$  ist regulär, und es existiert eine von  $n$  unabhängige Konstante  $S > 0$ , so daß

$$\|F'_n(p_n x^*)^{-1}\| \leq S \quad \forall n \in N \quad \text{gilt.}$$

- (iv)  $F'_n(u)$  ist gleichmäßig  $L$ -stetig auf  $S(p_n x^*, PR) \subset B_n$ , d.h. es existiert eine Konstante  $L \geq 0$ , so daß

$$\|F'_n(u) - F'_n(v)\| \leq L\|u - v\|$$

für alle  $u, v \in S(p_n x^*, PR)$  gilt.

- (v)  $F_n$  ist konsistent mit  $F$  auf  $x^*$ .

Dann ist  $F_n$  stabil auf  $p_n x^*$ , und es existieren Konstanten  $n_0 \in \mathbb{N}$  und  $r \in (0, R]$ , so daß für alle  $n \in N$ ,  $n \geq n_0$ , gilt:

(i) Die diskrete Gleichung  $F_n(u) = 0$  besitzt in der Kugelmenge  $S(p_n x^*, Pr)$  eine eindeutige Lösung  $u_n^*$ .

(ii)  $u_n^*$  konvergiert diskret gegen  $x^*$ .

BEWEIS: Mit den Voraussetzungen (ii), (iii) und (iv) ist die Stabilität von  $F_n$  auf  $u = p_n x^*$  wegen der Gültigkeit des Lemmas 9 garantiert. Stabilitätsschranke  $S_0$  und Stabilitätsschwelle  $r$  nehmen dabei die Werte

$$S_0 := \frac{S}{1 - SLP_r} \quad \text{und} \quad r := \min \left( \frac{\kappa}{SLP}, R \right), \quad \kappa \in (0, 1)$$

an. Für  $u^1, u^2 \in S(p_n x^*, Pr)$  gilt zudem die Stabilitätsungleichung (17) mit Stabilitätskonstante  $S_0$ .

Wir zeigen nun die Behauptung (i) und definieren dazu die Operatoren  $G_n : B_n \rightarrow B_n$  mit

$$G_n u \equiv u - F'_n(p_n x^*)^{-1} F_n u.$$

Man beweist leicht die gleichmäßige Kontraktivität von  $G_n$ . Denn für  $u, v \in S_r := S(p_n x^*, Pr)$  ergibt sich

$$\begin{aligned} \|G_n u - G_n v\| &= \|F'_n(p_n x^*)^{-1} [F'_n(p_n x^*)(u - v) - F_n u + F_n v]\| \\ &\leq S \frac{L}{2} (\|u - p_n x^*\| + \|v - p_n x^*\|) \cdot \|u - v\| \\ &\leq SLP_r \cdot \|u - v\| \\ \|G_n u - G_n v\| &\leq \kappa \cdot \|u - v\| \end{aligned} \quad (21)$$

mit  $\kappa \in (0, 1)$ , unabhängig von  $n \in \mathbb{N}$ . Zeigen wir nun, daß  $G_n$  die Kugel  $S_r$  in sich abbildet. Unter Benutzung von (21) und Voraussetzung (iii) erhält man für  $u \in S_r$  mit obiger Konstante  $\kappa \in (0, 1)$

$$\begin{aligned} \|G_n u - p_n x^*\| &\leq \|G_n u - G_n p_n x^*\| + \|G_n p_n x^* - p_n x^*\| \\ &\leq \kappa \|u - p_n x^*\| + S \|F_n p_n x^*\| \\ &\leq \kappa Pr + S \tau_n. \end{aligned}$$

Nach Voraussetzung (v) existiert ein  $n_0 \in \mathbb{N}$ , so daß für alle  $n \geq n_0$ ,  $n \in \mathbb{N}$  für den Diskretisierungsfehler  $\tau_n \leq (1 - \kappa)Pr/S$  gilt. Mit dieser Abschätzung erhalten wir schließlich aus obiger Ungleichung  $\|G_n u - p_n x^*\| \leq Pr$ , d.h.  $G_n u \in S_r$ . Nach dem Banachschen Fixpunktsatz existiert damit für jedes  $n \in \mathbb{N}$  mit  $n \geq n_0$  genau ein  $u^* \in S(p_n x^*, Pr)$  mit  $G_n u^* = u^*$ . Wegen der Regularität von  $F'_n(p_n x^*)$  ist die Fixpunktgleichung  $G_n u = u$  äquivalent zur Gleichung  $F_n u = 0$ , womit Behauptung (i) folgt.

Behauptung (ii) erhält man mit Voraussetzung (v) und Stabilitätsungleichung (17) für  $u^* \in S(p_n x^*, Pr)$ , woraus die Abschätzung

$$\|u - p_n x^*\| \leq S_0 \tau_n, \quad (22)$$

folgt und wegen  $\tau_n \rightarrow 0$  ( $n \rightarrow \infty$ ) die diskrete Konvergenz von  $u^*$ .  $\square$

Auch für diesen Satz ist leicht nachweisbar: Wenn der Operator  $F_n$  eine bestimmte Konsistenzordnung  $p \in \mathbb{N}$  hat, so besitzt das Verfahren bei entsprechend vorausgesetzter Glattheit auch die *Konvergenzordnung*  $p$ .

## 4 Zweipunkt-Randwertprobleme

Eine wesentliche Anwendung der eingeführten Begriffe stellen nichtlineare Randwertprobleme für gewöhnliche Differentialgleichungen dar. Wir beschränken uns auf die Klasse der Zweipunktprobleme mit allgemeinen Randbedingungen auf vorgegebenem endlichen Intervall  $I = [a, b]$

$$\dot{y} = f(t, y), \quad g(y(a), y(b)) = 0 \quad (23)$$

mit  $f : \mathbb{R} \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $g : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ . Zuerst notieren wir das Randwertproblem als Operatorgleichung  $F(y) = 0$  mit einem geeigneten Operator  $F$ , der neben den Differentialgleichungen auch die Randbedingungen enthält. Mit einer beliebigen, fest gewählten Norm  $\|\cdot\|_m$  in  $\mathbb{R}^m$  definiert man dazu die B-Räume  $B = C^1(I; \mathbb{R}^m)$  und  $B^0 = \mathbb{R}^m \times C(I; \mathbb{R}^m)$  zweckmäßig mit folgenden Normen:

$$\begin{aligned} \|y\|_B &:= \max_{t \in I} \|y(t)\|_m + \max_{t \in I} \|\dot{y}(t)\|_m, \\ \|z\|_{B^0} &:= \|z_0\|_m + \max_{t \in I} \|z_1(t)\|_m, \end{aligned} \quad (24)$$

wobei  $z = (z_0, z_1)^T \in B^0$  mit  $z_0 \in \mathbb{R}^m$  und  $z_1 \in C(I; \mathbb{R}^m)$  ist. Dann kann Problem (23) als Operatorgleichung

$$F(y) = 0, \quad y \in B \quad (25)$$

geschrieben werden, wenn  $F : B \rightarrow B^0$  durch

$$Fy(t) := \begin{pmatrix} g(y(a), y(b)) \\ \dot{y}(t) - f(t, y(t)) \end{pmatrix} \quad (26)$$

definiert wird. Auf  $I = [a, b]$  konstruieren wir zu jeder natürlichen Zahl  $n \in \mathbb{N}$  (d.h. die Diskretisierungsmenge  $N$  aus Abschnitt 1 ist gleich  $\mathbb{N}$ ) ein Gitter

$$I_n := \{t_j \in I \mid t_0 = a, t_n = b, t_j = t_{j-1} + h_j, j = 1(1)n\} \quad (27)$$

mit den Schritten  $h_j > 0, j = 1(1)n$ , und dem Maximalschritt

$$h := \max_{j=1(1)n} h_j. \quad (28)$$

Wir wollen des weiteren eine lineare Konvergenz der Gitterfolge  $I_n, n \in \mathbb{N}$ , fordern und geben dazu

**Definition 14** Die Gitterfolge  $\{I_n\}, n \in \mathbb{N}$ , konvergiert linear, wenn Konstanten  $c_1 > 0, c_2 > 0$  existieren, für die mit den Schritten  $h_j$  jedes Gitters  $I_n$  die Bedingung

$$c_1/n \leq h_j \leq c_2/n, \quad j = 1(1)n, \quad (29)$$

für alle  $n \in \mathbb{N}$  erfüllt ist.

Offenbar sind dann die Terme  $O(h_j^p)$ ,  $O(h^p)$  und  $O(n^{-p})$  asymptotisch äquivalent, weshalb Fehlerordnungen stets durch  $O(h^p)$  - wie dies für Differenzenverfahren üblich ist - angegeben werden sollen.

Zur Formulierung der diskreten Aufgabe  $F_n u = 0$  definieren wir diskrete Analoga  $B_n$  und  $B_n^0$ , die Räume von Gitterfunktionen darstellen. Seien  $B_n = C_n^1(I_n; \mathbb{R}^m)$  und  $B_n^0 = \mathbb{R}^m \times C_n(I_n; \mathbb{R}^m)$  derartige Räume von Gitterfunktionen  $u : I_n \rightarrow \mathbb{R}^m$  bzw.  $v = (v_0, v_1, \dots, v_m)^T$  mit  $v_j \in \mathbb{R}^m$ ,  $j = 0(1)n$ . Die Normen sind diskrete Analoga zu (24)

$$\begin{aligned} \|u\|_{B_n} &:= \max_{j=0(1)n} \|u_j\|_m + \max_{j=1(1)n} \|\partial u_j\|_m, \\ \|v\|_{B_n^0} &:= \|v_0\|_m + \max_{j=1(1)n} \|v_j\|_m, \end{aligned} \quad (30)$$

wobei  $\partial u_j = (u_j - u_{j-1})/h_j$  ist. Für die Restriktionsoperatoren  $p_n : B \rightarrow B_n$  und  $p_n^0 : B^0 \rightarrow B_n^0$  benutzt man im Hinblick auf die Definition von  $F_n$

$$\{p_n y\}_j := y(t_j), \quad j = 0(1)n \quad \text{und} \quad (31)$$

$$\{p_n^0 z\}_j := \begin{cases} z_0 & , \quad j = 0 \\ z_1(t_j - h_j/2) & , \quad j = 1(1)n. \end{cases} \quad (32)$$

Offensichtlich sind  $p_n$  und  $p_n^0$  normkonsistent und erfüllen wegen  $\|p_n\| \geq 1$  und  $\|p_n^0\|_0 \geq 1$  die Beziehung (8) mit  $P = P^0 = 1$ . Eine andere Wahl von  $p_n^0$  wäre

$$\{p_n^0 z\}_j := \begin{cases} z_0 & , \quad j = 0 \\ \frac{1}{2}[z_1(t_j) + z_1(t_{j-1})] & , \quad j = 1(1)n, \end{cases} \quad (33)$$

die offenbar denselben Beziehungen mit  $P = P^0 = 1$  genügt.

Zur Definition der Differenzenoperatoren  $F_n$  empfehlen sich Einschrittformeln, um spezielle Approximationen in Randnähe zu vermeiden und den Aufwand zur Lösung der entstehenden finiten Gleichungssysteme niedrig zu halten. Unter den Einschrittverfahren bieten sich für (26) wegen ihrer Konsistenzordnung 2 die *Trapezregel*

$$\{F_n u\}_j := \begin{cases} g(u_0, u_n) & , \quad j = 0 \\ \frac{1}{h_j}(u_j - u_{j-1}) - \frac{1}{2}(f(t_j, u_j) + f(t_{j-1}, u_{j-1})) & , \quad j = 1(1)n \end{cases} \quad (34)$$

und die *Mittelpunktregel*

$$\{F_n u\}_j := \begin{cases} g(u_0, u_n) & , \quad j = 0 \\ \frac{1}{h_j}(u_j - u_{j-1}) - f\left(\frac{t_j+t_{j-1}}{2}, \frac{u_j+u_{j-1}}{2}\right) & , \quad j = 1(1)n \end{cases} \quad (35)$$

an. Während KELLER [4] ursprünglich (35) benutzt, bauen LENTINI & PEREYRA [6] ihre Verfahren auf der Trapezregel (34) auf, die einfachere asymptotische Entwicklungen besitzt. In KELLER [5] wird eine vollständige Stabilitätstheorie für Differenzenverfahren zur Lösung von (23) gegeben, mit der die Stabilität von (34) und (35) im Sinne der Maximumnorm  $\|u\|_{B_n} = \max \|u_j\|_m$ ,  $j = 0(1)n$ , gezeigt wird. Da des weiteren jedoch die diskrete  $C^1$ -Norm (30) benutzt wird, geben wir folgendes

**Lemma 15** Besitze (23) die Lösung  $y^* \in B$ , und sei  $f \in C^1(I \times \mathbb{R}^m)$ , so gilt für alle  $u^1, u^2 \in S(p_n y^*, R)$  mit  $R > 0$

$$\max_{j=1(1)n} \|\partial(u_j^1 - u_j^2)\|_m \leq \|F_n u^1 - F_n u^2\|_{B_n^0} + K_0 \cdot \max_{j=0(1)n} \|u_j^1 - u_j^2\|_m, \quad (36)$$

wobei  $K_0 > 0$  konstant und  $F_n$  durch (34) oder (35) gegeben ist.

BEWEIS: Für (34) hat man mit  $j \in \{1, \dots, n\}$  und  $f \in C^1$  die Abschätzung

$$\|\partial(u_j^1 - u_j^2)\|_m \leq \|\{F_n u^1\}_j - \{F_n u^2\}_j\|_m + \frac{1}{2} \cdot K_0 \|u_j^1 - u_j^2\|_m + \frac{1}{2} \cdot K_0 \|u_{j-1}^1 - u_{j-1}^2\|_m,$$

woraus unmittelbar die Behauptung folgt. Analoges gilt für (35).  $\square$

Kann man nachweisen, daß die Aufgabe (23) eine isolierte Lösung  $y^* \in B$  besitzt, d.h. das homogene *Variationssystem*

$$\begin{aligned} \dot{e}(t) - \partial_2 f(t, y^*(t))e(t) &= 0 \\ \partial_1 g(y^*(a), y^*(b))e(a) + \partial_2 g(y^*(a), y^*(b))e(b) &= 0 \end{aligned} \quad (37)$$

nur die Lösung  $e(t) \equiv 0$  in  $B$  hat, so erhält man folgende Konsistenz- und Stabilitätsaussage:

**Lemma 16** Sei  $f \in C^3(I \times \mathbb{R}^m)$ ,  $g \in C^3(\mathbb{R}^m \times \mathbb{R}^m)$  und  $y^* \in B$  eine isolierte Lösung von (23). Dann existieren Konstanten  $h_0 > 0$  und  $R > 0$ , so daß für alle Gitter  $I_n$  mit  $0 < h \leq h_0$  gilt:

$$\|F_n p_n y^* - p_n^0 F y^*\|_{B_n^0} \leq C \cdot h^2 \quad \text{und} \quad (38)$$

$$\|u^1 - u^2\|_{B_n} \leq S \cdot \|F_n u^1 - F_n u^2\|_{B_n^0}, \quad u^1, u^2 \in S(p_n y^*, R) \quad (39)$$

mit Konstanten  $C > 0, S > 0$ , wobei  $F_n$  (34) oder (35) und  $p_n^0$  (31) oder (32) genügt.

BEWEIS: Behauptung (i) folgt durch Taylorabgleich unmittelbar. Für (35) zeigt damit KELLER [4] mit einer Stabilitätskonstanten  $S_0$ :

$$\max_{j=0(1)n} \|u_j^1 - u_j^2\| \leq S_0 \cdot \|F_n u^1 - F_n u^2\|_{B_n^0},$$

woraus mit Lemma 15 die  $(B_n, B_n^0)$ -Stabilität folgt. Für (34) verläuft der Beweis analog.  $\square$

Beide Differenzenverfahren sind also konsistent mit Ordnung 2 und zudem stabil in den Räumen  $B_n$  und  $B_n^0$ . Um die Existenz der diskreten Lösungen  $u^*$  und deren diskrete Konvergenz gegen die exakte Lösung  $y^*$  nachzuweisen, wenden wir nun den allgemeinen Satz 13 an:

**Satz 17** Sei  $f \in C^3(I \times \mathbb{R}^m)$ ,  $g \in C^3(\mathbb{R}^m \times \mathbb{R}^m)$  und  $y^* \in B$  eine isolierte Lösung von (23). Dann existieren Konstanten  $h_0 > 0$  und  $R > 0$ , so daß für alle Gitter  $I_n$  mit  $0 < h \leq h_0$  gilt:

(i) Das finite Gleichungssystem

$$\begin{aligned} g(u_0, u_n) &= 0, \quad j = 0 \\ \frac{1}{h_j}(u_j - u_{j-1}) - \frac{1}{2}(f(t_j, u_j) + f(t_{j-1}, u_{j-1})) &= 0, \quad j = 1(1)n \end{aligned} \quad (40)$$

besitzt eine eindeutige Lösung  $u^*$  in der Menge  $\|u - p_n y^*\|_{B_n} \leq R$ .

(ii)  $u^*$  konvergiert diskret gegen  $y^*$  mit  $u^* - p_n y^* = O(h^2)$ .

BEWEIS: Die Voraussetzungen (i) - (v) des Satzes 13 sind zu verifizieren. (i) und (ii) sind wegen der Annahmen erfüllt; (v) folgt aus Lemma 16. Die gleichmäßige L-Stetigkeit (iv) von  $F'_n(u)$  überprüft man leicht wegen der Annahme  $f \in C^3(I \times \mathbb{R}^m)$ ,  $g \in C^3(\mathbb{R}^m \times \mathbb{R}^m)$ . Es bleibt Voraussetzung (iii) nachzuweisen. Wir wenden die Stabilitätsungleichung (38) des Lemmas 16 auf die Werte  $u^1 = p_n y^* + v$  und  $u^2 = p_n y^*$  an

$$\begin{aligned} \|v\|_{B_n} &\leq S \cdot \|F_n(p_n y^* + v) - F_n(p_n y^*)\| \\ &= S \cdot \left\| \int_0^1 F'_n(p_n y^* + tv)v dt \right\| \\ &\leq S \cdot \|Lv\|, \end{aligned}$$

wobei  $L$  den Operator

$$L \equiv \int_0^1 F'_n(p_n y^* + tv) dt$$

bezeichnet. Damit existiert der inverse Operator und es gilt

$$\|L^{-1}\| \leq \left\| \left[ \int_0^1 F'_n(p_n y^* + tv) dt \right]^{-1} \right\| \leq S, \quad \forall \|v\| < R,$$

woraus für  $v = 0$  die Voraussetzung (iii) mit  $\|[F'_n(p_n y^*)]^{-1}\| \leq S$  folgt.  $\square$

Mittels des Lemmas 16 läßt sich analog nachweisen, daß die Behauptungen des Satzes 17 ebenfalls für die Mittelpunkregel

$$\begin{aligned} g(u_0, u_n) &= 0, \quad j = 0 \\ \frac{1}{h_j}(u_j - u_{j-1}) - f\left(\frac{t_j + t_{j-1}}{2}, \frac{u_j + u_{j-1}}{2}\right) &= 0, \quad j = 1(1)n \end{aligned} \quad (41)$$

gelten.

## 5 Quasilineare Systeme auf dem 2-Torus

Die Approximation invarianter Tori nichtlinearer dynamischer Systeme (vgl. [8]) führt auf Systeme quasilinearer partieller Differentialgleichungen auf dem Standardtorus  $\mathbb{T}^2$ . Wir betrachten vereinfachend als Grundgebiet den 2-dimensionalen Standardtorus

$$\mathbb{T}^2 = \{\theta \mid \theta = (\theta_1, \theta_2), \theta_i = \mathbb{R} \bmod 2\pi, i = 1, 2\} \quad (42)$$

und eine darauf definierte Funktion  $u : \mathbb{T}^2 \rightarrow \mathbb{R}^q$ . Die Ermittlung eines durch  $u(\theta) = (u_1(\theta), u_2(\theta))$  parametrisierten invarianten Torus führt auf die Torusgleichung

$$\omega_1(\theta, u) \frac{\partial u}{\partial \theta_1} + \omega_2(\theta, u) \frac{\partial u}{\partial \theta_2} = f(\theta, u), \quad \theta \in \mathbb{T}^2 \quad (43)$$

mit  $\omega_i : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}$  und  $f : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}^q$ . Dieses quasilineare System mit gleichem Hauptteil wird nun unter der Voraussetzung  $\omega_2(\theta, u) \neq 0$  auf  $\mathbb{T}^2 \times \mathbb{R}^q$  als zeitabhängiges Problem in  $(\theta_1, \theta_2) = (\theta, t) \in \mathbb{T}^2$  behandelt und mit stabilen Differenzenverfahren 1. Ordnung gelöst. Um hinreichende Genauigkeit zu erzielen, ist allerdings auf einem feinen Gitter

$$\mathbb{T}_h^2 = \{(\theta_j, t_h) \mid t_h = n \cdot \tau, \theta_j = j \cdot h, n = 0(1)N, j = 0(1)J\} \quad (44)$$

mit Schrittweiten  $\tau = 2\pi/N$  und  $h = 2\pi/J$  zu approximieren.

Sei  $B^0$  der Banachraum  $C^0(\mathbb{T}^2, \mathbb{R}^q)$  der auf  $\mathbb{T}^2$  stetigen Funktionen  $v$  mit der Norm  $\|v\|_0 = \max_{\mathbb{T}^2} \|v(t, \theta)\|_\infty$ . Wir betrachten nachfolgend die auf  $\mathbb{T}^2$  stetig differenzierbaren Funktionen (die damit bezüglich  $\theta$  und  $t$   $2\pi$ -periodisch sind), und definieren mit der Norm  $\|u\|_1 = \max\{\|u\|_0, \|u_t\|_0, \|u_\theta\|_0\}$  den Banachraum

$$B = \{u \mid u \in C^1(\mathbb{T}^2, \mathbb{R}^q), u(\theta, t) = u(\theta + 2\pi, t) = u(\theta, t + 2\pi), (\theta, t) \in \mathbb{T}^2\}. \quad (45)$$

Das Torusproblem (43) kann unter der getroffenen Voraussetzung mit  $u \in B$  nunmehr

$$\frac{\partial u}{\partial t} + \omega(\theta, t, u) \frac{\partial u}{\partial \theta} = f(\theta, t, u) \quad (46)$$

notiert werden, wobei die Funktionen  $\omega : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}$  und  $f : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}^q$  als hinreichend glatt vorausgesetzt werden. Definiert man den Operator  $F : B \rightarrow B^0$  durch

$$(Fu)(\theta, t) \equiv u_t(\theta, t) + \omega(\theta, t, u(\theta, t))u_\theta(\theta, t) - f(\theta, t, u(\theta, t)), \quad (47)$$

so lautet das Problem in Operatorschreibweise

$$Fu = 0, \quad u \in B. \quad (48)$$

Um die zu untersuchenden Differenzenverfahren in eine analoge Operatorform zu überführen, betrachten wir auf dem diskretisierten Torus  $\mathbb{T}_h^2$  gemäß (44) mit der Schrittweitenbedingung

$$\lambda := \tau/h = \text{const} \quad (49)$$

entsprechende Banach-Räume  $B_h$  und  $B_h^0$  von Gitterfunktionen  $u_h = \{u_j^n\}$  mit  $j = 0(1)J - 1$ ,  $n = 0(1)N - 1$ . Dabei approximiere  $u_j^n \sim u(\theta_j, t_n)$  auf  $\mathbb{T}_h^2$ .

Mit der Maximumnorm  $\|u_j^n\|_\infty$  des  $\mathbb{R}^q$  bezeichne

$$\|u_h\|_0 = \max_{\mathbb{T}_h^2} \|u_j^n\|_\infty$$

die diskrete  $C$ -Norm und  $B_h^0 = C_h^0(\mathbb{T}_h^2, \mathbb{R}^q)$  den entsprechenden Banach-Raum. Bezeichnet man mit  $\partial_t u$  und  $\partial_\theta u$  die Differenzenquotienten

$$\begin{aligned} \{\partial_t u_h\}_j^n &= \frac{1}{\tau}(u_j^n - u_j^{n-1}), \\ \{\partial_\theta u_h\}_j^n &= \frac{1}{h}(u_j^n - u_{j-1}^n), \end{aligned} \quad (50)$$

so erhält man die diskrete  $C^1$ -Norm

$$\|u_h\|_1 = \max\{\|u_h\|_0, \|\partial_t u_h\|_0, \|\partial_\theta u_h\|_0\}$$

und damit den Banach-Raum

$$B_h = \{u_h | u_h \in C_h^1(\mathbb{T}_h^2, \mathbb{R}^q), u_j^n = u_j^n \bmod J = u_j^n \bmod N, u_h \in \mathbb{T}_h^2\}. \quad (51)$$

Der Operator  $F_h : B_h \rightarrow B_h^0$  soll allgemein in der Form des 6-Punkt-Schemas

$$\{F_h u_h\}_j^n \equiv \frac{1}{\tau} \left\{ \sum_{\mu=-1}^1 S_\mu^*(\theta_j, t_h, u_j^n) u_{j+\mu}^{n+1} - \sum_{\mu=-1}^1 S_\mu(\theta_j, t_h, u_j^n) u_{j+\mu}^n \right\} - f(\theta_j, t_h, u_j^n)$$

(52)

mit  $j = 0(1)J - 1$ ,  $n = 0(1)N - 1$  definiert werden.  $S_\mu^*(\theta_j, t_h, u_j^n)$  und  $S_\mu(\theta_j, t_h, u_j^n)$  sind  $q$ -reihige Diagonalmatrizen für  $\mu = -1, 0, 1$ . Mit  $F_h$  lauten die betrachteten Verfahren in Operatorschreibweise

$$F_h u_h = 0, \quad u_h \in B_h.$$

(53)

In der allgemeinen Operatordarstellung (52) sind alle nachfolgenden expliziten und linear-impliziten 6-Punkt-Schemata enthalten.

### 1. Explizite Verfahren vom Upwind-Typ

Bei diesen Verfahren wird die Diagonalmatrix

$$C(\theta_j, t_n, u_j^n) = \text{diag}(c_1, c_2, \dots, c_q) = \omega(\theta_j, t_n, u_j^n) I_q, \quad I_q - \text{Einheitsmatrix}$$

in die Summe einer Matrix  $C^+(\theta_j, t_n, u_j^n)$  mit nur positiven Elementen und einer Matrix  $C^-(\theta_j, t_n, u_j^n)$  mit nur negativen Elementen zerlegt :

$$C(\theta_j, t_n, u_j^n) = C^+(\theta_j, t_n, u_j^n) + C^-(\theta_j, t_n, u_j^n) \quad . \quad (54)$$

(a) Das *explizite Courant-Isaacson-Rees Verfahren* (CIR) ergibt sich mit

$$\begin{aligned} C^+ &= \text{diag} \left( \frac{1}{2}(c_i + |c_i|) \right) \geq 0 \\ C^- &= \text{diag} \left( \frac{1}{2}(c_i - |c_i|) \right) \leq 0 \quad , \end{aligned} \quad (55)$$

(b) das *explizite glatte Upwind-Verfahren* mit der Wahl

$$\begin{aligned} C^+ &= \text{diag} \left( \frac{1}{2}(c_i + \Phi(c_i)) \right) \geq 0 \\ C^- &= \text{diag} \left( \frac{1}{2}(c_i - \Phi(c_i)) \right) \leq 0 \quad , \end{aligned} \quad (56)$$

mit  $\Phi(d) := \sqrt{\delta^2 + d^2}$  ,  $\delta \neq 0$  , *constant* .

Die Diagonalmatrizen  $S_\mu$  and  $S_\mu^*$  ,  $\mu = -1, 0, 1$ , besitzen für alle expliziten Verfahren vom Upwind-Typ die Form

$$\begin{aligned} S_{-1} &= \lambda C^+ & , & \quad S_{-1}^* = 0 \\ S_0 &= I - \lambda C^+ + \lambda C^- & , & \quad S_0^* = I \\ S_1 &= -\lambda C^- & , & \quad S_1^* = 0 . \end{aligned} \quad (57)$$

In der üblichen Notation läßt sich das explizite Upwind-Verfahren als

$$\begin{aligned} \{F_h u_h\}_j^n &\equiv \frac{1}{\tau}(u_j^{n+1} - u_j^n) + \omega(\theta_j, t_n, u_j^n) \frac{1}{2h}(u_{j+1}^n - u_{j-1}^n) - \\ &\quad - \Phi(\omega(\theta_j, t_n, u_j^n)) \frac{1}{2h}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) - \\ &\quad - r(\theta_j, t_n, u_j^n) = 0 \quad \text{mit} \quad (\theta_j, t_n) \in \mathbb{T}_h^2 \\ &\text{und} \quad \Phi(d) = \sqrt{\delta^2 + d^2}, \delta \neq 0, \text{const.} \end{aligned} \quad (58)$$

darstellen.

## 2. Explizites Friedrichs-Verfahren

In diesem Falle wird die Matrix  $C(\theta_j, t_n, u_j^n)$  nicht zerlegt, und die Diagonalmatrizen  $S_\mu$  and  $S_\mu^*$  ,  $\mu = -1, 0, 1$  besitzen nun die Form

$$\begin{aligned} S_{-1} &= \frac{1}{2}(I + \lambda C) & , & \quad S_{-1}^* = 0 \\ S_0 &= 0 & , & \quad S_0^* = I \\ S_1 &= \frac{1}{2}(I - \lambda C) & , & \quad S_1^* = 0 \quad . \end{aligned} \quad (59)$$

## 3. Linear implizite Verfahren

Analog zu den expliziten Verfahren vom Upwind-Typ wird die Matrix  $C(\theta_j, t_n, u_j^n)$  in

die beiden Matrizen  $C^+(\theta_j, t_n, u_j^n)$  und  $C^-(\theta_j, t_n, u_j^n)$  zerlegt, wobei im Falle des impliziten CIR-Verfahrens

$$\begin{aligned} C^+ &= \text{diag} \left( \frac{1}{2}(c_i + |c_i|) \right) \geq 0 \\ C^- &= \text{diag} \left( \frac{1}{2}(c_i - |c_i|) \right) \leq 0 \quad , \end{aligned} \quad (60)$$

und im Falle des impliziten glatten Upwind-Verfahrens

$$\begin{aligned} C^+ &= \text{diag} \left( \frac{1}{2}(c_i + \Phi(c_i)) \right) \geq 0 \\ C^- &= \text{diag} \left( \frac{1}{2}(c_i - \Phi(c_i)) \right) \leq 0 \quad , \end{aligned} \quad (61)$$

with  $\Phi(d) := \sqrt{\delta^2 + d^2}$  ,  $\delta \neq 0$  , *constant*

gilt. Für alle derartigen linear impliziten Verfahren sind die Diagonalmatrizen  $S_\mu$  und  $S_\mu^*$  ,  $\mu = -1, 0, 1$  durch

$$\begin{aligned} S_1 &= 0 \quad , \quad S_{-1}^* = -\lambda C^+ \\ S_0 &= I \quad , \quad S_0^* = I - \lambda C^- + \lambda C^+ \\ S_{-1} &= 0 \quad , \quad S_1^* = \lambda C^- \quad . \end{aligned} \quad (62)$$

gegeben.

In [11] wird die diskrete Konvergenz der Verfahren (53) mit Operatoren (52) detailliert nachgewiesen. Der folgende Konvergenzsatz läßt sich unter Benutzung des allgemeinen Satzes 13 verifizieren.

**Satz 18** *Der Differenzenoperator (52) für (46) genüge folgenden Voraussetzungen:*

- (i)  $\lambda := \frac{\tau}{h} = \text{const}$
- (ii)  $\omega(\theta, t, u) \in \mathcal{C}^2(\mathbb{T}^2 \times \mathbb{R}^q, \mathbb{R})$  ,  $f(\theta, t, u) \in \mathcal{C}^2(\mathbb{T}^2 \times \mathbb{R}^q, \mathbb{R}^q)$
- (iii) (48) besitzt eine lokal eindeutige Lösung  $u \in \mathcal{C}^2(\mathbb{T}^2)$
- (iv)  $\sum_{\mu=-1}^1 S_\mu(\theta, t, u) \equiv I$  ,  $\sum_{\mu=-1}^1 S_\mu^*(\theta, t, u) \equiv I$
- (v)  $\sum_{\mu=-1}^1 \mu(S_\mu^*(\theta, t, u) - S_\mu(\theta, t, u)) \equiv \lambda C$  , wobei  $C = (c_{ij})$   

$$c_{ij} = \begin{cases} \omega(\theta, t, u) & \text{für } i = j \\ 0 & \text{sonst} . \end{cases}$$

Dann ist jedes Verfahren (53) konsistent in  $h$  und  $\tau$  mit Ordnung 1. Gilt desweiteren

- (vi) (53) ist von positivem Typ, die Diagonalmatrizen  $S_\mu^*$  ,  $\mu = -1, 0, 1$  sind Lipschitz-stetig in  $u$  und zwei der drei Matrizen  $S_\mu^*$  besitzen Elemente  $(s_\mu^*)_{kl} \leq 0$  ,  $k, l = 1(1)q$

(vii) Für die Anfangswerte sei  $\|u_j^0 - g(\theta_j)\| \leq K_0 h \quad \forall j, h \leq h_0, K_0 > 0,$

so konvergiert jedes Verfahren (53) mit Ordnung 1 auf  $\mathbb{T}_h^2$ , d.h.

$$\|e_j^n\| = \|u_j^n - u(\theta_j, t_n)\|_\infty \leq Kh, \quad K > 0 \quad \forall j, n.$$

BEWEIS: Vgl. [11], S. 7 - 12.

Ein Differenzenoperator (52) ist dabei von *positivem Typ*, falls die 3 Matrizen  $S_\mu(\theta, t, u)$  mit  $\mu = -1, 0, 1$  für alle  $(\theta, t, u) \in \mathbb{T}^2 \times \mathbb{R}^q$ ,  $\|u\| \leq M$ ,  $M$  const, nur nichtnegative Elemente besitzen.

Im Falle der drei *expliziten Verfahren* sind die Voraussetzungen (iv) und (v) erfüllt. Der folgende Satz liefert eine hinreichende Bedingung für die Positivität (vi) der Verfahren.

**Satz 19** Sei auf der Menge  $G = \{(\theta, t, u) \mid (\theta, t) \in \mathbb{T}^2, \|u\| \leq M, M \in \mathbb{R}, \text{const}\}$  die *Schrittweitenbedingung*

$$\lambda \leq \frac{1}{D} \quad \text{mit} \quad D := \max_G \max_{i=1(1)q} |c_i(\theta, t, u)| \quad (63)$$

gegeben. Dann sind das *explizite CIR-Verfahren* und das *explizite Friedrichs-Verfahren* von *positivem Typ*. Ist jedoch

$$\lambda \leq \frac{1}{\sqrt{\delta^2 + D^2}}, \quad (64)$$

so ist auch das *explizite glatte Upwind-Verfahren* von *positivem Typ*.

BEWEIS: Vgl. [11], S. 13.

Die beiden *linear impliziten Verfahren* sind von *positivem Typ* für alle  $\lambda \in \mathbb{R}^+$ , und die Diagonalmatrizen  $S_\mu^*$ ,  $\mu = -1, 0, 1$ , sind Lipschitz-stetig in  $u$ . Desweiteren ergibt sich  $S_{-1}^* \leq 0$  and  $S_1^* \leq 0$  aus (60), (61) und (62). Damit sind beide Verfahren konsistent unter den Voraussetzungen (i)- (iv) und konvergent bei geeigneten Anfangswerten.

Um Approximationen höherer Ordnung und gleichzeitig asymptotische Fehlerschätzungen zu gewinnen, bieten sich insbesondere Extrapolationsverfahren und Defektkorrektur-Verfahren (deferred correction methods) an. Während die Richardson-Extrapolation gegenwärtig auch für partielle DGL-Systeme (vgl. [7]) häufiger genutzt wird, sind Defektkorrektur-Verfahren - vermutlich wegen ihres komplizierten theoretischen Hintergrunds - weniger bekannt. In [10] wird ein einheitlicher Zugang für alle betrachteten - expliziten und impliziten - Basis-Verfahren 1. Ordnung angegeben, der konvergente Lösungen 2. Ordnung und asymptotische Schätzungen des globalen Fehlers der Basislösung liefert. Im Unterschied zur Extrapolation ist bei der Defektkorrektur keine Gitterverfeinerung nötig. Werden die nichtlinearen Gleichungssysteme zudem mit einem Newton-ähnlichen Verfahren gelöst, so ist lediglich ein zusätzlicher Newtonschritt mit modifizierter rechter Seite erforderlich.

## Literatur

- [1] Ascher, U.M.; Mattheij, R.M.M.; Russell, R.D.: *Numerical solution of boundary value problems for ordinary differential equations*. SIAM Publications, Philadelphia 1988.
- [2] Bernet, K.; Vogt, W.: *Anwendung finiter Differenzenverfahren zur direkten Bestimmung invarianter Tori*. ZAMM 74 (1994), No. 6, T577-T579.
- [3] Keller, H.B.: *Approximation methods for nonlinear problems with applications to two-point boundary value problems*. Math. Comp. 29 (1975), 464 - 474.
- [4] Keller, H.B.: *Accurate difference methods for nonlinear two-point boundary value problems*. SIAM J. Numer. Anal. 11 (1974), 305 - 320.
- [5] Keller, H.B.: *Numerical solution of two-point boundary value problems*. SIAM Publications, Philadelphia 1976.
- [6] Lentini, M.; Pereyra, V.: *An adaptive difference solver for nonlinear two-point boundary value problems with mild boundary layers*. SIAM J. Numer. Anal. 14 (1977), 91 - 111.
- [7] Nowak, U.: *Adaptive Linienmethoden für nichtlineare parabolische Systeme in einer Raumdimension*. TR 93-14, Dez. 1993, ZIB.
- [8] Schilder, F.; Vogt, W.: *Semidiscretisation methods for quasi-periodic solutions*. Submitted to ZAMM, 2002.
- [9] Stetter, H.J.: *Analysis of discretization methods for ordinary differential equations*. Springer-Verlag, Berlin 1973.
- [10] Vogt, W.: *Zur Konstruktion von Differenzenverfahren 2. Ordnung für quasilineare hyperbolische Systeme auf dem Torus*. Preprint No. M 28/98, Technical University of Ilmenau, Department of Mathematics 1998.
- [11] Vogt, W.; Bernet, K.: *A Shooting Method for Invariant Tori*. Preprint No. M 3/95, Technical University of Ilmenau, Department of Mathematics 1995.