



KERNFORSCHUNGSANLAGE JÜLICH GmbH

Institut für Festkörperforschung

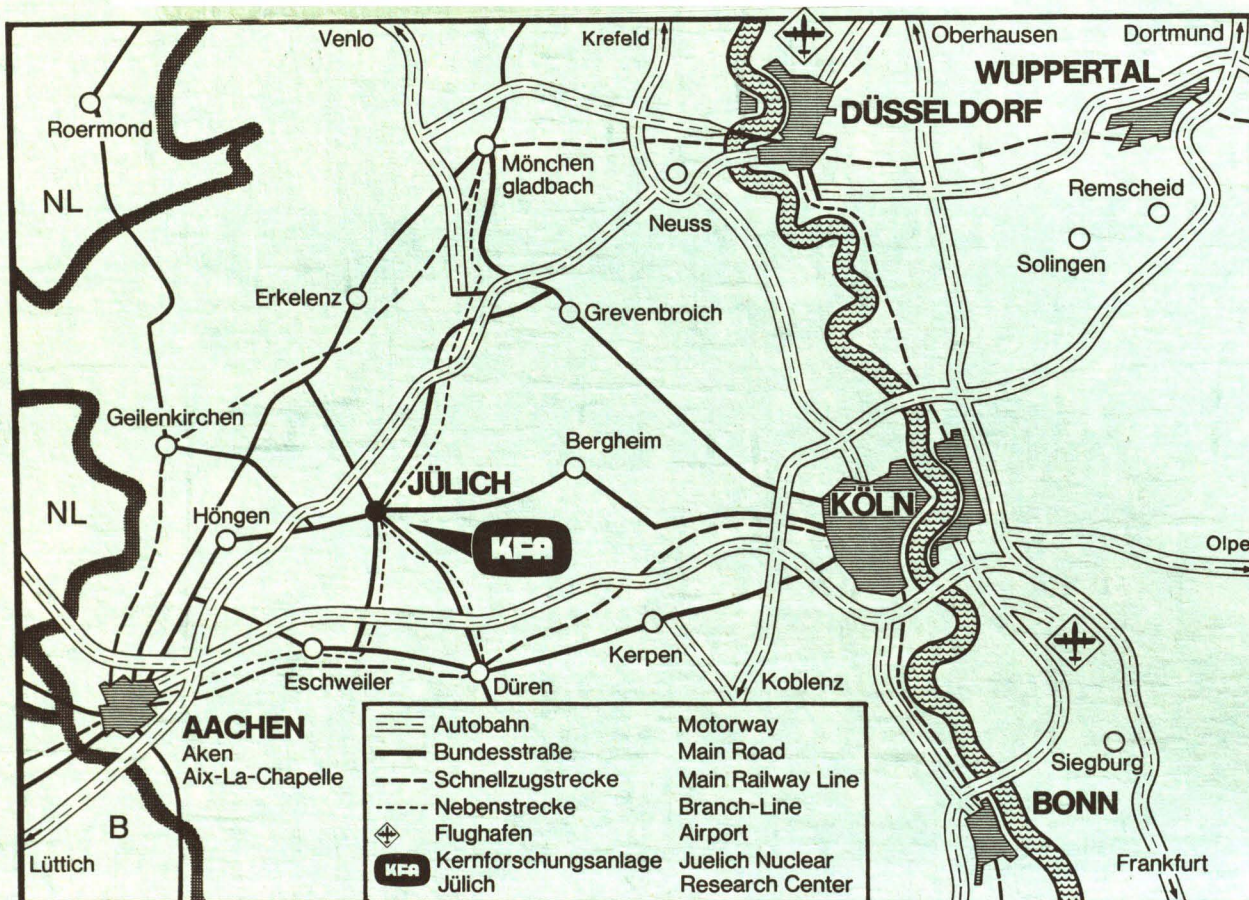
**Numerische Behandlung gewöhnlicher
Differentialgleichungen**

von

P. Albrecht

**JÜl - 1274
Februar 1976**

Als Manuskript gedruckt



Berichte der Kernforschungsanlage Jülich - Nr. 1274

Institut für Festkörperforschung Jül - 1274

Im Tausch zu beziehen durch: ZENTRALBIBLIOTHEK der Kernforschungsanlage Jülich GmbH,
Jülich, Bundesrepublik Deutschland

Numerische Behandlung gewöhnlicher Differentialgleichungen

von

P. Albrecht

Abstract

This report presents an introduction to the numerical treatment of ordinary differential equations (O.D.Es.) which, although containing the classical results, in many parts gives a new approach to the theory. The basic definitions and theorems are presented such that they can easily be extended to more general discretization problems (e.g. occurring in partial differential equations).

Large parts of the theory are based on a general stability theorem (theorem 3.1.), the application of which is not restricted to O.D.Es. This theorem also serves as starting point for a new and, in a sense, complete theory of cyclic multistep methods which originally have been presented by Donelson and Hansen; here, however, they are treated in a totally different way, showing an easier manner to calculate numerically stable M-cyclic k-step methods of maximum order. The underlying theory applies also to the more general case of multistage multistep methods as introduced by Stetter.

M-cyclic k-step methods of lower order $q < 2k$ have very favourable stability properties (section 10.2) and, thus, seem to be particularly suitable for stiff equations.

Vorwort

Dieser Bericht enthält eine leicht verständliche Einführung in die theoretischen und praktischen Aspekte der numerischen Behandlung von Anfangswertproblemen bei gewöhnlichen Differentialgleichungen. Er behandelt zwar auch den klassischen Stoff ausführlich, bringt in vielen Teilen jedoch neue Darstellungen und Ergebnisse.

Die grundlegenden Definitionen und Sätze in den Paragraphen 2 und 3 sind so angelegt, daß sie leicht auf allgemeinere Diskretisierungsprobleme (z.B. bei partiellen Differentialgleichungen) übertragbar sind. Große Teile der Theorie basieren auf einem allgemeinen Stabilitätssatz (Satz 3.1.), dessen Anwendung ebenfalls nicht auf die Diskretisierung von gewöhnlichen Differentialgleichungen beschränkt ist.

Er ist auch Ausgangspunkt für eine geschlossene Theorie über Verfahren der Konsistenzordnung q mit der Konvergenzordnung $(q+1)$ in Paragraph 5, die frühere Ergebnisse von Spijker and Donelson-Hansen als Spezialfälle enthält und zu einer neuen Theorie der zyklischen Mehrschrittverfahren führt. Diese werden hier auf völlig andere Weise als in der ursprünglichen Arbeit von Donelson und Hansen behandelt, was zu einem tieferen Verständnis dieser Methoden beiträgt und neue, einfacher zu handhabende Möglichkeiten ergibt, stabile M -zyklische k -Schritt-Verfahren maximaler Ordnung zu berechnen.

Es scheint, daß zyklische Verfahren auch sehr günstige Stabilitätseigenschaften haben können, was sie insbesondere für steife Differentialgleichungen geeignet und anderen Verfahren überlegen macht (vgl. § 10). Dieser Aspekt muß jedoch noch näher untersucht werden.

Abstract

This report presents an introduction to the numerical treatment of ordinary differential equations (O.D.Es.) which, although containing the classical results, in many parts gives a new approach to the theory. The basic definitions and theorems are presented such that they can easily be extended to more general discretization problems (e.g. occurring in partial differential equations).

Large parts of the theory are based on a general stability theorem (theorem 3.1.), the application of which is not restricted to O.D.Es. This theorem also serves as starting point for a new and, in a sense, complete theory of cyclic multistep methods which originally have been presented by Donelson and Hansen; here, however, they are treated in a totally different way, showing an easier manner to calculate numerically stable M-cyclic k-step methods of maximum order. The underlying theory applies also to the more general case of multistage multistep methods as introduced by Stetter.

M-cyclic k-step methods of lower order $q < 2k$ have very favourable stability properties (section 10.2) and, thus, seem to be particularly suitable for stiff equations.

Vorwort

Dieser Bericht enthält eine leicht verständliche Einführung in die theoretischen und praktischen Aspekte der numerischen Behandlung von Anfangswertproblemen bei gewöhnlichen Differentialgleichungen. Er behandelt zwar auch den klassischen Stoff ausführlich, bringt in vielen Teilen jedoch neue Darstellungen und Ergebnisse.

Die grundlegenden Definitionen und Sätze in den Paragraphen 2 und 3 sind so angelegt, daß sie leicht auf allgemeinere Diskretisierungsprobleme (z.B. bei partiellen Differentialgleichungen) übertragbar sind. Große Teile der Theorie basieren auf einem allgemeinen Stabilitätssatz (Satz 3.1.), dessen Anwendung ebenfalls nicht auf die Diskretisierung von gewöhnlichen Differentialgleichungen beschränkt ist.

Er ist auch Ausgangspunkt für eine geschlossene Theorie über Verfahren der Konsistenzordnung q mit der Konvergenzordnung $(q+1)$ in Paragraph 5, die frühere Ergebnisse von Spijker and Donelson-Hansen als Spezialfälle enthält und zu einer neuen Theorie der zyklischen Mehrschrittverfahren führt. Diese werden hier auf völlig andere Weise als in der ursprünglichen Arbeit von Donelson und Hansen behandelt, was zu einem tieferen Verständnis dieser Methoden beiträgt und neue, einfacher zu handhabende Möglichkeiten ergibt, stabile M -zyklische k -Schritt-Verfahren maximaler Ordnung zu berechnen.

Es scheint, daß zyklische Verfahren auch sehr günstige Stabilitätseigenschaften haben können, was sie insbesondere für steife Differentialgleichungen geeignet und anderen Verfahren überlegen macht (vgl. § 10). Dieser Aspekt muß jedoch noch näher untersucht werden.

Inhalt

Einleitung	1
1. Definitionen und Begriffe	2
2. Konsistenz	4
3. Stabilität und Konvergenz	12
3.1. Ein Beispiel	13
3.2. Definition der Stabilität	15
3.3. Ein allgemeines Stabilitätskriterium	19
3.4. Konvergenz	23
4. Anwendungen des Stabilitätssatzes	26
4.1. Stabilität von k-Schritt-Verfahren	26
4.2. Stabilität von Prädiktor-Korrektor-Verfahren	31
4.3. Stabilität zyklischer Verfahren	33
4.3.1. Definition zyklischer Verfahren	33
4.3.2. Stabilität	35
4.4. Stabilitätsbetrachtungen bei einem partiellen Differentialgleichungsproblem	38
5. Über die Ordnung zyklischer Verfahren	41
5.1. Eine Modifikation des Satzes 3.1.	41
5.2. Maximale Ordnung M-zyklischer k-Schritt-Verfahren	46
5.3. Vereinfachung der Ordnungsbedingung (5.27)	49
6. Stabilität eines Differentialgleichungsproblems	52
7. Stabilität bei fester Schrittweite h	54
7.1. Allgemeine Betrachtungen	54
7.2. Stabilitätsbereiche	55
7.3. Beispiele	61

8.	Einschrittverfahren	66
8.1.	Verfahren mit Ableitungen	67
8.2.	Explizite Runge-Kutta-Verfahren	68
8.2.1.	Explizite RK-Formeln mit $r=1,2,3$ Stufen	69
8.2.2.	Das klassische Runge-Kutta-Verfahren	74
8.2.3.	Runge-Kutta-Formeln höherer Ordnung ($q \geq 5$)	77
8.3.	Implizite Runge-Kutta-Verfahren	81
8.3.1.	Gauss-Formeln	83
8.3.2.	Radau- und Lobatto-Formeln	85
8.4.	Semi-implizite Runge-Kutta-Verfahren	87
8.5.	Bemerkungen zur praktischen Anwendung von RK-Verfahren	89
9.	Lineare Mehrschrittverfahren	92
9.1.	Klassische explizite Mehrschrittverfahren	93
9.1.1.	Die Verfahren von Adams-Bashforth	94
9.1.2.	Die Verfahren von Nyström	95
9.2.	Klassische implizite Mehrschrittverfahren	95
9.2.1.	Die Verfahren von Adams-Moulton und Milne-Simpson	96
9.3.	Allgemeine lineare Mehrschrittverfahren	97
9.4.	Instabile lineare Mehrschrittverfahren der Ordnung $(2k-1)$	99
9.5.	Prädiktor-Korrektor-Verfahren	102
9.5.1.	Konsistenzordnung von P-C-Verfahren	102
9.5.2.	Vergleich von $P(EC)^S E$ - und $P(EC)^S$ -Verfahren	103
9.5.3.	Stabilitätsbereiche von P-C-Verfahren	105
9.5.4.	Abschätzung des lokalen und globalen Fehlers	107
10.	Zyklische Verfahren	111
10.1.	Zur Konstruktion stabiler k -zyklischer k -Schritt-Verfahren maximaler Konvergenzordnung	111
10.2.	Bemerkungen zur Stabilität M -zyklischer k -Schritt-Verfahren	113
	Literatur	116
	Index	123

Einleitung

Im folgenden behandeln wir Methoden zur numerischen Lösung von Anfangswertaufgaben für Differentialgleichungssysteme 1. Ordnung der Form:

$$T[y](x) \equiv y'(x) - f(x, y(x)) = 0; \quad y(a) = \eta_0 \quad (1.1)$$

$$y: [a, b] \rightarrow \mathbb{R}^n; \quad f: [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

Bekanntlich hat (1.1) eine in $a \leq x \leq b$ eindeutige, einmal stetig differenzierbare Lösung y , wenn $f(x, y)$ im Gebiet $G := \{(x, y) : a \leq x \leq b; y \in \mathbb{R}^n\}$ stetig ist und dort einer Lipschitzbedingung:

$$\|f(x, y_1) - f(x, y_2)\| \leq L \|y_1 - y_2\|; \quad (1.2)$$

genügt mit einer Konstanten $L \geq 0$ und für alle $(x, y_1), (x, y_2) \in G$.

Im folgenden setzen wir stets voraus, daß $f(x, y)$ diese Eigenschaften hat; wenn nichts anderes gesagt wird, sind auftretende Normen stets Maximumnormen: $\|u\| = \max_{1 \leq j \leq n} |u_j|$.

Wir beschränken unsere Betrachtungen auf Anfangswertprobleme der Form (1.1), jedoch lassen sich die Definitionen und die meisten Sätze über Konsistenz, Stabilität und Konvergenz auf sehr viel allgemeinere Diskretisierungsverfahren übertragen, insbesondere auf Methoden zur numerischen Behandlung von Systemen höherer Ordnung und Randwertproblemen sowie von Anfangs- und Randwertproblemen partieller Differentialgleichungen. Wir werden diese Anwendungen an Beispielen erläutern, verzichten aber bewußt auf eine allgemeinere Darstellung, um die Betrachtungen nicht unnötig zu komplizieren.

1. Definitionen und Begriffe

Die Diskretisierungsverfahren zur numerischen Lösung von (1.1) bestehen im allgemeinen darin, y an den Stellen $x=x_j$ eines Punktegitters $I_h := \{x_j = a + jh \mid j=0(1)m; h=\frac{b-a}{m}\}$ (1.3) durch Näherungswerte¹⁾ y_j^* zu approximieren, indem man den Differentialoperator T in (1.1) in geeigneter Weise durch einen Differenzenoperator T_h ersetzt und die y_j^* aus dem resultierenden System von Differenzengleichungen bestimmt.

Approximiert man zum Beispiel $T[y](x) \equiv y'(x) - f(x, y(x))$, $y: \mathbb{R} \rightarrow \mathbb{R}$, durch

$$T_h[y](x) = \frac{y(x+h) - y(x)}{h} - f(x, y(x));$$

so geht das Problem (1.1) über in die Differenzengleichung

$$T_h[y^*](x) = 0; \quad y^*(a) = \eta_0$$

und mit $x=x_j$ ergeben sich die Näherungswerte y_j^* nacheinander aus:

$$y_0^* = \eta_0; \quad y_{j+1}^* - y_j^* - hf(x_j, y_j^*) = 0; \quad j=0(1)m-1 \quad (1.4)$$

(1.4) ist das Verfahren von Euler, das einfachste aller sogenannten "Einschrittverfahren".

Definition 1.1.

Ein Verfahren $T_h[y^*](x_j) = 0$ mit $hT_h[y^*](x_j) = F(x_j, y_j^*, y_{j+1}^*, \dots, y_{j+k}^*; h)$ zur Berechnung der Näherungen y_{j+k}^* ($j=0(1)m-k$) heißt k-Schritt-Verfahren. Bei $k=1$ spricht man auch von einem Einschrittverfahren, bei $k>1$ von einem Mehrschrittverfahren.

¹⁾ Im folgenden bezeichnet y stets die Lösung der Dgl. (1.1) und y^* ihre Näherung.

In Kapitel 8 werden wir Einschrittverfahren der Form

$$y_{j+1}^* - y_j^* - h\phi(x_j, y_j^*; h) = \mathcal{O} ; \quad y_0^* = \eta_0 \quad (1.5)$$

$$j = 0(1)m-1$$

ableiten und in Kapitel 9 "lineare Mehrschrittverfahren" der Form

$$\alpha_k y_{j+k}^* + \alpha_{k-1} y_{j+k-1}^* + \dots + \alpha_0 y_j^* - h(\beta_k f_{j+k}^* + \dots + \beta_0 f_j^*) = \mathcal{O} \quad (1.6)$$

$$j = 0(1)m-k$$

$$y_0^* = \eta_0 ; \quad y_1^* = \eta_1(h) ; \dots ; y_{k-1}^* = \eta_{k-1}(h) ;$$

$$f_j^* = f(x_j, y_j^*) ; \quad \alpha_k = 1 ; \quad |\alpha_0| + |\beta_0| \neq 0$$

wobei die $\eta_j(h)$, $j=1(1)k-1$, in einer Anlaufrechnung mittels anderer Verfahren (z.B. Einschrittverfahren) berechnet werden. Für $\beta_k = 0$ heißt das Verfahren explizit, für $\beta_k \neq 0$ implizit, denn y_{j+k}^* ist in diesem Fall noch in $f_{j+k}^* = f(x_{j+k}, y_{j+k}^*)$ enthalten. Wie wir sehen werden (Abschnitt 4.2.), ist bei impliziten Verfahren in jedem Rechenschritt j noch eine Iteration erforderlich.

Beispiele:

1.- Ersetzt man $y'(x)$ durch $\frac{1}{2h}[y(x+h)-y(x-h)]$ so erhält man aus (1.1) für $x=x_j$ das folgende 2-Schrittverfahren (midpoint-rule)

$$y_{j+1}^* = y_{j-1}^* + 2h f(x_j, y_j^*) ; \quad j = 1(1)m-1 \quad (1.7)$$

welches zwei Startwerte y_0^* und y_1^* erfordert; $y_0^* = \eta_0$ und (beispielsweise) $y_1^* = y_0^* + h f(x_0, y_0^*)$. Dabei ist

$$T_h[y](x) \equiv \frac{y(x+h)-y(x-h)}{2h} - f(x, y(x))$$

2.- Später werden wir das folgende, als modifiziertes Euler-Verfahren bekannte Einschrittverfahren kennenlernen:

$$y_0^* = \eta_0; \quad y_{j+1}^* = y_j^* + h f(x_j + \frac{h}{2}, y_j^* + \frac{h}{2} f(x_j, y_j^*)); \quad j = 0(1)m-1 \quad (1.8)$$

Hier ist: $T_h[y](x) \equiv \frac{y(x+h) - y(x)}{h} - f(x + \frac{h}{2}, y(x) + \frac{h}{2} f(x, y(x)))$

3.- (1.8) kann dazu benutzt werden, um den Anfangswert y_1^* für das Verfahren von Milne-Simpson zu berechnen:

$$y_{j+1}^* = y_{j-1}^* + \frac{h}{3} (f_{j+1}^* + 4 f_j^* + f_{j-1}^*); \quad j = 1(1)m-1 \quad (1.9)$$

$$y_0^* = \eta_0; \quad y_1^* = y_0^* + h f(x_0 + \frac{h}{2}, y_0^* + \frac{h}{2} f(x_0, y_0^*))$$

Dieses 2-Schrittverfahren ist implizit. Der zugehörige Differenzenoperator T_h lautet

$$T_h[y](x) \equiv \frac{y(x+h) - y(x-h)}{2h} - \frac{1}{6} [f(x+h, y(x+h)) + 4f(x, y(x)) + f(x-h, y(x-h))]$$

Einzelheiten über Ableitung und Anwendung dieser Verfahren werden später behandelt.

2. Konsistenz

Der Operator T_h kann als Näherung von T in einem Funktionenraum $\mathbb{F}$ angesehen werden, wenn für alle Funktionen z aus $\mathbb{F}$ und alle $x \in I_h'$ die Norm $\max_{x \in I_h'} \|T_h[z](x) - T[z](x)\|$ mit h gegen Null geht, wobei bei I_h' die Menge aller $x \in I_h$ bezeichnet, für die $T_h[y](x)$ definiert ist. Diese Bedingung ist bei nichtlinearem T_h (z.B. Mehrschrittverfahren der Form (1.6)) nur in Ausnahmefällen erfüllbar, daher begnügen wir uns mit der Forderung, daß sie wenigstens für $z(x) = y(x)$ gelten soll, wobei y wieder die Lösung von $T[y](x) = 0$ ist. Wir nennen insbesondere T_h mit T konsistent, wenn der lokale Ersetzungsfehler $T_h[y](x) - T[y](x) = T_h[y](x)$ mit h gleichmäßig gegen Null geht:

Definition 2.1.

Sei y Lösung von $T[y](x)=0$; dann heißt der Operator T_h und auch das durch ihn definierte Differenzenverfahren $T_h[y^*](x)=0$, $x \in I'_h$, konsistent (mit der Aufgabe $T[y](x)=0$)¹⁾, wenn für den Ersetzungsfehler gilt:

$$\lim_{h \rightarrow 0} \max_{x \in I'_h} \|T_h[y](x)\| = 0 \quad (2.1)$$

Existieren außerdem Konstanten $h_0 > 0$, $D > 0$ und $q > 0$, so daß für $h \leq h_0$ gilt:

$$\max_{x \in I'_h} \|T_h[y](x)\| \leq D h^q$$

so hat das Verfahren die Konsistenzordnung q .

Falls auch die Startwerte $\eta_j(h)$ eines k -Schrittverfahrens Näherungen der exakten Werte $y(x_j)$, $j=0(1)k-1$, sind, d.h. wenn gilt:

$$\lim_{h \rightarrow 0} \|\eta_j(h) - y(x_j)\| = 0; \quad j=0(1)k-1 \quad (2.2)$$

so wird man annehmen dürfen, daß für kleines h - möglicherweise unter zusätzlichen Voraussetzungen - die Lösung y_j eines konsistenten Differenzenverfahrens die Lösung $y(x_j)$ der Differentialgleichung (1.1) in den Punkten $x_j \in I_h$ approximiert. Wir werden später sehen (Satz 3.2.), daß das in der Tat der Fall ist. (2.2) wird oft Konsistenzbedingung für die Startwerte genannt.

¹⁾ dieser Zusatz wird weggelassen, wenn kein Zweifel besteht, auf welches Problem er sich bezieht.

Beispiele:

- 1.- Die midpoint-rule (1.7) hat für $y \in \mathcal{C}^3[a,b]$ die Konsistenzordnung 2, denn es gilt:

$$\begin{aligned} \|T_h[y](x) - y(x)\| &= \left\| \frac{1}{2h} (y(x+h) - y(x-h)) - f(x, y(x)) \right\| \\ &= \left\| \frac{1}{2h} (y(x+h) - y(x-h)) - y'(x) \right\| \\ &= \frac{h^2}{3!} \|y'''(\xi)\|; \quad \xi \in [a,b] \quad (\text{nach Taylor}) \end{aligned}$$

Auch der Startwert y_1^* ist konsistent, denn es gilt:

$$\lim_{h \rightarrow 0} \|y_0^* + hf(x_0, y_0^*) - y(x_1)\| = 0.$$

In gleicher Weise beweist man, daß T_h für $y \in \mathcal{C}^2[a,b]$ die Konsistenzordnung 1 hat.

Wegen $\|T_h[z](x) - T[z](x)\| = \frac{h^2}{3!} \|z'''(\xi)\|; \quad z \in \mathcal{C}^3[a,b]$

approximiert T_h für kleine h den Operator T sogar für alle $z \in \mathcal{C}^3[a,b]$ (und nicht nur für die Lösung y von (1.1)).

- 2.- Das Verfahren (1.9) von Milne-Simpson hat für $y \in \mathcal{C}^5[a,b]$ die Konsistenzordnung 4. Es gilt nämlich:

$$\begin{aligned} \|T_h[y](x) - y(x)\| &= \left\| \frac{y(x+h) - y(x-h)}{2h} - \frac{1}{6} [f(x+h, y(x+h)) + 4f(x, y(x)) + f(x-h, y(x-h))] \right\| \\ &= \left\| \frac{y(x+h) - y(x-h)}{2h} - \frac{1}{6} (y'(x+h) + 4y'(x) + y'(x-h)) \right\| \\ &= \left\| y'(x) + \frac{h^2}{3!} y'''(x) + \frac{h^4}{5!} y^{(5)}(\xi_1) - \frac{1}{6} [6y'(x) + h^2 y'''(x) + \right. \\ &\quad \left. + \frac{h^4}{12} y^{(5)}(\xi_2)] \right\| \leq C h^4 \end{aligned}$$

mit $C = \frac{1}{45} \max_{x \in [a,b]} \|y^{(5)}(x)\|$ und $x-h < \xi_1, \xi_2 < x+h$ (2.3)

3.- In der parabolischen partiellen Differentialgleichung (inhomogene Wärmeleitungsgleichung):

$$T[u](x,t) = u_t - \sigma u_{xx} - f(x,t,u) = 0; \quad \sigma > 0 \quad (2.4)$$

$$u(x,0) = \eta(x) \quad 0 \leq x \leq b$$

$$u(0,t) = u(b,t) = 0; \quad 0 \leq t \leq c$$

seien u_t bzw. u_{xx} ersetzt durch

a) Vorwärtsdifferenz bzw. zentrale Differenz an der Stelle t

b) Zentrale Differenz " " " " " " t

c) Vorwärtsdifferenz " " " " " " $(t+\Delta t)$

$$\begin{aligned} \text{a) } T_{\Delta x, \Delta t}^{(a)}[u](x,t) &= \frac{u(x,t+\Delta t) - u(x,t)}{\Delta t} - \sigma \frac{u(x+\Delta x,t) - 2u(x,t) + u(x-\Delta x,t)}{(\Delta x)^2} - \\ &= f(x,t,u) = 0 \end{aligned} \quad (2.5a)$$

$$\begin{aligned} \text{b) } T_{\Delta x, \Delta t}^{(b)}[u](x,t) &= \frac{u(x,t+\Delta t) - u(x,t-\Delta t)}{2\Delta t} - \sigma \frac{u(x+\Delta x,t) - 2u(x,t) + u(x-\Delta x,t)}{(\Delta x)^2} - \\ &= f(x,t,u) = 0 \end{aligned} \quad (2.5b)$$

$$\begin{aligned} \text{c) } T_{\Delta x, \Delta t}^{(c)}[u](x,t) &= \frac{u(x,t+\Delta t) - u(x,t)}{\Delta t} - \sigma \frac{u(x+\Delta x,t+\Delta t) - 2u(x,t+\Delta t) + u(x-\Delta x,t+\Delta t)}{(\Delta x)^2} - \\ &= f(x,t,u) = 0 \end{aligned} \quad (2.5c)$$

$$\text{mit } \Delta x = \frac{b}{m_1}; \quad \Delta t = \frac{c}{m_2}; \quad m_1, m_2 \in \mathbb{N};$$

dann ergibt sich mit

$$x_j = j\Delta x; \quad t_\ell = \ell\Delta t; \quad u^*(x_j, t_\ell) = u_j^{*(\ell)}; \quad f(x_j, t_\ell, u^*(x_j, t_\ell)) = f_j^{*(\ell)}$$

für $x = x_j$, $t = t_\ell$ und $j = 1(1)m_1 - 1$:

a) das explizite Verfahren

$$u_j^{*(\ell+1)} = \left(1 - \frac{2\sigma\Delta t}{(\Delta x)^2}\right) u_j^{*(\ell)} + \frac{\sigma\Delta t}{(\Delta x)^2} (u_{j+1}^{*(\ell)} + u_{j-1}^{*(\ell)}) + \Delta t f_j^{*(\ell)}; \quad (2.6a)$$

$$\ell = 0(1)m_2 - 1$$

b) das 2-Schrittverfahren

$$u_j^{*(\ell+1)} = u_j^{*(\ell-1)} + \frac{2\sigma\Delta t}{(\Delta x)^2} (u_{j+1}^{*(\ell)} - 2u_j^{*(\ell)} + u_{j-1}^{*(\ell)}) + 2\Delta t f_j^{*(\ell)}; \quad (2.6b)$$

$$\ell = 1(1)m_2 - 1$$

c) das implizite Verfahren (Crank-Nicolson)

$$\frac{-\sigma\Delta t}{(\Delta x)^2} u_{j+1}^{*(\ell+1)} + \left(1 + \frac{2\sigma\Delta t}{(\Delta x)^2}\right) u_j^{*(\ell+1)} - \frac{\sigma\Delta t}{(\Delta x)^2} u_{j-1}^{*(\ell+1)} = u_j^{*(\ell)} + \Delta t f_j^{*(\ell)} \quad (2.6c)$$

$$\ell = 0(1)m_2 - 1$$

In natürlicher Verallgemeinerung von Definition 2.1. nennen wir $T_{\Delta x, \Delta t}$ konsistent, wenn für die Lösung u von (2.4) gilt:

$$\lim_{\substack{\Delta x \rightarrow 0 \\ \Delta t \rightarrow 0}} \max_{x, t \in I'} \|T_{\Delta x, \Delta t}[u](x, t)\| = 0 \quad (2.7)$$

wobei I' die Menge aller x und t bezeichnet, für die $T_{\Delta x, \Delta t}$ definiert ist. $T_{\Delta x, \Delta t}$ hat die Konsistenzordnung $q = \min(r, s)$ wenn für die Lösung u von (2.4) gilt:

$$\max_{x, t \in I'} \|T_{\Delta x, \Delta t}[u](x, t)\| = \mathcal{O}((\Delta x)^r) + \mathcal{O}((\Delta t)^s)$$

Durch Taylorentwicklung erhält man:

$$T_{\Delta x, \Delta t}^{(a)}[u](x, t) = \frac{\Delta t}{2!} u_{tt}(x, t) - \sigma \frac{(\Delta x)^2}{12} u_{xxxx}(x, t) + \text{Glieder höh. Ordng. in } \Delta x, \Delta t$$

$$T_{\Delta x, \Delta t}^{(b)}[u](x, t) = \frac{(\Delta t)^2}{3!} u_{ttt}(x, t) - \sigma \frac{(\Delta x)^2}{12} u_{xxxx}(x, t) + \text{Glieder höh. Ordng. in } \Delta x, \Delta t$$

$$T_{\Delta x, \Delta t}^{(c)}[u](x, t) = \frac{\Delta t}{2} \frac{\partial}{\partial t} f(x, t, u(x, t)) + \frac{\sigma}{12} \frac{(\Delta t)^4}{(\Delta x)^2} u_{tttt}(x, t) +$$

$$+ \text{Glieder mit } (\Delta x)^r, (\Delta t)^r \text{ und } \frac{(\Delta t)^{r+2}}{(\Delta x)^2}; r=2, 3, \dots$$

Die Verfahren a) und b) haben daher die Konsistenzordnung 1 bzw. 2, wenn u hinreichend oft differenzierbar ist.

Für das Verfahren c) ist die Konsistenzbedingung (2.7) nur dann erfüllt, wenn zwischen Δx und Δt eine Beziehung besteht, so daß gilt:

$$\lim_{\Delta x, \Delta t \rightarrow 0} \frac{(\Delta t)^2}{\Delta x} = 0$$

In einem solchen Fall heißt das Verfahren bedingt konsistent. Für $\Delta t = \text{const.} \Delta x$, zum Beispiel, hat das Verfahren die Ordnung 2, falls $\frac{\partial}{\partial t} f(x, t, u(x, t)) \neq 0$, anderenfalls die Ordnung 1.

Untersuchen wir die Konsistenz von Ein- oder Mehrschrittverfahren der Form (1.5) und (1.6), so ergeben sich die folgenden Konsistenzsätze:

Satz 2.1.

Das Einschrittverfahren $y_{j+1}^* = y_j^* + h\phi(x_j, y_j^*; h)$; $y_0^* = \eta_0$ zur Approximation der Lösung $y(x_{j+1})$ von (1.1) ist konsistent, wenn

1. $\phi(x, y; 0) = f(x, y)$
2. $\phi(x, y; h)$ stetig in $G^* := \{(x, y, h) : a \leq x \leq b; y \in \mathbb{R}^n; 0 \leq h \leq h_0\}$ ist.

Sind darüber hinaus alle partiellen Ableitungen von der Ordnung q stetig, so hat es die Konsistenzordnung q, wenn gilt:

$$\phi(x, y; h) = f(x, y) + \frac{h}{2!} Df(x, y) + \dots + \frac{h^{q-1}}{q!} D^{q-1}f(x, y) + \mathcal{O}(h^q) \quad (2.8)$$

wobei $Df(x, y)$ definiert ist durch $Df(x, y) = (\frac{\partial}{\partial x} + f \frac{\partial}{\partial y})f(x, y)$; d.h. zum Beispiel

$$Df = f_x + f f_y$$

$$D^2 f = (\frac{\partial}{\partial x} + f \frac{\partial}{\partial y})(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y})$$

$$= f_{xx} + 2f f_{xy} + f^2 f_{yy} + f_x f_y + f f_y^2; \text{ etc.}$$

Beweis:

Für die Lösung y von (1.1) gilt mit $0 < \delta < 1$

$$\begin{aligned} \max_{x \in I'_h} \|T_h[y](x)\| &= \max_{x \in I'_h} \left\| \frac{y(x+h) - y(x)}{h} - \phi(x, y(x); h) \right\| \leq \\ &\leq \max_{x \in I'_h} \|y'(x+\delta h) - \phi(x, y(x); 0)\| + \max_{x \in I'_h} \|\phi(x, y(x); 0) - \phi(x, y(x); h)\|; \end{aligned}$$

Wegen $\phi(x, y(x); 0) = f(x, y(x)) = y'(x)$, $y \in C^1[a, b]$ und der gleichmäßigen Stetigkeit von $\phi(x, y(x); h)$ im abgeschlossenen Gebiet $a \leq x \leq b$; $0 \leq h \leq h_0$ geht dieser Ausdruck mit $h \rightarrow 0$ gegen Null; das Verfahren ist daher konsistent. Wenn alle partiellen Ableitungen von $f(x, y)$ der Ordnung q stetig sind, dann gilt $y \in C^{q+1}[a, b]$ und es folgt:

$$\begin{aligned} \|T_h[y](x)\| &= \left\| \frac{y(x+h) - y(x)}{h} - \phi(x, y(x); h) \right\| \\ &= \left\| y'(x) + \frac{h}{2!} y''(x) + \dots + \frac{h^q}{(q+1)!} y^{(q+1)}(\xi) - \phi(x, y(x); h) \right\|; \\ &= O(h^q) \quad x < \xi < x+h \end{aligned}$$

denn aus $y' = f(x, y(x))$ folgt
 $y^{(n+1)}(x) = D^n f(x, y(x)); \quad n = 1(1)q-1.$

Satz 2.2.

Das k -Schrittverfahren der Form (1.6):

$$\alpha_k y_{j+k}^* + \alpha_{k-1} y_{j+k-1}^* + \dots + \alpha_0 y_j^* - h(\beta_k f_{j+k}^* + \beta_{k-1} f_{j+k-1}^* + \dots + \beta_0 f_j^*) = 0;$$

$j = 0(1)m-k$

ist konsistent mit (1.1), wenn gilt

$$\sum_{i=0}^k \alpha_i = 0; \quad \sum_{i=1}^k i \cdot \alpha_i = \sum_{i=0}^k \beta_i \quad (2.9)$$

Für $y \in \mathbb{C}^{q+2}[a, b]$, $q \geq 1$, hat es die Konsistenzordnung q , wenn $c_0 = c_1 = \dots = c_q = 0$; $c_{q+1} \neq 0$, wobei die Konstanten c_ℓ , $\ell = 0(1)q+1$ definiert sind durch:

$$\begin{aligned} c_0 &= \sum_{i=0}^k \alpha_i; & c_1 &= \sum_{i=1}^k i \alpha_i - \sum_{i=0}^k \beta_i \\ c_\ell &= \frac{1}{\ell!} \sum_{i=1}^k i^\ell \alpha_i - \frac{1}{(\ell-1)!} \sum_{i=1}^k i^{\ell-1} \beta_i; & \ell &= 2(1)q+1. \end{aligned} \quad (2.10)$$

Beweis:

Wegen $y \in \mathbb{C}^1[a, b]$ gilt:

$$\begin{aligned} \|T_h[y](x)\| &= \left\| \frac{1}{h} \sum_{j=0}^k \alpha_j y(x+jh) - \sum_{j=0}^k \beta_j f(x+jh, y(x+jh)) \right\| \\ &= \left\| \frac{y(x)}{h} \sum_{j=0}^k \alpha_j + \sum_{j=0}^k y'(\xi_j) j \alpha_j - \sum_{j=0}^k \beta_j f(x+jh, y(x+jh)) \right\|; \\ &\hspace{25em} x < \xi_j < x+jh \end{aligned}$$

Wegen $\lim_{h \rightarrow 0} y'(\xi_j) = y'(x) = f(x, y(x))$; $j = 1(1)k$.

$$\lim_{h \rightarrow 0} \sum_{j=0}^k \beta_j f(x+jh, y(x+jh)) = f(x, y(x)) \sum_{j=0}^k \beta_j$$

geht $\|T_h[y](x)\|$ mit $h \rightarrow 0$ gegen Null, wenn die Gleichungen (2.9) erfüllt sind.

Ist $y \in \mathbb{C}^{q+2}[a, b]$, so folgt aus $f(x, y(x)) = y'(x)$ und einer Taylor-entwicklung:

$$\begin{aligned} \|T_h[y](x)\| &= \left\| \frac{1}{h} \sum_{j=0}^k \alpha_j y(x+jh) - \sum_{j=0}^k \beta_j f(x+jh, y(x+jh)) \right\| \\ &= \left\| \frac{c_0}{h} y(x) + c_1 y'(x) + c_2 h y''(x) + \dots \right. \\ &\quad \left. \dots + c_q h^{q-1} y^{(q)}(x) + c_{q+1} h^q y^{(q+1)}(x) + \mathcal{O}(h^{q+1}) \right\| \\ &= \mathcal{O}(h^q), \text{ falls } c_0 = c_1 = \dots = c_q = 0; \quad c_{q+1} \neq 0 \end{aligned} \quad (2.11)$$

Das beweist den zweiten Teil der Behauptung.

Definition 2.2.

Hat (1.6) die Konsistenzordnung q , so heißt c_{q+1} Fehlerkonstante des Verfahrens. Die Polynome

$$\begin{aligned}\rho(\mu) &= \alpha_k \mu^k + \alpha_{k-1} \mu^{k-1} + \dots + \alpha_1 \mu + \alpha_0 \\ \sigma(\mu) &= \beta_k \mu^k + \beta_{k-1} \mu^{k-1} + \dots + \beta_1 \mu + \beta_0\end{aligned}\tag{2.12}$$

heißen erzeugende Polynome des Verfahrens (1.6). Dieses wird daher häufig mit $M(\rho, \sigma)$ bezeichnet.

Mittels der erzeugenden Polynome lassen sich die Konsistenzbedingungen (2.9) kürzer schreiben, nämlich:

$$\rho(1) = 0 ; \quad \rho'(1) = \sigma(1) \tag{2.13}$$

Im folgenden wird stets $\alpha_k = 1$ gesetzt und, falls nichts anderes gesagt, werden ρ und σ teilerfremd vorausgesetzt.

3. Stabilität und Konvergenz

Die Konsistenz eines Operators gewährleistet eine sinnvolle Approximation der Gleichung $T[y](x) = \varphi$ durch eine Differenzengleichung $T_h[y^*](x) = \varphi$, was keineswegs einschließt, daß auch die Lösungen dieser Gleichungen einander approximieren!

Es gibt vielmehr Differenzengleichungen $T_h[y^*](x_j) = \varphi$, deren Lösung sich mit kleinem h beliebig stark verändert, wenn die Anfangswerte oder der Operator T_h auch nur ganz geringfügig verändert werden. Solche Gleichungen heißen instabil, und man wird nicht erwarten können, daß ihre Lösungen Näherungen für die Lösung y von $T[y](x) = \varphi$ sind, da sie ja nicht einmal die Lösung einer beliebig "ähnlichen" Differenzengleichung $\hat{T}_h[y^*](x) = \varphi$ approximieren.

3.1. Ein Beispiel

Zur Illustration approximieren wir die Differentialgleichung:

$$T[y](x) \equiv y'(x) - \lambda y(x) = 0; \quad y(0) = 1; \quad y: \mathbb{R} \rightarrow \mathbb{R} \quad (3.1)$$

durch:

$$T_h[y^*](x) \equiv \frac{-y^*(x+h) + 4y^*(x) - 3y^*(x-h)}{2h} - \lambda y^*(x) = 0; \quad (3.2)$$

$$y^*(0) = 1; \quad y^*(h) = \eta_1$$

Die Konsistenz von T_h folgt aus (2.13) mit $\rho(\mu) = -\mu^2 + 4\mu - 3$ und $\sigma(\mu) = 2\mu$. Für $x_j = jh$ erhält man das Verfahren

$$y_{j+1}^* + (2\lambda h - 4)y_j^* + 3y_{j-1}^* = 0; \quad y_0^* = 1; \quad y_1^* = \eta_1 \quad (3.3)$$

$$(j = 1(1)m-1)$$

Diese Differenzengleichung ist linear und daher analytisch lösbar. Der Ansatz $y_j^* = C\mu^j$ führt auf die charakteristische Gleichung:

$$\mu^2 + (2\lambda h - 4)\mu + 3 = 0 \quad (3.4)$$

und man erhält die allgemeine Lösung von (3.3) in der Form

$$y_j^* = C_1 \mu_1^j + C_2 \mu_2^j \quad (3.5)$$

wobei C_1 und C_2 Konstanten sind, die von den Anfangsbedingungen y_0^* und y_1^* sowie von h abhängen, und $\mu_1(h)$ und $\mu_2(h)$ die Nullstellen von (3.4):

$$\mu_1(h) := (2 - \lambda h) - \sqrt{1 - 4\lambda h + (\lambda h)^2} = e^{\lambda h} + \mathcal{O}(h^2)$$

$$\mu_2(h) := (2 - \lambda h) + \sqrt{1 - 4\lambda h + (\lambda h)^2} = 3e^{-\lambda h} + \mathcal{O}(h^2)$$

(den 2. Teil dieser Gleichung zeigt man durch Taylorentwicklung)

Aus den Anfangsbedingungen folgt

$$1 = C_1 + C_2 ; \quad \eta_1 = C_1 \mu_1 + C_2 \mu_2$$

$$C_1 = \frac{\mu_2 - \eta_1}{\mu_2 - \mu_1} ; \quad C_2 = \frac{\mu_1 - \eta_1}{\mu_1 - \mu_2}$$

μ_1^j approximiert die Lösung $y(x_j) = e^{\lambda h j}$ der Dgl. (3.1) mit der Ordnung $(q+1)$ in h ($q=1$ ist die Konsistenzordnung des Verfahrens (3.3)); eine Wurzel von (3.4) mit dieser Eigenschaft werden wir in Abschnitt 7 Hauptwurzel nennen.

Außer der Hauptwurzel tritt in der Näherung (3.5) noch eine weitere Wurzel μ_2 auf; im allgemeineren Fall eines k -Schritt-Verfahrens der Form (1.6) gibt es $(k-1)$ derartige weitere Wurzeln. Diese Nebenwurzeln haben nichts mit der Lösung der Differentialgleichung (3.1) zu tun; sie sind jedoch unvermeidlich und Folge der Approximation einer Differentialgleichung ersten Grades durch eine Differenzengleichung k -ten Grades.

Betrachten wir nun die Lösung von (3.3):

$$y_j^* = C_1 (e^{\lambda h} + \mathcal{O}(h^2))^j + C_2 (3e^{-\lambda h} + \mathcal{O}(h^2))^j$$

d.h. mit $jh = x_j$

$$y_j^* = C_1 e^{\lambda x_j} + C_2 3^{x_j/h} \cdot e^{-\lambda x_j} + \mathcal{O}(h^2)$$

Nehmen wir noch den günstigsten Fall an, daß der Startwert η_1 exakt ist, d.h. $\eta_1 = e^{\lambda h}$, so folgt:

$$y_j^* = (1 + \mathcal{O}(h^2)) e^{\lambda x_j} + \mathcal{O}(h^2) 3^{x_j/h} e^{-\lambda x_j} + \mathcal{O}(h^2) \quad (3.6)$$

Der erste Summand approximiert die exakte Lösung $y(x) = e^{\lambda x}$ der Differentialgleichung mit der Ordnung 2 in h . Der zweite Summand jedoch ist Folge der Nebenwurzel μ_2 und verfälscht das Ergebnis völlig, und zwar umso mehr, je kleiner h ist.

Dieser Effekt ist typisch für ein "instabiles" Verfahren und tritt offensichtlich immer dann auf, wenn eine Nullstelle der charakteristischen Gleichung (3.4) betragsmäßig größer ist als 1. Als weiteres, vorläufiges Ergebnis dieses Beispiels halten wir noch fest, daß die Nebenwurzeln μ_ℓ ($\ell=2(1)k$) auch betragsmäßig kleiner als die Hauptwurzel μ_1 sein sollten, um zu vermeiden, daß der durch sie entstandene Lösungsanteil bei großem j überwiegt.

3.2. Definition der Stabilität

Mit der Existenz eines konvergenten numerischen Verfahrens zur Lösung eines mathematischen Problems ist keineswegs gesichert, daß die Lösung (oder eine gute Approximation) mit dem Verfahren auch berechnet werden kann. Dieser scheinbare Widerspruch läßt sich sehr einfach am Beispiel des vorigen Abschnitts erläutern: Für $\eta_1=\mu_1$ konvergiert das Verfahren (3.3) mit $h \rightarrow 0$ gegen die Lösung $y(x)=e^{\lambda x}$ von (3.1); dennoch ist es selbst für $\eta_1=\mu_1$ nicht geeignet, gute Näherungen von $y(x)$ auch praktisch zu berechnen, denn diese werden für kleines h durch unvermeidliche Rundungsfehler völlig verfälscht. Praktikabel ist nämlich ein Diskretisierungsverfahren erst dann, wenn der Einfluß von Fehlern für alle Werte des Diskretisierungsparameters beschränkt bleibt. In diesem Fall nennen wir das Verfahren stabil. Stabilität ist somit die Eigenschaft, die ein Verfahren überhaupt erst für den praktischen Gebrauch geeignet macht!

Wir wollen nun den Begriff der Stabilität präzisieren: Seien $v: I_h \rightarrow \mathbb{R}^n$ und $w: I_h \rightarrow \mathbb{R}^n$ Lösungen der k -Schritt-Verfahren:

$$\begin{aligned} T_h[v](x_j) + S_h[v](x_j) &= 0 \quad \text{mit } v_j = \eta_j + \mu_j \quad (j=0(1)k-1) \\ &\quad j=0(1)m-k \end{aligned} \quad (3.7)$$

$$T_h[w](x_j) + R_h[w](x_j) = 0 \quad \text{mit } w_j = \eta_j + \omega_j \quad (j=0(1)k-1)$$

wobei $S_h, R_h, \mu_j, \omega_j$ als "Störungen" von

$$T_h[y^*](x_j) = 0; \quad y_j^* = \eta_j \quad (j = o(1)k-1) \quad (3.8)$$

interpretieren, die z.B. durch Ersetzungsfehler oder Rundungen zustande gekommen sein können. Unterscheiden sich S_h und R_h für alle $h \in (0, h_0]$ nur wenig:

$$\text{d.h. } \|S_h[v](x_j) - R_h[w](x_j)\| < \delta_1; \quad j = o(1)m-k; \quad 0 < \delta_1 \ll 1$$

und unterscheiden sich auch die Anfangswerte wenig:

$$\text{d.h. } \|\mu_j - \omega_j\| < \delta_2; \quad j = o(1)k-1; \quad 0 < \delta_2 \ll 1$$

so wollen wir (3.8) stabil nennen, wenn sich auch v und w wenig voneinander unterscheiden. Wir definieren daher:

Definition 3.1.

Ein k -Schritt-Verfahren $T_h[y^*](x_j) = 0; y_j^* = \eta_j \quad (j = o(1)k-1)$, $y^*: I_h \rightarrow \mathbb{R}^n$, heißt stabil, wenn es eine von h unabhängige Konstante C gibt, so daß für beliebige Funktionen $v, w: I_h \rightarrow \mathbb{R}^n$ gilt:

$$\|v_j - w_j\| \leq C \left\{ \max_{0 \leq i \leq k-1} \|v_i - w_i\| + \max_{0 \leq i \leq m-k} \|T_h[v](x_i) - T_h[w](x_i)\| \right\} \quad (3.9)$$

$j = o(1)m$

anderenfalls heißt das Verfahren instabil.

Ist T_h linear, so reduziert sich die Stabilitätsbedingung (3.9) mit $z = v - w$ auf:

$$\|z_j\| \leq C \left\{ \max_{0 \leq i \leq k-1} \|z_i\| + \max_{0 \leq i \leq m-k} \|T_h[z](x_i)\| \right\} \quad (3.10)$$

$j = o(1)m$

Beispiel

Betrachten wir das Verfahren von Euler:

$$hT_h[y^*](x_j) \equiv y_{j+1}^* - y_j^* - hf(x_j, y_j^*) = 0; \quad j = 0(1)m-1 \quad (3.11)$$

Für beliebige $v, w: I_h \rightarrow \mathbb{R}$ erhält man die Identitäten:

$$v_{j+1} = v_j + hf(x_j, v_j) + hT_h[v](x_j)$$

$$w_{j+1} = w_j + hf(x_j, w_j) + hT_h[w](x_j)$$

und mit der Lipschitzbedingung (1.2):

$$\|v_{j+1} - w_{j+1}\| \leq (1+hL)\|v_j - w_j\| + h \cdot \max_{0 \leq i \leq m-1} \|T_h[v](x_i) - T_h[w](x_i)\|; \\ j = 0(1)m-1$$

Wir setzen: $(1+hL) = A$; $\max_{0 \leq i \leq m-1} \|T_h[v](x_i) - T_h[w](x_i)\| = B$; dann gilt:

$$\|v_1 - w_1\| \leq A \|v_0 - w_0\| + hB$$

$$\|v_2 - w_2\| \leq A \|v_1 - w_1\| + hB \leq A^2 \|v_0 - w_0\| + hB(A+1)$$

$$\|v_3 - w_3\| \leq A \|v_2 - w_2\| + hB \leq A^3 \|v_0 - w_0\| + hB(A^2 + A + 1); \text{ etc.}$$

Durch Induktion erhält man:

$$\|v_j - w_j\| \leq A^j \|v_0 - w_0\| + hB(A^{j-1} + A^{j-2} + \dots + A + 1) \\ \leq A^j \|v_0 - w_0\| + hB \frac{A^j - 1}{A - 1}$$

Ersetzen von A und B ergibt:

$$\|v_j - w_j\| \leq (1+hL)^j \|v_0 - w_0\| + L^{-1}((1+hL)^j - 1) \max_{0 \leq i \leq m-1} \|T_h[v](x_i) - T_h[w](x_i)\|$$

Wegen¹⁾ $(1+hL)^j \leq e^{hjL} = e^{L(x_j-a)}$ folgt mit $C := e^{(b-a)L} \max(1, L^{-1})$

$$\|v_j - w_j\| \leq C(\|v_0 - w_0\| + \max_{0 \leq i \leq m-1} \|T_h[v](x_i) - T_h[w](x_i)\|)$$

und damit die Stabilität des Verfahrens.

Für $v_j = y(x_j)$ und $w_j = y_j^*$, erhält man die folgende Fehlerabschätzung für das Verfahren von Euler:

$$\|y(x_j) - y_j^*\| \leq \frac{(1+hL)^j - 1}{L} \max_{0 \leq i \leq m-1} \|T_h[y](x_i)\|; \quad j = o(1)m$$

Für $y \in C^2[a, b]$ und $K = \frac{1}{2} \max_{x \in [a, b]} \|y''(x)\|$ folgt aus

$$\|T_h[y](x_j)\| = \left\| \frac{y(x_j+h) - y(x_j)}{h} - f(x_j, y(x_j)) \right\| = \frac{h}{2} \|y''(\xi_j)\| \leq hK \quad (3.12)$$

$$x_j < \xi_j < x_j + h$$

die Ungleichung

$$\|y(x_j) - y_j^*\| \leq hKL^{-1} \{(1+hL)^j - 1\} \leq hKL^{-1} e^{L(x_j-a)}$$

y_j^* konvergiert somit gegen $y(x_j)$ wenigstens mit der Ordnung 1 in h .

Hinweis:

Tritt bei der Berechnung von y_{j+1}^* aus (3.11) in jedem Rechenschritt ein Rundungsfehler ϵ_j auf mit $\|\epsilon_j\| \leq \epsilon$ für alle j , d.h. haben wir statt des Operators T_h in (3.11) den Operator T_h^+ mit

$$hT_h^+[y^*](x_j) = y_{j+1}^* - y_j^* - hf(x_j, y_j^*) - \epsilon_j$$

¹⁾ $e^x = 1 + x + \frac{x^2}{2} e^{\eta x}$; $0 < \eta < 1$, daher $1 + x \leq e^x$ für bel. reelles x .

so ist T_h^+ nicht konsistent mit T und (3.12) ist zu ersetzen durch

$$\|T_h^+[y](x_j)\| \leq hK + \frac{\varepsilon}{h}$$

Der Ersetzungsfehler $T_h^+[y](x_j)$ ist somit in $(0, h_0]$ nicht be-
schränkt. Für den globalen Fehler an der Stelle x_j erhält man:

$$\|y(x_j) - y_j^*\| \leq (hK + \frac{\varepsilon}{h}) L^{-1} \{(1+hL)^j - 1\}$$

Wir sehen, daß der durch Rundung entstandene Anteil der Fehler-
schranke mit $h \rightarrow 0$ unbegrenzt wächst; dennoch ist das Verfahren
stabil im Sinne der Definition 3.1. Die Schrittweite h ist
optimal gewählt, wenn gilt: $Kh = \varepsilon h^{-1}$.

3.3. Ein allgemeines Stabilitätskriterium

Die Ergebnisse des vorigen Abschnitts lassen sich ganz wesent-
lich verallgemeinern durch folgenden Satz 3.1., der die Grund-
lage bildet für Stabilitäts- und Konvergenzbetrachtungen sowie
für Fehlerabschätzungen bei Diskretisierungsverfahren sehr all-
gemeiner Art.

Sei T_h definiert durch

$$hT_h[z](x) = z(x+h) - Az(x) - h\phi(x, z(x), z(x+h); h)$$

mit $A \in \mathbb{R}(k, k)$; $z: [a, b] \rightarrow \mathbb{R}^k$; $\phi: [a, b] \times \mathbb{R}^k \times \mathbb{R}^k \times [0, h_0] \rightarrow \mathbb{R}^k$

ϕ genüge für alle $y_1, z_1, y_2, z_2 \in \mathbb{R}^k$, $x \in [a, b]$, $h \in (0, h_0]$ der folgen-
den Lipschitzbedingung mit den Konstanten K_1 und K_2 :

$$\|\phi(x, y_1, z_1; h) - \phi(x, y_2, z_2; h)\| \leq K_1 \|y_1 - y_2\| + K_2 \|z_1 - z_2\| \quad (3.13)$$

und für die Konstante h_0 möge gelten: $h_0 K_2 < 1$.

Satz 3.1. ¹⁾

Die Differenzengleichung

$$z_{j+1}^* = Az_j^* + h\phi(x_j, z_j^*, z_{j+1}^*; h); z_0^* = \zeta_0(h) \quad (3.14)$$

$$j = o(1)m-k$$

ist für alle ϕ , die (3.13) erfüllen, genau dann stabil, wenn es eine Zahl $D \geq 1$ gibt mit $\|A^j\| \leq D$ für alle $j \in \mathbb{N}$. Für eine beliebige Funktion $v: I_h \rightarrow \mathbb{R}^k$ gilt dann:

$$\|v_{j+1} - z_{j+1}^*\| \leq \frac{(x_j - a)u}{1 - hK_2} \{ (1 + hK_1) \|v_0 - z_0^*\| + (x_{j+1} - a) \max_{0 \leq i \leq m-1} \|T_h[v](x_i)\| \}$$

$$\text{mit } u := D \frac{K_1 + K_2}{1 - hK_2} \quad j = o(1)m-k \quad (3.15)$$

Anmerkungen:

1. Nach dem Banachschen Fixpunktsatz ist z_{j+1}^* wegen $h_0 K_2 < 1$ eindeutig durch (3.14) bestimmt und läßt sich iterativ berechnen.
2. In der Matrix-Theorie wird gezeigt, daß die Zahl D existiert wenn:
 - a) für den Spektralradius von A gilt: $\rho(A) \leq 1$ und
 - b) zu jedem Eigenwert vom Betrag 1 (sofern vorhanden) nur lineare Elementarteiler gehören.

Der Beweis von Satz 3.1. ähnelt dem Stabilitätsbeweis für das Verfahren von Euler:

Für beliebige Funktionen $v, w: I_h \rightarrow \mathbb{R}^k$ folgt aus

$$v_{j+1} = Av_j + h\phi(x_j, v_j, v_{j+1}; h) + hT_h[v](x_j)$$

$$w_{j+1} = Aw_j + h\phi(x_j, w_j, w_{j+1}; h) + hT_h[w](x_j)$$

$$j = o(1)m-k$$

¹⁾ P. Albrecht: "Discretization Methods", 9. Coll. Bras. Mat., IMPA, Rio de Janeiro, 1973

die Gleichung

$$\begin{aligned} v_{j+1} - w_{j+1} = & A(v_j - w_j) + h(\phi(x_j, v_j, v_{j+1}; h) - \phi(x_j, w_j, w_{j+1}; h)) \\ & + h(T_h[v](x_j) - T_h[w](x_j)) \end{aligned}$$

Mit $q_j := v_j - w_j$; $\varphi_j := \phi(x_j, v_j, v_{j+1}; h) - \phi(x_j, w_j, w_{j+1}; h)$ ergibt sich

$$q_{j+1} = A q_j + h \varphi_j + h(T_h[v](x_j) - T_h[w](x_j)).$$

Durch Induktion findet man

$$q_{j+1} = A^{j+1} q_0 + h \sum_{\ell=0}^j A^{j-\ell} \varphi_\ell + h \sum_{\ell=0}^j A^{j-\ell} (T_h[v](x_\ell) - T_h[w](x_\ell)). \quad (3.16)$$

Mit $\Delta := \max_{0 \leq \ell \leq m-k} \|T_h[v](x_\ell) - T_h[w](x_\ell)\|$

und $\|A\| \leq D$ für alle $\ell \in \mathbb{N}$ folgt:

$$\|q_{j+1}\| \leq D \|q_0\| + h \sum_{\ell=0}^{j-1} D(K_1 \|q_\ell\| + K_2 \|q_{\ell+1}\|) + h(K_1 \|q_j\| + K_2 \|q_{j+1}\|) + h(j+1)D\Delta$$

Somit

$$\|q_1\| \leq \frac{D}{1-hK_2} \{(1+hK_1)\|q_0\| + h\Delta\} \quad (3.17)$$

$$\|q_{j+1}\| \leq \frac{D}{1-hK_2} \{(1+hK_1)\|q_0\| + h(j+1)\Delta + h(K_1+K_2) \sum_{\ell=1}^j \|q_\ell\|\} \quad (3.18)$$

Wir zeigen:

$$\|q_j\| \leq \frac{D}{1-hK_2} \{(1+hK_1)\|q_0\| + hj\Delta\} \cdot \left\{1 + h \frac{(K_1+K_2)D}{1-hK_2}\right\}^{j-1} \quad (3.19)$$

Beweis durch Induktion: (3.19) ist richtig für $j=1$, wie (3.17) zeigt. Wenn es auch für j richtig ist, so ergibt die Substitution von (3.19) in (3.18) mit $u:=(K_1+K_2)D/(1-hK_2)$:

$$\|q_{j+1}\| \leq \frac{D}{1-hK_2} \left\{ (1+hK_1)\|q_0\| + h(j+1)\Delta + hu \sum_{\ell=1}^j [(1+hK_1)\|q_0\| + h\ell\Delta] \cdot (1+hu)^{\ell-1} \right\}$$

$h\ell\Delta$ durch $h(j+1)\Delta$ ersetzt ergibt:

$$\begin{aligned} \|q_{j+1}\| &\leq \frac{D}{1-hK_2} \{ (1+hK_1)\|q_0\| + h(j+1)\Delta \} \{ 1+hu \sum_{\ell=1}^j (1+hu)^{\ell-1} \} = \\ &= \frac{D}{1-hK_2} [(1+hK_1)\|q_0\| + h(j+1)\Delta] \cdot (1+hu)^j \end{aligned} \quad (3.20)$$

Somit gilt (3.19) für alle $j \leq m-k+1$

Mit $(1+hu) \leq e^{hu}$; $jh = x_j - a$ folgt

$$\begin{aligned} \|v_{j+1} - w_{j+1}\| &\leq \frac{D}{1-hK_2} e^{(x_j-a)u} \left[(1+hK_1)\|v_0 - w_0\| + \right. \\ &\quad \left. + (x_{j+1}-a) \max_{0 \leq i \leq m-k} \|T_h[v](x_i) - T_h[w](x_i)\| \right] \end{aligned} \quad (3.21)$$

Wegen $(x_j-a) \leq b-a$ für $j=0(1)m$ ist die Stabilitätsbedingung (3.9) erfüllt mit:

$$C = \frac{De^{(b-a)u_0}}{1-h_0K_2} \max [(1+h_0K_1), (b-a)]; \quad u_0 := \frac{D(K_1+K_2)}{1-h_0K_2}$$

Für $w_j = z_j^*$ ergibt sich (3.15), denn $T_h[z^*](x_j) = 0$.

Ist umgekehrt (3.14) stabil für beliebiges ϕ , welches (3.13) erfüllt, so auch für $\phi \equiv 0$ und es gilt:

$$\|v_{j+1} - w_{j+1}\| = \|A^{j+1}(v_0 - w_0)\| \leq C \|v_0 - w_0\|$$

für alle j und beliebige v_0, w_0 mit $v_0 \neq w_0$. Das ist nur möglich, wenn der Spektralradius $\rho(A) \leq 1$ ist und wenn zu Eigenwerten vom Betrage 1 nur lineare Elementarteiler gehören. Damit ist Satz 3.1. bewiesen. $\triangleleft$

In Abschnitt 5.1. werden wir zeigen, daß die Fehlerformel (3.15) unter bestimmten Voraussetzungen erheblich verschärft werden kann.

3.4. Konvergenz

Definition 3.2.

Die Lösung $y^*(x)$, $x \in I_h$, eines Diskretisierungsverfahrens $T_h[y^*](x) = \mathcal{O}$ konvergiert (für $h \rightarrow 0$) gegen die Lösung $y(x)$, $x \in I_h$, von $T[y](x) = \mathcal{O}$, wenn gilt:

$$\lim_{h \rightarrow 0} \|y(x) - y^*(x)\| = 0$$

Ein Verfahren heißt konvergent mit der Konvergenzordnung q , wenn für alle $x \in I_h$ gilt:

$$\|y(x) - y^*(x)\| = \mathcal{O}(h^q) \quad (3.22)$$

Konsistenz und Stabilität sind hinreichende Bedingung für die Konvergenz von Diskretisierungsverfahren, wie der folgende Satz zeigt; sie sind nicht notwendig, wie schon Beispiel 3.1. verdeutlichte, jedoch sind inkonsistente und instabile Verfahren für die Praxis ohne Belang.

Satz 3.2.

Ein k -Schritt-Verfahren ($k \geq 1$)

$$T_h[y^*](x_j) = \mathcal{O} \quad (j = o(1)m - k) \text{ mit } y_j^* = \eta_j(h) \quad (j = o(1)k - 1)$$

ist konvergent, wenn es konsistent und stabil ist und wenn auch seine Startwerte konsistent sind,

$$\text{d.h.} \quad \lim_{h \rightarrow 0} \max_{0 \leq j \leq k-1} \|y(x_j) - \eta_j(h)\| = 0 \quad (3.23)$$

Beweis:

Die Stabilitätsdefinition (3.9) ergibt für $v_j := y_j$ und $w_j := y_j^*$ mit $T_h[y^*](x_j) = \mathcal{O}$:

$$\|y(x_j) - y_j^*\| \leq C \left\{ \max_{0 \leq j \leq k-1} \|y(x_j) - \eta_j\| + \max_{0 \leq j \leq m-k} \|T_h[y](x_j)\| \right\} \quad (3.24)$$

Die Behauptung folgt jetzt aus (3.23) und (2.1). $\triangleleft$

Aus (3.24) ergibt sich unmittelbar:

Korollar 3.2.

Ist (zusätzlich zu den Voraussetzungen in Satz 3.2.)

$$\|y(x_j) - \eta_j(h)\| = \mathcal{O}(h^q), \quad j=o(1)k-1$$

und hat das Verfahren $T_h[y^*](x_j) = \sigma$ die Konsistenzordnung q , so gilt für $j=o(1)m$:

$$\|y(x_j) - y_j^*\| = \mathcal{O}(h^q)$$

Die Konvergenzordnung eines stabilen Verfahrens mit geeigneten Startwerten ist somit (wenigstens) gleich seiner Konsistenzordnung. Bei bestimmten Verfahren kann die Konvergenzordnung größer sein als die Konsistenzordnung, wie in Kapitel 5 gezeigt werden wird (vgl. Satz 5.1.); bei linearen k -Schritt-Verfahren ist das jedoch nicht der Fall¹⁾.

Die Koeffizienten α_i und β_i , $i=o(1)k$, eines linearen Mehrschrittverfahrens der Konsistenzordnung q :

$$\alpha_k y_{j+k}^* + \alpha_{k-1} y_{j+k-1}^* + \dots + \alpha_o y_j^* - h(\beta_k f_{j+k}^* + \dots + \beta_o f_j^*) = \sigma; \quad \alpha_k = 1 \quad (3.25)$$

können nach Satz 2.2. aus dem linearen Gleichungssystem (2.10):

$$\begin{aligned} c_o &:= \sum_{i=0}^k \alpha_i = 0; & c_1 &:= \sum_{i=1}^k i\alpha_i - \sum_{i=0}^k \beta_i = 0 \\ c_\ell &:= \frac{1}{\ell!} \sum_{i=1}^k i^\ell \alpha_i - \frac{1}{(\ell-1)!} \sum_{i=1}^k i^{\ell-1} \beta_i = 0; & \ell &= 2(1)q \end{aligned} \quad (3.26)$$

bestimmt werden.

Theoretisch ist daher für implizite Verfahren ($\beta_k \neq 0$) die Konsistenzordnung $q=2k$ möglich und für explizite Verfahren ($\beta_k = 0$) die Konsistenzordnung $q=2k-1$, denn es sind $(2k+1)$ bzw. $2k$ Koeffizienten verfügbar, um die $q+1$ Gleichungen (9.2) zu erfüllen.

¹⁾ daher sprechen wir bei diesen oft nur von Ordnung und unterscheiden nicht zwischen Konsistenz- und Konvergenzordnung.

Es zeigt sich jedoch, daß (für festes k) alle Verfahren ab einer bestimmten Konsistenzordnung instabil sind; insbesondere gilt der folgende Satz von Dahlquist¹⁾:

Satz 3.3.

Ein stabiles, explizites k -Schrittverfahren ($\beta_k=0$) der Form (1.6) hat höchstens die Konvergenzordnung $q=k$; ein stabiles implizites k -Schrittverfahren ($\beta_k \neq 0$) hat höchstens die Konvergenzordnung $q=k+2$, wenn k gerade ist und $q=k+1$, wenn k ungerade ist.

Dieser Satz verdeutlicht nochmals, daß eine Approximation höherer Ordnung des Differentialoperators T durch Differenzoperatoren T_h nicht notwendig zu genaueren numerischen Verfahren führen muß.

Beispiele für stabile Verfahren maximaler Ordnung

1. Das explizite 3-Schrittverfahren

$$y_{j+1}^* = \frac{1}{3} (y_j^* + y_{j-1}^* + y_{j-2}^*) + \frac{h}{6} (13f_j^* - 4f_{j-1}^* + 3f_{j-2}^*)$$

ist stabil und hat für $y \in C^4[a,b]$ die Konsistenzordnung 3. Nach Korollar 3.2. hat es bei geeigneten Startwerten dann auch die Konvergenzordnung 3.

2. Das implizite 2-Schrittverfahren von Milne-Simpson:

$$y_{j+1}^* = y_{j-1}^* + \frac{h}{3} (f_{j+1}^* + 4f_j^* + f_{j-1}^*)$$

ist stabil und hat für $y \in C^5[a,b]$ die Konsistenzordnung 4 (wie im Beispiel 2 des Kapitels 2 gezeigt wurde). In Abschnitt 9.3. wird gezeigt, daß es das einzige stabile lineare 2-Schritt-Verfahren mit der Konsistenzordnung 4 ist.

¹⁾Dahlquist: Convergence and Stability in the Numerical Integration of O.D.E., Math. Scand. 4, 33-53 (1956)

4. Anwendungen des Stabilitätssatzes

Dieser Paragraph behandelt die vielfältigen Anwendungsmöglichkeiten des Satzes 3.1. und verdeutlicht die zentrale Stellung, die er für unsere Untersuchungen einnimmt.

In Abschnitt 4.1. wird gezeigt, daß Satz 3.1. nicht nur auf Einschrittverfahren anwendbar ist, wie es zunächst scheinen mag, sondern auch auf k-Schritt-Verfahren der Form (1.6). In Abschnitt 4.2. werden Stabilitäts- und Fehleraussagen für Prädiktor-Korrektor-Verfahren und in Abschnitt 4.3. für zyklische Verfahren gewonnen. Schließlich wird in Abschnitt 4.4. an einem Beispiel gezeigt, wie sich die Ergebnisse des Kapitels 3 auf partielle Differentialgleichungsprobleme übertragen lassen.

4.1. Stabilität von k-Schritt-Verfahren

Mehrschrittverfahren vom Typ $M(\rho, \sigma)$:

$$y_{j+k}^* = -\alpha_{k-1} y_{j+k-1}^* - \alpha_{k-2} y_{j+k-2}^* - \dots - \alpha_0 y_j^* + h(\beta_k f(x_{j+k}, y_{j+k}^*) + \dots + \beta_0 f(x_j, y_j^*)) \quad (4.1)$$

$$y_0^* = \eta_0; \quad y_1^* = \eta_1(h); \dots; y_{k-1}^* = \eta_{k-1}(h) \quad j=0(1)m-k$$

lassen sich sehr einfach auf Einschrittverfahren der Form (3.14) zurückführen; dazu definieren wir (zunächst für $y_j \in \mathbb{R}$):

$$z_j^* := \begin{pmatrix} y_j^* \\ y_{j+1}^* \\ \vdots \\ y_{j+k-1}^* \end{pmatrix}; \quad z_j := \begin{pmatrix} y_j \\ y_{j+1} \\ \vdots \\ y_{j+k-1} \end{pmatrix}; \quad Fz_j^* := \begin{pmatrix} f(x_j, y_j^*) \\ f(x_j+h, y_{j+1}^*) \\ \vdots \\ f(x_j+(k-1)h, y_{j+k-1}^*) \end{pmatrix} \quad (4.2)$$

$$A := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \dots & -\alpha_{k-1} \end{pmatrix}; \quad B := \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ \beta_0 & \beta_1 & \beta_2 & \dots & \beta_{k-1} \end{pmatrix}; \quad C := \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \beta_k \end{pmatrix}$$

Hiermit geht $M(\rho, \sigma)$ über in die Form (3.14):

$$z_{j+1}^* = Az_j^* + h\phi(x_j, z_j^*, z_{j+1}^*; h); \quad z_0^* = \begin{pmatrix} \eta_0 \\ \eta_1 \\ \vdots \\ \eta_{k-1} \end{pmatrix}; \quad j = 0(1)m-k$$

$$\text{mit } \phi(x_j, z_j^*, z_{j+1}^*; h) := BFz_j^* + CFz_{j+1}^*$$

$$\text{und } T_h[z](x_j) := h^{-1} \left[z_{j+1} - Az_j - h\phi(x_j, z_j, z_{j+1}; h) \right]$$

Hat (4.1) die Konsistenzordnung q , so folgt aus (2.11) für alle $x_j \in I_h'$:

$$T_h[z](x_j) = c_{q+1} h^q y^{(q+1)}(x_j) t + \mathcal{O}(h^{q+1}) \quad \text{mit } t := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^k \quad (4.3)$$

Genügt f der Lipschitzbedingung (1.2) mit der Lipschitzkonstanten L , so genügt ϕ der Lipschitzbedingung (3.13) mit

$$K_1 = L \sum_{j=0}^{k-1} |\beta_j| \quad \text{und} \quad K_2 = L |\beta_k| \quad (4.4)$$

Die Matrix A hat Frobeniusform, ihre Eigenwerte sind daher die Nullstellen des Polynoms

$$\rho(\mu) = \mu^k + \alpha_{k-1}\mu^{k-1} + \dots + \alpha_1\mu + \alpha_0$$

und zu mehrfachen Eigenwerten gehören nichtlineare Elementarteiler.

Im Falle $y_j^* \in \mathbb{R}^n$ ist A eine (kn, kn) -Matrix mit Block-Frobeniusform, d.h.

$$A = \begin{pmatrix} \sigma & I & \sigma & \sigma \dots \sigma \\ \sigma & \sigma & I & \sigma \dots \sigma \\ \cdot & \cdot & \cdot & \cdot \\ -\alpha_0 I & -\alpha_1 I & \dots & -\alpha_{k-1} I \end{pmatrix} ; I: (n, n)\text{-Einheitsmatrix}$$

und die Eigenwerte von A sind dann die Nullstellen von $\bar{\rho}(\mu) = (\rho(\mu))^n$. Auch in diesem Fall gehören zu mehrfachen Nullstellen von $\rho(\mu)$ nichtlineare Elementarteiler.

Definiert man die Norm von $(z(x_j) - z_j^*)$ in $\mathbb{R}^{nk}$ durch:

$$\|z(x_j) - z_j^*\| := \max_{0 \leq \ell \leq k-1} \|y(x_{j+\ell}) - y_{j+\ell}^*\|$$

so ist: $\|z(x_j) - z_j^*\| \geq \|y(x_{j+k-1}) - y_{j+k-1}^*\|$ und Satz 3.1. ergibt das folgende, sehr einfach zu handhabende Stabilitätskriterium für $M(\rho, \sigma)$:

Satz 4.1.

Genüge f der Lipschitzbedingung (1.2) und sei $h|\beta_k|L < 1$. Dann ist das Mehrschrittverfahren (1.6) genau dann stabil, wenn

1. keine Nullstelle von $\rho(\mu)$ betragsmäßig größer ist als 1,
2. die Nullstellen vom Betrag 1 einfach sind.

Außerdem gilt die Fehlerabschätzung:

$$\begin{aligned} \|y(x_{j+k}) - y_{j+k}^*\| \leq & \frac{De^{u(x_j-a)}}{1-hK_2} \{ (1+hK_1) \max_{0 \leq i \leq k-1} \|y(x_i) - \eta_i\| + \\ & + (x_{j+1}-a) \max_{0 \leq i \leq m-k} \|T_h[y](x_i)\| \} \end{aligned} \quad (4.5)$$

$j = o(1)m-k$

wobei K_1 und K_2 durch (4.4) gegeben sind und

$$D \geq \|A^j\|^{1)} \quad \text{für alle } j \in \mathbb{N}; \quad u := \frac{D(K_1+K_2)}{1-hK_2} = \frac{DL}{1-hL|\beta_k|} \sum_{j=0}^k |\beta_j|$$

$$\|T_h[y](x)\| = \left\| \frac{1}{h} \sum_{j=0}^k \alpha_j y(x+jh) - \sum_{j=0}^k \beta_j f(x+jh, y(x+jh)) \right\|; \quad \alpha_k = 1$$

Definition 4.1.

Sind die Bedingungen 1. und 2. von Satz 4.1. erfüllt, so sagen wir, daß die Nullstellen von ρ dem Wurzelkriterium genügen. Ist $M(\rho, \sigma)$ stabil und sind alle Nullstellen von ρ außer $\mu_1=1$ betragsmäßig kleiner als 1, so heißt $M(\rho, \sigma)$ stark stabil anderenfalls schwach stabil.

Schwach stabile Verfahren ergeben für "große" j im allgemeinen schlechtere Ergebnisse als stark stabile.

¹⁾ aus $\rho(1)=0$ folgt $\|A^j\| \geq 1$

²⁾ eine Möglichkeit, $\|T_h[y](x)\|$ näherungsweise zu bestimmen, ist im Abschnitt 9.5 dargestellt (vgl. Satz 9.5.).

Beispiel 1 (ein instabiles Verfahren):

Das in Abschnitt 3.1. betrachtete "Verfahren":

$$-y_{j+1}^* + 4y_j^* - 3y_{j-1}^* = 2h f(x_j, y_j^*); \quad y_0^* = \eta_0; \quad y_1^* = \eta_1$$

hat das erzeugende Polynom:

$$\rho(\mu) = -\mu^2 + 4\mu - 3 = -(\mu - 1)(\mu - 3)$$

mit den Nullstellen $\mu_1 = 1$ und $\mu_2 = 3$. Es ist daher instabil.

Dennoch hatten wir gesehen, daß es für die Differentialgleichung

$$y'(x) - \lambda y(x) = 0; \quad y(0) = 1$$

konvergiert, wenn man $\eta_0 = 1$ und $\eta_1 = (2 - \lambda h) - \sqrt{1 - 4\lambda h + (\lambda h)^2}$ wählt. In diesem Falle gilt nämlich

$$y_j^* = C_1 e^{\lambda x_j} + C_2 3^{x_j/h} \cdot e^{-\lambda x_j} + \mathcal{O}(h^2) \quad (4.6)$$

mit $C_1 = 1$ und $C_2 = 0$.

Trotzdem ist eine praktische Anwendung des Verfahrens auch bei dieser speziellen Wahl von η_1 nicht möglich, weil durch Rundungsfehler $C_2 \neq 0$ wird, was bei längerer Rechnung die Lösung $e^{\lambda x_j}$ völlig verfälscht.

Beispiel 2: (ein schwach stabiles Verfahren)

Im Falle des Verfahrens (1.9) von Milne-Simpson:

$$y_{j+1}^* - y_{j-1}^* = \frac{h}{3} (f_{j+1}^* + 4f_j^* + f_{j-1}^*); \quad y_0^* = \eta_0; \quad y_1^* = \eta_1(h)$$

ist $\rho(\mu) = \mu^2 - 1$; $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; $D = \sup_{j \in \mathbb{N}} \|A^j\| = 1$;

Die Eigenwerte von A sind $\mu_{1,2} = \pm 1$, das Verfahren ist daher schwach stabil. Wird der Startwert y_1^* mit dem modifizierten Euler-Verfahren (1.8) berechnet, so ist

$$\|y(x_1) - \eta_0 - hf(x_0 + \frac{h}{2}, \eta_0 + \frac{h}{2} f(x_0, \eta_0))\| \leq C_1 h^3; \text{ wegen (2.3) gilt außerdem}$$

$$\|T_h[y](x)\| \leq C_2 h^4 \text{ und mit } y_0^* = \eta_0, D=1, u=2L(1-\frac{hL}{3})^{-1}, \frac{Lh}{3} < 1 \text{ ergibt}$$

(4.5) die Fehlerschranke:

$$\|y(x_{j+2}) - y_{j+2}^*\| \leq \frac{e^{u(x_j - a)}}{1 - \frac{hL}{3}} \{ (1 + \frac{5}{3}hL)C_1 + (x_{j+1} - a)C_2 h \} h^3; \quad (4.7)$$

$j = 0(1)m-2$

In der Praxis würde man natürlich y_1^* mit einem Einschrittverfahren der Konvergenzordnung 4 berechnen; dann gilt

$$\|y(x_1) - y_1^*\| = \mathcal{O}(h^4) \text{ und die Fehlerschranke wäre von der Ordnung } h^4.$$

4.2. Stabilität von Prädiktor-Korrektor-Verfahren

Es wurde schon gesagt, daß bei impliziten Mehrschrittverfahren $M(\rho, \sigma)$ in jedem Rechenschritt eine Iteration erforderlich ist und daß diese Iteration für $|h\beta_k L| < 1$ konvergiert. Den Startwert $y_{j+k}^{*(0)}$ hierfür verschafft man sich im allgemeinen mittels einer expliziten Mehrschrittformel $M(\rho_1, \sigma_1)$. In diesem Zusammenhang nennt man die explizite Formel $M(\rho_1, \sigma_1)$ Prädiktor, die implizite Formel $M(\rho, \sigma)$ Korrektor; ein Prädiktor-Korrektor-Verfahren (P-C-Verfahren) hat somit die folgende Form:

$$z_{j+1}^{*(0)} = A_1 z_j^* + hB_1 Fz_j^*. \quad : \text{ Prädiktor } M(\rho_1, \sigma_1) \quad (4.8a)$$

$$z_{j+1}^{*(i)} = A z_j^* + hB Fz_j^* + hCFz_{j+1}^{*(i-1)} : \text{ Korrektor } M(\rho, \sigma) \quad (4.8b)$$

$i=1(1)s; \quad s: \text{ Anzahl der Auswertungen von } M(\rho, \sigma)$

$$z_{j+1}^* = z_{j+1}^{*(s)}$$

In der Praxis wird $s=1$ oder $s=2$ gewählt (siehe auch Abschnitt 9.5.).

Der englischsprachigen Literatur folgend wird ein solches Verfahren P(EC)^sE-Verfahren Prediction (Evaluation, Corrections)^s Evaluation genannt, denn mit dem Prädiktor wird z_{j+1}^* zunächst durch $z_{j+1}^{*(0)}$ angenähert, sodann wird F s-mal ausgewertet und die Näherung s-mal verbessert. Zum Schluß wird F (für den Gebrauch im nächsten Rechenschritt) mit dem zuletzt erhaltenen Wert $z_{j+1}^{*(s)}$ nochmals ausgewertet.

Die letzte Auswertung von F kann auch weggelassen werden; man benutzt dann im anschließenden Rechenschritt $F(z_{j+1}^{*(s-1)})$ statt $F(z_{j+1}^{*(s)})$ und spart so eine Funktionsauswertung pro Rechenschritt. Diese Variante der P-C-Verfahren wird P(EC)^s-Verfahren genannt.

Durch Substitution erkennt man sofort, daß die P-C-Verfahren (4.8a/b) alle die Form (3.14) haben,

$$\text{d.h.} \quad z_{j+1}^* = Az_j^* + h\phi(x_j, z_j^*, z_{j+1}^*; h);$$

$$\text{im Falle } s=1 \text{ ist} \quad \phi(x_j, z_j^*, z_{j+1}^*; h) = BFz_j^* + CF(A_1 z_j^* + hB_1 Fz_j^*); \quad (4.9)$$

für $s>1$ hat ϕ kompliziertere Gestalt.

Aus Satz 3.1. folgt daher:

Satz 4.2.

Ein Prädiktor-Korrektor-Verfahren ist genau dann stabil, wenn der Korrektor stabil ist.

Der Prädiktor darf somit auch instabil sein, d.h. $M(\rho_1, \sigma_1)$ braucht nicht das Wurzelkriterium zu erfüllen!

Im Falle (4.9) eines PECE-Verfahrens ergibt Satz 3.1. eine Fehler-schranke der Form (4.5), jedoch mit

$$K_1 = L \sum_{j=0}^{k-1} |\beta_j| + |\beta_k| L \left(\sum_{j=0}^{k-1} |a_j| + hL \sum_{j=0}^{k-1} |b_j| \right); \quad K_2 = 0;$$

Hierbei sind α_j bzw. β_j die Koeffizienten von ρ bzw. σ und a_j bzw. b_j die Koeffizienten von ρ_1 bzw. σ_1 .

Entsprechende Fehlerschranken lassen sich auch für $P(EC)^S E$ -Verfahren mit $s > 1$ angeben. Man beachte, daß Satz 3.1. die Berechnung von Fehlerschranken für das gesamte P-C-Verfahren gestattet und nicht, wie vielfach in der Literatur angegeben, einzeln für Prädiktor und Korrektor.

4.3. Stabilität zyklischer Verfahren

Bisher wurden die Näherungswerte y_j^* für $j=k(1)m$ stets mit demselben k -Schritt-Verfahren ($k \geq 1$) berechnet; im folgenden betrachten wir den Fall, daß die y_j^* mit verschiedenen Formeln berechnet werden, die in fester Reihenfolge zyklisch angewandt werden. Berechnet man beispielsweise y_j^* für gerades j mit der 2-Schritt-Formel (4.10a) von Milne-Simpson und für ungerades j mit der 2-Schritt-Formel (4.10b) von Adams-Moulton, so ergibt sich das folgende "2-zyklische 2-Schritt-Verfahren":

$$\begin{aligned} y_{2j+2}^* &= y_{2j}^* + \frac{h}{3} (f_{2j+2}^* + 4f_{2j+1}^* + f_{2j}^*) & j=o(1) \left[\frac{m}{2} - 1 \right]^{1)} \\ y_{2j+3}^* &= y_{2j+2}^* + \frac{h}{12} (5f_{2j+3}^* + 8f_{2j+2}^* - f_{2j+1}^*) \end{aligned} \quad (4.10a)$$

4.3.1. Definition zyklischer Verfahren

Definition 4.2.

Eine geordnete Menge von k_i -Schritt-Formeln, $i=o(1)M-1$ heißt M-zyklisches Verfahren, wenn die Formeln in vorgegebener Reihenfolge zyklisch zur Berechnung aufeinanderfolgender Näherungen y_j^* ($j=k(1)m$) angewendet werden. Die einzelnen Formeln eines M-zyklischen Verfahrens nennen wir Stufen.

¹⁾ $[a]$ bedeutet hier die größte natürliche Zahl $\leq a$.

(4.10a/b) beispielsweise ist ein 2-zyklisches Verfahren mit den Stufen (4.10a) und (4.10b). Theoretisch können die Stufen aus verschiedenartigen Verfahren bestehen, jedoch setzen wir im folgenden meistens voraus, daß sie alle lineare k-Schritt-Verfahren sind. Wir sprechen dann von M-zyklischen k-Schritt-Verfahren (z.B. für M=2, k=3 auch kürzer von M=2, k=3-Verfahren). In der Notation des Abschnitts 4.1. haben sie die Form:

$$\begin{aligned} z_{Mj+1}^* &= A^{(0)} z_{Mj}^* + h\phi^{(0)}(x_{Mj}, z_{Mj}^*, z_{Mj+1}^*; h) \\ z_{Mj+2}^* &= A^{(1)} z_{Mj+1}^* + h\phi^{(1)}(x_{Mj+1}, z_{Mj+1}^*, z_{Mj+2}^*; h) \\ &\vdots \\ z_{Mj+M}^* &= A^{(M-1)} z_{Mj+M-1}^* + h\phi^{(M-1)}(x_{Mj+M-1}, z_{Mj+M-1}^*, z_{Mj+M}^*; h) \quad j=o(1) \left[\frac{m}{M}-1 \right] \end{aligned} \quad (4.11)$$

wobei $A^{(i)} \in \mathbb{R}(k, k)$, $i=o(1)M-1$, Frobeniusmatrizen sind.

Von dieser speziellen Gestalt der $A^{(i)}$ wird jedoch an keiner Stelle der nachfolgenden Betrachtungen Gebrauch gemacht; es sei daher ausdrücklich angemerkt, daß alle Ergebnisse dieses und des nächsten Kapitels auf Verfahren vom Typ (4.11) mit allgemeineren $A^{(i)}$ anwendbar sind!

Man beachte, daß in jeder Stufe ein endgültiger Näherungswert y_j^* berechnet wird. Verfahren, bei denen das nicht notwendig der Fall ist, heißen nach Stetter [73] mehrstufige Mehrschrittverfahren. Zyklische Verfahren sind somit Spezialfälle dieser Verfahren, ebenso P-C-Verfahren (die keine zyklischen Verfahren gemäß Definition 4.1. sind, obwohl Prädiktor und Korrektor zyklisch benutzt werden!).

Einige der in Kapitel 8 behandelten Runge-Kutta-Verfahren lassen sich als zyklische Verfahren darstellen; das Verfahren (8.12), beispielsweise, läßt sich wie folgt als M-zyklisches Verfahren schreiben, wenn h durch $2h$ ersetzt und für die 2. Stufe der Prädiktor $\hat{y}_{2j+2}^* = y_{2j}^* + 2h(-f_{2j}^* + 2f_{2j+1}^*)$ gewählt wird:

$$\begin{aligned} y_{2j+1}^* &= y_{2j}^* + hf_{2j}^* \\ y_{2j+2}^* &= y_{2j}^* + \frac{h}{3} (f_{2j}^* + 4f_{2j+1}^* + \hat{f}_{2j+2}^*) \end{aligned} \quad j = 0(1) \left[\frac{m}{2} - 1 \right]$$

Auf diese Zusammenhänge soll hier jedoch nicht eingegangen werden; aus didaktischen Gründen beschränken wir unsere Betrachtungen vielmehr auf zyklische Mehrschrittverfahren, merken jedoch an, daß viele Ergebnisse sich ohne Schwierigkeiten auf mehrstufige Mehrschrittverfahren übertragen lassen.

4.3.2. Stabilität

Ein Stabilitätskriterium für zyklische Verfahren ergibt sich unmittelbar aus Satz 3.1. Wir verfahren dabei genau wie bei den Stabilitätsbetrachtungen für P-C-Verfahren in Abschnitt 4.2.

Dort wurde gezeigt, daß die Stabilität des P-C-Verfahrens

$$\hat{z}_{j+1}^* = A_1 z_j^* + h\phi_1(x_j, z_j^*; h) \quad (4.12)$$

$$z_{j+1}^* = A z_j^* + h\phi_2(x_j, z_j^*, \hat{z}_{j+1}^*; h)$$

nur von den Eigenwerten der Matrix A abhängt, indem wir (4.12) in der Form (3.14) darstellten, d.h.

$$z_{j+1}^* = A z_j^* + h\phi(x_j, z_j^*, z_{j+1}^*; h); \quad z_0^* = \zeta_0 \quad (4.13)$$

Ganz analog finden wir, daß die Stabilität des 2-zyklischen Verfahrens:

$$z_{2j+1}^* = A^{(0)} z_{2j}^* + h\phi^{(0)}(x_{2j}, z_{2j}^*, z_{2j+1}^*; h) \quad (4.14)$$

$$z_{2j+2}^* = A^{(1)} z_{2j+1}^* + h\phi^{(1)}(x_{2j+1}, z_{2j+1}^*, z_{2j+2}^*; h)$$

von den Eigenwerten der Matrix $A^{(1)} \cdot A^{(0)}$ abhängt, (welche die gleichen sind wie die der Matrix $A^{(0)} \cdot A^{(1)}$), weil es für hinreichend¹⁾ kleines $h > 0$ durch Substitution von z_{2j+1}^* bzw. z_{2j+2}^* übergeht in

$$z_{2j+2}^* = A^{(1)} A^{(0)} z_{2j}^* + h \hat{\phi}^{(0)}(x_{2j}, z_{2j}^*, z_{2j+2}^*; h) \quad (4.15a)$$

$$z_{2j+3}^* = A^{(0)} A^{(1)} z_{2j+1}^* + h \hat{\phi}^{(1)}(x_{2j+1}, z_{2j+1}^*, z_{2j+3}^*; h) \quad (4.15b)$$

und beide Gleichungen (nach Umindizierung $z_j^* := z_{2j}^*$ bzw. $z_j^* := z_{2j+1}^*$) die Gestalt (4.13) haben.

Die Verallgemeinerung auf beliebiges M ist offensichtlich. Damit ergibt Satz 3.1. das folgende Stabilitätskriterium für M-zyklische k-Schritt-Verfahren:

Satz 4.3.

M -zyklische Verfahren sind genau dann stabil, wenn

1. die Eigenwerte von $A := A^{(M-1)} A^{(M-2)} \dots A^{(1)} A^{(0)}$ betragsmäßig ≤ 1 sind und
2. zu den Eigenwerten vom Betrag 1 nur lineare Elementarteiler gehören.

Definition 4.3.

Erfüllen die Eigenwerte μ_i ($i=1(1)k$) einer Matrix $A \in \mathbb{R}(k, k)$ die Bedingung 1. und 2. von Satz 4.3., so sagen wir, daß sie dem erweiterten Wurzelkriterium genügen. Sie genügen dem engen Wurzelkriterium, wenn gilt: $\mu_1 = 1$, $|\mu_i| < 1$, $i=2(1)k$.

Genügen die Eigenwerte einer Matrix dem Wurzelkriterium (bezogen auf das charakteristische Polynom der Matrix), so erfüllen sie auch das erweiterte Wurzelkriterium; die Umkehrung gilt zwar für Frobeniusmatrizen, jedoch nicht für allgemeine Matrizen. Das rechtfertigt die Bezeichnung "erweitertes" Wurzelkriterium.

¹⁾ man beachte Anmerkung 1 auf Seite 20

Die Bedeutung zyklischer Verfahren liegt darin, daß die Eigenwerte von $A := A^{(M-1)} A^{(M-2)} \dots A^{(1)} A^{(0)}$ dem erweiterten Wurzelkriterium genügen können, obwohl die Eigenwerte von $A^{(r)}$ ($r=0(1)M-1$) ihm nicht genügen. Praktisch bedeutet das: Ein M-zyklisches Verfahren kann stabil sein, obwohl einige oder alle seine Stufen instabil sind. Verwendet man als Stufen insbesondere instabile lineare k-Schritt-Formeln hoher Konsistenzordnung, so lassen sich stabile Verfahren gewinnen, deren Konvergenzordnung größer ist, als die (durch Satz 3.3. bestimmte) maximale Ordnung stabiler linearer Mehrschrittverfahren.

Donelson und Hansen¹⁾ haben als erste systematisch zyklische Mehrschrittverfahren berechnet; von ihnen stammt das folgende Beispiel eines stabilen $M=3, k=3$ -Verfahrens:

$$33y_{3j+3}^* + 24y_{3j+2}^* - 57y_{3j+1}^* = h(10f_{3j+3}^* + 57f_{3j+2}^* + 24f_{3j+1}^* - f_{3j}^*) \quad (4.16a)$$

$$125y_{3j+4}^* - 144y_{3j+3}^* - 117y_{3j+2}^* + 136y_{3j+1}^* = h(42f_{3j+4}^* + 117f_{3j+3}^* - 144f_{3j+2}^* - 45f_{3j+1}^*) \quad (4.16b)$$

$$58y_{3j+5}^* + 531y_{3j+4}^* - 306y_{3j+3}^* - 283y_{3j+2}^* = h(9f_{3j+5}^* + 306f_{3j+4}^* + 531f_{3j+3}^* + 84f_{3j+2}^*) \quad (4.16c)$$

Alle drei Stufen sind stabil und haben für $y \in C^7[a, b]$ die Konsistenzordnung 5; dennoch konvergiert das Verfahren sogar mit der Ordnung 6 (wenn Startwerte und Prädiktoren ausreichende Ordnung haben), wie wir in Kapitel 5 zeigen werden.

¹⁾ Donelson, Hansen: Cyclic Composite Multistep Predictor-Corrector Methods, SIAM J. Numer. Anal. 8, 137-157 (1971).

4.4. Stabilitätsbetrachtungen bei einem partiellen Differentialgleichungsproblem

Ähnlich wie im Falle der Konsistenz, soll auch jetzt wieder an einem Beispiel gezeigt werden, daß der Stabilitätsbegriff des Abschnitts 3.1. und der Stabilitätssatz 3.1. auch auf andere Diskretisierungsprobleme als die gewöhnlicher Differentialgleichungen anwendbar ist. Hierzu betrachten wir wieder die parabolische Differentialgleichung (2.4), für die sich in Kapitel 2 durch Diskretisierung die folgenden Verfahren (2.6a-c) ergaben:

$$a) \quad u_j^{*(\ell+1)} = \left(1 - \frac{2\sigma\Delta t}{(\Delta x)^2}\right) u_j^{*(\ell)} + \frac{\sigma\Delta t}{(\Delta x)^2} (u_{j+1}^{*(\ell)} + u_{j-1}^{*(\ell)}) + \Delta t f_j^{*(\ell)};$$

$$j=1(1)m_1-1; \quad \ell=0(1)m_2-1$$

$$b) \quad u_j^{*(\ell+1)} = u_j^{*(\ell-1)} + \frac{2\sigma\Delta t}{(\Delta x)^2} (u_{j+1}^{*(\ell)} - 2u_j^{*(\ell)} + u_{j-1}^{*(\ell)}) + 2\Delta t f_j^{*(\ell)};$$

$$j=1(1)m_1-1; \quad \ell=1(1)m_2-1$$

$$c) \quad \frac{-\sigma\Delta t}{(\Delta x)^2} u_{j+1}^{*(\ell+1)} + \left(1 + \frac{2\sigma\Delta t}{(\Delta x)^2}\right) u_j^{*(\ell+1)} - \frac{\sigma\Delta t}{(\Delta x)^2} u_{j-1}^{*(\ell+1)} = u_j^{*(\ell)} + \Delta t f_j^{*(\ell)}$$

$$j=1(1)m_1-1; \quad \ell=0(1)m_2-1$$

Wir setzen: $a = \frac{\sigma\Delta t}{(\Delta x)^2}$ und

$$u_\ell^* := \begin{pmatrix} u_1^{*(\ell)} \\ \vdots \\ u_{m_1-1}^{*(\ell)} \end{pmatrix}; \quad u_0^* := \begin{pmatrix} \eta(\Delta t) \\ \vdots \\ \eta((m_1-1)\Delta t) \end{pmatrix}; \quad f_\ell^* := \begin{pmatrix} f_1^{*(\ell)} \\ \vdots \\ f_{m_1-1}^{*(\ell)} \end{pmatrix}; \quad A_1 = \begin{pmatrix} (1-2a) & a & 0 & \dots & 0 \\ a & (1-2a) & a & \dots & 0 \\ 0 & a & (1-2a) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a & (1-2a) \end{pmatrix}$$

$$C = \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ 0 & 1 & -2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 & -2 \end{pmatrix} ; \quad A_3 = \begin{pmatrix} (1+2a) & -a & 0 & \dots & 0 \\ -a & (1+2a) & -a & \dots & 0 \\ 0 & -a & (1+2a) & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & -a & (1+2a) \end{pmatrix}$$

Dann gehen a)-c) über in die Darstellungen:

$$a) \quad u_{\ell+1}^* = A_1 u_{\ell}^* + \Delta t \, f_{\ell}^* ; \quad \ell = 0(1)m_2 - 1 \quad (4.17a)$$

$$b) \quad u_{\ell+1}^* = 2a \, C \, u_{\ell}^* + u_{\ell-1}^* + 2\Delta t \, f_{\ell}^* ; \quad \ell = 1(1)m_2 - 1$$

mit $v_{\ell}^* = \begin{pmatrix} u_{\ell-1}^* \\ u_{\ell}^* \end{pmatrix}$ erhält man hieraus ein Einschrittverfahren

$$\text{der Form (3.14): } v_{\ell+1}^* = A_2 v_{\ell}^* + 2\Delta t \cdot g_{\ell}^* ; \quad v_1^* = \begin{pmatrix} u_0^* \\ u_1^* \end{pmatrix} ; \quad \ell = 1(1)m_2 - 1 \quad (4.17b)$$

$$\text{mit } A_2 = \begin{pmatrix} \sigma & I \\ I & 2aC \end{pmatrix} ; \quad g_{\ell}^* = \begin{pmatrix} \sigma \\ f_{\ell}^* \end{pmatrix} \in \mathbb{R}^{2(m_1-1)} ; \quad I \in \mathbb{R}(m_1-1, m_1-1)$$

$$c) \quad A_3 u_{\ell+1}^* = u_{\ell}^* + \Delta t \, f_{\ell}^* ; \\ u_{\ell+1}^* = A_3^{-1} u_{\ell}^* + \Delta t \, A_3^{-1} f_{\ell}^* ; \quad \ell = 0(1)m_2 - 1 \quad (4.17c)$$

Alle drei Verfahren sind damit auf die Form (3.14) zurückgeführt, ihre Stabilität ergibt sich daher nach Satz 3.1. aus den Eigenwerten der Matrizen A_1 , A_2 und A_3^{-1} :

a) Man kann zeigen, daß die Eigenwerte von A_1 gegeben sind durch:

$$\mu_j = 1 - 4a \sin^2 \left(\frac{j\pi}{2m_1} \right), \quad j=1(1)m_1-1; \text{ für } |a| \leq \frac{1}{2} \text{ gilt für den}$$

Spektralradius von A_1 : $\rho(A_1) < 1$; das explizite Verfahren (2.6a) ist daher stabil für $a = \frac{\sigma \Delta t}{(\Delta x)^2} \leq \frac{1}{2}$.

b) Ist $\mu \neq 0$ Eigenwert von A_2 , so folgt aus $\det(A_2 - \mu I) = 0$ durch

Multiplikation von links mit $\det \begin{pmatrix} I & \sigma \\ I & \mu I \end{pmatrix}$, $\mu \neq 0$:

$$\det (2a\mu C + (1 - \mu^2)I) = 0$$

$$\det \left(C + \frac{1 - \mu^2}{2a\mu} I \right) = 0$$

Ist λ Eigenwert von C , so gilt daher $\mu = a\lambda \pm \sqrt{a^2 \lambda^2 + 1}$. Da alle Eigenwerte von C von Null verschieden sind, existieren somit für beliebiges a Eigenwerte μ mit $|\mu| > 1$; das 2-Schrittverfahren (2.6b) ist somit für alle Δx und Δt instabil. Ersetzt man jedoch in (2.6b) den Term $2u_j^{*(\ell)}$ durch $(u_j^{*(\ell-1)} + u_j^{*(\ell+1)})$, so ergibt sich das Verfahren von Du Fort-Frankel, welches für alle Δx und Δt stabil ist.

c) Alle Eigenwerte von A_3 sind (vgl. A_1 mit $a := -a$) für $a > 0$ betragsmäßig größer als 1; daher gilt $\rho(A_3^{-1}) < 1$ für beliebiges $a > 0$. Das implizite Verfahren (2.6c) ist deshalb für alle Δx und Δt stabil.

5. Über die Ordnung zyklischer Verfahren

In Abschnitt 4.3. wurde bereits ein zyklisches Verfahren der Konvergenzordnung 6 angegeben, dessen Stufen lediglich die Konsistenzordnung 5 haben. Donelson und Hansen [71] gewannen derartige Verfahren, indem sie für $k=2,3,4$ k -stufige k -Schritt-Verfahren mit instabilen Stufen der Konsistenzordnung $(2k-1)$ ersetzten durch äquivalente k -stufige kM -Schritt-Verfahren mit stabilen Stufen der Konsistenzordnung $2k$. Die dabei erforderlichen Rechnungen sind jedoch sehr aufwendig und machen eine Untersuchung der Fälle $k>4$ aussichtslos.

Im folgenden wird ein völlig anderer Weg zur Gewinnung zyklischer Mehrschrittverfahren maximaler Ordnung beschritten. Ausgangspunkt ist dabei wieder Satz 3.1., aus dem sich - durch leichte Modifikation seines Beweises - eine allgemeine Theorie über Verfahren der Konsistenzordnung q mit der Konvergenzordnung $(q+1)$ ergibt. Diese Theorie führt zu einem tieferen Verständnis k -zyklischer k -Schritt-Verfahren der Ordnung $2k$, vereinfacht ihre Berechnung ganz erheblich und ist prinzipiell auch auf allgemeinere Verfahren anwendbar.

5.1. Eine Modifikation des Satzes 3.1.

Für Verfahren der allgemeinen Form (3.14):

$$z_{j+1}^* = Az_j^* + h\phi(x_j, z_j^*, z_{j+1}^*; h), \quad z_0^* = \zeta_0(h); \quad j=0(1)m-k \quad (5.1)$$

untersuchen wir im folgenden die Frage:

Wie kommt es, daß Verfahren mit der Konsistenzordnung q i. allg. auch die Konvergenzordnung q haben, und ist das immer der Fall?

Eine Antwort hierauf ergibt sich aus einer Analyse des Beweises von Satz 3.1.: Dort wurde (vgl. S. 21 mit $v_j = z_j$, $w_j = z_j^*$) die Norm des Terms

$$h \sum_{\ell=0}^j A^{j-\ell} T_h[z](x_\ell) \quad \text{mit} \quad T_h[z](x_\ell) = \mathcal{O}(h^q) \quad (5.2)$$

in Gleichung (3.16) nach oben abgeschätzt durch

$$h(j+1)D \|T_h[z](x_\ell)\| \quad \text{mit} \quad D \geq \|A^j\|, \quad j \in \mathbb{N}. \quad (5.3)$$

Wegen $h(j+1) = (x_{j+1} - a)$ hat (5.3) die Ordnung q in $h^{1)}$. Jeder Summand in (5.2) hat aber die Ordnung $(q+1)$, so daß unter bestimmten Voraussetzungen möglicherweise auch die Summe (5.2) für $j \rightarrow \infty$ die Ordnung $(q+1)$ in h hat. Verfahren, für welche das der Fall ist, hätten - wie der weitere Verlauf des Beweises von Satz 3.1. zeigt - bei passend gewähltem Startvektor $\zeta_0(h)$ die Konvergenzordnung $(q+1)$. Tatsächlich gelingt es mittels der folgenden Modifikation des Satzes 3.1., Verfahren der Konsistenzordnung q zu konstruieren, welche mit der Ordnung $(q+1)$ konvergieren (siehe Abschnitt 5.2. und Kapitel 10).

Lemma 5.1.

Seien $\mu_1 = 1$ und $\mu_i \neq 1$ ($i=2(1)k$) die (nicht notwendig voneinander verschiedenen) Eigenwerte von $A \in \mathbb{R}(k, k)$ und sei $\{u(\mu_i), i=1(1)k\}$ eine Normalbasis²⁾ in $\mathbb{R}^k$; dann sind die folgenden beiden Bedingungen äquivalent:

$$t = \sum_{i=1}^k d_i u(\mu_i) \quad \text{mit} \quad d_1 = 0 \quad (5.4)$$

$$p^T t = 0 \quad \text{mit} \quad A^T p = p; \quad p \neq 0 \quad (5.5)$$

¹⁾ im folgenden werden stets ausreichende Differenzierbarkeits-eigenschaften von y vorausgesetzt.

²⁾ vgl. z.B. Faddejew-Faddejewa, S. 90

Beweis:

Sei r_i die Stufe des Vektors $u(\mu_i)$, so gilt $(A - \mu_i I)^{r_i} u(\mu_i) = 0$ und aus $p^T A^j = p^T$ für alle $j \in \mathbb{N}$

$$\text{folgt: } p^T \left[\sum_{\ell=0}^{r_i} \binom{r_i}{\ell} (-\mu_i)^{r_i-\ell} A^\ell \right] u(\mu_i) = (1 - \mu_i)^{r_i} p^T u(\mu_i) = 0$$

$$\text{Für } i \neq 1 \text{ ist daher } p^T u(\mu_i) = 0. \quad (5.6)$$

$$\text{Aus der linearen Unabhängigkeit der } u(\mu_i), i=1(1)k, \text{ und } p \neq 0 \text{ folgt: } p^T u(1) \neq 0 \quad (5.7)$$

Die Behauptung folgt nun aus $p^T t = d_1 p^T u(1)$. $\triangleleft$

Voraussetzungen:

(1) Die Eigenwerte μ_i , $i=1(1)k$, der Matrix $A \in \mathbb{R}(k,k)$ mögen dem engen Wurzelkriterium genügen.

(2) Das Verfahren (5.1) habe einen "lokalen Ersetzungsfehler"

$$T_h[z](x_\ell) := h^{-1} (z_{j+1} - A z_j - h \phi(x_j, z_j, z_{j+1}; h)) \quad (5.8)$$

der Form:

$$T_h[z](x_\ell) = h^q g_\ell t + \mathcal{O}(h^{q+1}) \quad (5.9)$$

mit z_j wie in (4.2), beschränktem $g: [a,b] \rightarrow \mathbb{R}$, $g_\ell := g(x_\ell)$ und konstantem $t \in \mathbb{R}^k$.

$$(3a) \text{ Es sei } \underline{\text{entweder}}: p^T t = 0 \text{ mit } A^T p = p, p \neq 0 \quad (5.10)$$

$$(3b) \quad \underline{\text{oder}}: \left| \sum_{\ell=0}^j g_\ell \right| \text{ für alle } j \in \mathbb{N} \text{ beschränkt} \quad (5.11)$$

(4) Der Startvektor $z_0(h)$ habe (wenigstens) die Konvergenzordnung $(q+1)$:

$$\|z_0 - z_0(h)\| = \mathcal{O}(h^{q+1}) \quad (5.12)$$

Satz 5.1.

Unter den Voraussetzungen (1)-(4) konvergiert der Verfahren (5.1) mit der Ordnung $(q+1)$.

Anmerkungen:

1. Eine Abschwächung der Voraussetzung (1) ist möglich, jedoch ohne praktische Bedeutung.

2. Beim linearen k-Schritt-Verfahren (4.1) hat $T_h[z](x_\ell)$ die Gestalt (5.4) mit $g_\ell = c_{q+1} y^{(q+1)}(x_\ell)$, wie (4.3) zeigt; im Falle eines Systems von n Differentialgleichungen ist daher $g_\ell \in \mathbb{R}^n$. Eine Verallgemeinerung der Voraussetzungen auf den Fall $n > 1$ wäre somit sinnvoll und ist - analog zum Vorgehen in Abschnitt 4.1.¹⁾ - auch möglich. Aus Gründen der Verständlichkeit wird hierauf aber verzichtet und nur der Fall $n=1$ dargestellt.

Beweis von Satz 5.1.

(5.2) ergibt mit (5.9) und $t = \sum_{i=1}^k d_i u(\mu_i)$

$$h \sum_{\ell=0}^j A^{j-\ell} T_h[z](x_\ell) = h^{q+1} S^{(j)} + \mathcal{O}(h^{q+1}) \quad (5.13)$$

$$\text{mit } S^{(j)} := \sum_{\ell=0}^j g_\ell A^{j-\ell} t = d_1 \left(\sum_{\ell=0}^j g_\ell \right) u(1) + \sum_{i=2}^k d_i \sum_{\ell=0}^j g_\ell \left(\sum_{r=0}^{r_i-1} P_{ir}(\ell) A^r \right) \mu_i^{j-\ell} u(\mu_i) \quad (5.14)$$

wobei die $P_{ir}(\ell)$ Polynome vom Grad (r_i-1) sind. $S^{(j)}$ ist somit für alle $j \in \mathbb{N}$ beschränkt, wenn die Voraussetzungen (1), (2) und (3) gelten.

(5.13) hat dann die Ordnung $(q+1)$ in h und durch entsprechende Modifikation im 2. Teil des Beweises von Satz 3.1. ergibt sich statt (3.21) die Ungleichung:

$$\|v_{j+1} - w_{j+1}\| \leq \frac{(x_j - a)u}{1 - hK_2} \{ (1 + hK_1) \|v_0 - w_0\| + \max_{0 \leq i \leq j+1} h \sum_{\ell=0}^i A^{i-\ell} (T_h[v](x_\ell) - T_h[w](x_\ell)) \} \quad (5.15)$$

Mit $K_4 := \frac{(1 + h_0 K_1)}{(1 - h_0 K_2)} \text{De}^{(b-a)u_0}$; $v_j = z_j, w_j = z_j^*, T_h[z^*](x_\ell) = \sigma$ folgt aus (5.13) und (5.15):

$$\|z_j - z_j^*\| \leq K_4 \|z_0 - \zeta_0(h)\| + \mathcal{O}(h^{q+1})$$

Wegen Voraussetzung (4) hat das Verfahren daher die Konvergenzordnung $(q+1)$. $\triangleleft$

¹⁾ man ersetzt im folgenden Beweis A durch ihre Block-Frobeniusform $A \in \mathbb{R}(nk, nk)$, $g_\ell t$ durch $g_\ell \otimes t := \begin{pmatrix} g(x_\ell) t_1 \\ \vdots \\ g(x_\ell) t_k \end{pmatrix} \in \mathbb{R}^{nk}$ und beweist die Beschränktheit von $\| \sum_{\ell=0}^j A^{j-\ell} (g(x_\ell) \otimes t) \|_{\mathbb{R}^{nk}}$ mit $\|g_\ell \otimes t\|_{\mathbb{R}^{nk}} := \max_{0 \leq i \leq k} \|t_i g(x_\ell)\|_{\mathbb{R}^n}$ unter Beachtung von $A(g_\ell \otimes t) = g_\ell \otimes At$, $g_\ell \in \mathbb{R}^n$, $t \in \mathbb{R}^k$, $A \in \mathbb{R}(k, k)$.

Beispiel zum Satz 5.1.:

Für das Verfahren

$$\begin{aligned} y_{2j+1}^* &= y_{2j}^* + h f_{2j}^* \\ y_{2j+2}^* &= y_{2j+1}^* + h f_{2j+2}^* \end{aligned} \quad j=0(1)\left[\frac{m}{2}-1\right] \quad (5.16)$$

gilt:

$$\begin{aligned} hT_h[y](x_{2j}) &= y_{2j+1} - y_{2j} - hy'_{2j} = \frac{h^2}{2} y''(x_{2j}) + \mathcal{O}(h^3) \\ hT_h[y](x_{2j+1}) &= y_{2j+2} - y_{2j+1} - hy'_{2j+2} = -\frac{h^2}{2} y''(x_{2j}) + \mathcal{O}(h^3) \end{aligned}$$

Daher ist (vgl. 5.9): $T_h[y](x_\ell) = hg_\ell t + \mathcal{O}(h^2)$ mit $t=1$ und

$$g_\ell = \begin{cases} \frac{1}{2} y''(x_{2j}) & \text{für } \ell=2j \\ -\frac{1}{2} y''(x_{2j}) & \text{für } \ell=2j+1 \end{cases}$$

Wegen $\left| \sum_{\ell=0}^j g_\ell \right| \leq \frac{1}{2} \cdot \max_{x \in [a,b]} |y''(x)|$ ist (5.11) erfüllt. Das Verfahren

konvergiert daher nach Satz 5.1. mit der Ordnung 2, obwohl jede der beiden Stufen nur die Konsistenzordnung $q=1$ hat. $\triangleleft$

Das obige Beispiel wurde in anderem theoretischen Zusammenhang bereits von Spijker [68] angegeben. Spijker zeigt, daß sich bei Wahl einer anderen Stabilitätsdefinition in gewissen Fällen eine höhere Konvergenzordnung nachweisen läßt und gibt als Beispiel dafür zyklische Verfahren an, die allerdings sehr spezieller Art sind. Alle Verfahren, die Voraussetzung (3b) erfüllen, haben auch nach Spijker die Ordnung $(q+1)$. Wesentlich allgemeiner und wichtiger für die Praxis ist jedoch Voraussetzung (3a); die hiermit gewonnenen Verfahren der Konvergenzordnung $(q+1)$ lassen sich nicht aus der Spijkerschen Theorie ableiten¹⁾. Beispiele für die Anwendung der Voraussetzung (3a) werden am Ende von Abschnitt 5.3. angegeben.

¹⁾ Es ist jedoch wahrscheinlich, daß sich diese Theorie erweitern läßt, wenn man etwa die Beschränktheit von (5.13) in die Stabilitätsdefinition aufnimmt.

5.2. Maximale Ordnung M-zyklischer k-Schritt-Verfahren

Wir zeigen nun, wie man mittels Satz 5.1. k-zyklische k-Schritt-Verfahren der Konvergenzordnung $2k$ konstruieren kann und betrachten dazu den Fall $M=3$, $k=3$ mit Stufen der Ordnung 5. Dann gilt (in der Notation (4.2)) für die Vektoren z_j der exakten Lösung bei hinreichenden Differentiationseigenschaften von y :

$$z_{3j+1} = A^{(0)} z_{3j} + h\phi^{(0)}(x_{3j}, z_{3j}, z_{3j+1}; h) + \begin{pmatrix} 0 \\ 0_{(0)} \\ c_6^{(0)} \end{pmatrix} y^{(6)}(x_{3j}) h^6 + \mathcal{O}(h^7) \quad (5.17a)$$

$$z_{3j+2} = A^{(1)} z_{3j+1} + h\phi^{(1)}(x_{3j+1}, z_{3j+1}, z_{3j+2}; h) + \begin{pmatrix} 0 \\ 0_{(1)} \\ c_6^{(1)} \end{pmatrix} y^{(6)}(x_{3j}) h^6 + \mathcal{O}(h^7) \quad (5.17b)$$

$$z_{3j+3} = A^{(2)} z_{3j+2} + h\phi^{(2)}(x_{3j+2}, z_{3j+2}, z_{3j+3}; h) + \begin{pmatrix} 0 \\ 0_{(2)} \\ c_6^{(2)} \end{pmatrix} y^{(6)}(x_{3j}) h^6 + \mathcal{O}(h^7) \quad (5.17c)$$

wobei $c_6^{(r)}$ die Fehlerkonstanten der Stufen sind.

Für genügend kleines h sind (5.17a-c) auflösbar; durch Substitution folgt:

$$z_{3j+3} = A^{(2)} A^{(1)} A^{(0)} z_{3j} + h\hat{\phi}(x_{3j}, z_{3j}, z_{3(j+1)}; h) + h\hat{T}_h[z](x_{3j}) \quad (5.18)$$

$$\text{mit } \hat{T}_h^{(0)}[z](x_{3j}) := h^5 y^{(6)}(x_{3j}) t_0 + \mathcal{O}(h^6) \quad (5.19)$$

$$t_0 := \begin{pmatrix} 0 \\ 0_{(2)} \\ c_6^{(2)} \end{pmatrix} + A^{(2)} \begin{pmatrix} 0 \\ 0_{(1)} \\ c_6^{(1)} \end{pmatrix} + A^{(2)} A^{(1)} \begin{pmatrix} 0 \\ 0_{(0)} \\ c_6^{(0)} \end{pmatrix} = \begin{pmatrix} c_6^{(0)} \\ c_6^{(1)} - \alpha_2^{(1)} c_6^{(0)} \\ c_6^{(2)} - \alpha_2^{(2)} c_6^{(1)} - (\alpha_1^{(2)} - \alpha_2^{(2)} \alpha_2^{(1)}) c_6^{(0)} \end{pmatrix} \quad (5.20)$$

d.h. die Näherungen z_j^* mit $j=0(\text{mod } 3)$ lassen sich aus einem Einschrittverfahren der Form (5.1) gewinnen, nämlich aus

$$z_{3(j+1)}^* = \hat{A}_0 z_{3j}^* + h\hat{\phi}_0(x_{3j}, z_{3j}^*, z_{3(j+1)}^*; h) \quad (5.21)$$

mit $\hat{A}_0 := A^{(2)} A^{(1)} A^{(0)}$ und dem lokalen Ersetzungsfehler (5.19).

Für die Näherungen $z_j^* (j=r(\bmod 3), r=1,2)$ gilt entsprechend:

$$z_{3(j+1)+r}^* = \hat{A}_r z_{3j+r}^* + h \phi_r(x_{3j+r}, z_{3j+r}^*, z_{3(j+1)+r}^*; h) \quad (5.22)$$

$$\text{mit } ^1) \quad \hat{A}_r := A^{(2+r)} A^{(1+r)} A^{(r)} \quad (5.23)$$

$$\text{und} \quad \hat{T}_h^{(r)}[z](x_{3j+r}) := t_r y^{(6)}(x_{3j+r}) h^5 + \mathcal{O}(h^6) \quad (5.24)$$

$$t_r := c^{(2+r)} + A^{(2+r)} c^{(1+r)} + A^{(2+r)} A^{(1+r)} c^{(r)}; \quad c^{(r)} := (0, 0, c_6^{(r)})^T \quad (5.25)$$

Genügen $\hat{A}_0$ und t_0 den Voraussetzungen des Satzes 5.1., so haben die Näherungen z_{3j}^* die Konvergenzordnung 6.

Lemma 5.2.

Gilt $^2) \quad p_r^T t_r = 0$ mit $p_r^T \hat{A}_r = p_r^T, p_r^T \neq 0$, so gibt es ein $p_{r+1}^T \neq 0$ mit $p_{r+1}^T t_{r+1} = 0$ und $p_{r+1}^T \hat{A}_{r+1} = p_{r+1}^T$.

Beweis:

$\mu_1 = 1$ ist einfacher Eigenwert von $\hat{A}_{r+1}$, daher existiert ein $p_{r+1}^T \neq 0$ mit $p_{r+1}^T \hat{A}_{r+1} = p_{r+1}^T$ und mit (5.23) folgt: $p_{r+1}^T A^{(r)} \neq 0$. Nach Multiplikation von rechts mit $A^{(r)}$ gilt:

$$(p_{r+1}^T A^{(r)}) \hat{A}_r = (p_{r+1}^T A^{(r)}), \text{ denn } \hat{A}_{r+1} A^{(r)} = A^{(r)} \hat{A}_r.$$

Aber auch p_r ist Eigenvektor zum einfachen Eigenwert $\mu_1 = 1$ von $\hat{A}_r^T$; daher ist: $p_{r+1}^T A^{(r)} = a p_r^T$; $a = \text{const.} \neq 0$ (5.26)

Aus $t_{r+1} = A^{(r)} t_r + (I - \hat{A}_{r+1}) c^{(r)}$ und (5.26) folgt nun $p_{r+1}^T t_{r+1} = a p_r^T t_r$ und hieraus die Behauptung. $\triangleleft$

Lemma 5.2. besagt: Genügen $\hat{A}_0$ und t_0 den Voraussetzungen des Satzes 5.1., so haben alle Näherungen z_{3j}^*, z_{3j+1}^* und z_{3j+2}^* die Konvergenzordnung 6.

¹⁾ obere Indizes sind mod 3 zu nehmen!

²⁾ alle Indizes im Lemma und seinem Beweis sind mod 3 zu nehmen.

Völlig analog folgt für beliebiges M aus Satz 5.1.:

Satz 5.2.

Ein M-zyklisches k-Schritt-Verfahren mit Stufen der Konsistenzordnung $q_r \geq q$ und Startwerten der Konvergenzordnung $\geq (q+1)$ konvergiert für $y \in C^{q+2}[a,b]$ mit der Ordnung $(q+1)$, wenn gilt:

- (1) die Eigenwerte $\mu_i, i=1(1)k$, von $A := A^{(M-1)} A^{(M-2)} \dots A^{(1)} A^{(0)}$ ($A^{(r)} \in \mathbb{R}(k,k)$ Frobeniusmatrix der r-ten Stufe) genügen dem engen Wurzelkriterium.

$$(2) p^T t = 0 \text{ mit } A^T p = p, p \neq 0 \quad (5.27)$$

$$t := c^{(M-1)} + A^{(M-1)} c^{(M-2)} + A^{(M-1)} A^{(M-2)} c^{(M-3)} + \dots + A^{(M-1)} \dots A^{(1)} c^{(0)} \quad (5.28)$$

$$c^{(r)} := (0, 0, \dots, 0, c_{q+1}^{(r)})^T \in \mathbb{R}^k \quad (5.29)$$

$c_{q+1}^{(r)}$: Fehlerkonstante der r-ten Stufe, falls $q_r = q$;
sonst 0.

An keiner Stelle des Beweises von Satz 5.2. wurde benutzt, daß $A^{(r)}$ Frobenius-Matrizen sind; er läßt sich daher prinzipiell auch auf allgemeinere zyklische Verfahren ausdehnen¹⁾.

¹⁾ eine Arbeit hierüber ist in Vorbereitung.

5.3. Vereinfachung der Ordnungsbedingung (5.27)

Die Bedingung (5.27) läßt sich ohne jede Rechnung unmittelbar angeben, wenn man berücksichtigt, daß die $A^{(r)}$ Frobenius-Matrizen sind. Diese Möglichkeit wird im folgenden wieder am Beispiel $k=M=3$ erläutert:

Sei als $k=M=3$, $\hat{A} := A^{(2)} A^{(1)} A^{(0)}$ und $A^{(i)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\alpha_0^{(i)} & -\alpha_1^{(i)} & -\alpha_2^{(i)} \end{pmatrix}$,
 $i=0,1,2$.

Mit $L := \begin{pmatrix} 1 & 0 & 0 \\ \alpha_2^{(1)} & 1 & 0 \\ \alpha_1^{(2)} & \alpha_2^{(2)} & 1 \end{pmatrix}$ gilt: $L\hat{A} = - \begin{pmatrix} \alpha_0^{(0)} & \alpha_1^{(0)} & \alpha_2^{(0)} \\ 0 & \alpha_0^{(1)} & \alpha_1^{(1)} \\ 0 & 0 & \alpha_0^{(2)} \end{pmatrix}$ und $Lt = \begin{pmatrix} c_6^{(0)} \\ c_6^{(1)} \\ c_6^{(2)} \end{pmatrix} := c$

wobei $c_6^{(i)}$ ($i=0,1,2$) wieder die Fehlerkonstanten der Stufen sind. Definiert man q durch $p = L^T q$, so geht Voraussetzung (5.27) über in die einfachere Form:

$$q^T c = 0 \quad \text{mit} \quad q^T (L - L\hat{A}) = 0 \quad (5.30)$$

Dieses lineare System für die Komponenten von q hat eine Lösung $q \neq 0$, wenn die Systemmatrix den Rang $\leq (k-1)$ hat. Wegen der Singularität (als Folge der Konsistenz der Stufen) von

$$L - L\hat{A} = \begin{pmatrix} 1+\alpha_0^{(0)} & \alpha_1^{(0)} & \alpha_2^{(0)} \\ \alpha_2^{(1)} & 1+\alpha_0^{(1)} & \alpha_1^{(1)} \\ \alpha_1^{(2)} & \alpha_2^{(2)} & 1+\alpha_0^{(2)} \end{pmatrix}$$

ist das der Fall, wenn gilt:

$$\det \begin{pmatrix} 1+\alpha_0^{(0)} & \alpha_1^{(0)} & c_6^{(0)} \\ \alpha_2^{(1)} & 1+\alpha_0^{(1)} & c_6^{(1)} \\ \alpha_1^{(2)} & \alpha_2^{(2)} & c_6^{(2)} \end{pmatrix} = 0 \quad (5.31a)$$

Berücksichtigt man noch $c_{2k}^{(i)} = \text{const.} \cdot (1 + \alpha_o^{(i)})$, $i=0,1,2$, so kann Bedingung (5.27) für $k=M=3$ ersetzt werden durch:

$$\det \begin{pmatrix} 1 + \alpha_o^{(0)} & \alpha_1^{(0)} & (1 + \alpha_o^{(0)}) \\ \alpha_2^{(1)} & 1 + \alpha_o^{(1)} & (1 + \alpha_o^{(1)}) \\ \alpha_1^{(2)} & \alpha_2^{(2)} & (1 + \alpha_o^{(2)}) \end{pmatrix} = 0 \quad (5.31b)$$

Die Verallgemeinerung auf beliebiges $M=k$ ist offensichtlich und ergibt das folgende Korollar:

Korollar 5.2.

Ist $M=k$ und haben die Matrizen $A^{(r)}$, $r=0(1)k-1$, Frobeniusgestalt, so kann die Ordnungsbedingung (5.27) ersetzt werden durch

$$\det \begin{pmatrix} 1 + \alpha_o^{(0)} & \alpha_1^{(0)} & \dots & \alpha_{k-2}^{(0)} & (1 + \alpha_o^{(0)}) \\ \alpha_{k-1}^{(1)} & 1 + \alpha_o^{(1)} & \dots & \alpha_{k-3}^{(1)} & (1 + \alpha_o^{(1)}) \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_2^{(k-2)} & \alpha_3^{(k-2)} & \dots & 1 + \alpha_o^{(k-2)} & (1 + \alpha_o^{(k-2)}) \\ \alpha_1^{(k-1)} & \alpha_2^{(k-1)} & \dots & \alpha_{k-1}^{(k-1)} & (1 + \alpha_o^{(k-1)}) \end{pmatrix} = 0 \quad (5.32)$$

und das Verfahren ist stabil, wenn für $(k-1)$ Zahlen μ_i , $i=2(1)k$, mit $|\mu_i| < 1$ gilt:

$$\det \begin{pmatrix} \mu_i + \alpha_o^{(0)} & \alpha_1^{(0)} & \dots & \alpha_{k-2}^{(0)} & \alpha_{k-1}^{(0)} \\ \mu_i \alpha_{k-1}^{(1)} & \mu_i + \alpha_o^{(1)} & \dots & \alpha_{k-3}^{(1)} & \alpha_{k-2}^{(1)} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_i \alpha_2^{(k+2)} & \mu_i \alpha_3^{(k-2)} & \dots & \mu_i + \alpha_o^{(k-2)} & \alpha_1^{(k-2)} \\ \mu_i \alpha_1^{(k-1)} & \mu_i \alpha_2^{(k-2)} & \dots & \mu_i \alpha_{k-1}^{(k-1)} & \mu_i + \alpha_o^{(k-1)} \end{pmatrix} = 0 \quad (5.33)$$

Mit Satz 5.2. und seinem Korollar können durch zyklische Kombination von k (instabilen) k -Schritt-Formeln der Konsistenzordnung $(2k-1)$ stabile Verfahren mit der maximalen Konvergenzordnung $2k$ gewonnen werden (siehe Kapitel 10).

Beispiele:

1. Für das Verfahren (4.16a-c) rechnet man leicht nach, daß Gleichung (5.31b) erfüllt ist.
2. Die erste Stufe des $M=2, k=2$ -Verfahrens (4.10a/b) hat die Konsistenzordnung 4 und die zweite die Ordnung 3; für die Fehlerkonstanten gilt daher: $c_4^{(0)}=0$, $c_4^{(1)} \neq 0$. Hiermit lautet die dem Fall $M=k=2$ entsprechende Gleichung (5.31a):

$$\det \begin{pmatrix} 1+\alpha_0^{(0)} & c_4^{(0)} \\ \alpha_1^{(1)} & c_4^{(1)} \end{pmatrix} = \det \begin{pmatrix} 0 & 0 \\ -1 & c_4^{(1)} \end{pmatrix} = 0 \quad (5.34)$$

Sie ist für beliebiges $c_4^{(1)}$ erfüllt, d.h. jede beliebige lineare 2-Schritt-Formel mit der Konsistenzordnung 3 erfüllt in Verbindung mit (4.10a) die Voraussetzung (2) von Satz 5.2.

Zur Kontrolle der Voraussetzung (1) berechnen wir $\hat{A} := A^{(1)} A^{(0)}$ mit

$$A^{(0)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad A^{(1)} = \begin{pmatrix} 0 & 1 \\ -\alpha_0^{(1)} & -\alpha_1^{(1)} \end{pmatrix}; \quad \hat{A} = \begin{pmatrix} 1 & 0 \\ -\alpha_1^{(1)} & -\alpha_0^{(1)} \end{pmatrix}$$

Für $|\alpha_0^{(1)}| < 1$ ist damit auch Voraussetzung (1) erfüllt und aus Satz 5.2. folgt:

Die zyklische Kombination von (4.10a) mit einer beliebigen linearen 2-Schritt-Formel mit der Konsistenzordnung 3 und $|\alpha_0^{(1)}| < 1$ ergibt ein Verfahren der Konvergenzordnung 4, wenn die Startwerte die Ordnung 4 haben.

§6. Stabilität eines Differentialgleichungsproblems

In den vorigen Abschnitten wurde die Stabilität numerischer Verfahren definiert und untersucht. Im folgenden fragen wir nach der Stabilität des gestellten Differentialgleichungspro-
blems (1.1):

$$y' - f(x, y) = 0; \quad y(a) = \eta_0 \quad \text{im Intervall} \quad x \in [a, \infty) \quad (6.1)$$

Betrachten wir hierzu das Problem

$$\bar{y}' - f(x, \bar{y}) = 0; \quad \bar{y}(a) = \eta_0 + \delta_0$$

mit der "Störung" δ_0 in der Anfangsbedingung. Für den Fehler $e(x) = \bar{y}(x) - y(x)$ gilt:

$$e' - g(x, e) = 0; \quad e(a) = \delta_0 \quad (6.2)$$

mit $g(x, e(x)) = f(x, y(x) + e(x)) - f(x, y(x))$

Die Lösung von (6.2) bezeichnen wir mit $e(x; \delta_0, a)$. Analog zu den Betrachtungen in § 3.2. nennen wir (6.1) stabil, wenn "kleine" Störungen in der Anfangsbedingung auch nur "kleine" Änderungen der Lösung $y(x)$ in $x \in [a, \infty)$ bewirken. Genauer ausgedrückt definieren wir:

Definition 6.1.

Das Problem (6.1) heißt stabil, wenn es zu jedem $\epsilon > 0$ eine positive Zahl δ gibt, so daß $\|e(x; \delta_0, a)\| < \epsilon$ gilt für alle $x \in [a, \infty)$ wenn nur $\|\delta_0\| < \delta$ ist; anderenfalls heißt es instabil.

Beispiel

Die Dgl. $y' = 10y - 11\exp(-x); \quad y(0) = 1$ (6.3)

hat die Lösung $y(x) = \exp(-x)$. (6.2) ergibt:

$$e' - 10e = 0; \quad e(0) = \delta_0$$

d.h. $e(x; \delta_0, 0) = \delta_0 \cdot \exp(10x)$

Wie klein auch immer δ_0 ist, stets existiert ein x_0 so daß $\|e(x; \delta_0, 0)\| > \epsilon$ für $x > x_0$. Die Dgl. ist daher instabil.

Allgemeiner gilt der folgende Satz:

Satz 6.1.

Das lineare Differentialgleichungssystem mit konstanten Koeffizienten $y' = Ay, y(a) = \eta_0, y: \mathbb{R} \rightarrow \mathbb{R}^n$; ist genau dann stabil, wenn für die Eigenwerte λ_j der Matrix A gilt: $\max_{1 \leq j \leq n} \operatorname{Re} \lambda_j \leq 0$.

Der Beweis folgt aus $e(x; \delta_0, a) = \exp((x-a)A)\delta_0; \delta_0 \in \mathbb{R}^n$.

Instabile Differentialgleichungsprobleme lassen sich über "kleine" Intervalle ohne Schwierigkeit lösen; bei "großen" Intervallen versagt jedoch jedes numerische Verfahren.

§ 7. Stabilität bei fester Schrittweite h

7.1. Allgemeine Betrachtungen

Bisher betrachteten wir stets ein endliches Integrationsintervall $[a, b]$ mit $b-a=m \cdot h$ und untersuchten die Stabilität eines Verfahrens für $m \rightarrow \infty$, d.h. für $h \rightarrow 0$. Im folgenden sei $[a, \infty)$ das Integrationsintervall und wir fragen nun nach der Stabilität eines Verfahrens für $m \rightarrow \infty$ bei fester Schrittweite $h > 0$. Insbesondere interessiert, wie klein h wenigstens gewählt werden muß, um Stabilität zu erreichen.

Eine Antwort hierauf ist auch bei großen endlichen Integrationsintervallen von praktischem Interesse (wobei wir ein Intervall $[a, b]$ schon "groß" nennen wollen, wenn $\frac{b-a}{h} > 100$ ist).

Bezeichne $M(\rho, \sigma)$ wieder ein k -Schritt-Verfahren der Form (1.6) mit den erzeugenden Polynomen ρ und σ . Dann gilt mit $E^r y_j := y_{j+r}$:

$$\rho(E)y_j^* - h\sigma(E)f_j^* = 0$$

und für die (exakte) Lösung y der Dgl. (1.1) gilt

$$\rho(E)y_j - h\sigma(E)f_j = h T_h[y](x_j)$$

Hieraus folgt

$$\rho(E)(y_j - y_j^*) - h\sigma(E)(f_j - f_j^*) = h T_h[y](x_j) \quad (7.1)$$

Mit $f(x_j, y_j) - f(x_j, y_j^*) = J_f(x_j, \tilde{y}_j)(y_j - y_j^*)$

J_f : Funktionalmatrix von f ; $\tilde{y}_j = y_j^* + \delta_j(y_j - y_j^*)$; $\|\delta_j\| < 1$

und den Abkürzungen

$$e_j := y_j - y_j^*; \quad J_f(x_j, \tilde{y}_j) := J_j; \quad T_h[y](x_j) := T_j \quad (7.2)$$

ergibt sich die folgende Differenzengleichung für den globalen Fehler e_j :

$$\rho(E)e_j - h\sigma(E)J_j e_j = hT_j \quad (7.3)$$

Sind die Startfehler e_j ($j=0(1)k-1$) bekannt und kennt man auch J_j und T_j , so läßt sich (7.3) lösen und der globale Fehler e_j für jedes j berechnen. Im allgemeinen sind J_j und T_j jedoch nicht bekannt. Daher sind exakte Stabilitätsaussagen nur für den sehr einfachen Fall der linearen Differentialgleichung $y'=\lambda y$, $y:\mathbb{R}\rightarrow\mathbb{R}$ möglich, auf den wir uns im folgenden beschränken, sowie für lineare Differentialgleichungssysteme mit konstanten Koeffizienten: $y'=Ay$, $A\in\mathbb{R}(n,n)$, $y:\mathbb{R}\rightarrow\mathbb{R}^n$.

7.2. Stabilitätsbereiche

Im Fall der Dgl. $y'=\lambda y$, $y(a)=\eta_0$, $y:\mathbb{R}\rightarrow\mathbb{R}$, $\lambda\in\mathbb{C}$ ist $J_j\in\mathbb{C}$ und von j unabhängig und es gilt $J_j=\lambda$. (7.3) lautet dann

$$(\rho(E)-H\sigma(E))e_j = hT_j \quad \text{mit} \quad H:=\lambda h \quad (7.4)$$

Die homogene Gleichung (7.4) hat dann die Lösung

$$\bar{e}_j = \sum_{i=1}^k C_i (\mu_i)^j$$

falls die Nullstellen $\mu_i(H)$, $i=1(1)k$, des Polynoms

$$Q(\mu, H) = \rho(\mu) - H\sigma(\mu) \quad (7.5)$$

einfach sind. Bei (beispielsweise) r -facher Nullstelle μ_1 lautet sie

$$\bar{e}_j = (C_1 + C_2 j + \dots + C_r j^{r-1}) \mu_1^j + C_{r+1} \mu_{r+1}^j + \dots + C_k \mu_k^j \quad (7.6)$$

Daher muß h wenigstens so klein gewählt werden, daß für $H:=\lambda h$ gilt:

1. $|\mu_i(H)| \leq 1, \quad i=1(1)k$ (7.7)
2. die $\mu_i(H)$ vom Betrag 1 müssen einfach sein,

anderenfalls wächst der globale Fehler e_j unbeschränkt mit j . Die Lösung e_j der inhomogenen Gleichung (7.4) kann mit $j \rightarrow \infty$ auch dann unbeschränkt wachsen, wenn (7.7) erfüllt ist, jedoch nicht im Falle $\operatorname{Re} \lambda < 0$.

Anmerkungen:

1. Für $h \rightarrow 0$ geht die Bedingung (7.7) über in das Stabilitätskriterium des Satzes 4.1. (Seite 28), denn die $\mu_i(0), i=1(1)k$, sind gerade die Nullstellen des Polynoms p .
2. Nach Satz 6.1. ist $y'=\lambda y$ für $\operatorname{Re} \lambda > 0$ instabil, d.h. die Lösung y_j ist für $j \rightarrow \infty$ nicht beschränkt. Daher braucht (7.7) für $\operatorname{Re} \lambda > 0$ nicht erfüllt zu sein, jedoch sollte der Fehler e_j für $j \rightarrow \infty$ nicht schneller wachsen als y_j , d.h. der relative Fehler sollte beschränkt sein (siehe hierzu Def. 7.2).

Wir fassen in etwas allgemeinerer Formulierung zusammen: Wendet man ein Diskretisierungsverfahren M an auf die (skalare) Differentialgleichung $y'=\lambda y$, so ergibt sich eine lineare Differenzengleichung mit konstanten Koeffizienten. Das charakteristische Polynom $Q(\mu;H)$ dieser Differenzengleichung enthält $H:=\lambda h$ als Parameter und von seinen Nullstellen $\mu_i(H)$ hängt das Verhalten des globalen Fehlers e_j für $j \rightarrow \infty$ ab.

Definition 7.1.

$Q(\mu;H)$ heißt charakteristisches Polynom des Verfahrens M .

Beispiele:

1. Im Fall des k-Schritt-Verfahrens $M(\rho, \sigma)$ hat, wie wir gesehen haben, die Differenzengleichung die Form (vgl. (7.1.))

$$\rho(E)y_j^* - \lambda h \sigma(E)y_j^* = 0 \quad (7.8)$$

und das charakteristische Polynom von $M(\rho, \sigma)$ lautet

$$Q(\mu; H) = \rho(\mu) - H\sigma(\mu)$$

2. Das charakteristische Polynom des modifizierten Euler-Verfahrens (1.8) lautet:

$$Q(\mu; H) = \mu - (1 + H + \frac{1}{2}H^2)$$

Die Forderungen (7.7) und Bemerkung 2 legen die folgende Definition nahe:

Definition 7.2.

Ein Verfahren M der Ordnung q heißt absolut stabil für $H \in \mathbb{C}$, wenn für die Nullstellen $\mu_i(H)$, $i=1(1)k$, seines charakteristischen Polynoms $Q(\mu; H)$ gilt:

$$|\mu_i(H)| \leq 1 \quad (7.9)$$

Es heißt relativ stabil für H , wenn gilt:

$$|\mu_i(H)| \leq |\mu_1(H)|, \quad i=2(1)k \quad (7.10)$$

wobei μ_1 diejenige Wurzel ist, für die gilt:

$$\mu_1 = e^H + \mathcal{O}(h^{q+1})$$

Diese Wurzel heißt Hauptwurzel.

Man beachte, daß ein für H absolut stabiles Verfahren nicht notwendig auch relativ stabil zu sein braucht. Bei einem für H absolut, jedoch nicht relativ stabilen Verfahren geht im Falle der (skalaren) Dgl. $y' = \lambda y$ zwar der absolute Fehler e_j mit $j \rightarrow \infty$ gegen Null nicht aber der relative Fehler e_j/y_j .

Konsistente Einschrittverfahren sind stets relativ stabil.

Beim Beispiel des Abschnitts 3.1. (Seite 13ff) ist

$$|\mu_2| > 1 \quad \text{für alle } \lambda h < 0$$

es ist daher für kein $H < 0$ absolut stabil; wegen

$$|\mu_2| > |\mu_1| \quad \text{für alle } \lambda h < 0$$

ist es auch für kein $H < 0$ relativ stabil.

Definition 7.3.

Seien $\mu_i(H)$, $i=1(1)k$, die Nullstellen des charakteristischen Polynoms $Q(\mu, H)$ des Verfahrens M . Dann heißt die Menge

$$H := \{H \in \mathbb{C} : |\mu_i(H)| < 1, \quad i=1(1)k\} \quad (7.11)$$

Bereich absoluter Stabilität (kürzer: Stabilitätsbereich) von M .

H ist also die Menge derjenigen $H = \lambda h$, für welche die Methode M , angewandt auf $y' = \lambda y$, (absolut) stabil ist.

Diese Aussage läßt sich wie folgt auf Systeme linearer Differentialgleichungen mit konstanten Koeffizienten der Form

$$y' = Ay ; \quad y: \mathbb{R} \rightarrow \mathbb{R}^n; \quad A \in \mathbb{R}(n, n) \quad (7.12)$$

verallgemeinern, wobei auch deutlich wird, warum komplexwertige λ zugelassen wurden:

Satz 7.1.

Die Methode $M(\rho, \sigma)$, angewandt auf (7.12), ist genau dann absolut stabil, wenn gilt: $h\lambda_j \in H$, $j=1(1)n$, wobei $\lambda_j \in \mathbb{C}$ die Eigenwerte der Matrix A sind.

Beweis:

$M(\rho, \sigma)$ auf $y' = Ay$ angewandt ergibt:

$$\rho(E)y_j^* - hA\sigma(E)y_j^* = 0 \quad (7.13)$$

Sei $N = S^{-1}AS$ die Jordan'sche Normalform von A , so folgt aus (7.13) durch Multiplikation von links mit S^{-1} und $z := S^{-1}y^*$:

$$N = \begin{pmatrix} \lambda_1 & 1 & & & \\ & \ddots & \ddots & & \\ & & \lambda_1 & 1 & \\ & & & \ddots & \ddots \\ 0 & & & & \lambda_r & 1 \\ & & & & & \ddots \\ & & & & & & \lambda_r \end{pmatrix}$$

$$\rho(E)z_j - hN\sigma(E)z_j = 0 \quad (7.14)$$

Diese Gleichung zerfällt in r (skalare) homogene Differenzengleichungen

$$\rho(E)z_j^{(i)} - h\lambda_i\sigma(E)z_j^{(i)} = 0; \quad i=1(1)r \quad (7.15a)$$

die je einem der r Elementarteiler von A zugeordnet sind, und - falls einige der Elementarteiler nicht linear sind - in $(n-r)$ inhomogene Differenzengleichungen

$$\rho(E)z_j^{(i)} - h\lambda_i\sigma(E)z_j^{(i)} = h\sigma(E)z_j^{(i+1)}; \quad i=r+1(1)n \quad (7.15b)$$

Damit ist Satz 7.1. durch Zurückführung auf den skalaren Fall $y' = \lambda_j y$ bewiesen, denn auch die partikuläre Lösung der inhomogenen Differenzengleichung (7.15b) ist beschränkt, wenn (7.15a) stabil ist.

Bevor die Definitionen und Ergebnisse dieses Abschnitts an Beispielen erläutert werden noch eine Definition:

Definition 7.4.

Verfahren mit $H \supset \{H: \operatorname{Re} H < 0\}$ heißen A-stabil (unbeschränkt absolut stabil); sie heißen A(α)-stabil, wenn $H \supset \{H: |\arg(-H)| < \alpha; H \neq 0\}$

A($\frac{\pi}{2}$)-stabile Verfahren sind offensichtlich A-stabil. A-stabile Verfahren haben den Vorzug, daß h unabhängig von Stabilitäts-erwägungen (und nur mit Blick auf die Genauigkeitsanforderungen) gewählt werden kann. Das ist z.B. wichtig, wenn gleichzeitig betragsmäßig sehr große und sehr kleine λ_i -Werte auftreten (vgl. das erste Beispiel im nächsten Abschnitt). Differentialgleichungen, bei denen das der Fall ist, heißen steif.

Schwach stabile Verfahren (s. Seite 29) haben einen leeren Stabilitätsbereich. Das Beispiel in Abschnitt 9.5.3. zeigt jedoch, daß ein Prädiktor-Korrektor-Verfahren mit schwach stabilem Korrektor durchaus einen nicht-leeren Stabilitätsbereich haben kann, wenn der Prädiktor entsprechend gewählt wird.

Für A-stabile lineare k-Schritt-Verfahren bewies Dahlquist:

Satz 7.2.

Kein explizites Verfahren der Form (1.6) ist A-stabil; A-stabile, implizite Verfahren der Form (1.6) haben höchstens die Ordnung 2.

Der folgende Satz über A(α)-stabile Verfahren stammt von Widlund¹⁾:

Satz 7.3.

Kein explizites Verfahren der Form (1.6) ist A(0)-stabil; zu jedem $\alpha \in [0, \frac{\pi}{2})$ gibt es A(α)-stabile (implizite) k-Schritt-Verfahren der Ordnung q mit $k=q=3$ und $k=q=4$.

Beide Sätze gelten nicht für zyklische Verfahren, insbesondere gibt es A(α)-stabile $k=3, M=2$ -Verfahren der Ordnung 4 und $k=3, M=3$ -Verfahren der Ordnung 5 (vgl. Abschnitt 10.2.).

¹⁾Widlund: A Note on Unconditionally Stable Linear Multistep Methods, BIT 7, 65-70 (1967)

Die Definitionen und Ergebnisse dieses Paragraphen erscheinen insofern unbefriedigend, als sie genaue Stabilitätsaussagen nur für den trivialen Fall linearer Differentialgleichungssysteme mit konstanten Koeffizienten zulassen.

Dennoch ist es sinnvoll, die Stabilitätsbereiche von Diskretisierungsmethoden zu untersuchen, denn durch Linearisierung (entweder der gegebenen Differentialgleichung oder der Fehlergleichung (7.3)) lassen sich die Ergebnisse dieses Abschnitts (in begrenztem Umfang) auch auf allgemeinere Differentialgleichungsprobleme übertragen; Ansätze hierfür findet man in Stetter [73], S. 178ff.

7.3. Beispiele

Beispiel 1:

$$\text{Das System } y' = Ay \text{ mit } y(0) = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}; A = \begin{pmatrix} -1 & -24 & 0 \\ 0 & -25 & 0 \\ 0 & 125 & -150 \end{pmatrix} \quad (7.16)$$

$$\begin{aligned} \text{hat die Lösung: } y_1 &= e^{-x} + e^{-25x} \\ y_2 &= e^{-25x} \\ y_3 &= e^{-25x} - e^{-150x} \end{aligned}$$

Die Eigenwerte von A sind $\lambda_1 = -1$; $\lambda_2 = -25$; $\lambda_3 = -150$; mit der Schrittweite $h = 0,024$ ergibt sich daher:

$$H_1 := \lambda_1 h = -0,024; H_2 := \lambda_2 h = -0,6; H_3 := \lambda_3 h = -3,6.$$

Das (hier nicht näher angegebene) Verfahren M zur numerischen Behandlung der Dgl. hat ein charakteristisches Polynom mit den Nullstellen $\mu_1(H)$ und $\mu_2(H)$, für die gilt:

H	$\mu_1(H)$	$\mu_2(H)$
H_1	0,976	-0,703
H_2	0,549	-0,867
H_3	0,004	0,851

H_1, H_2, H_3 liegen im Stabilitätsbereich von M, (denn die zugehörigen Nullstellen μ_1, μ_2 des charakteristischen Polynoms sind betragsmäßig kleiner als 1) das Verfahren ist daher absolut stabil für die gewählte Schrittweite h.

Es ist $e^{H_1} = 0,976$; $e^{H_2} = 0,549$; $e^{H_3} = 0,027$; die ersten beiden Werte stimmen gut mit den Hauptwurzeln $\mu_1(H_1)$ und $\mu_1(H_2)$ überein, das Verfahren M wird daher gute Näherungen für die Terme $e^{\lambda_1 x}$ und $e^{\lambda_2 x}$ ergeben. Die schlechte Übereinstimmung von e^{H_3} mit $\mu_1(H_3)$ läßt eine ungenaue Approximation des Terms $e^{\lambda_3 x}$ erwarten; da dieser Term jedoch rasch abklingt, wird die Näherung y^* hierdurch nicht nennenswert verfälscht. Gravierender wird sich auswirken, daß M für H_2 und H_3 relativ instabil ist. Das hat zur Folge, daß die Lösungsterme $e^{\lambda_2 x}$ und $e^{\lambda_3 x}$ mit großem relativen Fehler berechnet werden. Das fällt bei y_1 wegen des viel größeren Terms $e^{\lambda_1 x}$ nicht ins Gewicht, wohl aber bei y_2 und y_3 , für die man einen großen relativen Fehler erwarten muß.

Tatsächlich werden diese theoretischen Voraussagen von den numerischen Ergebnissen bestätigt. Das Verfahren M ergab in den Punkten x_{10} , x_{30} und x_{50} die folgenden Fehler $e_i(x_j) = y_i(x_j) - y_i^*(x_j)$; ($j=10, 30, 50$; $i=1,2,3$):

j	10	30	50	
x_j	0,24	0,72	1,20	
$e_1(x_j)$	$0,784 \cdot 10^{-5}$	$0,194 \cdot 10^{-7}$	$0,342 \cdot 10^{-9}$	kl. absol. u. rel. Fehler große <u>relative</u> Fehler " " "
$e_2(x_j)$	$0,784 \cdot 10^{-5}$	$0,192 \cdot 10^{-7}$	$0,942 \cdot 10^{-10}$	
$e_3(x_j)$	$-0,437 \cdot 10^{-3}$	$-0,176 \cdot 10^{-4}$	$-0,700 \cdot 10^{-6}$	
$y_1(x_j)$	0,780	0,487	0,301	
$y_2(x_j)$	$0,248 \cdot 10^{-2}$	$0,152 \cdot 10^{-7}$	$0,936 \cdot 10^{-13}$	
$y_3(x_j)$	$0,248 \cdot 10^{-2}$	$0,152 \cdot 10^{-7}$	$0,936 \cdot 10^{-13}$	

Der Beitrag $e^{\lambda_3 x}$ des Eigenwertes $\lambda_3 = -150$ ist für $x > 0,1$ praktisch ohne Bedeutung für die Lösung der Dgl. Dennoch bestimmt λ_3 wesentlich die Wahl von h : Wäre h größer gewählt worden (was bei der unnötig hohen Genauigkeit der numerischen Lösung sinnvoll erscheint) so hätte $H_3 = \lambda_3 h$ nicht mehr im Stabilitätsbereich von M gelegen und das Verfahren hätte divergiert. Wegen der betragsmäßig sehr unterschiedlichen Eigenwerte von A ist man gezwungen, die Stabilität durch (vom Standpunkt der Genauigkeit) unnötig kleine Schrittweiten h zu sichern. Diese Situation ist typisch für Differentialgleichungen, die steif sind.

Beispiel 2:

Das implizite Euler-Verfahren

$$y_{j+1}^* = y_j^* + h f_{j+1}^*$$

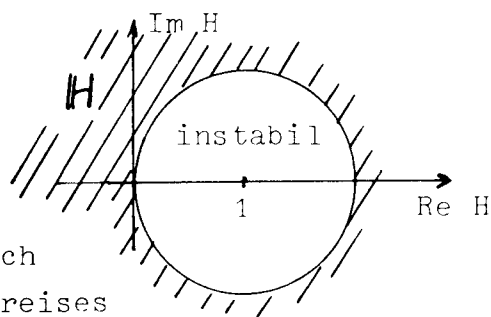
ergibt für die Dgl. $y' = \lambda y$ die Differenzengleichung

$$y_{j+1}^* = y_j^* + H y_{j+1}^* ; \quad H := \lambda h$$

Sein charakteristisches Polynom lautet daher

$$Q(\mu, H) = (1-H)\mu - 1$$

und es ist $\mu_1 = (1-H)^{-1}$



Die Stabilitätsgrenze ist die durch

$H = \frac{\mu-1}{\mu}$ vermittelte Abbildung des Kreises

$\mu(\psi) = e^{i\psi}$, d.h. der Kreis $H(\psi) = 1 - e^{-i\psi}$. Das Verfahren ist A-stabil.

Beispiel 3:

Das Trapezverfahren

$$y_{j+1}^* = y_j^* + \frac{h}{2} (f_j^* + f_{j+1}^*)$$

ergibt für die Dgl. $y' = \lambda y$ die Differenzengleichung

$$y_{j+1}^* = y_j^* + \frac{H}{2} y_j^* + \frac{H}{2} y_{j+1}^*$$

Das charakteristische Polynom ist $Q(\mu, H) = (1 - \frac{H}{2})\mu - (1 + \frac{H}{2})$. Es ist

$$|\mu_1| = \left| \frac{1 + \frac{H}{2}}{1 - \frac{H}{2}} \right| < 1 \quad (7.17)$$

für alle H mit $\text{Re } H < 0$; das Verfahren ist daher ebenfalls A-stabil.

Von allen A-stabilen linearen k -Schritt-Verfahren hat es den kleinsten Ersetzungsfehler; zugleich ist es das einzige A(o)-stabile lineare k -Schritt-Verfahren mit der Ordnung $q \geq k+1$.

Beispiel 4:

Das 2-Schritt-Verfahren von Milne-Simpson:

$$y_{j+2}^* - y_j^* = \frac{h}{3} (f_{j+2}^* + 4f_{j+1}^* + f_j^*)$$

hat das charakteristische Polynom

$$Q(\mu, H) = (1 - \frac{H}{3})\mu^2 - \frac{4}{3}H\mu - (1 + \frac{H}{3})$$

Da $Q(\mu, H) = 0$ den Kreis $\mu(\psi) = e^{i\psi}$ auf $\{H: \operatorname{Re} H = 0, -\sqrt{3} \leq \operatorname{Im} H \leq \sqrt{3}\}$ abbildet, ist H leer (Das Verfahren ist schwach stabil).

Beispiel 5:

Das Runge-Kutta-Verfahren 3. Ordnung:

$$y_{j+1}^* = y_j^* + \frac{1}{6} (k_1 + 4k_2 + k_3)$$

$$k_1 = h f(x_j, y_j^*)$$

$$k_2 = h f(x_j + \frac{h}{2}, y_j^* + \frac{k_1}{2})$$

$$k_3 = h f(x_j + h, y_j^* - k_1 + 2k_2)$$

ergibt für die Dgl. $y' = \lambda y$ die Differenzengleichung:

$$y_{j+1}^* = (1 + H + \frac{1}{2} H^2 + \frac{1}{6} H^3) y_j^*$$

$$\text{d.h.} \quad Q(\mu, H) = \mu - (1 + H + \frac{1}{2!} H^2 + \frac{1}{3!} H^3) \quad (7.18)$$

Für reelle H ist es stabil, wenn $-2,51 \leq H \leq 0$.

Beispiel 6:

Die implizite_Gauss-Formel 4. Ordnung

$$y_{j+1}^* = y_j^* + \frac{1}{2}(k_1 + k_2)$$

$$k_1 = h f(x_j + \frac{h}{6}(3 - \sqrt{3}), y_j^* + \frac{1}{4} k_1 + \frac{1}{12} (3 - 2\sqrt{3})k_2)$$

$$k_2 = h f(x_j + \frac{h}{6}(3 + \sqrt{3}), y_j^* + \frac{1}{12} (3 + 2\sqrt{3})k_1 + \frac{1}{4} k_2)$$

ergibt für die Dgl. $y' = \lambda y$:

$$y_{j+1}^* = y_j^* + \frac{1}{2} (k_1 + k_2) \quad (7.19)$$

$$k_1 = h\lambda (y_j^* + \frac{1}{4} k_1 + \frac{1}{12} (3 - 2\sqrt{3})k_2)$$

$$k_2 = h\lambda (y_j^* + \frac{1}{12} (3 + 2\sqrt{3})k_1 + \frac{1}{4} k_2) \quad (7.20)$$

Löst man die Gleichungen (7.20) nach k_1 und k_2 auf, so ergibt (7.19):

$$y_{j+1}^* = \mu_1 y_j^* \quad \text{mit} \quad \mu_1 = \frac{1 + \frac{H}{2} + \frac{H^2}{12}}{1 - \frac{H}{2} + \frac{H^2}{12}}$$

das charakteristische Polynom lautet daher:

$$Q(\mu, H) = \mu - \mu_1$$

Man kann zeigen, daß $|\mu_1| < 1$ für alle H mit $\text{Re } H < 0$; das Verfahren ist daher A-stabil.

Hinweis: μ_1 ist Padé-Approximation von e^H , das gleiche gilt für μ_1 in (7.17). Zu allen Padé-Approximationen $\mu_1(H)$ von e^H lassen sich Einschrittverfahren mit $Q(\mu, H) = \mu - \mu_1$ angeben.

§ 8 Einschrittverfahren

Wendet man Satz 3.2. (Seite 23) mit $k=1$ auf Einschritt-Verfahren der Form (1.5) an und berücksichtigt man dabei auch die früheren Ergebnisse über Konsistenz (Satz 2.1.) und Stabilität (Satz 3.1.), so ergibt sich der folgende Konvergenzsatz für Einschritt-Verfahren:

Satz 8.1.

Sei y Lösung der Anfangswertaufgabe (1.1) und

$$y_{j+1}^* = y_j^* + h\phi(x_j, y_j^*; h); y_0^* = \eta_0; \quad j = 0(1)m-1$$

ein Einschrittverfahren zur Approximation von $y(x_{j+1})$, $x_{j+1} \in I_h$. Dieses Verfahren konvergiert, wenn gilt:

1. $\phi(x, y; h)$ ist stetig in $G := \{(x, y, h) : a \leq x \leq b; y \in \mathbb{R}^n; 0 \leq h \leq h_0\}$ und genügt dort der Lipschitzbedingung

$$\|\phi(x, y_1; h) - \phi(x, y_2; h)\| \leq K \|y_1 - y_2\|$$

2. $\phi(x, y; 0) \equiv f(x, y)$

Wenn außerdem

3. alle partiellen Ableitungen von f der Ordnung q ($q \geq 1$) stetig sind und
4. $\phi(x, y; h) = f(x, y) + \frac{h}{2!} Df(x, y) + \dots + \frac{h^{q-1}}{q!} D^{q-1}f(x, y) + O(h^q)$

so konvergiert das Verfahren mit der Ordnung q und es gilt die Fehlerabschätzung:

$$\text{mit } \|y(x_j) - y_j^*\| \leq K^{-1} (e^{(x_j - a)K} - 1) C h^q; \quad j = 0(1)m$$

$$\|y(x+h) - y(x) - h\phi(x, y(x); h)\| \leq C h^{q+1}$$

Es stellt sich nun die Frage, wie man Einschritt-Verfahren der Form (1.5) findet. Hiermit beschäftigen sich die nächsten Abschnitte.

8.1. Verfahren mit Ableitungen:

Sind alle partiellen Ableitungen von f bis zur Ordnung q stetig in G , so ergibt sich nach Satz 4.2. sofort ein (konsistentes und stabiles, daher auch konvergentes) Einschrittverfahren der Ordnung q zur Berechnung von Näherungen y_j^* von $y(x)$, $x \in I_h$:

$$y_0^* = \eta_0; \quad y_{j+1}^* = y_j^* + h \cdot \hat{\phi}(x_j, y_j^*; h) \quad j=0(1)m-1 \quad (8.1)$$

$$\text{mit} \quad \hat{\phi}(x_j, y_j; h) = \sum_{i=0}^{q-1} \frac{h^i}{(i+1)!} D^i f(x_j, y_j). \quad (8.2)$$

Dieses Verfahren mit den Ableitungen $D^i f(x_j, y_j) = (\frac{\partial}{\partial x} + f \frac{\partial}{\partial y})^i f(x_j, y_j)$ ist allerdings nur dann für den praktischen Gebrauch geeignet, wenn sich die Ausdrücke $D^i f(x_j, y_j)$, etwa durch Rekursionsformeln, leicht berechnen lassen; das ist aber schon bei einfachen Funktionen f nur selten der Fall¹⁾. Der Vorteil dieses sogenannten Taylor-Verfahrens (8.1) besteht darin, daß sich, bei hinreichend oft stetig differenzierbarer rechter Seite f , eine beliebig hohe Ordnung q erzielen läßt.

Beispiele:

1. Für $q=1$ ergibt (8.1) wieder das bekannte Euler-Verfahren.
2. Wendet man (8.1) für $q=2$ auf das Anfangswertproblem

$$y'(x) = -xy^2(x); \quad y(1) = 2,$$

an, so erhält man mit $Df = 2x^2y^3 - y^2$

$$y_0^* = 2; \quad y_{j+1}^* = y_j^* + h y_j^{*2} (-x_j + h x_j^2 y_j^* - \frac{h}{2}); \quad j=0(1)m-1 \quad (8.3)$$

¹⁾ Diese Schwierigkeit wurde in neuerer Zeit durch Entwicklung spezieller Programmiersprachen zur Behandlung analytischer Ausdrücke, z.B. FORMAC, überwunden.

8.2. Explizite Runge-Kutta-Verfahren:

Die Idee Runges beruht darauf, die komplizierte Berechnung der Ausdrücke $D^i f(x_j, y_j)$ unter Beibehaltung der erstrebten Ordnung q zu vermeiden, indem man $D^i f(x_j, y_j)$ durch Linearkombinationen von Funktionswerten $f(x, y)$ ersetzt, genommen an r Zwischenstellen im Intervall von (x_j, y_j) bis (x_{j+1}, y_{j+1}) .

Dazu setzt man

$$\phi_r(x, y; h) = \sum_{i=1}^r \gamma_i k_i, \quad (8.4)$$

mit

$$\begin{aligned} k_1 &= f(x, y) \\ k_2 &= f(x + \alpha_2 h, y + \beta_{21} k_1 h) \\ k_3 &= f(x + \alpha_3 h, y + \beta_{31} k_1 h + \beta_{32} k_2 h) \\ &\dots\dots\dots \\ k_r &= f(x + \alpha_r h, y + h \sum_{i=1}^{r-1} \beta_{ri} k_i). \end{aligned} \quad (8.5)$$

Die Parameter α_i ($\alpha_i \in [0, 1]$), β_{ij} , γ_i ($\gamma_i \geq 0$) werden dann so bestimmt, daß für alle hinreichend oft stetig differenzierbaren Funktionen $f \in \mathcal{C}^q(G)$ gilt:

$$\phi_r(x, y; h) - \hat{\phi}(x, y; h) = \mathcal{O}(h^q), \quad (8.6)$$

so daß das so erhaltene r-stufige explizite Verfahren von Runge-Kutta

$$y_0^* = \eta_0; \quad y_{j+1}^* = y_j^* + h \phi_r(x_j, y_j^*; h); \quad j = 0(1)m-1$$

mit der Ordnung q konvergiert (ϕ_r ist mit f Lipschitz-stetig). Aus (8.6) erhält man durch Taylorabgleich ein nichtlineares und im allgemeinen nicht eindeutig lösbares Gleichungssystem zur Bestimmung der Parameter α_i , β_{ij} , γ_i ; daher gibt es für $r > 1$ unendlich viele ϕ_r , die (8.6) erfüllen. Es ist üblich, ein r -stufiges Runge-Kutta-Verfahren ("RK-Verfahren") durch folgendes, von Butcher stammendes, Parameter-Schema zu charakterisieren:

$$\begin{array}{c|cccc}
 \alpha_2 & \beta_{21} & & & \\
 \alpha_3 & \beta_{31} & \beta_{32} & & \\
 \vdots & & & & \\
 \alpha_r & \beta_{r1} & \beta_{r2} & \dots\dots\dots \beta_{r,r-1} & \\
 \hline
 & \gamma_1 & \gamma_2 & \dots\dots\dots \gamma_{r-1} & \gamma_r
 \end{array} \tag{8.7}$$

RK-Verfahren, deren Parameter-Schema die Dreiecksgestalt (8.7) aufweist, heißen explizit, denn für die Berechnung von k_i werden nur die bekannten Größen $k_1, \dots, k_{i-1}$ herangezogen. In Abschnitt 8.3. werden wir auch noch implizite RK-Verfahren kennenlernen, in deren Parameterschema $\beta_{ij} \neq 0$ mit $i \leq j$ auftreten; bei diesen ist in jedem Rechenschritt zur Berechnung der k_i ein nichtlineares Gleichungssystem zu lösen.

Für alle RK-Verfahren gilt:

$$\sum_{i=1}^r \gamma_i = 1 \text{ und} \tag{8.8}$$

$$\alpha_i = \sum_{j=1}^{i-1} \beta_{ij} ; \quad i=2(1)r. \tag{8.9}$$

8.2.1. Explizite RK-Formeln mit $r=1,2,3$ Stufen:

In diesem Abschnitt werden Spezialfälle von expliziten RK-Verfahren mit $r=1,2$ und 3 Stufen betrachtet. Dabei werden die Ordnung q und die Anzahl r der Stufen eines Verfahrens durch ein geordnetes Zahlenpaar (q,r) angegeben.

1. Die einzige (1,1)-Formel ist die Eulersche Formel:

$$y_0^* = \eta_0 ; \quad y_{j+1}^* = y_j^* + hf(x_j, y_j^*), \quad j=0(1)m-1.$$

Für $f \in \mathcal{C}^1(G)$ hat das Verfahren die Ordnung $q=1$.

2. (2,2)-Formeln:

Für $r=2$ und $f \in \mathcal{C}^2(G)$ ergibt sich:

$$k_1(x,y) = f(x,y)$$

$$\begin{aligned} k_2(x,y) &= f(x + \alpha_2 h, y + \beta_{21} k_1(x,y)h) \\ &= f(x,y) + \alpha_2 h f_x(x,y) + \beta_{21} h f(x,y) f_y(x,y) + \mathcal{O}(h^2) \end{aligned}$$

Daraus folgt:

$$\phi_2 - \hat{\phi} = \gamma_1 f + \gamma_2 (f + \alpha_2 h f_x + \beta_{21} h f f_y) - f - \frac{h}{2} f_x - \frac{h}{2} f f_y + \mathcal{O}(h^2)$$

Diese Differenz ist von der Ordnung 2 in h , wenn

$$\gamma_1 + \gamma_2 - 1 = 0; \quad \gamma_2 \alpha_2 - \frac{1}{2} = 0; \quad \gamma_2 \beta_{21} - \frac{1}{2} = 0.$$

Für $\alpha_2 \neq 0$ erhält man: $\beta_{21} = \alpha_2$; $\gamma_1 = 1 - \frac{1}{2\alpha_2}$; $\gamma_2 = \frac{1}{2\alpha_2}$

Daraus folgt die allgemeine Form eines 2-stufigen RK-Verfahrens der Ordnung 2:

$$y_0^* = \eta_0; \quad y_{j+1}^* = y_j^* + h \left(1 - \frac{1}{2\alpha_2}\right) f(x_j, y_j^*) + \frac{h}{2\alpha_2} f(x_j + \alpha_2 h, y_j^* + \alpha_2 h f(x_j, y_j^*)) \quad (8.10)$$

Für jede Wahl des freien Parameters α_2 erhält man ein anderes (2,2)-Verfahren.

Spezialfälle:

a) Wählt man $\alpha_2 = \frac{1}{2}$, so erhält man das modifizierte Euler-Verfahren:

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 0 \quad 1 \end{array}$$

Für $f(x,y) \equiv f(x)$ geht dieses Verfahren über in die Tangenten-trapezregel zur numerischen Quadratur von f im Intervall $[x_j, x_{j+1}]$.

b) Für $\alpha_2 = \frac{2}{3}$ ergibt sich das Verfahren von Heun:

$$\begin{array}{c|cc} \frac{2}{3} & & \frac{2}{3} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}$$

c) Schließlich hat man für $\alpha_2 = 1$ das verbesserte Euler-Verfahren:

$$\begin{array}{c|cc} 1 & & 1 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Dieses Verfahren entspricht im Falle $f(x,y) \equiv f(x)$ der Quadratur von f in $[x_j, x_{j+1}]$ nach der Sehnentrapezregel.

Alle (2,2)-RK-Verfahren der Form (8.10) haben einen beschränkten Stabilitätsbereich (vgl. Abschnitt 7.2., Seite 46) und bei reellem H ist ihre Stabilität für

$$-2.0 < H < 0$$

gewährleistet.

Beispiel:

Das verbesserte Verfahren von Euler auf

$$y'(x) = -xy^2(x); \quad y(1) = 2$$

angewandt ergibt:

$$y_0^* = \eta_0; \quad y_{j+1}^* = y_j^* - \frac{h}{2} x_j y_j^{*2} - \frac{h}{2} (x_j + h) (y_j^* - h x_j y_j^{*2})^2, \quad j = 0(1)m-1.$$

Für $h=0.1$ und $x_j=1+0.1j$, $j=0(1)10$ erhält man hierfür die Fehler $e_E=y_j^*-y(x_j)$, die in der folgenden Tabelle den Fehlern e_A des Algorithmus (8.3) (Seite 57) gegenübergestellt sind:

x_j	e_E	e_A
1.0	0.0000	0.0000
1.1	0.0063	0.0071
1.2	0.0085	0.0096
1.3	0.0089	0.0099
1.4	0.0084	0.0094
1.5	0.0077	0.0096
1.6	0.0069	0.0076
1.7	0.0061	0.0067
1.8	0.0053	0.0059
1.9	0.0047	0.0052
2.0	0.0041	0.0045

3. (3,3)-Formeln:

Die Herleitung einer allgemeinen (3,3)-Formel erfolgt wieder nach dem oben angegebenen Prinzip:

Durch Taylorentwicklung der Ordnung 3 ergibt sich für $f \in \mathcal{C}^3(G)$:

$$k_1 = f$$

$$k_2 = f + h(\alpha_2 f_x + \beta_{21} f f_y) + h^2 \left(\frac{\alpha_2^2}{2} f_{xx} + \alpha_2 \beta_{21} f f_{xy} + \frac{1}{2} \beta_{21}^2 f^2 f_{yy} \right) + \mathcal{O}(h^3)$$

$$k_3 = f + h(\alpha_3 f_x + \beta_{31} f f_y + \beta_{32} f f_y) + h^2 \left[\frac{\alpha_3^2}{2} f_{xx} + \alpha_3 (\beta_{31} + \beta_{32}) f f_{xy} + \frac{1}{2} (\beta_{31} + \beta_{32})^2 f^2 f_{yy} + \right. \\ \left. + \beta_{32} (\alpha_2 f_x + \beta_{21} f f_y) f_y \right] + \mathcal{O}(h^3)$$

und

$$\hat{\phi} = f + \frac{h}{2} (f_x + f f_y) + \frac{h^2}{6} (f_{xx} + 2f f_{xy} + f^2 f_{yy} + f_y f_x + f f_y^2) + \mathcal{O}(h^3)$$

Aus der Bedingung $\phi_3(x,y;h) - \hat{\phi}(x,y;h) = O(h^3)$ erhält man dann das Gleichungssystem:

$$\gamma_1 + \gamma_2 + \gamma_3 = 1 \quad (8.11a)$$

$$\beta_{21} \gamma_2 + (\beta_{31} + \beta_{32}) \gamma_3 = \frac{1}{2} \quad (b)$$

$$\alpha_2 \beta_{21} \gamma_2 + \alpha_3 (\beta_{31} + \beta_{32}) \gamma_3 = \frac{1}{3} \quad (c)$$

$$\alpha_2 \beta_{32} \gamma_3 = \frac{1}{6} \quad (d)$$

$$\alpha_2 \gamma_2 + \alpha_3 \gamma_3 = \frac{1}{2} \quad (e)$$

$$\frac{1}{2} \alpha_2^2 \gamma_2 + \frac{1}{2} \alpha_3^2 \gamma_3 = \frac{1}{6} \quad (f)$$

$$\frac{1}{2} \beta_{21}^2 \gamma_2 + \frac{1}{2} (\beta_{31} + \beta_{32})^2 \gamma_3 = \frac{1}{6} \quad (g)$$

$$\beta_{21} \beta_{32} \gamma_3 = \frac{1}{6} \quad (h)$$

Aus (8.11c) und (8.11f) sowie (8.11d) und (8.11h) folgt unmittelbar (8.9), nämlich

$$\alpha_2 = \beta_{21}, \quad \alpha_3 = \beta_{31} + \beta_{32}.$$

Man hat also insgesamt vier unabhängige Gleichungen für sechs Unbekannte. Je nach Wahl der beiden freien Parameter erhält man verschiedene (3,3)-Formeln, z.B. das klassische (3,3)-Verfahren:

$$\begin{aligned} k_1 &= f(x,y) ; \quad k_2 = f(x + \frac{h}{2}, y + \frac{h}{2} k_1) \\ k_3 &= f(x+h, y - h k_1 + 2 h k_2) \\ y_{j+1}^* &= y_j^* + \frac{h}{6} (k_1(x_j, y_j^*) + 4 k_2(x_j, y_j^*) + k_3(x_j, y_j^*)) \end{aligned} \quad (8.12)$$

mit dem Parameter-Schema:

$\frac{1}{2}$	$\frac{1}{2}$
1	-1 2
$\frac{1}{6}$	$\frac{4}{6}$ $\frac{1}{6}$

Alle expliziten (3,3)-RK-Verfahren haben den reellen Stabilitätsbereich $-2.5 < H < 0$.

8.2.2. Das klassische Runge-Kutta-Verfahren:

Für $r=4$ ergibt sich zur Bestimmung der Parameter $\alpha_i, \beta_{ij}, \gamma_i$ aus der Bedingung

$$\phi_4(x,y;h) - \hat{\phi}(x,y;h) = \mathcal{O}(h^4),$$

(Voraussetzung: $f \in \mathcal{C}^4(G)$) unter Berücksichtigung der allgemein gültigen Beziehungen (8.9)

$$\beta_{21} = \alpha_2, \quad \beta_{31} = \alpha_3 - \beta_{32}, \quad \beta_{41} = \alpha_4 - \beta_{42} - \beta_{43},$$

das folgende nichtlineare Gleichungssystem:

$$\begin{aligned} \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 &= 1 \\ \alpha_2 \gamma_2 + \alpha_3 \gamma_3 + \alpha_4 \gamma_4 &= \frac{1}{2} \\ \alpha_2^2 \gamma_2 + \alpha_3^2 \gamma_3 + \alpha_4^2 \gamma_4 &= \frac{1}{3} \\ \alpha_2^3 \gamma_2 + \alpha_3^3 \gamma_3 + \alpha_4^3 \gamma_4 &= \frac{1}{4} \\ \alpha_2 \beta_{32} \gamma_3 + (\alpha_2 \beta_{42} + \alpha_3 \beta_{43}) \gamma_4 &= \frac{1}{6} \\ \alpha_2 \alpha_3 \beta_{32} \gamma_3 + (\alpha_2 \beta_{42} + \alpha_3 \beta_{43}) \gamma_4 \alpha_4 &= \frac{1}{8} \\ \alpha_2^2 \beta_{32} \gamma_3 + (\alpha_2^2 \beta_{42} + \alpha_3^2 \beta_{43}) \gamma_4 &= \frac{1}{12} \\ \alpha_2 \beta_{32} \beta_{43} \gamma_4 &= \frac{1}{24} \end{aligned} \quad (8.13)$$

Jede Lösung dieses Systems definiert eine Formel der Ordnung $q=4$ von Runge-Kutta; insbesondere ergeben sich die folgenden wichtigen Spezialfälle:

a) Wählt man als freie Parameter $\alpha_2 = \alpha_3 = \frac{1}{2}$, so erhält man das klassische RK-Verfahren, der Ordnung 4:

$$\begin{array}{c|ccc} \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \quad (8.14)$$

Für $f(x,y) \equiv f(x)$ geht dieses Verfahren über in die Quadraturformel von Simpson.

Zur Schrittweitensteuerung nimmt man in der Praxis die Zahl $\|hf_y(x_j, y_j^*)\|$, die erfahrungsgemäß bei 10^{-1} liegen soll.

Im Fall des klassischen (4,4)-RK-Verfahrens gilt nach dem Mittelwertsatz

$$k_3(x,y) - k_2(x,y) = f(x + \frac{h}{2}, y + \frac{h}{2} k_2) - f(x + \frac{h}{2}, y + \frac{h}{2} k_1) = \frac{h}{2} f_y(x + \frac{h}{2}, \tilde{y}) (k_2 - k_1),$$

mit einem Zwischenwert $\tilde{y}$. Daher kann hier die Schrittweite nach der Faustregel von Collatz

$$\left\| \frac{k_3 - k_2}{k_2 - k_1} \right\| \approx 0.05$$

kontrolliert werden.

Der reelle Stabilitätsbereich des klassischen Runge-Kutta-Verfahrens liegt etwa bei

$$- 2.785 < H < 0.$$

b) Bei der Wahl $\alpha_2 = \frac{1}{3}$, $\alpha_3 = \frac{2}{3}$ erhält man das Verfahren von Kutta oder 3/8-Verfahren

$\frac{1}{3}$	$\frac{1}{3}$			
$\frac{2}{3}$	$-\frac{1}{3}$	1		
1	1	-1	1	
	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

(8.15)

b) Für $\alpha_2 = \alpha_3 = \frac{1}{2}$ ergibt sich das Verfahren von Gill

$$\begin{array}{c|ccc}
 \frac{1}{2} & \frac{1}{2} & & \\
 \frac{1}{2} & (\sqrt{2}-1)/2 & (2-\sqrt{2})/2 & \\
 1 & 0 & -\sqrt{2}/2 & 1+\sqrt{2}/2 \\
 \hline
 & \frac{1}{6} & (2-\sqrt{2})/6 & (2+\sqrt{2})/6 \quad \frac{1}{6}
 \end{array} \quad (8.16)$$

Gill wählte die Parameter so, daß die Anzahl der pro Rechenschritt zu speichernden Werte und damit der Rundungsfehler minimiert wird.

Die beiden letztgenannten (4,4)-RK-Verfahren haben etwa denselben reellen Stabilitätsbereich wie die klassische Runge-Kutta-Formel.

Beispiel:

Es soll das klassische (4,4)-RK-Verfahren auf ein Differentialgleichungssystem mit drei Gleichungen angewandt werden. Das Anfangswertproblem lautet ausgeschrieben:

$$\begin{aligned}
 (y')^1 &= f^1(x, y^1, y^2, y^3) ; & y^1(a) &= \eta_0^1 \\
 (y')^2 &= f^2(x, y^1, y^2, y^3) ; & y^2(a) &= \eta_0^2 \\
 (y')^3 &= f^3(x, y^1, y^2, y^3) ; & y^3(a) &= \eta_0^3
 \end{aligned}$$

Dann sind die k_i , $1 \leq i \leq 4$, ebenfalls drei-dimensionale Vektoren, und die 4 Stufen des j-ten Rechenschritts (am j-ten Gitterpunkt) bestehen aus der Berechnung von:

$$\begin{aligned}
 k_1 &= f(x, y^1, y^2, y^3) \text{ für } x=x_j, y^1=y_j^1, y^2=y_j^2, y^3=y_j^3 ; \\
 k_2 &= f(x, y^1, y^2, y^3) \text{ für } x=x_j+\frac{h}{2}, y^1=y_j^1+\frac{h}{2}k_1^1, y^2=y_j^2+\frac{h}{2}k_1^2, y^3=y_j^3+\frac{h}{2}k_1^3 ; \\
 k_3 &= f(x, y^1, y^2, y^3) \text{ für } x=x_j+\frac{h}{2}, y^1=y_j^1+\frac{h}{2}k_2^1, y^2=y_j^2+\frac{h}{2}k_2^2, y^3=y_j^3+\frac{h}{2}k_2^3 ; \\
 k_4 &= f(x, y^1, y^2, y^3) \text{ für } x=x_j+h, y^1=y_j^1+hk_3^1, y^2=y_j^2+hk_3^2, y^3=y_j^3+hk_3^3
 \end{aligned}$$

Der neue Vektor y_{j+1}^* am Gitterpunkt x_{j+1} ergibt sich dann als

$$y_{j+1}^* = y_j^* + \frac{h}{6} (k_1(x_j, y_j^*) + 2k_2(x_j, y_j^*) + 2k_3(x_j, y_j^*) + k_4(x_j, y_j^*)) + \frac{h}{6} s(x_j, y_j^*)$$

Dieser Rechenablauf ist in dem folgenden Schema nochmals zusammengefaßt:

x	y^1	y^2	y^3	$f^1(x,y)$	$f^2(x,y)$	$f^3(x,y)$
x_j	y_j^1	y_j^2	y_j^3	k_1^1	k_1^2	k_1^3
$x_j + \frac{h}{2}$	$y_j^1 + \frac{h}{2} k_1^1$	$y_j^2 + \frac{h}{2} k_1^2$	$y_j^3 + \frac{h}{2} k_1^3$	$2k_2^1$	$2k_2^2$	$2k_2^3$
$x_j + \frac{h}{2}$	$y_j^1 + \frac{h}{2} k_2^1$	$y_j^2 + \frac{h}{2} k_2^2$	$y_j^3 + \frac{h}{2} k_2^3$	$2k_3^1$	$2k_3^2$	$2k_3^3$
$x_j + h$	$y_j^1 + h k_3^1$	$y_j^2 + h k_3^2$	$y_j^3 + h k_3^3$	k_4^1	k_4^2	k_4^3
				+	+	+
				s^1	s^2	s^3
x_{j+1}	$y_{j+1}^1 = y_j^1 + h s^1$	$y_{j+1}^2 = y_j^2 + h s^2$	$y_{j+1}^3 = y_j^3 + h s^3$			

8.2.3. Runge-Kutta-Formeln höherer Ordnung ($q \geq 5$)

In den bisher betrachteten Fällen war es stets möglich, Formeln der Ordnung q zu konstruieren, die mit $r=q$ Stufen auskamen, d.h. mit q Berechnungen von $f(x,y)$ pro Rechenschritt. Dies ist für $q \geq 5$ nicht mehr möglich! Butcher¹⁾ zeigt, daß kein RK-Verfahren der Ordnung $q=5$ mit $r=5$ Stufen existiert.

Spezielle wichtige RK-Verfahren mit $q \geq 5$, $r \geq 6$ sind:

I. (5,6)-Formeln:

Nyström-Formel:

$$\begin{array}{c|cccccc}
 \frac{1}{3} & \frac{1}{3} & & & & & \\
 \frac{2}{5} & \frac{4}{25} & \frac{6}{25} & & & & \\
 1 & \frac{1}{4} & -\frac{12}{4} & \frac{15}{4} & & & \\
 \frac{2}{3} & \frac{6}{81} & \frac{90}{81} & -\frac{50}{81} & \frac{8}{81} & & \\
 \frac{4}{5} & \frac{6}{75} & \frac{36}{75} & \frac{10}{75} & \frac{8}{75} & 0 & \\
 \hline
 & \frac{23}{192} & 0 & \frac{125}{192} & 0 & -\frac{81}{192} & \frac{125}{192}
 \end{array} \quad (8.17)$$

¹⁾ Butcher, J.C.: On RK-Processes of High Order, J. Austral. Math. Soc. 4 (1964)

Luther-Formel:

1	1					
1	$\frac{1}{2}$	$\frac{1}{2}$				
$\frac{1}{4}$	$\frac{14}{64}$	$\frac{5}{64}$	$-\frac{3}{64}$			
$\frac{1}{2}$	$-\frac{12}{96}$	$-\frac{12}{96}$	$\frac{8}{96}$	$\frac{64}{96}$		
$\frac{3}{4}$	0	$-\frac{9}{64}$	$\frac{5}{64}$	$\frac{16}{64}$	$\frac{36}{64}$	
	$\frac{7}{90}$	0	$\frac{7}{90}$	$\frac{32}{90}$	$\frac{12}{90}$	$\frac{32}{90}$

(8.18)

Dieses Verfahren entspricht für $f(x,y) \equiv f(x)$ einer Newton-Cotes-Quadratur von f mit fünf Stützstellen.

Butcher-Formel:

$\frac{1}{4}$	$\frac{1}{4}$					
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$				
$\frac{1}{2}$	0	$-\frac{1}{2}$	1			
$\frac{3}{4}$	$\frac{3}{16}$	0	0	$\frac{9}{16}$		
1	$-\frac{3}{7}$	$\frac{2}{7}$	$\frac{12}{7}$	$-\frac{12}{7}$	$\frac{8}{7}$	
	$\frac{7}{90}$	0	$\frac{32}{90}$	$\frac{12}{90}$	$\frac{32}{90}$	
					$\frac{7}{90}$	

(8.19)

Auch dieses Verfahren entspricht im Falle $f(x,y) \equiv f(x)$ einer Newton-Cotes-Quadratur mit fünf Stützstellen. Es hat - nach dem Verfahren (8.23) von Lawson - den zweitgrößten reellen Stabilitätsbereich der hier vorgestellten Verfahren, nämlich

$$- 3.2 < H < 0$$

Luther-Formel:

$\frac{4}{11}$	$\frac{4}{11}$					
$\frac{2}{5}$	$\frac{9}{50}$	$\frac{11}{50}$				
1	0	$-\frac{11}{4}$	$\frac{15}{4}$			
$\frac{6-\sqrt{6}}{10}$	$\frac{81+9\sqrt{6}}{600}$	0	$\frac{255-55\sqrt{6}}{600}$	$\frac{24-14\sqrt{6}}{600}$		
$\frac{6+\sqrt{6}}{10}$	$\frac{81-9\sqrt{6}}{600}$	0	$\frac{255+55\sqrt{6}}{600}$	$\frac{24+14\sqrt{6}}{600}$	0	
	$\frac{4}{36}$	0	0	0	$\frac{16+\sqrt{6}}{36}$	$\frac{16-\sqrt{6}}{36}$

(8.20)

Dieses Verfahren entspricht, im Unterschied zur Luther-Formel (8.18), für $f(x,y) \equiv f(x)$ einer Radauschen Quadraturformel. Es ist numerisch günstig, da es aufgrund der vielen zu Null verschwindenden Parameter β_{ij} weniger Speicherplatz als andere Verfahren braucht.

Sarafyan-Formel:

$\frac{1}{2}$	$\frac{1}{2}$					
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$				
1	0	-1	2			
$\frac{2}{3}$	$\frac{7}{27}$	$\frac{10}{27}$	0	$\frac{1}{27}$		
$\frac{2}{10}$	$\frac{28}{625}$	$-\frac{125}{625}$	$\frac{546}{625}$	$\frac{54}{625}$	$-\frac{378}{625}$	
	$\frac{14}{336}$	0	0	$\frac{35}{336}$	$\frac{162}{336}$	$\frac{125}{336}$

(8.21)

Fehlberg-Formel:

$\frac{1}{6}$	$\frac{1}{6}$					
$\frac{4}{15}$	$\frac{4}{75}$	$\frac{16}{75}$				
$\frac{2}{3}$	$\frac{5}{6}$	$-\frac{8}{3}$	$\frac{5}{2}$			
$\frac{4}{5}$	$-\frac{8}{5}$	$\frac{144}{25}$	-4	$\frac{16}{25}$		
1	$\frac{361}{320}$	$-\frac{18}{5}$	$\frac{407}{128}$	$-\frac{11}{80}$	$\frac{55}{128}$	
	$\frac{31}{384}$	0	$\frac{1125}{2816}$	$\frac{9}{32}$	$\frac{125}{768}$	$\frac{5}{66}$

(8.22)

Lawson-Formel:

$\frac{1}{2}$	$\frac{1}{2}$					
$\frac{1}{4}$	$\frac{3}{16}$	$\frac{1}{16}$				
$\frac{1}{2}$	0	0	$\frac{1}{2}$			
$\frac{3}{4}$	0	$-\frac{3}{16}$	$\frac{6}{16}$	$\frac{9}{16}$		
1	$\frac{1}{7}$	$\frac{4}{7}$	$\frac{6}{7}$	$-\frac{12}{7}$	$\frac{8}{7}$	
	$\frac{7}{90}$	0	$\frac{32}{90}$	$\frac{12}{90}$	$\frac{32}{90}$	$\frac{7}{90}$

(8.23)

Dieses explizite (5,6)-RK-Verfahren von Lawson ergibt für $f(x,y) \equiv f(x)$ wieder eine Newton-Cotes-Quadraturformel mit fünf Stützstellen. Von allen betrachteten Verfahren hat es den größten reellen Stabilitätsbereich, nämlich

$$-5.7 < H < 0.$$

II. (6,7)-Formeln:

Diese und höherstufige Runge-Kutta-Verfahren sind wegen der hohen Rechenzeiten, die sie benötigen, für die Praxis weniger bedeutend. Daher soll nur ein von Butcher stammendes Verfahren der Ordnung $q=6$ mit $r=7$ Stufen angegeben werden:

$$\begin{array}{c|cccccc}
 \frac{1}{3} & \frac{1}{3} & & & & & \\
 \frac{2}{3} & 0 & \frac{2}{3} & & & & \\
 \frac{1}{3} & \frac{1}{12} & \frac{1}{3} & -\frac{1}{12} & & & \\
 \frac{1}{2} & -\frac{1}{16} & \frac{9}{8} & -\frac{3}{16} & -\frac{3}{8} & & \\
 \frac{1}{2} & 0 & \frac{9}{8} & -\frac{3}{8} & -\frac{3}{4} & \frac{1}{2} & \\
 1 & \frac{9}{44} & -\frac{9}{11} & \frac{63}{44} & \frac{18}{11} & 0 & -\frac{16}{11} \\
 \hline
 & \frac{11}{120} & 0 & \frac{27}{40} & \frac{27}{40} & -\frac{4}{15} & -\frac{4}{15} & \frac{11}{120}
 \end{array} \quad (8.24)$$

8.3. Implizite Runge-Kutta-Verfahren

Die allgemeine Form der r -stufigen RK-Verfahren ergibt sich aus dem Ansatz

$$y_{j+1}^* = y_j^* + h \sum_{i=1}^r \gamma_i k_i(x_j, y_j^*), \quad j=0(1)m-1 \quad (8.25)$$

$$k_i(x, y) = f(x + \alpha_i h, y + h \sum_{j=1}^r \beta_{ij} k_j), \quad i=1(1)r. \quad (8.26)$$

Bei dem im letzten Abschnitt betrachteten expliziten RK-Verfahren galt: $\beta_{ij} = 0, i \leq j$. Im folgenden werden implizite RK-Verfahren behandelt, für die es $\beta_{ij} \neq 0$ mit $i \leq j$ gibt, und semi-implizite RK-Verfahren mit $\beta_{ij}=0$ für $i < j$, aber $\beta_{ii} \neq 0$ für wenigstens ein i .

Bei impliziten RK-Verfahren erhöht sich also die Anzahl der frei zu wählenden Parameter, und es besteht die Hoffnung, durch geschickte Wahl bei r -stufigen impliziten Formeln eine wesentlich höhere Ordnung als bei expliziten Verfahren zu erzielen. In der Tat hat Butcher¹⁾ gezeigt, daß es zu jedem r ein r -stufiges implizites RK-Verfahren der Ordnung $q=2r$ gibt. Explizite, implizite und semi-implizite Verfahren lassen sich durch folgendes Schema beschreiben:

RK-Verfahren	Anzahl der frei wählb. Parameter	Erläuterungen
explizit	$\frac{r(r+1)}{2}$	k_i berechnet sich nur aus $k_1, \dots, k_{i-1}$, Summation in (8.26) nur bis $i-1$; $\beta_{ij}=0$ für $i \leq j$
semi-implizit	$\frac{r(r+3)}{2}$	k_i berechnet sich aus $k_1, \dots, k_i$, Summation in (8.26) bis i ; $\beta_{ij}=0$ für $i < j$
implizit	$r(r+1)$	k_i berechnet sich aus $k_1, \dots, k_r$, Summation in (8.26) bis r ; keine Beschränkungen der β_{ij} .

Im Spezialfall $f(x,y) \equiv f(x)$ entspricht das implizite RK-Verfahren einer Quadraturformel der Form

$$y_{j+1}^* = y_j^* + \int_{x_j}^{x_{j+1}} f(x) dx =: y_j^* + I_j, \quad (8.27)$$

mit

$$I_j = h \sum_{i=1}^r \gamma_i f(x_j + \alpha_i h).$$

¹⁾ Butcher, J.C.: Implicit RK-Processes, Math. Comp. 18 (1964)

Offensichtlich gehört zu jedem Runge-Kutta-Verfahren eine Quadraturformel, charakterisiert durch die γ_i und α_i . In den folgenden Abschnitten wird in Stichworten gezeigt werden, wie man diese Tatsache zur Konstruktion allgemeiner impliziter RK-Verfahren benutzen kann.

Eine kompliziertere Konstruktionsmöglichkeit liegt, wie bei den expliziten RK-Verfahren, in der Methode der Taylorentwicklung bis zur erstrebten Ordnung q und anschließenden Koeffizientenvergleich.

8.3.1. Gauss-Formeln

r-stufige implizite RK-Formeln lassen sich nach Butcher¹⁾ wie folgt aus den Gauss-Quadratur-Formeln gewinnen:

- Die α_i , $i=1(1)r$, werden als Wurzeln von $P_r(2\alpha-1)$ gewählt, wobei P_r das Legendre-Polynom vom Grad r ist.
- Die β_{ij} , $i,j=1(1)r$, sind Lösung des Gleichungssystems

$$\sum_{j=1}^r \beta_{ij} \alpha_j^{k-1} = \frac{1}{k} \alpha_i^k ; \quad i=1(1)r; k=1(1)r$$

- Die γ_j , $j=1(1)r$, werden aus dem Gleichungssystem

$$\sum_{j=1}^r \gamma_j \alpha_j^{k-1} = \frac{1}{k} ; \quad k=1(1)r$$

bestimmt.

Für die so berechneten Parameterwerte gilt: $\alpha_i \in (0,1)$ und $\gamma_j > 0$

¹⁾Butcher, J.C.: Implicit R.K. Processes, Math. Comp. 18 (1964)

Die so gewonnenen Formeln heißen r-stufige implizite Gauss-Formeln; sie haben die Ordnung $q=2r$ und sind alle A-stabil; ihre charakteristische Wurzel $\mu_1(H)$ ist eine (r,r) -Padé-Approximation¹⁾ von e^H .

Spezialfälle:

a)
$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array} \quad \text{implizite (2,1)-Gauss-Formel} \quad (8.28)$$

b)
$$\begin{array}{cc|cc} (3-\sqrt{3})/6 & & \frac{1}{4} & (3-2\sqrt{3})/12 \\ (3+\sqrt{3})/6 & (3+2\sqrt{3})/12 & & \frac{1}{4} \\ \hline & & \frac{1}{2} & \frac{1}{2} \end{array} \quad \text{implizite (4,2)-Gauss-Formel} \quad (8.29)$$

c)
$$\begin{array}{ccc|ccc} (5-\sqrt{15})/10 & & \frac{5}{36} & & (10-3\sqrt{15})/45 & (25-6\sqrt{15})/180 \\ \frac{1}{2} & & (10+3\sqrt{15})/72 & & \frac{2}{9} & (10-3\sqrt{15})/72 \\ (5+\sqrt{15})/10 & (25+6\sqrt{15})/180 & & (10+3\sqrt{15})/45 & & \frac{5}{36} \\ \hline & & \frac{5}{18} & & \frac{4}{9} & \frac{5}{18} \end{array} \quad (8.30)$$

implizite (6,3)-Gauss-Formel

¹⁾d.h. $\mu_1(H)$ ist diejenige rationale Funktion $\frac{P_r(H)}{Q_r(H)}$, deren Taylorentwicklung in $H=0$ mit der Entwicklung von e^H am besten übereinstimmt.

8.3.2. Radau- und Lobatto-Formeln

Weitere implizite RK-Verfahren ergeben sich, wenn man anstelle von Gauss-Formeln von Quadraturformeln nach Radau und Lobatto¹⁾ ausgeht. Als Spezialfälle seien die folgenden Verfahren genannt:

a) Implizite (3,2)-Radau-Formeln:

$$\begin{array}{c|cc}
 0 & 0 & 0 \\
 \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\
 \hline
 & \frac{1}{4} & \frac{3}{4}
 \end{array}
 \qquad
 \begin{array}{c|cc}
 \frac{1}{3} & \frac{1}{3} & 0 \\
 1 & 1 & 0 \\
 \hline
 & \frac{3}{4} & \frac{1}{4}
 \end{array}
 \qquad (8.31a/b)$$

b) Implizite (5,3)-Radau-Formeln

$$\begin{array}{c|ccc}
 0 & 0 & 0 & 0 \\
 (6-\sqrt{6})/10 & (9+\sqrt{6})/75 & (24+\sqrt{6})/120 & (168-73\sqrt{6})/600 \\
 (6+\sqrt{6})/10 & (9-\sqrt{6})/75 & (168+73\sqrt{6})/600 & (24-\sqrt{6})/120 \\
 \hline
 & \frac{1}{9} & (16+\sqrt{6})/36 & (16-\sqrt{6})/36
 \end{array}
 \qquad (8.32a)$$

$$\begin{array}{c|ccc}
 (4-\sqrt{6})/10 & (24-\sqrt{6})/120 & (24-11\sqrt{6})/120 & 0 \\
 (4+\sqrt{6})/10 & (24+11\sqrt{6})/120 & (24+\sqrt{6})/120 & 0 \\
 1 & (6-\sqrt{6})/12 & (6+\sqrt{6})/12 & 0 \\
 \hline
 & (16-\sqrt{6})/36 & (16+\sqrt{6})/36 & \frac{1}{9}
 \end{array}
 \qquad (8.32b)$$

c) Implizite (2,2)-Lobatto-Formeln

$$\begin{array}{c|cc}
 0 & 0 & 0 \\
 1 & 1 & 0 \\
 \hline
 & \frac{1}{2} & \frac{1}{2}
 \end{array}
 \qquad (8.33)$$

¹⁾ s. Fußnote 2, Seite 73

d) Die implizite (4,3)-Lobatto-Formel von Milne:

0	0	0	0	(8.34)
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0	
1	0	1	0	
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$	

e) Die implizite (6,4)-Lobatto-Formel:

0	0	0	0	(8.35)	
$(5-\sqrt{5})/10$	$(5+\sqrt{5})/60$	$\frac{1}{6}$	$(15-7\sqrt{5})/60$		0
$(5+\sqrt{5})/10$	$(5-\sqrt{5})/60$	$(15+7\sqrt{5})/60$	$\frac{1}{6}$		0
1	$\frac{1}{6}$	$(5-\sqrt{5})/12$	$(5+\sqrt{5})/12$		0
	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$		$\frac{1}{12}$

f) Die implizite (8,5)-Lobatto-Formel:

0	0	0	0	0	0	(8.36)
$(7-\sqrt{21})/14$	$\frac{1}{14}$	$\frac{1}{9}$	$(13-3\sqrt{21})/63$	$(14-3\sqrt{21})/126$	0	
$\frac{1}{2}$	$\frac{1}{32}$	$(91+21\sqrt{21})/576$	$\frac{11}{72}$	$(91-21\sqrt{21})/576$	0	
$(7+\sqrt{21})/14$	$\frac{1}{14}$	$(14+3\sqrt{21})/126$	$(13+3\sqrt{21})/63$	$\frac{1}{9}$	0	
1	0	$\frac{7}{18}$	$\frac{2}{9}$	$\frac{7}{18}$	0	
	$\frac{1}{20}$	$\frac{49}{180}$	$\frac{16}{45}$	$\frac{49}{180}$	$\frac{1}{20}$	

Alle diese Formeln haben einen beschränkten Stabilitätsbereich; die Hauptwurzel $\mu_1(H)$ jedes dieser Verfahren ist eine Padé-Approximation $P_m(H)/Q_n(H)$ von e^H mit $m+n$; z.B. haben die Verfahren (8.31a/b) die Hauptwurzel:

$$\mu_1(H) = \frac{1 + \frac{2}{3}H + \frac{1}{6}H^2}{1 - \frac{1}{3}H}$$

für das Verfahren (8.34) gilt:

$$\mu_1(H) = \frac{1 + \frac{3}{4}H + \frac{1}{4}H^2 + \frac{1}{24}H^3}{1 - \frac{1}{4}H}$$

8.4. Semi-implizite Runge-Kutta-Verfahren

Die in diesem Abschnitt behandelten semi-impliziten RK-Verfahren wurden von Rosenbrock²⁾ eingeführt und von Haines³⁾ verfeinert. O.B.d.A. werde in diesem Paragraphen angenommen, daß die rechte Seite der zu behandelnden Dgl. nur implizit über y von x abhängt; es sei also

$$\begin{aligned} y'(x) &= f(y(x)), \quad x \in [a, b], \\ y(a) &= \eta_0 \end{aligned}$$

das zu lösende Anfangswertproblem¹⁾.

¹⁾ Es kann nämlich jedes Anfangswertproblem mit explizit von x abhängiger rechter Seite f durch Vermehrung der Gleichungen (d.h. Anheben der Dimension des Problems) auf die oben angegebene Form gebracht werden. Beispielsweise ist das skalare Anfangswertproblem

$$y' = f(x, y); \quad y(a) = \eta_0,$$

mit dem zwei-dimensionalen Problem

$$\begin{aligned} y_1' &= 1; \quad y_1(a) = a \\ y_2' &= f(y_1, y_2); \quad y_2(a) = \eta_0. \end{aligned}$$

äquivalent.

²⁾ Rosenbrock, H.H.: Some General Implicit Processes for the Numerical Solution of Differential Equations, Comput. J. 5 (1963)

³⁾ Haines, C.F.: Implicit Integration Processes with Error Estimation of the Numerical Solution of Differential Equations Comput. J. 12 (1969)

Semi-implizite RK-Verfahren sind dann spezielle implizite RK-Methoden, bei denen zur Berechnung von k_i , $i=1(1)r$, keine Iteration nötig ist, da aufgrund der einfachen Gestalt der Bestimmungsgleichungen leicht nach k_i aufgelöst werden kann. Dies soll für den Fall von $r=3$ Stufen kurz skizziert werden:

Die Grundgleichungen eines 3-stufigen semi-impliziten RK-Verfahrens sind hier:

$$y_{j+1}^* = y_j^* + h(\gamma_1 k_1(y_j^*) + \gamma_2 k_2(y_j^*) + \gamma_3 k_3(y_j^*)) \quad j=0(1)m-1 \quad (8.37)$$

wobei

$$\begin{aligned} k_1 &= k_1(y_j^*) = f(y_j^*) + \beta_{11} J(y_j^*) k_1 \\ k_2 &= k_2(y_j^*) = f(y_j^* + \beta_{21} h k_1) + \beta_{22} J(y_j^* + \gamma_{21} h k_1) k_2 \\ k_3 &= k_3(y_j^*) = f(y_j^* + \beta_{31} h k_1 + \beta_{32} h k_2) + \beta_{33} J(y_j^* + \gamma_{31} h k_1 + \gamma_{32} h k_2) k_3 \end{aligned} \quad (8.38)$$

Dabei ist $J(y) = f_y(y)$, die Jakobi-Matrix von f und die Parameter γ_j , β_{ij} , γ_{ij} werden wieder durch die Methode der Taylorentwicklung bis zu gewünschter Ordnung q bestimmt. Man erhält:

$$\begin{aligned} k_1 &= (I - \beta_{11} J(y_j^*))^{-1} f(y_j^*) \\ k_2 &= (I - \beta_{22} J(y_j^* + \gamma_{21} h k_1))^{-1} f(y_j^* + \beta_{21} h k_1) \\ k_3 &= (I - \beta_{33} J(y_j^* + \gamma_{31} h k_1 + \gamma_{32} h k_2))^{-1} f(y_j^* + \beta_{31} h k_1 + \beta_{32} h k_2) \end{aligned} \quad (8.39)$$

Als Spezialfälle dieser semi-impliziten RK-Formeln seien die folgenden Verfahren angegeben:

a) Die semi-implizite (3,2)-Rosenbrock-Formel:

$$\begin{aligned} \beta_{11} &= 1 + \sqrt{6}/6; & \gamma_1 &= -0.41315432, \\ \beta_{22} &= 1 - \sqrt{6}/6; & \gamma_2 &= 1.41315432, \\ \beta_{21} &= \gamma_{21} = (-6 - \sqrt{6} + \sqrt{58 + 20\sqrt{6}}) / (6 + 2\sqrt{6}) = 0.17378667... \end{aligned} \quad (8.40)$$

b) Die semi-implizite (2,2)-Rosenbrock-Formel:

$$\begin{aligned} \beta_{11} &= \beta_{22} = 1 - \sqrt{2}/2; & \beta_{21} &= (\sqrt{2} - 1)/2, \\ \gamma_{21} &= 0; & \gamma_1 &= 0; & \gamma_2 &= 1 \end{aligned} \quad (8.41)$$

c) Die semi-implizite (3,4) Haines-Formel:

$$\begin{aligned} \beta_{11}=1; \beta_{22}=1; \beta_{33}=1; \beta_{44}=\frac{2}{3}; \\ \beta_{21}=1; \beta_{31}=\frac{1}{2}; \beta_{41}=\frac{2}{99}; \\ \gamma_{21}=1; \gamma_{31}=\frac{1}{2}; \beta_{32}=\frac{1}{2}; \beta_{42}=\frac{95}{99}; \\ \gamma_{32}=\frac{1}{2}; \beta_{43}=\frac{2}{99}; \gamma_{41}=\gamma_{42}=\gamma_{43}=0; \\ \gamma_1=\frac{19}{9}; \gamma_2=-\frac{43}{18}; \gamma_3=\frac{28}{9}; \gamma_4=-\frac{11}{6}. \end{aligned} \quad (8.42)$$

d) Die semi-implizite (3,2) Calahan-Formel:

$$\begin{aligned} \beta_{11}=\beta_{22}=0.788675134; \gamma_{21}=0 \\ \beta_{21}=-1.15470054; \gamma_1=\frac{3}{4}; \gamma_2=\frac{1}{4}. \end{aligned} \quad (8.43)$$

Alle vier genannten Formeln sind A-stabil.

8.5. Bemerkungen zur praktischen Anwendung der RK-Verfahren

Die Leistungsfähigkeit eines Verfahrens hängt weitgehend vom zu lösenden Differentialgleichungsproblem ab; daher sind Aussagen darüber, welches Verfahren "das beste" ist, beim gegenwärtigen Stand der Theorie nicht möglich und auch in Zukunft wohl nicht zu erwarten. Bei der Auswahl eines Verfahrens kann man sich daher nur auf recht allgemeine Kriterien wie Ordnung, Stabilitätsbereich, Rechenaufwand etc. stützen, sowie auf Erfahrungen, die mit den einzelnen Verfahren gemacht wurden. In den letzten Jahren wurden ausführliche Untersuchungen zur Praxis von Integrationsmethoden angestellt. Zu erwähnen ist insbesondere das an der Universität von Toronto entwickelte Programm DETEST¹⁾, welches einen ausgewogenen Vergleich numerischer Verfahren zur Lösung von Anfangswertproblemen ermöglicht²⁾.

¹⁾ HALL, ENRIGHT, HULL, SEDGWICK: Detest: A Program for Comparing Numerical Methods for Ordinary Differential Equations, Technical Report No. 60, Univ. of Toronto, Dec. 73.

²⁾ ENRIGHT, BEDET, FARKAS, HULL: Test Results on Initial Value Methods for Non-stiff Ordinary Differential Equations, Technical Report No. 68, Univ. of Toronto, May 74.

Bei speziellen Differentialgleichungsproblemen ist es oft ratsam, statt der behandelten allgemeinen Verfahren, spezielle Methoden zu verwenden. Auch hierüber sind zahlreiche Veröffentlichungen erschienen, und die Theorie der numerischen Behandlung von Differentialgleichungen entwickelt sich zunehmend in diese Richtung.

Die folgenden allgemeinen Bemerkungen können bei der Auswahl eines RK-Verfahrens hilfreich sein:

- Alle expliziten RK-Verfahren haben einen endlichen Stabilitätsbereich H , der mit der Ordnung wächst.
- Die Stabilitätsbereiche impliziter RK-Verfahren sind i. allg. größer als die expliziter Methoden.
- Insbesondere sind die impliziten Gauss-Formeln (Abschnitt 8.3.1.) und die semi-impliziten Verfahren des Abschnitts 8.4. A-stabil; ihre Schrittweite h kann daher ausschließlich mit Blick auf die Genauigkeitsanforderungen gewählt werden, was für stiff-Systeme von Bedeutung ist. Ihre Anwendung auf andere Probleme erscheint wegen des hohen Rechenaufwands nicht sinnvoll.
- Verfahren höherer Ordnung sind deutlich genauer als solche niedrigerer Ordnung, solange der Ersetzungsfehler dominiert; bei kleiner Schrittweite h ist jedoch der Rundungsfehler größer als der Ersetzungsfehler, und die Genauigkeitsunterschiede von Verfahren verschiedener Ordnung verwischen sich.
- Das Rechnen mit doppelter Genauigkeit vermindert den Einfluß der Rundungsfehler und erfordert nur etwa 35% mehr Rechenzeit als einfache Genauigkeit.
- Explizite RK-Verfahren der Ordnung 5 sind besser als die Verfahren der Ordnung 3 und 4, was das Verhältnis Genauig-

keit zu Rechenaufwand betrifft. Insbesondere benötigt das Verfahren (8.19) von Butcher die geringste Rechenzeit und hat den zweitgrößten reellen Stabilitätsbereich der vorgestellten Verfahren mit der Ordnung 5. Das Verfahren (8.23) von Lawson hat den größten reellen Stabilitätsbereich bei vergleichbarem Rechenaufwand. Das Verfahren (8.24) von Butcher mit der Ordnung 6 ist vergleichsweise aufwendiger.

- Die impliziten Radau-Formeln (8.32a/b) ergeben i.allg. etwa dieselbe Genauigkeit wie die expliziten RK-Verfahren der Ordnung 5. Da sie nur 3 Stufen haben, benötigen sie pro Rechenschritt weniger Funktionsauswertungen. Dieser Vorteil wird jedoch durch die erforderliche Iteration zur Berechnung der k_i gemindert.
- Im Vergleich zu den - im folgenden Paragraphen diskutierten - Mehrschrittverfahren haben die RK-Formeln den Vorzug, keine Startprozedur zu erfordern (d.h. sie sind selbststartend) und den Nachteil, daß sie pro Rechenschritt mehr Funktionsauswertungen benötigen, was zu längeren Rechenzeiten führt.

§ 9 Lineare Mehrschrittverfahren

Wendet man Satz 3.2. (Seite 23) auf lineare k-Schritt-Verfahren (1.6) an und berücksichtigt man dabei die früheren Ergebnisse über Konsistenz (Satz 2.2.) und Stabilität (Satz 3.1.), so erhält man den folgenden Konvergenzsatz für lineare k-Schritt-Verfahren:

Satz 9.1.

Sei y Lösung der Differentialgleichung (1.1), L die Lipschitzkonstante in (1.2) und

$$\alpha_k y_{j+k}^* + \alpha_{k-1} y_{j+k-1}^* + \dots + \alpha_0 y_j^* - h(\beta_k f_{j+k}^* + \beta_{k-1} f_{j+k-1}^* + \dots + \beta_0 f_j^*) = 0 \quad (9.1)$$

$j = o(1)m-k$

mit $y_j^* = \eta_j$; $j = o(1)k-1$; $\alpha_k = 1$; $f_j^* = f(x_j, y_j^*)$; $h|\beta_k|L < 1$

ein k-Schritt-Verfahren zur Approximation von $y(x_j)$, $x_j \in I_h$. Bezeichnen $\rho(u)$ und $\sigma(u)$ die erzeugenden Polynome (2.12), so ist das Verfahren konvergent, wenn

1. $\rho(1) = 0$; $\rho'(1) = \sigma(1)$
2. keine Nullstelle von $\rho(u)$ betragsmäßig größer ist als 1,
3. alle Nullstellen von $\rho(u)$ mit dem Betrag 1 einfach sind.

Ist außerdem

4. $y \in C^{q+2}[a, b]$, $q \geq 1$
5. $c_0 = c_1 = \dots = c_q = 0$; $c_{q+1} \neq 0$ (c_j gemäß (2.10))
6. $\|y(x_j) - \eta_j\| = O(h^q)$, $j = o(1)k-1$

so hat das Verfahren die Konvergenzordnung q und es gilt die Fehlerabschätzung:

$$\|y(x_{j+k}) - y_{j+k}^*\| \leq \frac{u(x_j - a)}{1 - hK_2} \left\{ (1 + hK_1) \max_{0 \leq i \leq k-1} \|y(x_i) - \eta_i\| + \right. \\ \left. + (x_{j+1} - a)h^q \max_{x \in [a, b]} \|c_{q+1} y^{(q+1)}(x)\| \right\} \quad (9.2)$$

mit $K_1 = L \sum_{j=0}^{k-1} |\beta_j|$, $K_2 = L|\beta_k|$,

$j = o(1)m-k$

$D \geq \|A^j\|^{1^1}$ für alle $j \in \mathbb{N}$, $u = \frac{DL}{1 - hL|\beta_k|} \cdot \sum_{j=0}^k |\beta_j|$

In den nachfolgenden Abschnitten werden lineare k-Schritt-Verfahren hergeleitet.

¹⁾ aus $\rho(1) = 0$ folgt $\|A^j\| \geq 1$

9.1. Klassische explizite Mehrschrittverfahren

Der klassische Weg zur Gewinnung stabiler Mehrschrittverfahren beliebiger Ordnung besteht darin, $y'(x)=f(x,y(x))$ durch Integration auf ein Quadraturproblem zurückzuführen

$$y(x_{j+1})=y(x_{j-r+1})+\int_{x_{j-r+1}}^{x_{j+1}} f(x,y(x)) dx; \quad r \in \mathbb{N}; \quad r \leq k \quad (9.3)$$

und das Integral für $j=k-1(1)m-1$ mit den offenen oder geschlossenen Quadraturformeln von Newton-Cotes¹⁾ auszuwerten.

Für verschiedene $r \leq k$ ²⁾ erhält man so eine Vielzahl von k -Schritt-Verfahren, von denen insbesondere die expliziten Mehrschrittverfahren von Adams-Bashforth und Nyström, sowie die impliziten Verfahren von Adams-Moulton und Milne-Simpson noch viel benutzt werden. Alle diese Verfahren sind stabil, denn ihr charakteristisches Polynom $\rho(u)=u^k-u^{k-r}$ erfüllt das Wurzelkriterium (vgl. S. 29).

¹⁾ Zur Erinnerung: Im Falle offener Newton-Cotes-Formeln wird $f(x,y(x))$ ersetzt durch das Interpolationspolynom an den Stellen $x_i, i=j-k+1(1)j$; $P_{k-1}(x)=\sum_{\ell=0}^{k-1} (-1)^\ell \binom{-u}{\ell} \nabla^\ell f_j$; $u=\frac{x-x_j}{h}$ und bei geschlossenen Formeln durch das Interpolationspolynom an den Stellen $x_i, i=j-k+1(1)j+1$; $P_k(x)=\sum_{\ell=0}^k (-1)^\ell \binom{-u}{\ell} \nabla^\ell f_{j+1}$; $u=\frac{x-x_{j+1}}{h}$

²⁾ Für $r > k$ wäre r und nicht k die Schrittzahl des Verfahrens.

Werten wir das Integral in (9.3) mit offenen Newton-Cotes-Formeln aus, und bezeichnen wir die so gewonnenen (mit einem Quadraturfehler behafteten) Näherungswerte von $y(x_j)$ wieder mit y_j^* sowie $f(x_j, y^*(x_j))$ mit f_j^* , so erhalten wir die folgende Klasse expliziter Mehrschrittverfahren:

$$y_{j+1}^* = y_{j-r+1}^* + h \sum_{\ell=0}^{k-1} c_{\ell}^* (r-1) \nabla^{\ell} f_j^* \quad \text{mit} \quad c_{\ell}^*(r) := (-1)^{\ell} \int_{-1}^r \binom{u}{\ell} du \quad (9.4)$$

Aus Satz 9.1. folgt die Konvergenz dieser Verfahren für beliebige natürliche Zahlen k und $r \leq k$; in der Praxis wird $r=1$ oder $r=2$ gewählt.

9.1.1. Die Verfahren von Adams-Bashforth

Für $r=1$ ergibt (9.4) die k-Schrittverfahren von Adams-Bashforth:

$$y_{j+1}^* = y_j^* + h \left(f_j^* + \frac{1}{2} \nabla f_j^* + \frac{5}{12} \nabla^2 f_j^* + \frac{3}{8} \nabla^3 f_j^* + \frac{251}{720} \nabla^4 f_j^* + \dots + c_{k-1}^*(0) \nabla^{k-1} f_j^* \right); \quad (9.5)$$

sie haben die Ordnung k^1 .

Diese Verfahren können in der Form (9.5) programmiert werden, vielfach ist es jedoch zweckmäßig, sie auf die Form (9.1) umzurechnen.

Für die Koeffizienten α_i, β_i erhält man die folgenden Werte:

$\alpha_k = 1$; $\alpha_{k-1} = -1$; $\alpha_i = 0$ für $i = 0(1)k-2$; $\beta_k = 0$ und

k	β_{k-1}	β_{k-2}	β_{k-3}	β_{k-4}	β_{k-5}	β_{k-6}
1	1					
2	3/2	-1/				
3	23/12	-16/	5/			
4	55/24	-59/	37/	-9/		
5	1901/720	-2774/	2616/	-1274/	251/	
6	4277/1440	-7923/	9982/	-7298/	2877/	-475/

(Alle Zahlen einer Zeile haben denselben Nenner)

¹⁾ s. Fußnote Seite 24

9.1.2. Die Verfahren von Nyström

Für $r=2$, $k \geq 2$ ergibt (9.4) die k-Schrittverfahren von Nyström:

$$y_{j+1}^* = y_{j-1}^* + h \left(2f_j^* + \frac{1}{3} \nabla^2 f_j^* + \frac{1}{3} \nabla^3 f_j^* + \frac{29}{90} \nabla^4 f_j^* + \dots + c_{k-1}^* (1) \nabla^{k-1} f_j^* \right) \quad (9.6a)$$

Sie haben die Ordnung k. Für $k=2$ erhält man insbesondere die midpoint-rule (vgl. (1.7)) mit der Ordnung 2:

$$y_{j+1}^* = y_{j-1}^* + 2hf_j^* ; j=1(1) m-1 \quad (9.6b)$$

Umrechnung auf die Form (9.1) ergibt die folgenden Koeffizienten:

$\alpha_k=1$; $\alpha_{k-1}=0$; $\alpha_{k-2}=-1$; $\alpha_i=0$ für $i=0(1)k-3$; $\beta_k=0$ sowie

k	β_{k-1}	β_{k-2}	β_{k-3}	β_{k-4}	β_{k-5}
2	2	0			
3	7/3	-2/	1/		
4	8/3	-5/	4/	-1/	
5	269/90	-266/	294/	-146/	29/

9.2. Klassische implizite Mehrschrittverfahren

Approximieren wir das Integral (9.4) mit geschlossenen Newton-Cotes-Formeln, so ergibt sich eine Klasse impliziter Integrationsformeln ("Korrektoren") der Form:

$$y_{j+1}^* = y_{j-r+1}^* + h \sum_{\ell=0}^k a_{\ell}^*(r) \nabla^{\ell} f_{j+1}^* ; \quad a_{\ell}^*(r) = (-1)^{\ell} \int_0^r \binom{u}{\ell} du ; \quad (9.7)$$

Nach Satz 9.1. konvergieren diese Formeln für beliebige natürliche Zahlen k und ¹⁾ $r \leq k$;

¹⁾ siehe Fußnote 2 auf Seite 93

Bezeichnet g_j die Summe aller Terme, die nicht von y_{j+1}^* abhängen, und berücksichtigt man die Relation:

$$\sum_{\ell=0}^k a_{\ell}^* (r) = c_k^*(r-1),$$

so geht (9.7) über in

$$y_{j+1}^* = hc_k^*(r-1)f(x_{j+1}, y_{j+1}^*) + g_j \quad (9.8)$$

Diese Gleichung für y_{j+1}^* wird in jedem Rechenschritt j , $j=k-1(1)m-1$, durch Iteration gelöst:

$$y_{j+1}^{*(i+1)} = hc_k^*(r-1)f(x_{j+1}, y_{j+1}^{*(i)}) + g_j; \quad i=0(1)s$$

wobei $y_{j+1}^{*(0)}$ mit einem Prädiktor (z.B. (9.4)), berechnet wird. Die Iteration konvergiert nach dem Fixpunktsatz, wenn gilt: $|hc_k^*(r-1)L| < 1$, wobei L die Lipschitzkonstante von $f(x, y)$ ist. Diese Bedingung¹⁾ ergab sich bereits früher und ist stets erfüllbar, wenn h genügend klein gewählt wird.

9.2.1. Die Verfahren von Adams-Moulton und Milne-Simpson

Für $r=1$ ergeben sich aus (9.7) die k-Schrittverfahren von Adams-Moulton mit der Ordnung $q=k+1$:

$$y_{j+1}^* = y_j^* + h \left(f_{j+1}^* - \frac{1}{2} \nabla f_{j+1}^* - \frac{1}{12} \nabla^2 f_{j+1}^* - \frac{1}{24} \nabla^3 f_{j+1}^* - \frac{19}{720} \nabla^4 f_{j+1}^* + \dots + a_k^*(1) \nabla^k f_{j+1}^* \right) \quad (9.9)$$

In der Form (9.1) hat es die Koeffizienten: $\alpha_k=1; \alpha_{k-1}=-1; \alpha_i=0$ für $i=0(1)k-2$ und

k	β_k	β_{k-1}	β_{k-2}	β_{k-3}	β_{k-4}	β_{k-5}
1	1/2	1/				
2	5/12	8/	-1/			
3	9/24	19/	-5/	1/		
4	251/720	646/	-264/	106/	-19/	
5	475/1440	1427/	-798/	482/	-173/	27/

¹⁾ Man beachte, daß aus (9.8) folgt: $\beta_k = c_k^*(r-1)$

Für $r=2$, $k \geq 2$ ergibt (9.7) die Verfahren von Milne-Simpson:

$$y_{j+1}^* = y_{j-1}^* + h(2f_{j+1}^* - 2\sqrt{f_{j+1}^*} + \frac{1}{3}\sqrt[3]{f_{j+1}^*} - \frac{1}{90}\sqrt[4]{f_{j+1}^*} - \frac{1}{90}\sqrt[5]{f_{j+1}^*} + \dots + a_k^* (2)\sqrt[k]{f_{j+1}^*}) \quad (9.10)$$

die für $k \geq 2$ die Ordnung $(k+1)$ haben.

Für $k=2$ erhält man das Verfahren (1.9) mit der Ordnung 4 (vgl. Seite 6).

Anmerkung:

Die Darstellung aller dieser Verfahren mit Newtonschen Differenzen hat programmiertechnisch den Vorteil, daß man sehr einfach zu Verfahren höherer oder niedrigerer Ordnung überwechseln kann.

9.3. Allgemeine lineare Mehrschrittverfahren

Man gewinnt stabile lineare Mehrschrittverfahren beliebig hoher Konsistenzordnung q , indem man k genügend groß wählt und die Parameter α_j und β_j so bestimmt, daß gilt:

1. $c_i = 0$ ($i=0(1)q$) und $c_{q+1} \neq 0$ (c_i siehe (2.10), Seite 11)
2. die Nullstellen von $\rho(\mu) = \sum_{i=0}^k \alpha_i \mu^i$ erfüllen das Wurzelkriterium (Seite 29).

Nach Satz 3.3. muß dabei $k+1 \geq q$ oder $k+2 \geq q$ gewählt werden, bei expliziten Formeln sogar $k \geq q$. Diese Art der Berechnung von Mehrschrittverfahren ist mühsam; es soll daher noch eine andere Möglichkeit betrachtet werden.

Aus Satz 9.1. folgt: Ein konvergentes k -Schriftverfahren der Ordnung q ist exakt für Polynome P_n vom Grad $0 \leq n \leq q$ wenn die Startwerte exakt gewählt werden. Wir betrachten speziell die Polynome

$$P_0(x) \equiv 1; P_n(x) = \int_0^x \binom{t}{n-1} dt \quad n=1(1)q$$

Wegen $P_n'(\ell) = 0$ für $\ell=0(1)n-2$ und $P_n'(n-1)=1$

ergibt das Verfahren (9.1) mit $h=1$ und $j=0$, angewandt auf die Polynome $P_n(x)$ ($n=0(1)k+2$) die Gleichungen:

$$n=k+2 : \sum_{j=1}^k \alpha_j P_{k+2}(j) = 0 \quad (9.11)$$

$$n=k+1 : \sum_{j=1}^k \alpha_j P_{k+1}(j) = \beta_k$$

$$\begin{aligned} n=k : \sum_{j=1}^k \alpha_j P_k(j) &= \beta_k P'_k(k) + \beta_{k-1} \\ \vdots & \\ n=2 : \sum_{j=1}^k \alpha_j P_2(j) &= \beta_k P'_2(k) + \dots + \beta_2 P'_2(2) + \beta_1 \end{aligned} \quad (9.12)$$

$$n=1 : \sum_{j=1}^k \alpha_j P_1(j) = \beta_k P'_1(k) + \dots + \beta_1 P'_1(1) + \beta_0$$

$$n=0 : \sum_{j=0}^k \alpha_j = 0 \quad (9.13)$$

Man erhält beliebig viele stabile Verfahren der Ordnung¹⁾ $k+1$, wenn die α_j so gewählt werden, daß (9.13) und das Wurzelkriterium erfüllt sind, und anschließend die β_j sukzessiv aus (9.12) berechnet werden.

Stabile Verfahren der (optimalen) Ordnung $k+2$ ergeben sich, wenn außerdem noch Gleichung (9.11) erfüllt werden kann. Nach Satz 3.3. ist das jedoch nur für gerades k möglich und man kann zeigen, daß die so gewonnenen Verfahren auch nur schwach stabil sind.

Beispiel: $k=2$

Es ist: $P_1(x) = x$; $P_2(x) = \frac{1}{2} x^2$

$$P_3(x) = \frac{x^2}{12} (2x-3); P_4(x) = \frac{x^2}{24} (x-2)^2$$

Dann lautet (9.11): $\frac{1}{24} \alpha_1 = 0$

Mit $\alpha_2=1$ ergibt (9.13): $\alpha_0=-1$ und (9.12): $\beta_2=\frac{1}{3}$; $\beta_1=\frac{4}{3}$; $\beta_0=\frac{1}{3}$.

Wir erhalten damit

$$y_{j+2}^* - y_j^* = \frac{h}{3} (f_{j+2}^* + 4f_{j+1}^* + f_j^*)$$

¹⁾ s. Fußnote Seite 24

Das ist das 2-Schrittverfahren (1.9) von Milne-Simpson, von dem wir bereits wissen (siehe S. 6), daß es die Ordnung 4 hat. Zugleich ist gezeigt, daß dieses das einzige stabile 2-Schrittverfahren der Ordnung 4 ist.

Auf dieselbe Weise lassen sich auch Verfahren optimaler Ordnung mit nichtäquidistantem Punktegitter konstruieren, die u.a. zur Änderung der Schrittweite verwendet werden können.

9.4. Instabile lineare Mehrschrittverfahren der Ordnung $2k-1$

Lineare k -Schritt-Verfahren der Form (9.1) mit der Ordnung $q > k+2$ sind nach Satz 3.3. notwendig instabil. Nun wurde bereits im Abschnitt 4.2. bemerkt, daß Prädiktoren instabil sein dürfen, und in Abschnitt 4.3.2. wurde sogar gezeigt, daß eine zyklische Kombination instabiler Korrektoren ein stabiles Verfahren ergeben kann. Es ist daher sinnvoll, auch instabile k -Schritt-Formeln vom Typ (9.1) zu betrachten.

Naturgemäß interessieren insbesondere Verfahren maximaler Konsistenzordnung, für die aus den Betrachtungen des vorigen Abschnitts folgt:

Satz 9.2.

Es gibt genau ein¹⁾ explizites k -Schritt-Verfahren (9.1) der Konsistenzordnung $(2k-1)$ und genau ein implizites k -Schritt-Verfahren (9.1) der Konsistenzordnung $2k$, sowie eine einparametrische Schar impliziter k -Schritt-Verfahren der Konsistenzordnung $(2k-1)$.

¹⁾ wenn o.B.d.A. $\alpha_k = 1$ gewählt wird

Diese Verfahren lassen sich mit der Methode des vorigen Abschnitts verhältnismäßig einfach berechnen; für $k=2(1)6$ ergeben sich so die folgenden¹⁾ Parametrisierungen von Verfahren der Ordnung $(2k-1)$:

$$\begin{aligned} \underline{k=2:} \quad \alpha_0 &= -5\alpha_2 + 12\beta_2 & ; & \quad \beta_0 = 2\alpha_2 - 5\beta_2 \\ \alpha_1 &= 4\alpha_2 - 12\beta_2 & ; & \quad \beta_1 = 4\alpha_2 - 8\beta_2 \end{aligned} \quad (a)$$

$$\begin{aligned} \underline{k=3:} \quad \alpha_0 &= -10\alpha_3 + 33\beta_3 & ; & \quad \beta_0 = 3\alpha_3 - 10\beta_3 \\ \alpha_1 &= -9\alpha_3 + 24\beta_3 & ; & \quad \beta_1 = 18\alpha_3 - 57\beta_3 \\ \alpha_2 &= 18\alpha_3 - 57\beta_3 & ; & \quad \beta_2 = 9\alpha_3 - 24\beta_3 \end{aligned} \quad (b)$$

$$\begin{aligned} \underline{k=4:} \quad \alpha_0 &= (-141\alpha_4 + 550\beta_4)/9 & \quad \beta_0 &= (12\alpha_4 - 47\beta_4)/3 \\ \alpha_1 &= -64\alpha_4 + 240\beta_4 & \quad \beta_1 &= 48\alpha_4 - 184\beta_4 \\ \alpha_2 &= 36\alpha_4 - 150\beta_4 & \quad \beta_2 &= 72\alpha_4 - 264\beta_4 \\ \alpha_3 &= (384\alpha_4 - 1360\beta_4)/9 & \quad \beta_3 &= (48\alpha_4 - 152\beta_4)/3 \end{aligned} \quad (c)$$

(9.14)

$$\begin{aligned} \underline{k=5:} \quad \alpha_0 &= (-786\alpha_5 + 3425\beta_5)/36 & \quad \beta_0 &= (30\alpha_5 - 131\beta_5)/6 \\ \alpha_1 &= (-1725\alpha_5 + 7390\beta_5)/9 & \quad \beta_1 &= (300\alpha_5 - 1295\beta_5)/3 \\ \alpha_2 &= -100\alpha_5 + 390\beta_5 & \quad \beta_2 &= 300\alpha_5 - 1270\beta_5 \\ \alpha_3 &= (2100\alpha_5 - 8990\beta_5)/9 & \quad \beta_3 &= (1800\alpha_5 - 7320\beta_5)/9 \\ \alpha_4 &= (2850\alpha_5 - 11065\beta_5)/36 & \quad \beta_4 &= (150\alpha_5 - 535\beta_5)/6 \end{aligned} \quad (d)$$

$$\begin{aligned} \underline{k=6:} \quad \alpha_0 &= (-1420\alpha_6 + 6713\beta_6)/50 & \quad \beta_0 &= (30\alpha_6 - 142\beta_6)/5 \\ \alpha_1 &= -426\alpha_6 + 1995\beta_6 & \quad \beta_1 &= 180\alpha_6 - 846\beta_6 \\ \alpha_2 &= -825\alpha_6 + 3780\beta_6 & \quad \beta_2 &= 900\alpha_6 - 4185\beta_6 \\ \alpha_3 &= 400\alpha_6 - 1960\beta_6 & \quad \beta_3 &= 1200\alpha_6 - 5480\beta_6 \\ \alpha_4 &= (1500\alpha_6 - 6825\beta_6)/2 & \quad \beta_4 &= 450\alpha_6 - 1980\beta_6 \\ \alpha_5 &= (3210\alpha_6 - 13419\beta_6)/25 & \quad \beta_5 &= (180\alpha_6 - 702\beta_6)/5 \end{aligned} \quad (e)$$

i. allg. wird $\alpha_k=1$ gesetzt

1) Donelson, Hansen: Cyclic Composite Multistep Predictor-Corrector-
Methods, SIAM J. Numer. Anal. 8, 137-157 (1971)

Entsprechend findet man Parametrisierungen der Ordnung $2k-2$ (die natürlich zwei freie Parameter - etwa β_k und β_{k-1} - enthalten);

Beispiel: 5-Schritt-Verfahren haben die Ordnung 8,
wenn gilt:

$$\begin{aligned}\alpha_0 &= \frac{1}{18} (1482\alpha_5 - 75\beta_4 - 4975\beta_5) \\ \alpha_1 &= \frac{1}{3} (1425\alpha_5 - 80\beta_4 - 4670\beta_5) \\ \alpha_2 &= -100\alpha_5 + 390\beta_5 \\ \alpha_3 &= \frac{1}{9} (-3900\alpha_5 + 240\beta_4 + 12410\beta_5) \\ \alpha_4 &= \frac{1}{6} (-150\alpha_5 + 25\beta_4 + 385\beta_5) \\ \beta_0 &= \frac{1}{3} (-60\alpha_5 + 3\beta_4 + 202\beta_5) \\ \beta_1 &= -300\alpha_5 + 16\beta_4 + 995\beta_5 \\ \beta_2 &= -600\alpha_5 + 36\beta_4 + 1940\beta_5 \\ \beta_3 &= \frac{1}{3} (-600\alpha_5 + 48\beta_4 + 1840\beta_5)\end{aligned}\tag{9.15}$$

9.5. Prädiktor-Korrektor-Verfahren

9.5.1. Konsistenzordnung von P-C-Verfahren

In Abschnitt 4.2. wurde bewiesen (Satz 4.2.), daß die Stabilität eines P-C-Verfahrens nur vom Korrektor abhängt. Das könnte uns veranlassen, instabile Formeln der maximalen Konsistenzordnung $(2k-1)$ als Prädiktoren zu verwenden. Die folgenden Betrachtungen zeigen jedoch, daß die Konsistenzordnung eines $P(EC)^SE$ -Verfahrens hierdurch i.allg. nicht erhöht werden kann, weil sie durch die Ordnung des Korrektors begrenzt ist. Es gilt nämlich

Satz 9.3.

Sei q die Konsistenzordnung des Korrektors und $\hat{q}$ die des Prädiktors, dann hat das $P(EC)^SE$ -Verfahren die Konsistenzordnung $\min(q, \hat{q}+s)$.

Beweis:

Bezeichnen $T_h^{(1)}[z](x_j)$ bzw. $T_h^{(2)}[z](x_j)$ die lokalen Ersetzungsfehler

$$\text{des Prädiktors: } z_{j+1}^{*(o)} = A_1 z_j^* + hB_1 Fz_j^* \quad (9.16a)$$

$$\text{bzw. des Korrektors: } z_{j+1}^* = A z_j^* + hB Fz_j^* + hCFz_{j+1}^* \quad (9.16b)$$

so gilt:

$$z(x_{j+1}) = A_1 z(x_j) + hB_1 Fz(x_j) + hT_h^{(1)}[z](x_j) \quad (9.17a)$$

$$z(x_{j+1}) = Az(x_j) + hBFz(x_j) + hCFz(x_{j+1}) + hT_h^{(2)}[z](x_j) \quad (9.17b)$$

$$\text{mit } \max_{x \in I'_h} \|T_h^{(1)}[z](x)\| \leq D_1 h^{\hat{q}}; \quad \max_{x \in I'_h} \|T_h^{(2)}[z](x)\| \leq D_2 h^q \quad (9.18)$$

und für den Ersetzungsfehler $T_h[z](x)$ des PECE-Verfahrens erhalten wir damit:

$$z(x_{j+1}) = Az(x_j) + hBFz(x_j) + hCF\hat{z}(x_{j+1}) + hT_h[z](x_j) \quad (9.19)$$

$$\text{mit } \hat{z}(x_{j+1}) := A_1 z(x_j) + hB_1 Fz(x_j) = z(x_{j+1}) - hT_h^{(1)}[z](x_j) \quad (9.20)$$

(9.17b) und (9.19) voneinander abgezogen ergibt:

$$T_h[z](x_j) = T_h^{(2)}[z](x_j) + C(Fz(x_{j+1}) - F\hat{z}(x_{j+1})) \quad (9.21)$$

Mit $\|C(Fz - F\hat{z})\| \leq |\beta_k| L \|z - \hat{z}\|$ erhält man aus (9.20)/(9.21):

$$\|T_h[z](x_j)\| \leq \|T_h^{(2)}[z](x_j)\| + |\beta_k| L h \|T_h^{(1)}[z](x_j)\|$$

und somit, unter Berücksichtigung von (9.18):

$$\max_{x \in I'_h} \|T_h[z](x)\| \leq D_2 h^q + D_1^* h^{\hat{q}+1}; \quad D_1^* = |\beta_k| L D_1$$

Entsprechend beweist man für $s > 1$:

$$\max_{x \in I'_h} \|T_h[z](x)\| \leq D_2 h^q + \hat{D}_1 h^{\hat{q}+s}. \quad \triangleleft$$

Die Konsistenz eines P-C-Verfahrens ist nach Satz 9.3. von der Anzahl s der Korrektorschritte unabhängig, wenn die Ordnung des Prädiktors höchstens um 1 geringer ist als die des Korrektors; in diesem Fall kann s klein gewählt werden, in der Praxis wählt man $s=1$ oder $s=2$. Aus Stabilitätsgründen (siehe 9.5.3.) wird in Ausnahmefällen auch $s>2$ gewählt. Der im Vergleich zu expliziten Verfahren höhere Rechenaufwand bei P-C-Verfahren wird i.allg. durch größere Genauigkeit kompensiert.

9.5.2. Vergleich von $P(EC)^{SE}$ - und $P(EC)^S$ -Verfahren

Die $P(EC)^S$ -Verfahren benötigen pro Rechenschritt $(s+1)$ Auswertungen von f , die $P(EC)^{SE}$ -Verfahren nur s Auswertungen, d.h. für $s \geq 1$ ungefähr halb soviel Rechenzeit wie ein PECE-Verfahren. Das bedeutet, daß der Rechenaufwand eines PEC-Verfahrens mit der Schrittweite $h/2$ etwa der gleiche ist, wie bei einem PECE-Verfahren mit der Schrittweite h . Da die Konvergenzordnung eines P-C-Verfahrens bei geeigneten Startwerten davon unabhängig ist, ob es als $P(EC)^{SE}$ oder $P(EC)^S$ -Verfahren angewendet wird, ist die Genauigkeit des PEC-Verfahrens mit $h/2$ i.allg. größer. Dieser Vorteil wird jedoch dadurch erkauft, daß die Stabilitätsbereiche von $P(EC)^S$ -Verfahren kleiner sind als von $P(EC)^{SE}$ -Verfahren. Eine solche Wechselwirkung zwischen Genauigkeit und Stabilität ist bei allen Mehrschrittverfahren zu beobachten.

Beispiel (Krogh¹⁾; Klopfenstein und Millman²⁾):

Mit dem Adams-Bashforth Prädiktor der Ordnung 4 und dem Adams-Moulton Korrektor der Ordnung 4 lautet der PECE-Algorithmus:

$$\hat{y}_{j+1}^* = y_j^* + \frac{h}{24} (55f_j^* - 59f_{j-1}^* + 37f_{j-2}^* - 9f_{j-3}^*) \quad (9.22)$$

$$y_{j+1}^* = y_j^* + \frac{h}{24} (9\hat{f}_{j+1}^* + 19f_j^* - 5f_{j-1}^* + f_{j-2}^*)$$

und der PEC-Algorithmus

$$\hat{y}_{j+1}^* = y_j^* + \frac{h}{24} (55\hat{f}_j^* - 59\hat{f}_{j-1}^* + 37\hat{f}_{j-2}^* - 9\hat{f}_{j-3}^*) \quad (9.23)$$

$$y_{j+1}^* = y_j^* + \frac{h}{24} (9\hat{f}_{j+1}^* + 19\hat{f}_j^* - 5\hat{f}_{j-1}^* + \hat{f}_{j-2}^*)$$

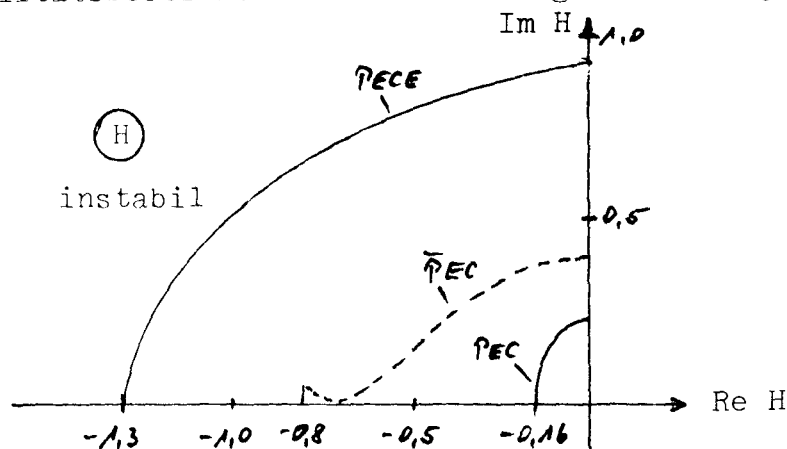
Für (9.22) ergibt sich ein charakteristisches Polynom 4. Ordnung:

$$Q(\mu; H) = \mu^4 - (1 + \frac{7}{6}H + \frac{55}{64}H^2)\mu^3 + (\frac{5}{24}H + \frac{59}{64}H^2)\mu^2 - (\frac{H}{24} + \frac{37}{64}H^2)\mu + \frac{9}{64}H^2$$

und für (9.23) ein Polynom 5. Ordnung:

$$Q(\mu; H) = (\mu^5 - (1 + \frac{8}{3}H)\mu^4 + \frac{95}{24}H\mu^3 - \frac{91}{24}H\mu^2 + \frac{15}{8}H\mu - \frac{3}{8}H)\mu^3$$

Die Stabilitätsbereiche haben die in Fig. 9.1. dargestellte Gestalt:



Figur 9.1.

¹⁾ Krogh: Predictor-Corrector-Methods of High Order with Improved Stability Characteristics, JACM 13, 374 (1966)

²⁾ Klopfenstein, Millmann: Num. Stability of a One-Evaluation P-C-Algorithm for Num. Sol. of O.D.E., Math. Comp. 22, 557 (1968)

Kombiniert man den Adams-Moulton-Korrektor des obigen Beispiels mit einem anderen Prädiktor, so lassen sich u.U. größere Stabilitätsbereiche erzielen. Mit dem folgenden Prädiktor, beispielsweise, hat das PEC-Verfahren die in Fig. 9.1. gebrochen gezeichnete Stabilitätsgrenze:

$$\begin{aligned} \hat{y}_{j+1}^* = & -0,29y_j^* - 15,39y_{j-1}^* + 12,13y_{j-2}^* + 4,55y_{j-3}^* \\ & + h(2,27\hat{f}_j^* + 6,65\hat{f}_{j-1}^* + 13,91\hat{f}_{j-2}^* + 0,69\hat{f}_{j-3}^*) \end{aligned}$$

Lambert¹⁾ hat gezeigt, daß der größte erreichbare Stabilitätsbereich eines $P(EC)^S$ -Verfahrens mit festem Korrektor, aber beliebigem Prädiktor und beliebigem s , nicht größer sein kann als der größte Stabilitätsbereich eines $P(EC)^SE$ -Verfahrens mit demselben Korrektor.

9.5.3. Stabilitätsbereiche von P-C-Verfahren

Das charakteristische Polynom - und damit auch der Stabilitätsbereich - eines P-C-Verfahrens mit dem Prädiktor $\hat{M}(\hat{\rho}, \hat{\sigma})$ und dem Korrektor $M(\rho, \sigma)$ hängt nicht nur von $\hat{M}$ und M sondern auch noch von der Anzahl s der Korrektorschritte ab.

Für $P(EC)^SE$ -Verfahren mit $\alpha_k = \hat{\alpha}_k = 1$ gilt:

$$Q_S(\mu; H) = (\rho(\mu) - H\sigma(\mu)) \sum_{\ell=0}^{s-1} (\beta_k H)^\ell + (\hat{\rho}(\mu) - H\hat{\sigma}(\mu)) (\beta_k H)^s \quad (9.24)$$

und für $P(EC)^S$ -Verfahren mit $\alpha_k = \hat{\alpha}_k = 1$:

$$Q_S(\mu; H) = \mu^k (\rho(\mu) - H\sigma(\mu)) \sum_{\ell=0}^{s-1} (\beta_k H)^\ell + H(\hat{\rho}(\mu)\sigma(\mu) - \rho(\mu)\hat{\sigma}(\mu)) (\beta_k H)^{s-1} \quad (9.25)$$

Hierbei ist o.B.d.A. vorausgesetzt, daß Prädiktor und Korrektor die-
selbe Schrittzahl k haben, was durch Umindizierung formal immer erreicht werden kann (die in § 2, Seite 12, getroffene Vereinbarung teilerfremder ρ und σ sei dabei aufgehoben).

¹⁾ Lambert: Predictor-Corrector Methods with Identical Regions of Stability, SIAM J. Num. Anal. 8, 337-344 (1971).

Sind $\mu_i^{(s)}(H)$ die Nullstellen von $Q_s(\mu; H)$ und $\mu_i(H)$ die Nullstellen des charakteristischen Polynoms des Korrektors, so gilt nur dann

$$\lim_{s \rightarrow \infty} \mu_i^{(s)}(H) = \mu_i(H), \quad i=1(1)k$$

$$\lim_{s \rightarrow \infty} \mu_i^{(s)}(H) = 0, \quad i=k+1(1)2k \text{ (bei PEC-Verfahren)}$$

wenn (für $\alpha_k = \hat{\alpha}_k = 1$) $|\beta_k H| < 1$ ist. Das heißt: Selbst wenn unendlich oft mit dem Korrektor iteriert wird, braucht der Stabilitätsbereich des P-C-Verfahrens nicht mit dem Stabilitätsbereich H des Korrektors identisch zu sein.

Satz 9.4.

Der Stabilitätsbereich eines $P(EC)^s E$ - oder $P(EC)^s$ -Verfahrens ist für $s \rightarrow \infty$ nur dann mit dem Stabilitätsbereich H des Korrektors identisch, wenn (für $\alpha_k = 1$) gilt:

$$H \subseteq \{H: |\beta_k H| < 1\}$$

Bei den üblichen Verfahren mit $\alpha_k = 1$ liegt β_k in der Größenordnung von 2, so daß große Stabilitätsbereiche des Korrektors für das Stabilitätsverhalten des P-C-Verfahrens bedeutungslos sind. Andererseits kann der Stabilitätsbereich eines Korrektors durch einen geeignet gewählten Prädiktor vergrößert werden.

Beispiel¹⁾:

Der schwach stabile Korrektor:

$$y_{j+2}^* = y_j^* + \frac{h}{3} (f_j^* + 4f_{j+1}^* + f_{j+2}^*)$$

hat einen leeren Stabilitätsbereich (vgl. Abschnitt 7.3., Beispiel 4). Das PECE-Verfahren mit diesem Korrektor und dem folgenden instabilen Prädiktor

$$y_{j+2}^* = -4y_{j+1}^* + 5y_j^* + h (2f_j^* + 4f_{j+1}^*)$$

hat das charakteristische Polynom

$$Q(\mu; H) := \mu^2 - \frac{4}{3}H^2\mu - (1 + 2H + \frac{2}{3}H^2)$$

und den reellen Stabilitätsbereich: $-1 < H < 0$.

¹⁾Krogh: A Test for Instability in the Numerical Solution of Ordinary Differential Equations, J. ACM 14, 350-362 (1967).

9.5.4. Abschätzung des lokalen und globalen Fehlers¹⁾

Die Fehlerformel (3.24) in Abschnitt 3.4 enthält den lokalen Ersetzungsfehler $T_h[y](x)$, der im Falle eines Verfahrens der Ordnung q bei hinreichend glattem y gegeben ist durch (siehe § 2, Seite 11)

$$T_h[y](x) = c_{q+1} y^{(q+1)}(x) h^q + \mathcal{O}(h^{q+1})$$

$T_h[y](x)$ ist i.allg. schwer abzuschätzen, weil dabei Schranken für die q -ten partiellen Ableitungen von f benutzt werden müßten (denn $y(x)$ ist ja unbekannt). Bei P-C-Verfahren läßt sich nun wenigstens der Term $c_{q+1} y^{(q+1)}(x) h^q$ in einfacher Weise und ohne zusätzlichen Rechenaufwand berechnen mit Hilfe des folgenden Satzes:

Satz 9.5.

Für den lokalen Fehler $T_h[y](x_j)$ eines P-C-Verfahrens gilt:

$$h T_h[y](x_j) = \frac{c_{q+1} \hat{\alpha}_k}{\hat{c}_{q+1} - c_{q+1}} (y_j^* - \hat{y}_j^*) + \mathcal{O}(h^{q+2}) \quad (9.26)$$

($\hat{y}_j^*$: Prädiktor-Näherung; y_j^* : Näherung nach $s \geq 1$ Korrektorschritten)

wenn:

1. der Prädiktor $\hat{M}(\hat{\rho}, \hat{\sigma})$ wenigstens die Konsistenzordnung q des Korrektors $M(\rho, \sigma)$ hat, $\hat{q} \geq q$,
2. für die Wurzeln μ_i , $i=1(1)r$, von ρ mit dem Betrag 1 gilt: $\hat{\rho}(\mu_i) = 0$
3. die Startwerte des Verfahrens im Falle $r=1$ die Konvergenzordnung q , anderenfalls die Ordnung $q+1$ haben,
4. $\hat{c}_{q+1} \neq c_{q+1}$ ($\hat{c}_{q+1} = 0$ falls $\hat{q} > q$).

¹⁾ die Ergebnisse und Beispiele dieses Abschnitts sind dem Buch von Stetter [73] entnommen.

Den Beweis dieses Satzes findet man bei Stetter [73], S. 261, oder in spezieller Form bei Henrici [62], S. 256ff.

Man benutzt (9.26), um damit den globalen Fehler mittels Satz 3.1. abzuschätzen oder die Schrittweite h zu kontrollieren.

Beispiel:

Der Prädiktor sei die midpoint-rule (9.6b):

$$y_{j+1}^* = y_{j-1}^* + 2hf_j^* \quad \text{mit} \quad T_h[y](x_j) = \frac{h^2}{3} y^{(3)}(x_j) + \mathcal{O}(h^3)$$

der Korrektor die Sehnentrapezregel:

$$y_{j+1}^* = y_j^* + \frac{h}{2}(f_{j+1}^* + f_j^*) \quad \text{mit} \quad T_h[y](x_j) = -\frac{h^2}{12} y^{(3)}(x_j) + \mathcal{O}(h^3)$$

Beide Verfahren haben für $y \in \mathcal{C}^4[a, b]$ die Konsistenzordnung $q=2$ und es ist $\hat{c}_{q+1} = \frac{1}{3}$ und $c_{q+1} = \frac{-1}{12}$; $\rho(\mu)$ hat nur die Wurzel $\mu_1=1$. (9.26) ergibt somit:

$$hT_h[y](x_j) = -\frac{1}{5}(y_j^* - \hat{y}_j^*) + \mathcal{O}(h^4)$$

Der folgende Satz enthält eine Möglichkeit zur Abschätzung des globalen Fehlers, seinen Beweis findet man ebenfalls bei Stetter [73] (Seite 297):

Satz 9.6.

Das PECE- bzw. PEC-Verfahren der Konsistenzordnung q habe einen Prädiktor der Ordnung $q-1$ und es

$$\text{gelte} \quad \beta^* \hat{c}_q + \hat{\alpha}_k c_{q+1} = 0 \quad \text{und}^{1)} \quad \sum_{\ell=0}^k \beta_\ell = 1 \quad (9.27)$$

mit $\beta^* = \beta_k$ bzw. $\beta^* = 1$

Genügen außerdem die Startwerte der (sehr speziellen) Bedingung:

$$y_j^* = y(x_j) - c_{q+1} y^{(q)}(x_j) h^q + \mathcal{O}(h^{q+1}); \quad j = 0(1)k-1 \quad (9.28)$$

und bei PEC-Verfahren zusätzlich

$$\hat{y}_j^* = y(x_j) + \mathcal{O}(h^q); \quad j = 0(1)k-1$$

dann gilt für den globalen Fehler:

$$y_j^* - y(x_j) = \beta^* (y_j^* - \hat{y}_j^*) + \mathcal{O}(h^{q+1}) \quad (9.29)$$

wobei $\hat{y}_j^*$ wieder die Prädiktor-Näherungen bezeichnet.

Für eine PEC-Methode mit $\beta^* = 1$ bedeutet (9.29):

$$\hat{y}_j^* = y(x_j) + \mathcal{O}(h^{q+1})$$

d.h. die Prädiktor-Näherungen haben die Konvergenzordnung $(q+1)$,
(während der Korrektor nur Näherungen der Ordnung q ergibt!).

Auch bei einem PECE-Verfahren kann die Konvergenzordnung $(q+1)$ erreicht werden, indem man den in jedem Rechenschritt errechneten Wert y_j^* ersetzt durch den korrigierten Wert

$$y_j^{**} = y_j^* - \beta_k (y_j^* - \hat{y}_j^*) \quad (9.30)$$

¹⁾ diese Normalisierung tritt hier an die Stelle von $\alpha_k = 1$.

Man hat somit die Wahl, entweder die höhere Konvergenzordnung $(q+1)$ zu erzielen ohne Kontrolle des globalen Fehlers oder ein Ergebnis der Konvergenzordnung q , dieses jedoch mit Fehlerabschätzung.

Die beschriebene Erhöhung der Ordnung bzw. Fehlerabschätzung ist stets möglich, denn zu jedem Korrektor der Konsistenzordnung $q \geq 2$ existiert ein Prädiktor der Ordnung $(q-1)$, für den (9.27) gilt (Beweis s. Stetter [73], Seite 299).

Beispiel (Stetter S. 300):

Das PECE-Verfahren mit

$$\hat{y}_{j+2}^* = y_{j+1}^* + \frac{h}{3} (4f_{j+1}^* - f_j^*) \quad \text{Prädiktor (Konsistenzordnung 1)}$$

$$y_{j+2}^* = y_{j+1}^* + \frac{h}{2} (\hat{f}_{j+2}^* + f_{j+1}^*) \quad \begin{array}{l} \text{Korrektor (Sehnentrapezregel)} \\ \text{(Konsistenzordnung 2)} \end{array}$$

erfüllt mit $\hat{c}_q = \frac{1}{6}$; $c_{q+1} = -\frac{1}{12}$; $\beta^* = \beta_k = \frac{1}{2}$; $\hat{\alpha}_k = 1$ die Bedingung (9.27).
Es erlaubt die Abschätzung

$$y_j^* - y(x_j) = \frac{1}{2} (y_j^* - \hat{y}_j^*) + \mathcal{O}(h^3)$$

falls für die Startwerte gilt (vgl. (9.28)):

$$y_j^* = y(x_j) + \frac{1}{12} y^{(2)}(x_j) h^2 + \mathcal{O}(h^3) ; \quad j=0,1. \quad (9.31)$$

Da y_0^* nur in f_j^* des Prädiktors auftritt, braucht (9.31) nur für $j=1$ zu gelten, was z.B. für

$$y_1^* = \eta_0 + \frac{h}{12} \left[7f(x_0+h, \eta_0+hf_0) + 5f(x_0, \eta_0) \right]$$

der Fall ist. $\triangleleft$

Einzelheiten und weitere Ergebnisse über Fehlerabschätzungen findet man in C.W. Gear: Estimation of Errors and Derivatives in Ordinary Differential Equations, 1974, IFIP-Kongreßberichte.

10. Zyklische Verfahren

Mit Satz 5.2. und seinem Korollar lassen sich stabile zyklische Verfahren sowohl der (maximalen) Konvergenzordnung $2k$, als auch niedriger Ordnung konstruieren; beides ist sinnvoll und die Entscheidung für die eine oder andere Möglichkeit hängt von Stabilitätsbetrachtungen ab (siehe 10.2.).

10.1. Zur Konstruktion stabiler k-zyklischer k-Schritt-Verfahren maximaler Konvergenzordnung

Stabile k-zyklische k-Schritt-Verfahren der Konvergenzordnung $2k$ gewinnt man wie folgt:

- Man berechnet das allgemeine lineare k-Schritt-Verfahren der Konsistenzordnung $(2k-1)$, welches einen Parameter - beispielsweise β_k - enthält und parametrisiert die α_ℓ und β_ℓ , $\ell=o(1)k-1$, als Funktionen von β_k . Diese Parametrisierung ist für $k=2(1)6$ in (9.14) bereits ausgerechnet.
- Mit k derart parametrisierten k-Schritt-Verfahren (mit den Parametern $\beta_k^{(r)}$, $r=o(1)k-1$) wird ein k-zyklisches k-Schritt-Verfahren gebildet.
- Sodann werden $(k-1)$ Zahlen μ_i , $i=2(1)k$, mit $|\mu_i|<1$ gewählt und die Parameter $\beta_k^{(r)}$ aus dem (nichtlinearen) Gleichungssystem (5.32)/(5.33) berechnet.

Nach Satz 5.2. und seinem Korollar (Seite 48/50) charakterisiert jede Lösung dieses Systems ein stabiles k-zyklisches k-Schritt-Verfahren der Konvergenzordnung $2k$, obwohl jede seiner Stufen i.allg. nur die Ordnung $(2k-1)$ hat.

Die Existenz von Lösungen konnte bisher noch nicht für alle meines k bewiesen werden, wird jedoch vermutet (vgl. Stetter [73], S. 227). Donelson und Hansen haben (auf völlig andere Weise) Verfahren für $k=2,3,4$ angegeben; Verfahren mit $k=5$ und $k=6$ konnten mittels (5.32)/(5.33) ausgerechnet werden.

Als Beispiel¹⁾ betrachten wir den Fall $M=k=3$. Gleichung (5.31b) ergibt (nach Kürzung mit dem Faktor 243):

$$1001 \beta_3^{(0)} \beta_3^{(1)} \beta_3^{(2)} - 300(\beta_3^{(0)} \beta_3^{(1)} + \beta_3^{(1)} \beta_3^{(2)} + \beta_3^{(0)} \beta_3^{(2)}) + 90(\beta_3^{(0)} + \beta_3^{(1)} + \beta_3^{(2)}) - 27 = 0 \quad (10.1a)$$

und aus (5.33) folgt:

$$\mu_i^2 + a\mu_i + b = 0 \quad i=2,3. \quad (10.1b/c)$$

$$\begin{aligned} \text{mit } a = & -185193\beta_3^{(0)}\beta_3^{(1)}\beta_3^{(2)} + 59850(\beta_3^{(0)}\beta_3^{(1)} + \beta_3^{(1)}\beta_3^{(2)} + \beta_3^{(0)}\beta_3^{(2)}) \\ & - 19380(\beta_3^{(0)} + \beta_3^{(1)} + \beta_3^{(2)}) + 6289 \\ b = & -35937\beta_3^{(0)}\beta_3^{(1)}\beta_3^{(2)} + 10890(\beta_3^{(0)}\beta_3^{(1)} + \beta_3^{(1)}\beta_3^{(2)} + \beta_3^{(0)}\beta_3^{(2)}) \\ & - 3300(\beta_3^{(0)} + \beta_3^{(1)} + \beta_3^{(2)}) + 1000 \end{aligned}$$

$$\begin{aligned} \text{Setzt man } x := \beta_3^{(0)}; \quad y := \beta_3^{(1)}; \quad z := \beta_3^{(2)} \quad \text{und} \\ A := x+y+z; \quad B := xy+yz+zx; \quad C := xyz \end{aligned} \quad (10.3)$$

so reduziert sich (10.1a-c) auf das folgende lineare Gleichungssystem:

$$\begin{aligned} 19380A - 59850B + 185193C &= 6289 + (\mu_2 + \mu_3) \\ 3300A - 10890B + 35937C &= 1000 - \mu_2\mu_3 \\ 90A - 300B + 1001C &= 27 \end{aligned}$$

mit der Lösung

$$\begin{aligned} A &= \frac{1}{144} \left(\frac{12577}{110} + \frac{121}{190} (\mu_2 + \mu_3) + \frac{509}{22} \mu_2 \mu_3 \right) \\ B &= \frac{1}{144} \left(\frac{1909}{66} + \frac{11}{30} (\mu_2 + \mu_3) + \frac{4793}{330} \mu_2 \mu_3 \right) \\ C &= \frac{1}{144} \left(\frac{25}{11} + \frac{1}{19} (\mu_2 + \mu_3) + \frac{25}{11} \mu_2 \mu_3 \right) \end{aligned}$$

¹⁾ diese rechnerische Behandlung des Falles $M=k=3$ stammt von K. Mika.

Die weitere Rechnung ergibt, daß x , y und z aus

$$\begin{aligned} y^3 - Ay^2 + By - C &= 0 \\ x &= \frac{1}{2} (u + \sqrt{u^2 - 4v}); \\ z &= \frac{1}{2} (u - \sqrt{u^2 - 4v}) \end{aligned} \quad (10.4)$$

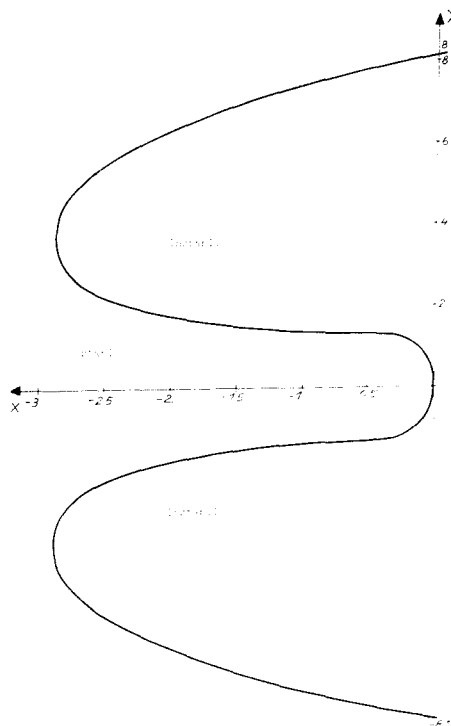
mit $u = \frac{By - C}{y^2}; \quad v = B - uy$

berechnet werden können. Für jedes Wertepaar (μ_2, μ_3) läßt sich so eine Lösung $(\beta_3^{(0)}, \beta_3^{(1)}, \beta_3^{(2)})$ gewinnen und die Abhängigkeit des zugehörigen Verfahrens von μ_2 und μ_3 untersuchen.

10.2. Bemerkungen zur Stabilität M-zyklischer k-Schritt-Verfahren

Untersuchungen¹⁾ zeigen, daß die zyklischen Verfahren mit $M=k \geq 3$ und maximaler Konvergenzordnung $2k$ nur sehr kleine Stabilitätsbereiche besitzen und daher schon für schwach steife Differentialgleichungen ungeeignet sind. Es ist daher sinnvoll, zyklische Verfahren geringerer Ordnung zu betrachten und zu versuchen, mittels freier Parameter ihren Stabilitätsbereich zu optimieren.

M. Mihelčić²⁾ gewann dabei Verfahren mit $M=2$ und $k=3$ der Ordnung 4 und Verfahren mit $M=k=3$ der Ordnung 5, die in der ganzen Halbebene $H = \{\lambda h \mid \operatorname{Re} \lambda h < 0\}$ mit Ausnahme endlicher Bereiche nahe der imaginären Achse absolut stabil sind. Die Ausnahmereiche sind klein und hängen u.a. von der Wahl der Eigenwerte μ_i ($i=2(1)k$) von $\hat{A} = A^{(M-1)} \dots A^{(1)} A^{(0)}$ ab.



Figur 10.1

¹⁾ von Th. Chaves/PUC-Rio de Janeiro (z.Zt. KFA-Jülich)

²⁾ eine Arbeit hierüber ist in Vorbereitung.

Das folgende $M=2, k=3$ -Verfahren, beispielsweise, hat die Ordnung 4 und den in Figur 10.1. skizzierten Stabilitätsbereich:

$$\begin{aligned} 200y_{2j+3}^* - 1143y_{2j+2}^* + 816y_{2j+1}^* + 127y_{2j}^* = h(109f_{2j+3}^* - 326f_{2j+2}^* - 653f_{2j+1}^*) \\ 127000y_{2j+4}^* - 120978y_{2j+3}^* + 36576y_{2j+2}^* - 42598y_{2j+1}^* = h(49149f_{2j+4}^* + 94929f_{2j+3}^* \\ + 10527f_{2j+2}^* + 21015f_{2j+1}^*) \end{aligned} \quad (10.5)$$

Lediglich zur Illustration des Vorgehens im Falle $M < k$ und ohne Optimierung des Stabilitätsbereichs wird abschließend noch ein Beispiel für die Berechnung eines stabilen 2-zyklischen 5-Schritt-Verfahrens der Ordnung 8 gegeben:

Die Parametrisierung (9.15) von (instabilen) 5-Schritt-Formeln der Ordnung 8 enthält zwei freie Parameter β_4 und β_5 . Kombiniert man $M=2$ dieser Formeln zyklisch, so hat man vier freie Parameter zur Verfügung, um das zyklische Verfahren zu stabilisieren, d.h. um zu erreichen, daß die vier von $\mu_1=1$ verschiedenen Eigenwerte $\mu_i, i=2(1)5$, von $\hat{A}=A^{(1)}A^{(0)}$ dem erweiterten Wurzelkriterium genügen (wegen der Konsistenz der beiden Stufen ist $\mu_1=1$ stets Eigenwert von $\hat{A}$).

Hierzu berechnet man das charakteristische Polynom $Q(\mu)$ von $A^{(1)}A^{(0)}$ (die Elemente der letzten Zeile der Frobenius-Matrizen $A^{(0)}$ bzw. $A^{(1)}$ werden mit α_i bzw. a_i bezeichnet):

$$\begin{aligned} Q(\mu) = \mu^5 + (\alpha_3 + a_3 - \alpha_4 a_4) \mu^4 + (\alpha_1 + a_1 + \alpha_3 a_3 - \alpha_2 a_4 - \alpha_4 a_2) \mu^3 + \\ + (\alpha_3 a_1 + \alpha_1 a_3 - \alpha_0 a_4 - \alpha_4 a_0 - \alpha_2 a_2) \mu^2 \\ + (\alpha_1 a_1 - \alpha_0 a_2 - \alpha_2 a_0) \mu - \alpha_0 a_0, \end{aligned} \quad (10.6)$$

substituiert die Parametrisierung (9.15) und berechnet die Parameter $\beta_4, \beta_5, b_4, b_5$ so, daß die Nullstellen μ_i von $Q(\mu)$ das erweiterte Wurzelkriterium erfüllen, etwa indem man setzt:

$$Q(\mu) = \mu^4 (\mu - 1). \quad (10.7)$$

Man erhält dann ein System von 4 nichtlinearen Gleichungen, im Falle (10.7) beispielsweise¹⁾:

$$\begin{aligned}
 \alpha_3 + a_3 - \alpha_4 a_4 &= -1 \\
 \alpha_1 + a_1 + \alpha_3 a_3 - \alpha_2 a_4 - \alpha_4 a_2 &= 0 \\
 \alpha_3 a_1 + \alpha_1 a_3 - \alpha_0 a_4 - \alpha_4 a_0 - \alpha_2 a_2 &= 0 \\
 \alpha_1 a_1 - \alpha_0 a_2 - \alpha_2 a_0 &= 0
 \end{aligned}
 \tag{10.8}$$

(Da $\mu_1=1$ automatisch Nullstelle von (10.6) ist, ist $\alpha_0 a_0=0$ dann von selbst erfüllt).

(10.8) besitzt eine Lösung und zwar (gerundet auf 8 Stellen):

$$\begin{aligned}
 \beta_4 &= 2,98233879 ; & b_4 &= 0,27734958 \\
 \beta_5 &= 0,25292957 ; & b_5 &= 0,29957694
 \end{aligned}$$

Die restlichen Koeffizienten des 2-zyklischen Verfahrens ergeben sich hiermit aus (9.15).

In ähnlicher Weise läßt sich durch zyklische Kombination von $(k-1)$ k -Schritt-Formeln der Ordnung $(2k-1)$ ein $(k-1)$ -zyklisches k -Schritt-Verfahren der Konvergenzordnung $(2k-1)$ konstruieren, falls das dabei auftretende System aus $(k-1)$ nichtlinearen Gleichungen eine Lösung besitzt (für $k=2(1)6$ ist das der Fall).

¹⁾ die α_i und a_i sind Funktionen von $\beta_4, \beta_5, b_4, b_5$ gemäß (9.15)!

Literatur

- 1) Albrecht, P., "Discretization Methods", 9. Coll. Bras. Mat., Inst. de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 1973.
- 2) Ames, W.F., "Nonlinear ordinary differential equations in transport processes", Academic Press, New York, 1968.
- 3) Barton D.; Willers, I.M.; and Zahar, R.V.M., "Taylor series methods for ordinary differential equations - An evaluation", in Rice [71, 13], pp. 369-390.
- 4) Bettis, D.G. (Ed.), Proceedings of the conference on the numerical solution of ordinary differential equations, Vol. 362, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1974.
- 5) Bjurel, G.; Dahlquist, G.; Lindberg, B.; Linde, S.; and Odén, L., "Survey of stiff ordinary differential equations", Report NA70.11, Dept. of Information Processing, Royal Inst. of Technology, Stockholm, Sweden, 1970.
- 6) Butcher, J.C., "On the convergence of numerical solutions to ordinary differential equations", Math. Comp., 20, 93 (Jan. 1966), pp. 1-10.
- 7) Butcher, J.C., "Implicit Runge-Kutta processes", Math. Comp., 18, 85 (Jan. 1964), pp. 50-64.
- 8) Butcher, J.C., "A modified multistep method for the numerical integration of ordinary differential equations", J. ACM, 12, 1 (Jan. 1965), pp. 124-135.
- 9) Butcher, J.C., "On Runge-Kutta processes of high order", J. Austral. Math. Soc., 4, Part 2 (May 1964), pp. 179-194.
- 10) Butcher, J.C., "The effective order of Runge-Kutta methods", Lecture Notes in Mathematics, Vol. 109, Springer-Verlag, Berlin, 1969, pp. 133-139.
- 11) Butcher, J.C., "Integration Processes based on Radau Quadrature Formula", Math. Comp. 18 (1964).

- 12) Byrne, G.D.; and Hindmarsh, A.C., "A polyalgorithm for the numerical solution of ordinary differential equations", Report UCRL-75652, Lawrence Livermore Lab., Livermore, Calif., April 1974.
- 13) Ceschino, F.; and Kuntzmann, J., "Numerical solution of initial value problems", Prentice-Hall, Englewood Cliffs, N.J., 1966.
- 14) Collatz, L., "The numerical treatment of differential equations", (3d ed.), Springer-Verlag, Berlin, 1960.
- 15) Dahlquist, G.G., "Convergence and stability in the numerical integration of ordinary differential equations", Math. Scand., Vol. 4, 1956, pp. 33-53.
- 16) Daniel, J.W.; and Moore, R.E., "Computation and theory in ordinary differential equations", W.H. Freeman, San Francisco, Calif., 1970.
- 17) Donelson, J.; and Hansen, E., "Cyclic Composite Multistep Predictor-Corrector-Methods", SIAM J. Numer. Anal. 8, 1971, pp. 137-157.
- 18) Enright, W., "Studies in the numerical solution of stiff ordinary differential equations", Tech. Report No. 46, Dept. of Computer Science, Univ. of Toronto, Ont., Canada, Oct. 1972.
- 19) Enright, W.H.; Bedet, R.; Farkas, I.; and Hull, T.E., "Test results on initial value methods for non-stiff ordinary differential equations", Tech. Report No. 68, Univ. of Toronto, May 1974.
- 20) Fehlberg, E., "Classical fifth-sixth-seventh- and eighth-order Runge-Kutta formulas with step-size control", NASA Tech. Report, NASA TR R-287, George C. Marshall Space Flight Center, Huntsville, Ala., Oct. 1968.
- 21) Fehlberg, E., "Low-order classical Runge-Kutta formulas with step-size control and their application to some heat transfer problems", NASA Tech. Report, NASA TR R-315, George C. Marshall Space Flight Center, Huntsville, Ala., July 1969; or German version in Computing, 6, 1-2 (1970), pp. 61-71.

- 22) Fowler, M.E.; and Warten, R.M., "A numerical integration technique for ordinary differential equations with widely separated eigenvalues", IBM J. Res. Develop., 11, 5 (sept. 1967), pp. 537-543.
- 23) Fyfe, D.J., "Economical Evaluation of Runge-Kutta formulae", Math. Comp., 20, 95 (July 1966), pp. 392-398.
- 24) Gear, C.W., "The automatic integration of large systems of ordinary differential equations", Digest Record of the Joint Conference on Mathematical Aids to Design, Anaheim, Calif., sponsored by ACM/SIAM/IEEE, Oct. 1969, pp. 27-58.
- 25) Gear, C.W., "The automatic integration of ordinary differential equations", Comm. ACM, 14, 3 (March 1971), pp. 176-179.
- 26) Gear, C.W., "The automatic integration of stiff ordinary differential equations", Proc., Information Processing 68, Vol. 1, (A.J.H. Morrell, Ed.), North-Holland Publ., Amsterdam, The Netherlands, 1968, pp. 187-193.
- 27) Gear, C.W., "Estimation of Errors and Derivatives in Ordinary Differential Equations", Report of the IFIP-Congress, 1974.
- 28) Gear, C.W., "Numerical initial value problems in ordinary differential equations", Prentice-Hall, Englewood Cliff, N.J., 1971.
- 29) Gear, C.W., "The numerical integration of stiff differential equations", Report No. 221, Dept. of Computer Science, Univ. of Illinois, Urbana, III., Jan. 1967.
- 30) Gear, C.W.; and Tu, K.W., "The effect of variable mesh size on the stability of multistep methods", Report UIUCDCS-R-73-570, Computer Science Dept., Univ. of Illinois, Urbana, III., April 1973.
- 31) Gear, C.W.; Tu, K.W.; and Watanabe, D.S., "The stability of automatic programs for numerical problems", in Willoughby [74,7], pp. 111-121.

- 32) Gear, C.W.; and Watanabe, D.S., "Stability and convergence of variable order multistep methods", Report UIUCDCS-R-73-571, Computer Science Dept., Univ. of Illinois, Urbana, III., May 1973.
- 33) Gragg, W.B.; and Stetter, H.J., "Generalized multistep predictor-corrector methods", J. ACM, 11, 2 (April 1964), pp. 188-209.
- 34) Haines, C.F., "Implicit integration processes with error estimation of the numerical solution of differential equations", Comput. J. 12 (1969).
- 35) Hall, G.; Enright, W.H.; Hull, T.E.; and Sedgwick, A.E., "Detest: A program for comparing numerical methods for ordinary differential equations", Tech. Report No. 60, Univ. of Toronto, Dec. 1973.
- 36) Henrici, P., "Discrete variable methods in ordinary differential equations", John Wiley, New York, 1962.
- 37) Hull, T.E., "The effectiveness of numerical methods for ordinary differential equations", in Studies in numerical analysis 2, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1968, pp. 114-121.
- 38) Hull, T.E., "The numerical integration of ordinary differential equations", in Information Processing 68, Vol. 1, (A.J.H. Morrell, Ed.), North-Holland Publ., Amsterdam, The Netherlands, 1969, pp. 40-53.
- 39) Hull, T.E., "A search for optimum methods for the numerical integration of ordinary differential equations", SIAM Rev., 9, 4 (Oct. 1967), pp. 647-654.
- 40) Hull, T.E.; and Creemer, A.L., "Efficiency of predictor-corrector procedures", J. ACM, 10, 3 (July 1963), pp. 291-301.
- 41) Hull, T.E.; Enright, W.H.; and Sedgwick, A.E., "Memorandum on problems and methods to apply to stiff systems of ordinary differential equations", Computer Science Dept., Univ. of Toronto, Toronto, Canada, March 1972.

- 42) Hull, T.E.; and Johnston, R.L., "Optimum Runge-Kutta methods", Math. Comp., 18, 86 (April 1964), pp. 306-310.
- 43) Hall, G.; Enright, W.H.; Hull, T.E.; and Sedgwick, A.E., "Detest: A program for comparing numerical methods for ordinary differential equations", Tech. Report No. 60, Dept. of Computer Science, Univ. of Toronto, Ont. Canada, Dec. 1973.
- 44) Krogh, F.T., "Predictor-corrector-methods of higher order with improved stability characteristics, JACM 13, (1966) p. 374.
- 45) Krogh, F.T., "Algorithms for changing the stepsize used by a multistep method", Tech. Memo. No. 275, Jet Propulsion Lab., Pasadena, Calif., May 1971; or SIAM J. Numer. Anal., 10, 5 (Oct. 1973), pp. 949-965.
- 46) Krogh, F.T., "A test for instability in the numerical solution of ordinary differential equations", J. ACM, 14, 2 (April 1967), pp. 351-354.
- 47) Krogh, F.T., "A variable step, variable order multistep method for the numerical solution of ordinary differential equations", Proc., Information Processing 68, Vol. 1, (A.J.H. Morrel, Ed.), North-Holland Publ., Amsterdam, The Netherlands, pp. 194-199.
- 48) Lambert, J.D., "Predictor-corrector methods with identical regions of stability", SIAM J. Num. Anal. 8 (1971), pp. 337-344.
- 49) Lambert, J.D., "Computational methods in ordinary differential equations", John Wiley, New York, 1973.
- 50) Lapidus.L.; and Seinfeld, J., "Numerical solution of ordinary differential equations", Academic Press, New York, 1971.
- 51) Lindberg, B., "Optimal stepsize sequences and requirements for the local error for methods for (stiff) differential equations", Tech. Report No. 67, Dept. of Computer Science, Univ. of Toronto, Ont., Canada, May 1974.
- 52) Liniger, W., "A criterion for A-stability of linear multistep integration formulae", Computing 3, 4 (1968), pp. 280-285.

- 53) Liniger, W.; and Willoughby, R.A., "Efficient integration methods for stiff systems of ordinary differential equations", SIAM J. Numer. Anal. 7, 1 (March 1970), pp. 47-66.
- 54) Meyer, G.H., "Initial value methods for boundary value problems", Academic Press, New York, 1973.
- 55) Morris, J.L. (Ed.), Conference on the numerical solution of differential equations, Vol. 109, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1969.
- 56) Rosenbrock, H.H., "Some general implicit processes for the numerical solution of differential equations", Comput. J. 5 (1963).
- 57) Rosser, J.B., "A Runge-Kutta for all seasons", SIAM Rev., 9, 3 (July 1967), pp. 417-452.
- 58) Sandberg, I.W.; and Shichman, H., "Numerical integration of systems of stiff nonlinear differential equations", Bell Systems Tech. J., 47, 4 (April 1968), pp. 511-527.
- 59) Sedgwick, A., "An efficient variable order variable step Adams method", Tech. Report No. 53, Dept. of Computer Science, Univ. of Toronto, Ont., Canada, May 1973.
- 60) Shampine, L.F., "Stability region for extrapolated Runge-Kutta and Adams methods", Research Report SC-RR-72 0223, Applied Math. Div., Sandia Labs., Albuquerque, N.M., March 1972.
- 61) Shampine, L.F.; and Gordon, M.K., "Computer solution of ordinary differential equations: The initial value problem", Freeman, San Francisco, Calif., 1974.
- 62) Spijker, M.N., "Stability and convergence of finite-difference methods", Thesis, University of Leiden, 1968.
- 63) Stetter, H.J., "Analysis of discretization methods for ordinary differential equations", Springer-Verlag, Berlin, 1973.

- 64) Widlund, O.B., "A note on unconditionally stable linear multistep methods", BIT 7, (1967) pp. 65-70.
- 65) Willoughby, R.A. (Ed.), "Stiff differential systems", Plenum Press, New York, 1974.

Index

A-stabil	63,65,84,89,	explizites Mehrschritt- verfahren	31,93,94,
A(o)-stabil	60,63,	explizites Runge-Kutta- Verfahren	68,69,81,90,
absolut stabil	57,59	explizites Verfahren	3,
Adams-Bashfort, Verf. von	94	Faddejew-Faddejewa	42,
Adams-Moulton, Verf. von	96	Farkas	89,
Albrecht	20	Fehlberg-Formel	80,
Änderung der Schrittweite	99	Fehler, globaler	55,108
Bedet	89	lokaler	4,43,107
bedingt konsistent	9	-abschätzung	66,92,
Bereich absoluter Stabi- lität	58	-konstante	12,46,48
Butcher	77,81,82,83,	Fixpunktsatz	96
-Formel,	91,	FORMAC	67
,Verfahren von	78 91	Frobeniusmatrix	28,34,36,48, 49
Calahan-Formel	89	Gauss-Formel	83,90,
charakteristische Gleichung	13	,implizite	65,84,
charakteristisches Polynom	56,58,93,	Gear	110
Chaves	113	Gill, Verfahren von	76
Collatz, Faustregel von	75	globaler Fehler	55,108
Dahlquist	20,25	Haines	87
DETEST	89	Haines-Formel	89
Donelson	37,41,100,	Hall	89
Du Fort-Frankel, Ver- fahren von	40	Hansen	37,41,100
Einschrittverfahren	2,66,	Hauptwurzel	14,57,86
enges Wurzelkriterium	43,48,	Henrici	108
Enright	89,	Heun, Verfahren von	71
Ersetzungsfehler, lokaler	4,43	Hull	89
erzeugendes Polynom	12,92,	implizite-Gauss-Formeln	84
Euler-Verfahren	2,67,	implizites Mehrschritt- verfahren	31,95
,modifiziertes-	4,57,70,	implizites Runge-Kutta- Verfahren	69,81,83,90
,implizites-	63,	implizites Verfahren	3
,verbessertes-	71,	instabil	16,52,53,99
		instabile lineare Mehr- schrittverfahren	99
		Iteration	96

k-Schritt-Verfahren	2,92	Milne-Simpson, Verf. von	4,6,64,93,96
,M-zyklisches-	34	modifiziertes Euler-Verf.	4
k-zyklisches k-Schritt-Verfahren	41,46,111	Nebenwurzel	14
Klopfenstein	104	Newton-Cotes-Formeln	93
konsistent	5,8	,offene	94
,bedingt	9	,geschlossene	95
Konsistenzbedingung	12	Normalbasis	42
Konsistenzordnung	5,24,41,97,99,102	Normalform, Jordan'sche	59
Konsistenzsatz	9	Nyström, Verfahren von	77,93,95
Konvergenz	23	Ordnung	24,25,41,43,48,66,92,102
Konvergenzordnung	24,24,25,43,46,48,92	Ordnungsbedingung	49,50
, maximale	51,111	Padé-Approximation	65,84,86
Konvergenzsatz für Ein-schritt-Verfahren	66	Parametrisierung	100
für lineare k-Schritt-Verfahren	92	P-C-Verfahren	31,102,105,109
Korrektor	31,95	PEC-Verfahren	32,103,109
Krogh	106	PECE-Verfahren	32,103,109
Lambert	105	Prädiktor	31,96,99
Lawson-Formel	80,91	Prädiktor-Korrektor-Verfahren	31,32,102,105,109
lineares Mehrschritt-verfahren	3,92,97	Quadraturformel von Newton-Cotes	93
Lobatto-Formel	85,86	Radau-Formel	85,91
lokaler Ersetzungsfehler	4,43,46	relativ stabil	57
Luther-Formel	78,79	Rosenbrock-Formel	87,77
M=2,k=3-Verfahren	34	Runge-Kutta, Verfahren von	68
M-zyklisches-Verfahren	33	,explizites-	68,69
M-zyklisches k-Schritt-Verfahren	34,48,113	,höherer Ordnung	77
Mehrschrittverfahren	2	,implizites-	69,81
,explizites-	93,94	,klassisches-	74,75
,implizites-	31,95	,r-stufiges-	70,72
,instabiles-	99	,semi-implizites-	81,87
,lineares-	3,92,97	Sarafyan-Formel	79
,mehrstufiges-	34	Schrittweitensteuerung	75
midpoint-rule	3,6	schwach stabil	29,98
Mihelcić	113	Sedgwick	89
Mika	112	Sehnentrapezregel	71
Millmann	104	selbststartend	91

Semi-implizites		Verfahren Runge-Kutta-	68
Runge-Kutta-Verfahren	81,87,88	,Trapez-	63
Spijker	45	,zyklisches-	48
stabil	16,20,36,52,	Wärmeleitungsgleichung	7
	99	Widlund	60
,absolut	57,59,61	Wurzelkriterium	29
,relativ	57	,erweitertes-	36
,schwach	29,60	,enges-	36
,stark	29		
Stabilität	15		
einer Differential-		zyklisches Verfahren	33,48,111
gleichung	52		
für festes h	54		
zyklischer Verfahren	113		
Stabilitätsbereich	58,91,113		
von P-C-Verfahren	105		
Stabilitätskriterium			
allgemeines	19		
für k-Schritt-Verf.	28		
für zykl. Verfahren	36		
stark stabil	29		
steif	60,62		
Stetter	61,107,108,		
	110		
Stufe eines zyklischen			
Verfahrens	33		
Tangententrapezregel	70		
Taylor-Verfahren	67		
Trapezverfahren	63		
Verfahren, Adams-Bashfort-	94		
,Adams-Moulton-	96		
,Butcher-	91		
,Du Fort-Frankel-	40		
,Einschritt-	2		
,Euler-	2		
,explizites-	3,60		
,Gill-	76		
,Heun-	71		
,implizites-	60		
,k-Schritt-	2		
,3/8-	75		
,Lawson-	91		
, Mehrrschritt-	2		
,Milne-Simpson-	4,6,96		
mit Ableitungen	67		
,Nyström-	95		
,Prädiktor-Korrektor-	102		
,r-stufiges	68		

Abschließend danke ich meinen Mitarbeitern Frau C. Schröck-Pauli, G. Honig, M. Mihelčić und K. Mika für wertvolle Diskussionen und Beiträge, insbesondere Frau Schröck-Pauli für die Ausarbeitung des Paragraphen 8 und zahlreiche Verbesserungsvorschläge und Herrn Mika für seine ungemein sorgfältige Durchsicht des größeren Teils des Manuskripts. Von meinen Studenten waren Frl. A. Dapper, Frl. R. Osthoff und Herr Th. Kellermann an der Durchsicht des Manuskripts beteiligt. Meiner Sekretärin, Frau Ch. Lackes, danke ich für die viele Mühe, die mit dieser Arbeit verbunden war.