Numerik
gewöhnlicher Differentialgleichungen

Skript zur Vorlesung
Wintersemester 2008/09

Hans Joachim Oberle

Inhalt

1. Einige analytische Lösungmethoden

A. Allgemeines.

Wir beginnen mit einigen grundlegenden Begriffen und Klassifikationen im Zusammenhang mit gewöhnlichen Differentialgleichungen.

Definition (1.1)

a) Eine Gleichung bzw. ein Gleichungssystem der Form

$$F(t, y(t), y'(t), \ldots, y^{(p)}(t)) \;=\; 0 \;\in\; \mathbb{R}^n, \tag{1.2}$$

für eine unbekannte, p–fach stetig differenzierbare Funktion $y : \mathbb{R} \supset I \to \mathbb{R}^n$, also $t \in \mathbb{R}$ (Zeit), $y(t) \in \mathbb{R}^n$ (Zustand), heißt ein *System gewöhnlicher Differentialgleichungen (DGL).*

Kommt die p-te Ableitung $y^{(p)}(t) = (y_1^{(p)}(t), \ldots, y_n^{(p)}(t))^{\mathrm{T}}$ explizit in (1.2) vor, so spricht man von einer DGL der *Ordnung* p.

b) Ist die DGL (1.2) von der Form

$$y^{(p)}(t) \;=\; f(t, y(t), y'(t), \ldots, y^{(p-1)}(t)), \tag{1.3}$$

so heißt sie *explizit*, andernfalls *implizit*.

c) Hängt die DGL nicht explizit von der Zeit t ab, so heißt sie *autonom*.

d) Die DGL (1.2) bzw. (1.3) heißt *linear*, falls sie affin-linear in der abhängigen Variablen $y(t)$ ist; im impliziten Fall lautet eine lineare DGL also

$$\sum_{k=0}^{p} A_k(t) y^{(k)}(t) \;=\; b(t). \tag{1.4}$$

Die $A_k(t) \in \mathbb{R}^{(n,n)}$ sind dabei ev. zeitabhängige, reelle $n \times n$ Matrizen. $b(t) \in \mathbb{R}^n$ heißt die *Inhomogenität* der linearen DGL. (1.4). Die lineare DGL heißt *homogen*, falls $b = 0$ ist, andernfalls *inhomogen*. Im expliziten Fall gilt in (1.4) $A_p(t) = I_n$ (Einheitsmatrix).

Bemerkungen (1.5)

a) Im Allgemeinen werden die in Definition (1.1) vorgegebenen Funktionen F, f bzw. A_k und b als hinreichend glatt, d.h. hinreichend oft stetig differenzierbar, vorausgesetzt.

b) Ist die Matrix $F_{y^{(p)}}(t_0, y_0, y_0', \ldots, y_0^{(p)}) \in \mathbb{R}^{(n,n)}$ zu einem Zeitpunkt t_0 und für vorgegebene Werte $y_0, y_0', \ldots, y_0^{(p)} \in \mathbb{R}^n$ regulär und gilt zudem $F(t_0, y_0, y_0', \ldots, y_0^{(p)}) = 0$, so lässt sich die DGL (1.2) nach dem Satz über implizite Funktionen lokal eindeutig nach $y^{(p)}(t)$ auflösen und man erhält eine explizite DGL (1.3).

c) Ist $A_p(t)$ in (1.3) für alle betrachteten $t \in I$ regulär, so lässt sich (1.3) eindeutig nach $y^{(p)}(t)$ auflösen und man erhält (nach Umbenennung) die (äquivalente) *explizite Form einer linearen Differentialgleichung*

$$y^{(p)}(t) \; + \; \sum_{k=0}^{p-1} A_k(t) y^{(k)}(t) \;\; = \;\; b(t). \tag{1.6}$$

Es ist klar, dass eine vorgegebene DGL i. Allg. unendlich viele Lösungen y besitzen wird.

Beispiel (1.7)

Die skalare DGL $y^{(m+1)}(t) = 0$ besitzt gerade als Lösungsraum den Polynomraum Π_m aller (reellen) Polynomfunktionen vom Grad kleiner gleich m. Der Lösungsraum ist also ein $(m+1)$ dimensionaler (reeller) linearer Teilraum von $\mathrm{C}^{m+1}(\mathbb{R})$.

Beispiel (1.8)

Die *allgemeine Lösung* der skalaren, linearen und homogenen DGL $y'(t) = y(t)$ lautet $y(t) = C\,\mathrm{e}^t$, $C = \text{const}$. Der Lösungsraum ist also ein eindimensionaler linearer Raum.

Beispiel (1.9)

Die allgemeine Lösung der skalaren, linearen und homogenen DGL $y''(t) = -y(t)$ lautet $y(t) = C_1 \cos t + C_2 \sin t$, $C_i = \text{const}$. Der Lösungsraum ist also einen zweidimensionaler linearer Teilraum von $\mathrm{C}^{m+1}(\mathbb{R})$.

Um Eindeutigkeit zu erzielen, kann man zusätzlich zur DGL gewisse Daten der gesuchten Lösung $y \in \mathrm{C}^p(I, \mathbb{R}^n)$ vorschreiben. Welche Daten man dazu vorzugeben hat, ist natürlich nicht beliebig und hängt vom konkreten Zusammenhang und auch von der DGL selbst ab.

Schreibt man alle Daten $y^{(k)}(t)$, $k = 0, \ldots, p$ zu einem festen Zeitpunkt t_0 vor, so spricht man von einer *Anfangswertaufgabe (AWA).*

Für die explizite DGL (1.3) lautet die allgemeine AWA

$$\begin{aligned} y^{(p)}(t) &= f(t, y(t), y'(t), \ldots, y^{(p-1)}(t)), \\ y^{(k)}(t_0) &= y_0^k, \qquad k = 0, \ldots, p-1. \end{aligned} \tag{1.10}$$

Man beachte, dass $y^{(p)}(t_0)$ aus der DGL selbst berechnet werden kann. Die übrigen Anfangswerte $y_0^k \in \mathbb{R}^n,\ k = 0, \ldots, p-1$ können hierbei beliebig vorgegeben werden.

Beispiel (1.11) Die AWA

$$y''(t) \ = \ -y(t), \quad y(0) \ = \ 1, \ y'(0) \ = \ -1,$$

besitzt die eindeutig bestimmte Lösung $y(t) \ = \ \cos t \ - \ \sin t.$

Für implizite DGLen sind konsistente Anfangswerte $y_0^k \in \mathbb{R}^n,\ k = 0, \ldots, p,$ i. Allg. nicht leicht zu finden, da diese die DGL in t_0 erfüllen müssen. Es muss also gelten $F(t_0, y_0, \ldots, y_0^p) = 0$. Damit läuft die Bestimmung konsistenter Anfangswerte i. Allg. auf die Lösung eines nichtlinearen Gleichungssystems hinaus.

Eine andere Möglichkeit Eindeutigkeit zu erzielen, besteht darin gewisse Daten von y an *mehreren Zeitpunkten* $t_0 < \ldots < t_m$ vorzuschreiben. Man spricht dann von einer *Randwertaufgabe (RWA)*.

Beispiel (1.12) Die *Zweipunkt-RWA*

$$y''(t) \ = \ -y(t), \quad y(0) \ = \ 1, \quad y(3\pi/2) \ = \ 1,$$

besitzt die eindeutig bestimmte Lösung $y(t) \ = \ \cos t \ - \ \sin t.$

Beispiel (1.13)

Die Standard-Interpolationsaufgabe für Polynome

$$y^{(m+1)}(t) = 0, \quad y(t_0) = y_0, \ \ldots, \ y(t_m) = y_m, \tag{1.14}$$

bei vorgegebenen Stützstellen (t_k, y_k) kann als Beispiel für eine *Mehrpunkt-RWA* angesehen werden.

Die Standard-Formulierung einer *Zweipunkt-RWA für ein explizites DGL-system erster Ordnung* lautet

$$y'(t) \ = \ f(t, y(t)), \quad r(y(a), \, y(b)) \ = \ 0. \tag{1.15}$$

Dabei ist $a < b$ und $r : \ \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ eine Funktion, die ev. auch nichtlineare Randbedingungen beschreibt. Interpretation: Genau n (unabhängige) Daten müssen vorgegeben werden, um die Lösung von $y' = f(t, y),\ y(t) \in \mathbb{R}^n,$ eindeutig festzulegen.

B. Zwei Beispiele.

Wir betrachten zwei praxisnahe Beispiele für Anfangswertaufgaben, wobei das erste Beispiel auf eine implizite, das zweite auf eine explizite DGL führt. Beide Beispiel sind nichtlinear.

Beispiel (1.16) (Elektrischer Schaltkreis[1])

Als ein relativ einfaches Beispiel aus der Simulation elektrischer Schaltkreise betrachten wir einen Verstärker, der aus Ohmschen Widerständen, Kondensatoren und zwei Transistoren besteht, vgl. das Schaltdiagramm in Abb. 1.1.

$U_e(t)$ ist die zeitabhängige Eingangsspannung, U_b eine vorgegebene konstante Betriebsspannung.

Für die Ohmschen Widerstände gilt das Ohmsche Gesetz $U = R\,I,$ wobei U die anliegende Spannung, R den Widerstand und I die Stromstärke bezeichnet. Für die Kondensatoren hat man die Regel $I = C\,\dot{U},$ $\dot{U}$ ist die zeitliche Ableitung von $U,$ und für die Transistoren wird das Strom/Spannungsverhalten durch eine Kennlinie der Form $I = f(U)$ beschrieben. Konkret wird im Beispiel $f(U) := \beta\,(\exp(U/U_F) - 1)$ gewählt.

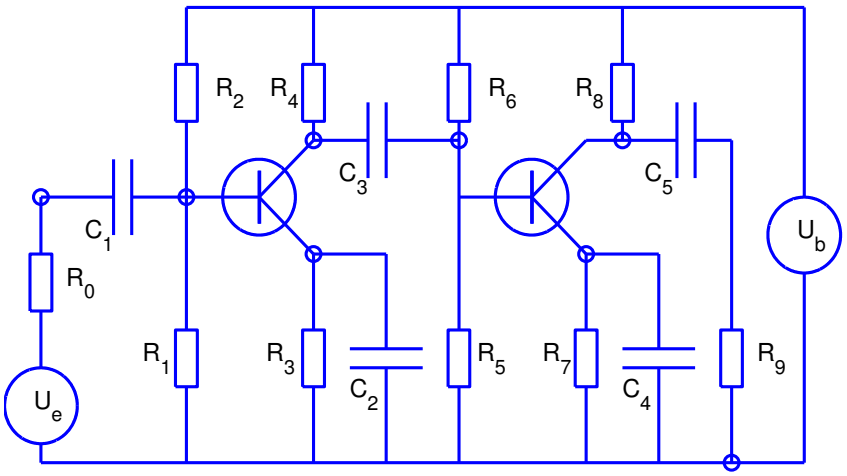

Abb. 1.1 Elektrischer Verstärker.

Zur Bestimmung des zeitlichen Verlaufs aller auftretenden Spannungen und Ströme wendet man die *Kirchhoffsche Knotenregel* an, nach der an jedem Knoten die Summe der in diesen Knoten einlaufenden vorzeichenbehafteten Ströme verschwinden muss.

Im konkreten Fall ergibt sich mit den in Abb.1.1 eingezeichneten Koten die folgenden Relationen für die Knotenspannungen $U_j,\ j = 1, \ldots, 8.$

[1] Nach Rentrop, Roche, Steinebach: Numer. Math. 55, 545–563 (1989).

$$
\begin{aligned}
&U_e/R_0 \;-\; U_1/R_0 \;+\; (\dot{U}_2 - \dot{U}_1)C_1 \;=\; 0.\\
&(\dot{U}_1 - \dot{U}_2)C_1 \;-\; U_2(1/R_1 + 1/R_2) \;+\; U_b/R_2 \;-\; f(U_2 - U_3) \;+\; \alpha f(U_2 - U_3) \;=\; 0,\\
&f(U_2 - U_3) \;-\; U_3/R_3 \;-\; \dot{U}_3 C_2 \;=\; 0,\\
&-\alpha f(U_2 - U_3) \;-\; U_4/R_4 \;+\; U_b/R_4 \;-\; (\dot{U}_4 - \dot{U}_3)C_3 \;=\; 0,\\
&(\dot{U}_4 - \dot{U}_5)C_3 \;-\; U_5(1/R_5 + 1/R_6) \;+\; U_b/R_6 \;-\; f(U_5 - U_6) \;+\; \alpha f(U_5 - U_6) \;=\; 0,\\
&f(U_5 - U_6) \;-\; U_6/R_7 \;-\; \dot{U}_6 C_4 \;=\; 0,\\
&-\alpha f(U_5 - U_6) \;-\; U_7/R_8 \;+\; U_b/R_8 \;-\; (\dot{U}_7 - \dot{U}_8)C_5 \;=\; 0,\\
&(\dot{U}_7 - \dot{U}_8)C_5 \;-\; U_8/R_9 \;=\; 0.
\end{aligned}
$$

Fasst man die ableitungsfreien Terme zu einer Funktion $F(U)$, $U := (U_1, \ldots, U_8)^{\mathrm{T}}$, zusammen, so erhält das obige DGL-System die Form

$$
C\ \dot{U}(t) \;-\; F(U(t)) \;=\; 0. \tag{1.17}
$$

Die Kapazitätsmatrix $C \in \mathbb{R}^{(8,8)}$ ist hierbei eine konstante, aber singuläre Matrix.

$$
C \;=\; \begin{pmatrix}
-C_1 & C_1 & & & & & & \mathbf{0} \\
C_1 & -C_1 & & & & & & \\
 & & -C_2 & & & & & \\
 & & & -C_3 & C_3 & & & \\
 & & & C_3 & -C_3 & & & \\
 & & & & & -C_4 & & \\
 & & & & & & -C_5 & C_5 \\
\mathbf{0} & & & & & & C_5 & -C_5
\end{pmatrix}.
$$

Das DGL System (1.17) ist daher implizit vom Typ (1.2) und besitzt die Ordnung $p = 1$. Es lässt sich jedoch nicht nach $\dot{U}$ auflösen. Zudem ist die DGL nichtlinear aufgrund der nichtlinearen Kennlinie der Transistoren.

Man sieht unmittelbar, dass C den $\mathrm{Rang}(C) = 5$ besitzt. Somit besteht das DGLsystem (1.17) eigentlich aus fünf skalaren DGLen und drei algebraischen Gleichungen. Man spricht bei derartigen Systemen von *differentiell–algebraischen Gleichungen (DAEs)*.

Der Vollständigkeit halber seien noch die (dimensionslosen) Modellparameter des Problems angegeben:

$$
\begin{aligned}
&U_b \;=\; 6, \quad U_F \;=\; 0.026, \quad \alpha \;=\; 0.99, \quad \beta \;=\; 10^{-6},\\
&R_0 \;=\; 1000, \quad R_i \;=\; 9000 \;\; (i = 1, \ldots, 9), \quad C_j \;=\; j \cdot 10^{-6} \;\; (j = 1, \ldots, 5)
\end{aligned}
$$

Beispiel (1.18) (Restringiertes Dreikörper Problem[2])

Zur Beschreibung der ebenen Bewegung eines Satelliten im Kraftfeld von Erde und Mond betrachtet man ein rotierendes kartesisches Koordinatensystem, dessen x-Achse durch die Zentren von Erde und Mond gehen und dessen y-Achse durch den gemeinsamen Schwerpunkt von Erde und Mond geht.

Die Position (x, y) eines Massenpunktes (Satelliten) in dem von Erde und Mond aufgebautem Gravitationsfeld genügt dann dem folgenden DGL-System

$$\begin{aligned} \ddot{x} &= x + 2\,\dot{y} - \hat{\mu}\,\frac{x+\mu}{[(x+\mu)^2+y^2]^{3/2}} - \mu\,\frac{x-\hat{\mu}}{[(x-\hat{\mu})^2+y^2]^{3/2}} \\ \ddot{y} &= y - 2\,\dot{x} - \hat{\mu}\,\frac{y}{[(x+\mu)^2+y^2]^{3/2}} - \mu\,\frac{y}{[(x-\hat{\mu})^2+y^2]^{3/2}}. \end{aligned} \tag{1.19}$$

Hierbei bezeichnet $\mu = 1/82.45$ das Massenverhältnis vom Mond zur Erde und es ist $\hat{\mu} := 1-\mu$. Die Skalierung des Modells ist so gewählt, dass der Abstand 1 in der x, y-Ebene gerade dem (als konstant angenommenen) Abstand von der Erde zum Mond entspricht.

Bei (1.19) handelt es sich um ein explizites DGLsystem vom Typ (1.3) mit zwei Zustandsgrößen x und y, also $n = 2$, und der Ordnung $p = 2$. Um eine Anfangswertaufgabe zu erhalten hat man also gemäß (1.10) die Daten $x(0)$, $y(0)$, $\dot{x}(0)$ und $\dot{y}(0)$ vorzugeben.

Die Anfangsposition des Satelliten sei nun

$$x(0) = 1.2, \quad y(0) = 0.$$

Die Anfangsgeschwindigkeit wird senkrecht zur x-Achse gewählt, also $\dot{x}(0) = 0$. Ferner wird $\dot{y}(0)$ so gewählt, dass sich eine periodische Satellitenbahn mit einer Periode $T > 0$ einstellt. Man erhält z.B.

$$\dot{x}(0) = 0, \quad \dot{y}(0) = -1.049357510, \quad T = 6.192169331$$

Zur numerischen Lösung dieser AWA mit einem Standard-Integrator aus der MATLAB-Programmbibliothek (z.B. ode45) muss man die DGL (1.19) zunächst in ein DGLsystem erster Ordnung transformieren. Dazu setzt man

$$y_1 := x, \quad y_2 := y, \quad y_3 := \dot{x}, \quad y_4 := \dot{y}.$$

Man erhält dann die folgende AWA

$$\begin{aligned} y_1' &= y_3 \\ y_2' &= y_4 \\ y_3' &= y_1 + 2\,y_4 - \mu'\,(y_1+\mu)/r_1 - \mu\,(y_1-\mu')/r_2 \\ y_4' &= y_2 - 2\,y_3 - \mu'\,y_2/r_1 - \mu\,y_2/r_2 \end{aligned} \tag{1.20}$$

[2]nach Bulirsch, Stoer: Numer. Math. 8, 1–13 (1966).

$$
\begin{aligned}
r_1 &= [(y_1+\mu)^2 + y_2^2]^{3/2} \\
r_2 &= [(y_1-\mu')^2 + y_2^2]^{3/2} \\
y_1(0) &= 1.2, \quad y_2(0) = 0, \\
y_3(0) &= 0, \quad y_4(0) = -1.049357510
\end{aligned}
\tag{1.20}
$$

Numerisch integriert wird diese AWA im Intervall $[0, T]$, wobei T die oben angegebene Periode bezeichnet. Anhand der Abweichungen $y_j(T) - y_j(0)$ lässt sich dann auf die Genauigkeit der numerischen Integration schließen. Die mit dem Programm ode45 und einer moderaten Genauigkeitforderung von $\mathrm{TOL} = 10^{-5}$ berechnete Bahn ist in der Abbildung 1.2 dargestellt.

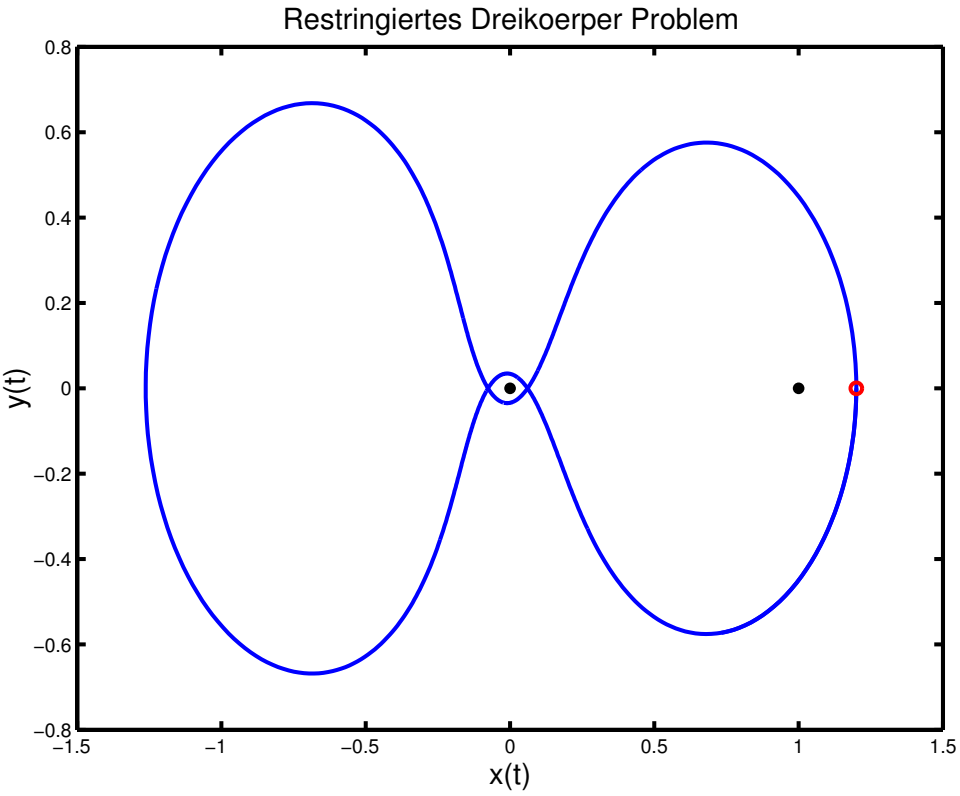

Abb.1.2 Periodische Satellitenbahn um Erde und Mond.

Bemerkung (1.21)

Zur Bestimmung von Anfangsbedingungen, die auf eine periodische Bahn führen, hat man eigentlich eine Randwertaufgabe zu lösen. Im vorliegenden Fall kann dies etwa mittels der folgenden Randbedingungen erfolgen.

$$\begin{aligned} x(0) &= 1.2, \quad \dot{x}(0) = 0, \qquad \dot{x}(T) = 0, \\ y(0) &= 0, \qquad x(T) = 1.2. \end{aligned} \tag{1.22}$$

Man beachte, dass die Endzeit T dieser RWA selbst unbekannt ist und mitbestimmt werden muss. Daher hat man die vier DGLn erster Ordnung aus (1.20), den unbekannten Parameter T und die fünf Randbedingungen (1.22).

Man spricht hierbei von einer *Randwertaufgabe mit freier Endzeit*.

C. Skalare DGL erster Ordnung.

Wir gehen auf einige (wenige) Standardmethoden zur analytischen Lösung expliziter, skalarer DLGen erster Ordnung ein, also DGLen vom Typ $y'(t) = f(t, y(t))$, mit $t \in I \subset \mathbb{R}$, $y(t) \in \mathbb{R}$.

C1. Trennung der Variablen.

Lassen sich die Variablen t und $y = y(t)$ multiplikativ trennen, also

$$y'(t) = h(t) \cdot g(y(t)), \tag{1.23}$$

und ist $g(y(t)) \neq 0$, $t \in I$, so lässt sich (1.22) durch $g(y(t))$ dividieren.

Mittels Integration über t und der Substitution $y = y(t)$ für die linke Seite folgt

$$\int_{y_0}^{y} \frac{dy}{g(y)} = \int_{t_0}^{t} h(\tau)\, d\tau.$$

Dabei ist $(t_0, y_0) \in I \times \mathbb{R}$ ein beliebig vorgegebener Anfangspunkt. Die Berechnung der Integrale und die Auflösung der resultierenden Gleichung nach $y = y(t)$ ergibt dann die Lösung der zugehörigen AWA. Durch die beliebige der Anfabgswerte erhält man alle Lösungen der DGL, für die $g(y) \neq 0$ ist.

Neben diesen Lösungen kann es aber noch weitere, so genannte *singuläre* Lösungen geben. Dies sind konstante Lösungen der Form $y(t) := y_0$ wobei $g(y_0) = 0$ ist.

Beispiel (1.24) $\qquad y' = -t/y.$

Trennung der Variablen ergibt $y\, y' = -t$. Integration liefert die Lösungsdarstellung

$$y^2 + t^2 = y_0^2 + t_0^2 =: r^2,$$

d.h. die Lösungen sind Ursprungskreise $y(t) = \pm\sqrt{r^2 - t^2}$.

Natürlich gehören die Punkte auf der t–Achse (formal) nicht zu den Lösungen der Differentialgleichung. Dort ist ja die rechte Seite der Differentialgleichung nicht definiert, bzw. $y' = \pm\infty$. Jeder Kreis besteht also aus *zwei* Lösungen, nämlich der mit $y > 0$ und der mit $y < 0$. Wir stellen zugleich fest, dass die Lösungen der Differentialgleichung (anders als in den bisherigen Beispielen) nur auf *beschränkten*, offene Intervallen definiert sind.

Über die Lösungen eine skalaren Differentialgleichung kann man sich durch Skizzierung des *Richtungsfeldes* (t, y, y') etwa auf einem geeigneten Gitter in der (t, y)–Ebene einen qualitativen Einblick verschaffen.

Für das obige Beispiel erhält man das in Abb. 1.3 dargestellte Bild.

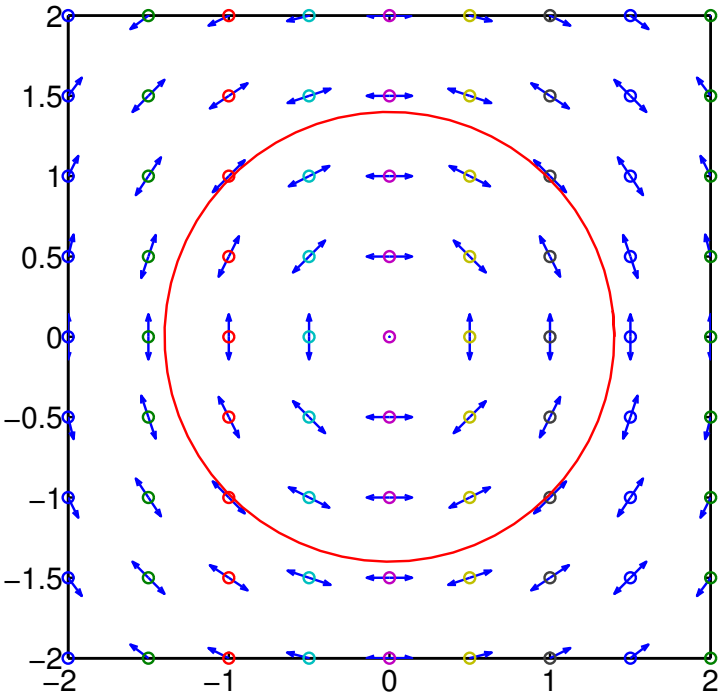

Abb.1.3 Richtungsfeld der Differentialgleichung $y' = -t/y$.

C2. Lineare DGL erster Ordnung.

Eine skalare, lineare DGL erster Ordnung hat nach (1.6) die Form

$$y'(t) + a(t)\, y(t) = b(t) \tag{1.25}$$

mit glatten (zumindest stetigen) Funktionen a und b.

Analog zur Lösungsdarstellung für lineare Gleichungssysteme kann man auch für die allgemeine Lösung y von (1.25) die Darstellung zeigen:

$$y(t) \;=\; y_p(t) \;+\; y_h(t). \tag{1.26}$$

Hierbei ist y_p eine (beliebige) *spezielle oder partikuläre Lösung* von (1.25) und y_h die *allgemeine Lösung der zugehörigen homogenen DGL* $y_h' + a\, y_h = 0$. Der Zusatz "allgemein" bedeutet hier und im Folgenden, dass sich jede Lösung der DGL in der angegebenen Form schreiben lässt.

(i) Die homogen DGL.

$y_h' + a\, y_h = 0$ lässt sich durch Trennung der Variablen lösen. Man erhält

$$y_h(t) \;=\; C\, \exp\left(-\int_{t_0}^{t} a(\tau)\, d\tau \right). \tag{1.27}$$

$C \in \mathbb{R}$ ist eine Integrationskonstante.
Die singuläre Lösung $y_h = 0$ ist in der Lösungsschar (1.27) mit $C = 0$ enthalten.

(ii) Eine partikuläre Lösung.

Eine partikuläre Lösung y_p lässt sich mit einem auf Joseph Louis Lagrange (1736 – 1813) zurückgehenden Ansatz erhalten. Man setzt

$$y_p(t) \;=\; C(t)\, \exp\left(-\int_{t_0}^{t} a(\tau)\, d\tau \right), \tag{1.28}$$

verwendet also die gleiche Lösungsformel wie für die homogene DGL, allerding mit zeitabhängigem $C(t)$. Der Lagrangesche Ansatz heißt daher auch *Variation der Konstanten.*

Setzen wir (1.28) in die inhomogene DGL ein, so folgt

$$C'(t)\, \exp\left(-\int_{t_0}^{t} a(\tau)\, d\tau \right) \;-\; a(t)\, y_p(t) \;+\; a(t)\, y_p(t) \;=\; b(t)\,.$$

Diese Gleichung lässt sich nach $C'(t)$ auflösen und hieraus lässt sich $C(t)$ durch eine Quadratur gewinnen

$$C(t) \;=\; \int_{t_0}^{t} b(\tau)\, \exp\left(\int_{t_0}^{\tau} a(\xi)\, d\xi \right)\, d\tau.$$

Insgesamt erhält man die partikuläre Lösung

$$y_p(t) \;=\; \int_{t_0}^{t} b(\tau)\, \exp\left(-\int_{\tau}^{t} a(\xi)\, d\xi \right)\, d\tau. \tag{1.29}$$

Bei festem (aber beliebigem) t_0 ist somit durch (1.26), (1.27) und (1.29) eine allgemeine Lösungdarstellung für die lineare DGL (1.26) gegeben.

(iii) Konstante Koeffizienten.

Im Fall eines konstanten Koeffizienten $a(t) = a = \text{const.}$ vereinfacht sich die Darstellung erheblich. Man erhält

$$y_h(t) \;=\; C\,\mathrm{e}^{-a(t-t_0)}, \qquad y_p(t) \;=\; \int_{t_0}^{t} \mathrm{e}^{-a(t-\tau)}\, b(\tau)\, d\tau. \tag{1.30}$$

Beispiel (1.31) (Newtonsche Abkühlung)

Die (räumlich gemittelte) Temperatur $T(t)$ eines homogenen Körpers lässt sich vereinfacht durch die folgende lineare Differentialgleichung beschreiben:

$$\frac{dT}{dt} \;=\; \frac{k\,F}{c\,m}\,(\,T_a(t) - T(t)). \tag{1.32}$$

Dabei bezeichnet $T(t)$ die Temperatur des Körpers zur Zeit t, $T_a(t)$ die Außentemperatur, m die Masse des Körpers, F die Oberfläche, c die spezifische Wärme und k einen Proportionalitätsfaktor.

Mit (1.30) erhält man die Lösungsdarstellung

$$T(t) \;=\; T(t_0)\,\mathrm{e}^{-\lambda(t-t_0)} + \lambda \int_{t_0}^{t} T_a(\tau)\,\mathrm{e}^{\lambda(\tau-t)} d\tau, \quad \lambda := (kF)/(cm).$$

Im Fall eines konstanten Koeffizienten a ist die Inhomogenität $b(t)$ mitunter von einer speziellen Form. In diesem Fall kann es vorteilhaft sein, eine partikuläre Lösung y_p mit einem *spezieller Ansatz* zu ermitteln. In der Tabelle 1.1 sind für einige Inhomogenitäten (polynomial, trigonometrisch, exponentiell) solche Ansätze angegeben.

Tabelle 1.1: Spezielle Ansätze für partikuläre Lösungen

$b(t)$	$y_p(t)$
$\displaystyle\sum_{k=0}^{m} b_k\, t^k$	$\displaystyle\sum_{k=0}^{m} C_k\, t^k$
$b_1\ \cos(\omega t) + b_2 \sin(\omega t)$	$C\ \sin(\omega t - \gamma)$
$b\,\mathrm{e}^{\lambda t}$	$C\,\mathrm{e}^{\lambda t}$, falls $\lambda \neq -a$ $C\,t\,\mathrm{e}^{\lambda t}$, falls $\lambda = -a$

D. Ebene autonome DGL.

Wir betrachten eine ebene ($n = 2$) autonome AWA

$$\begin{array}{rclcl} x'(t) & = & f(x(t), y(t)), & & x(0) = x_0 \\ y'(t) & = & g(x(t), y(t)), & & y(0) = y_0. \end{array} \tag{1.33}$$

Ist der Anfangswert (x_0, y_0) kein *Gleichgewichtspunkt* des DGLsystems, d.h. gilt $(f(x_0, y_0), g(x_0, y_0)) \neq 0$, so wird o.E.d.A. angenommen, dass $f(x_0, y_0) \neq 0$ gilt. (Anderfalls Vertauschung von x, y.)

Die Funktion x ist dann lokal bei $t = 0$ streng monoton und somit auch umkehrbar. Setzt man $Y(x) := y(t(x))$, so genügt Y der so genannten *Phasendifferentialgleichung*

$$f(x, Y)\, Y'(x) \; - \; g(x, Y) \;\; = \;\; 0. \tag{1.34}$$

Diese DGL erster Ordnung beschreibt die Gestalt der Kurve $(x(t), y(t))$, jedoch nicht ihre zeitliche Durchlaufung.

Beispiel (1.35) (Schwingungsgleichung)

Durch

$$x''(t) \;\; = \;\; -\omega^2\, x(t), \qquad \omega > 0, \tag{1.36}$$

ist eine lineare DGL zweiter Ordnung mit konstanten Koeffizienten gegeben.

Mit Hilfe der Standardtransformation $y(t) := x'(t)$ lässt sich diese DGL in ein äquivalentes System erster Ordnung transformieren:

$$\begin{array}{rclcl} x'(t) & = & y(t), & & x(0) = x_0 \\ y'(t) & = & -\omega^2\, x(t), & & y(0) = y_0. \end{array}$$

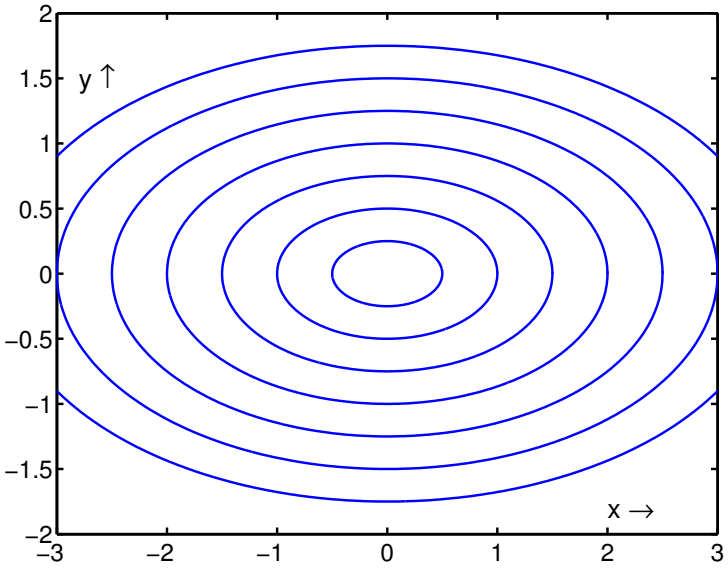

Abb. 1.4 Phasenkurven zur DGL aus Beispiel (1.35).

Die zugehörige Phasendifferentialgleichung lautet nun

$$Y(x)\,Y'(x) \;=\; -\omega^2\,x.$$

Sie besitzt die Lösung

$$Y^2 \,+\, \omega^2\,x^2 \;=\; y_0^2 \,+\, \omega^2\,x_0^2,$$

vgl. auch (1.24). Die Phasenkurven sind also Ellipsen, bei denen die Halbachsen a und b im festen Verhältnise $a : b \;=\; 1 : \omega$ stehen.

E. Skalare DGLen zweiter Ordnung.

(i) Autonome DGLn.

Die Technik aus dem Beispiel (1.35) lässt sich auf beliebige autonome DGL zweiter Ordnung anwenden

$$x''(t) \;=\; f(x, x'). \tag{1.37}$$

Vermöge $y(t) := x'(t)$ wird (1.37) in ein System erster Ordnung transformiert

$$x' \;=\; y, \quad y' \;=\; f(x, y),$$

für das die Phasen-DGL gegeben ist durch $Y'(x) \;=\; f(x, Y)/Y$. Ist hieraus $Y(x)$ bestimmbar, so lässt sich die Funktion $t = t(x)$ durch eine Quadratur ermitteln

$$\frac{d\,t}{d\,x} \;=\; \frac{1}{x'(t)} \;=\; \frac{1}{Y(x)} \quad\Longrightarrow\quad t - t_0 \;=\; \int_{x_0}^{x} \frac{d\,x}{Y(x)}.$$

Schließlich erhält man dann durch Bildung der Umkehrfunktion $x(t)$ und damit auch $y(t) = Y(x(t))$.

(ii) Erster Spezialfall.

Wir sehen uns im Folgenden zwei einfachere Spezialfälle einer Differentialgleichung zweiter Ordnung an. Zunächst:

$$x''(t) \;=\; f(t, x'). \tag{1.38}$$

Mit der Definition $y(t) := x'(t)$ erhält man das System erster Ordung

$$x' \;=\; y, \quad y' \;=\; f(t, y).$$

Dieses System ist nun aber *entkoppelt*, d.h. wir können die zweite Differentialgleichung zunächst (unabhängig von x) lösen und danach die Erste durch eine Quadratur.

Beispiel (1.39)

Die Gestalt einer Kette wird durch folgende DGL beschrieben

$$y''(x) \;=\; k\,\sqrt{1 \,+\, y'(x)^2}.$$

Mit $z(x) := y'(x)$ ergibt sich das entkoppelte System

$$y' \;=\; z, \quad z' \;=\; k\,\sqrt{1 \,+\, z^2}.$$

Die zweite DGL lässt sich mit Variablentrennung lösen: $z(x) = \sinh(kx + C_1)$. Hieraus liefert die erste DGL per Quadratur die so genannte *Kettenlinie*

$$y(x) \;=\; \frac{1}{k}\,\cosh(kx + C_1) \,+\, C_2.$$

Die Integrationskonstanten C_1 und C_2 werden durch den beiden Randbedingungen $y(a) = y_a$ und $y(b) = y_b$ (Aufhängung der Kette) festgelegt.

(iii) Zweiter Spezialfall.

Wir betrachten autonome DGLen der Form

$$x''(t) \;=\; f(x(t)). \tag{1.40}$$

Multiplizieren wir diese Gleichung mit x' und integrieren anschließend, so folgt:

$$\begin{aligned} x'\;x'' \quad &=\quad f(x)\,x' \\ \Rightarrow\quad \frac{1}{2}\;(x')^2 \quad &=\quad \int f(x)\,dx \;=:\; F(x) \,+\, C \\ \Rightarrow\quad\quad x' \quad &=\quad \pm\sqrt{2\,(F(x) \,+\, C)}. \end{aligned}$$

Nehmen wir an, dass $x' \neq 0$ ist, so lässt sich diese DGL wiederum durch Trennung der Variablen lösen. Wir erhalten

$$t \;=\; t(x) \;=\; \pm \int \frac{dx}{\sqrt{2\,(F(x) \,+\, C)}}.$$

Die Durchführung dieser Integration und die Invertierung der resultierenden Beziehung, d.h. die Auflösung nach x, liefert sodann die Lösung der DGL (1.40).

Beispiel (1.41) (Fluchtgeschwindigkeit einer Rakete)

Die Bewegung einer (antriebslosen) Rakete außerhalb der Erdatmosphäre ist durch das Gravitationsgesetz bestimmt. Vernachlässigt man den Einfluss anderer Himmelskörper, und nimmt man eine geradlinige, eindimensionale Bewegung an, so gilt für den Abstand Erde – Rakete die Differentialgleichung

$$\ddot{r}(t) \;=\; -\,\gamma\,M_E \cdot \frac{1}{r^2}, \quad r(0) \;=\; r_0, \quad \dot{r}(0) \;=\; v_0.$$

Dabei bezeichnet γ die Gravitationskonstante $(\gamma \doteq 6.67 \cdot 10^{-11}\mathrm{Nm^2kg^{-2}})$ und M_E die Masse der Erde $(M_E \doteq 5.95 \cdot 10^{24}\mathrm{kg})$.

Wir bestimmen die kleinste Anfangsgeschwindigkeit v_0, die die Rakete besitzen muss, um den Anziehungsbereich der Erde verlassen zu können, die so genannte *Fluchtgeschwindigkeit*.

Dazu multipliziert wir die Differentialgleichung mit $\dot{r}$ und integrieren:

$$\ddot{r}\,\dot{r} \;=\; -\,\gamma\,M_E\,\frac{\dot{r}}{r^2}$$

$$\Rightarrow \qquad \dot{r}^2 \;=\; \frac{2\,\gamma\,M_E}{r} \;+\; C, \quad C \;=\; \text{const.}$$

Hierin setzen wir die Anfangswerte $r(0) = r_0$ und $v(0) = v_0$ ein, und finden:

$$\dot{r}^2 \;=\; 2\,\gamma\,M_E\,\left(\frac{1}{r} \;-\; \frac{1}{r_0}\right) \;+\; v_0^2.$$

Die gesuchte Fluchtgeschwindigkeit v_0 ist nun die kleinste Anfangsgeschwindigkeit, für die $\dot{r}(t)$ stets positiv bleibt. Damit folgt $v_0 \;=\; \sqrt{2\,\gamma\,M_E/r_0}$.

Setzt man für r_0 den Erdradius ($r_0 \;\doteq\; 6.36 \cdot 10^6$ m) ein, so erhält man $v_0 \;\approx\; 11.2$ km/s.

2. Lineare Differentialgleichungen

A. Zeitabhängige Systeme erster Ordnung.

Wir betrachten explizite lineare DGLsysteme erster Ordnung

$$y'(t) \quad = \quad A(t)\, y(t) \; + \; b(t)\,. \tag{2.41}$$

wobei vorausgesetzt wird, dass die Koeffizientenmatrix $A : \mathbb{R} \to \mathbb{R}^{(n,n)}$ sowie die Inhomogenität $b : \mathbb{R} \to \mathbb{R}^n$ stetige Funktionen der Zeit $t \in \mathbb{R}$ sind.

Wie wir im nächten Abschnitt sehen werden, besitzt die zugehörige AWA mit Anfangswert $(t_0, y_0) \in \mathbb{R}^{n+1}$ stets eine eindeutig bestimmte Lösung, die für alle $t \in \mathbb{R}$ erklärt ist. Wir bezeichnen diese mit $y(t; t_0, y_0)$.

Aufgrund der Linearität gilt für die allgemeine Lösung von (2.1) analog zum skalaren Fall die folgenden Strukturaussage.

Satz (2.2)

Die allgemeine Lösung der linearen DGL (2.1) besitzt die Darstellung

$$y(t) \quad = \quad y_p(t) \; + \; y_h(t)\,. \tag{2.3}$$

Dabei ist y_p eine (beliebige) partikuläre Lösung der inhomogenen DGL und y_h die allgemeine Lösung der zugehörigen homogenen DGL $y' = A\, y$.

Beweis: Sind y_p und y_h wie oben gegeben, so ist $y := y_p + y_h$ offenbar eine Lösung der inhomogenen DGL. Umgekehrt: Sind y und y_p Lösungen der inhomogenen DGL, so erfüllt $y - y_p$ offensichtlich die homogene DGL. □

Die homogene Differentialgleichung.

Die Lösungen der homogenen linearen DGL

$$y'(t) \quad = \quad A(t)\, y(t) \tag{2.4}$$

bilden einen endlichdimensionalen linearen Teilraum des Vektorraums $\mathrm{C}^1(\mathbb{R}, \mathbb{R}^n)$ aller stetig differenzierbaren Funktionen $\mathbb{R} \to \mathbb{R}^n$. Zur Aufstellung der allgemeinen Lösung genügt es daher, eine Basis des Lösungsraumes zu ermitteln.

Eine solche Basis lässt sich folgendermaßen konstruieren:

a) Man wähle $t_0 \in \mathbb{R}$ und eine Basis $(v^1, \ldots, v^n)$ des $\mathbb{R}^n$.

b) Man löse die folgenden n AWA (für $k = 1, \ldots, n$):

$$\frac{d}{dt}\, y^k(t) \quad = \quad A(t)\, y^k(t), \qquad y^k(t_0) \;=\; v^k\,.$$

Die Lösungen $y^k(t)$, $k = 1, \ldots, n$, werden zu einer Matrix

$$Y(t) \;:=\; (y^1(t), \ldots, y^n(t)) \;\in\; \mathbb{R}^{(n,n)} \tag{2.5}$$

zusammengefasst. Diese heißt eine *Fundamentalmatrix* oder ein *Fundamentalsystem* der DGL (2.1) bzw. (2.4). Offenbar ist Y zugleich eine Lösung der Matrix–AWA

$$Y'(t) \;=\; A(t)\, Y(t), \qquad Y(t_0) \;=\; (v^1, \ldots, v^n)\,. \tag{2.6}$$

Im folgenden Satz zeigen wir, dass Y tatsächlich eine Basis des Lösungsraums bildet.

Satz (2.7)

Es sei $Y : \mathbb{R} \to \mathbb{R}^{(n,n)}$ ein beliebiges Fundamentalsystem. Dann gelten:

a) Die allgemeine Lösung der homogenen DGL lautet

$$y_h(t) \;=\; Y(t) \cdot c \;=\; \sum_{k=1}^{n} c_k\, y^k(t), \qquad c \in \mathbb{R}^n\,. \tag{2.8}$$

b) Die Fundamentalmatrix $Y(t)$ ist für alle $t \in \mathbb{R}$ regulär.

Beweis:
zu a) Nach Konstruktion ist $Y(t_0)$ regulär; ferner ist klar, dass $y(t) := Y(t)\, c$ für jedes $c \in \mathbb{R}^n$ eine Lösung der homogenen DGL ist.

Umgekehrt: Ist y^* eine Lösung der homogenen DGL, so setze man $c^* := Y(t_0)^{-1} y^*(t_0)$. Damit ist sowohl y^* als auch $y := Y(t)c^*$ eine Lösungen der AWA $y' = Ay$, $y(t_0) = y^*(t_0)$. Aufgrund der eindeutigen Lösbarkeit folgt $y^* = Y\, c^*$. Damit ist a) gezeigt.

zu b) Es bleibt zu zeigen, dass $Y(t)$ auch für $t = t_1 \neq t_0$ regulär ist. Dazu zeigen wir:

$$\forall\, y^1 \in \mathbb{R}^n :\; \exists\, c \in \mathbb{R}^n :\; Y(t_1)\, c \;=\; y^1\,.$$

Für vorgegebenes $t_1 \neq t_0$ und $y^1 \in \mathbb{R}^n$ besitzt die AWA $y' = A\, y$, $y(t_1) = y^1$ eine eindeutig bestimmte Lösung y. Nach a) existiert daher ein c mit $y(t) = Y(t)c$. Speziell für $t = t_1$ ergibt sich hiermit die Behauptung. □

Bemerkung (2.9) Die C^1–Funktion $W(t) := \det(Y(t))$ heißt die *Wronski–Determinante* des Fundamentalsystems Y, benannt nach Josef–Maria Hoene–Wronski (1778–1853).

Ohne Beweis sei erwähnt, dass W der folgenden linearen homogenen DGL genügt

$$W'(t) \quad = \quad \text{Spur}\,(A(t)) \cdot W(t). \tag{2.10}$$

Damit ergibt sich mit (1.27) die folgende Darstellung für die Wronski-Determinante

$$W(t) \quad = \quad W(t_0)\ \exp\left(\int_{t_0}^{t} \text{Spur}\,(A(\tau))\, d\tau \right) . \tag{2.11}$$

Die inhomogene Differentialgleichung.

Sei $Y(t)$ ein Fundamentalsystem der zugehörigen homogenen DGL, also

$$y_h(t) \quad = \quad Y(t)\, c, \qquad c \in \mathbb{R}^n\,.$$

Zur Bestimmung einer partikulären Lösung der inhomogenen DGL verwenden wir wie im skalaren Fall den Ansatz (*Variation der Konstanten*)

$$y(t) \quad := \quad Y(t)\, c(t)\,. \tag{2.12}$$

Differentiation ergibt

$$\begin{aligned} y'(t) &= Y'(t)\, c(t) \ + \ Y(t)\, c'(t) \\ &= A(t)\, Y(t)\, c(t) \ + \ Y(t)\, c'(t) \\ &= A(t)\, y(t) \ + \ Y(t)\, c'(t)\,. \end{aligned}$$

y löst also die inhomogene DGL, falls gilt:

$$Y(t)\, c'(t) \ = \ b(t), \quad \text{oder} \quad c(t) \ = \ c_0 \ + \ \int_{t_0}^{t} Y(\tau)^{-1}\, b(\tau)\, d\tau\,.$$

Insgesamt haben wir damit gezeigt:

Satz (2.13)

a) Die allgemeine Lösung der inhomogenen DGL gegeben durch

$$y(t) \ = \ Y(t) \left[c_0 \ + \ \int_{t_0}^{t} Y(\tau)^{-1}\, b(\tau)\, d\tau \right] , \qquad c_0 \in \mathbb{R}^n\,.$$

b) Für $c_0 := Y(t_0)^{-1}\, y_0$ erfüllt y die Anfangsbedingung $y(t_0) = y_0$.

B. Systeme mit konstanten Koeffizienten.

Wir betrachten ein homogenes DGLsystem mit konstanter Koeffizientenmatrix

$$y'(t) \quad = \quad A\ y(t), \qquad A\ \in\ \mathbb{R}^{(n,n)}\,. \tag{2.14}$$

Zur Bestimmung eines Fundamentalsystems verwenden wir – wiederum in Analogie zum eindimensionalen Fall – den *Ansatz*

$$y(t) \quad = \quad \mathrm{e}^{\lambda t}\,v, \qquad \lambda \in \mathbb{R}/\mathbb{C}, \qquad v \in \mathbb{R}^n/\mathbb{C}^n\,. \tag{2.15}$$

Setzt man diesen Ansatz in die DGL ein, so folgt

$$y' \quad = \quad A\ y \qquad \Leftrightarrow \qquad A\,v \quad = \quad \lambda\,v\,,$$

d.h., durch (2.15) ist genau dann eine nichttriviale Lösung der DGL gegeben, falls λ ein Eigenwert von A und v ein zugehöriger Eigenvektor ist.

Fall 1: Alle Eigenwerte von A sind reell und es existiert eine Basis aus Eigenvektoren.

In diesem Fall ist durch

$$Y(t) \quad = \quad \left(\mathrm{e}^{\lambda_1 t}\,v^1, \ldots, \mathrm{e}^{\lambda_n t}\,v^n\right) \tag{2.16}$$

ein (reelles) Fundamentalsystem der homogenen DGL gegeben und die allgemeine Lösung lautet somit:

$$y_h(t) \quad = \quad \sum_{k=1}^{n} C_k\,\mathrm{e}^{\lambda_k t}\,v^k, \qquad C_k \in \mathbb{R}. \tag{2.17}$$

Fall 2: A ist diagonalisierbar.

Es gibt dann eine Basis von $\mathbb{C}^n$ aus Eigenvektoren $v^1, \ldots, v^n$. Die zugehörigen Eigenwerte $\lambda_1, \ldots, \lambda_n$ müssen dabei aber weder einfach noch reell sein. Dieser Fall trifft für alle normalen, insbesondere also auch für alle symmetrischen Matrizen zu.

Wie im ersten Fall (allerdings mit Rechnung in $\mathbb{C}$ statt in $\mathbb{R}$) lautet die allgemeine Lösung der homogenen DGL

$$y_h(t) \quad = \quad \sum_{k=1}^{n} C_k\,\mathrm{e}^{\lambda_k t}\,v^k, \qquad C_k \in \mathbb{C}. \tag{2.18}$$

Wir sind jedoch daran interessiert, ein *reelles* Fundamentalsystem zu finden. Hierzu beachten wir, dass mit $\lambda \in \mathbb{C} \setminus \mathbb{R}$ auch stets der konjugiert komplexe Wert $\overline{\lambda}$ ein Eigenwert der *reellen* Matrix A ist und ferner mit v (Eigenvektor zum Eigenwert λ) auch $\overline{v}$ ein Eigenvektor (zum Eigenwert $\overline{\lambda}$) ist.

Nichtreelle Eigenwerte und Eigenvektoren treten also stets paarweise auf, und man erhält die zugehörigen reellen Lösungen gemäß

$$\begin{aligned} y^1(t) \;&=\; \mathrm{Re}\,\left(\mathrm{e}^{\lambda t}\,v\right) \;=\; \frac{1}{2}\,\left(\mathrm{e}^{\lambda t}\,v + \mathrm{e}^{\overline{\lambda} t}\,\overline{v}\right) \\ y^2(t) \;&=\; \mathrm{Im}\,\left(\mathrm{e}^{\lambda t}\,v\right) \;=\; \frac{1}{2i}\,\left(\mathrm{e}^{\lambda t}\,v - \mathrm{e}^{\overline{\lambda} t}\,\overline{v}\right)\,. \end{aligned} \tag{2.19}$$

Beispiel (2.20)

Für das DGLsystem $\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

erhält man die folgenden Eigenwerte und Eigenvektoren

$$\lambda_1 = 1+2i\,, \quad v^1 = \begin{pmatrix} 1 \\ -2i \end{pmatrix}, \quad \lambda_2 = 1-2i\,, \quad v^2 = \begin{pmatrix} 1 \\ 2i \end{pmatrix}.$$

Ein komplexes Fundamentalsystem ist daher gegeben durch

$$z^1(t) = \mathrm{e}^{(1+2i)t} \begin{pmatrix} 1 \\ -2i \end{pmatrix}, \qquad z^2(t) = \mathrm{e}^{(1-2i)t} \begin{pmatrix} 1 \\ 2i \end{pmatrix}.$$

Die Umrechnung in ein reelles Fundamentalsystem ergibt

$$y^1(t) = \mathrm{e}^t \begin{pmatrix} \cos(2t) \\ 2\sin(2t) \end{pmatrix}, \qquad y^2(t) = \mathrm{e}^t \begin{pmatrix} \sin(2t) \\ -2\cos(2t) \end{pmatrix}.$$

Schließlich hat man die allgemeine (reelle) Lösung:

$$y_h(t) = C_1\,\mathrm{e}^t \begin{pmatrix} \cos(2t) \\ 2\sin(2t) \end{pmatrix} + C_2\,\mathrm{e}^t \begin{pmatrix} \sin(2t) \\ -2\cos(2t) \end{pmatrix}, \qquad C_1,\, C_2 \in \mathbb{R}.$$

Fall 3: A ist nicht diagonalisierbar.

In diesem Fall ermittelt man die Jordansche Normalform J der Matrix A einschließlich einer zugehörigen Transformationsmatrix S, die A auf Jordansche Normalform transformiert. Es gelte also:

$$\begin{aligned}
J &= S^{-1}\,A\,S \\
J &= \begin{pmatrix} J_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & J_m \end{pmatrix}; \quad J_j \in \mathbb{C}^{(r_j,r_j)} : \text{ Jordan–Kästchen} \\
S &= (v^{11},\ldots,v^{1r_1}\,|\,v^{21},\ldots,v^{2r_2}|\,\ldots\,|\,v^{m1},\ldots,v^{mr_m}) \qquad (2.21) \\
v^{j1} &: \text{ Eigenvektor zum Eigenwert } \lambda_j\,,\ j=1,\ldots,m \\
v^{jk} &: \text{ Hauptvektor der Stufe } (k-1), \quad k=2,\ldots,r_j \\
(A-\lambda_j\,I_n)\,v^{j,k} &= v^{j,k-1}, \quad k=2,\ldots,r_j \quad (\textit{Kettenbedingung}).
\end{aligned}$$

Setzt man nun $z(t) := S^{-1}\,y(t) \in \mathbb{R}^n$, so ergibt sich für z die DGL

$$z'(t) = S^{-1}\,y'(t) = S^{-1}\,A\,y(t) = S^{-1}\,A\,S\;z(t),$$

also $\quad z'(t) \;=\; J\,z(t)\;.$

Kennt man nun ein Fundamentalsystem $Z(t)$ der transformierten DGL $z' = J\,z,$ so erhält man hieraus ein Fundamentalsystem für die vorgegebene DGL durch Rücktransformation $Y(t) := S\,Z(t)$.

Das transformierte DGLsystem $z' = J\,z$ zerfällt in die einzelnen Jordan–Blöcke. Es genügt daher, die zu einem einzelnen Jordan–Kästchen (o.E. dem ersten) gehörigen DGLen zu betrachten

$$\frac{d}{dt}\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{pmatrix} \;=\; \begin{pmatrix} \lambda_1 & 1 & & 0 \\ & \lambda_1 & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_r \end{pmatrix}\;. \tag{2.22}$$

(2.22) ist ein gestaffeltes System linearer inhomogenen DGLen, die - beginnend bei der letzten Gleichung für z_r - rekursiv für $k = r, \ldots, 1$ mittels Variation der Konstanten gelöst werden können. Man erhält so das folgende Fundamentalsystem (in $\mathbb{C}^r$) für (2.22); hierbei sind nur die zu diesem Jordan–Kästchen gehörigen Koordinaten $z_1, \ldots, z_r$ angegeben (die anderen Koordinaten sind jeweils Null zu setzen):

$$\mathrm{e}^{\lambda_1 t}\begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{pmatrix},\; \mathrm{e}^{\lambda_1 t}\begin{pmatrix} t/1! \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix},\; \mathrm{e}^{\lambda_1 t}\begin{pmatrix} t^2/2! \\ t/1! \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\; \ldots,\; \mathrm{e}^{\lambda_1 t}\begin{pmatrix} t^{r-1}/(r-1)! \\ \vdots \\ \vdots \\ \vdots \\ t/1! \\ 1 \end{pmatrix}\;. \tag{2.23}$$

Ist nun $(v^{11}, \ldots, v^{1r})$ ein zugehöriges System aus Eigenvektor v^{11} und Hauptvektoren v^{12}, …, v^{1r} in $\mathbb{C}^n$, so liefert die Rücktransformation den folgenden Anteil für das Fundamentalsystem der Ausgangsgleichung $y' = A\,y$:

$$\begin{aligned} y^{11}(t) &= \mathrm{e}^{\lambda_1 t}\,v^{11} \\ y^{12}(t) &= \mathrm{e}^{\lambda_1 t}\left[\frac{t}{1!}\,v^{11} + v^{12}\right] \\ &\vdots \\ y^{1r}(t) &= \mathrm{e}^{\lambda_1 t}\left[\frac{t^{r-1}}{(r-1)!}\,v^{11} + \ldots + \tfrac{t}{1!}\,v^{1,r-1} + v^{1r}\right]\;. \end{aligned} \tag{2.24}$$

Behandelt man nun alle Jordan–Kästchen auf diese Weise, so erhält man insgesamt ein Fundamentalsystem für die DGL (2.14).

Beispiel (2.25)

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 0 & -1 & -1 \\ 0 & 4 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Für das charakteristische Polynom der Koeffizientenmatrix ergibt sich:

$$p_A(\lambda) = \det(A - \lambda I_3) = (1-\lambda)^3,$$

$\lambda = 1$ ist also dreifacher Eigenwert.

Eigenvektoren:

$$\left(\begin{array}{ccc|c} 0 & -2 & 1 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 4 & 2 & 0 \end{array}\right) \Rightarrow v^1 = \begin{pmatrix} 16 \\ 0 \\ 0 \end{pmatrix}.$$

Der zu $\lambda = 1$ gehörige Eigenraum ist eindimensional, die geometrische Vielfachheit des Eigenwerts also $g_A(\lambda) = 1$.

Hauptvektoren:

$$\left(\begin{array}{ccc|c} 0 & -2 & 1 & 16 \\ 0 & -2 & -1 & 0 \\ 0 & 4 & 2 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 0 & 0 & 2 & 16 \\ 0 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow v^2 = \begin{pmatrix} 0 \\ -4 \\ 8 \end{pmatrix}$$

$$\left(\begin{array}{ccc|c} 0 & -2 & 1 & 0 \\ 0 & -2 & -1 & -4 \\ 0 & 4 & 2 & 8 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 0 & 0 & 2 & 4 \\ 0 & -2 & -1 & -4 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow v^3 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}.$$

Damit erhält man das folgende Fundamentalsystem:

$$y^1(t) = \mathrm{e}^t \begin{pmatrix} 16 \\ 0 \\ 0 \end{pmatrix}, \; y^2(t) = \mathrm{e}^t \begin{pmatrix} 16t \\ -4 \\ 8 \end{pmatrix}, \; y^3(t) = \mathrm{e}^t \begin{pmatrix} 8t^2 \\ -4t+1 \\ 8t+2 \end{pmatrix},$$

und die allgemeine Lösung lautet:

$$y_h(t) = C_1\, y^1(t) + C_2\, y^2(t) + C_2\, y^3(t), \qquad C_k \in \mathbb{R}.$$

Beispiel (2.26)

$$\begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Wieder ist $\lambda = 1$ dreifacher Eigenwert der Koeffizientenmatrix A, allerdings mit der geometrischen Vielfachheit $g_A(\lambda) = 2$.

Eigenvektoren:

$$\left(\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow v^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Hauptvektor:

Es gilt: $(A - \lambda I_3)^2 = 0$. Gesucht ist daher ein von v^1, v^2 linear unabhängiger Vektor v^{22}. Wählt man etwa $v^{22} = (0,0,1)^{\mathrm{T}}$, so folgt mit der Kettenbedingung

$$v^{21} = (A - \lambda I_3)\, v^{22} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} .$$

Man hat damit das folgende System von Eigen– bzw. Hauptvektoren

$$v^{11} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \quad v^{21} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} , \quad v^{22} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

Hiermit bestätigt man: $S^{-1} A S = J$ mit

$$S = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} , \qquad J = \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 1 & 1 \\ 0 & 0 & 1 \end{array}\right) .$$

Ein Fundamentalsystem der Differentialgleichung lautet somit:

$$y^1(t) = \mathrm{e}^t \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} , \quad y^2(t) = \mathrm{e}^t \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} , \quad y^3(t) = \mathrm{e}^t \begin{pmatrix} t \\ t \\ 1 \end{pmatrix} .$$

Beispiel (2.27)

Betrachtet werden zwei ungedämpft gekoppelte Pendel. Sind x, y die Ausschläge der Pendel aus der Ruhelage (Winkel), so gelten unter vereinfachten Annahmen die folgenden Differentialgleichungen:

$$m\ddot{x} = -\frac{mg}{\ell}\, x - k(x - y)$$

$$m\ddot{y} = -\frac{mg}{\ell}\, y - k(y - x) .$$

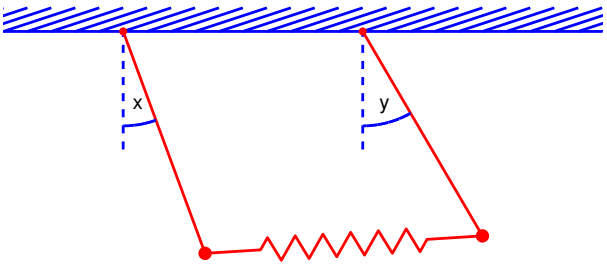

Abb. 2.1. Gekoppelte Pendel

Mit der üblichen Transformation $p := \dot{x},\ \ q := \dot{y}$ erhält man das folgende homogene Differentialgleichungssystem erster Ordnung:

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ p \\ q \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -(\omega_0^2 + k_0) & k_0 & 0 & 0 \\ k_0 & -(\omega_0^2 + k_0) & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ p \\ q \end{pmatrix}$$

mit $\omega_0 := \sqrt{g/\ell}\,,\ \ k_0 := k/m.$

Eigenwerte:

$$\lambda_{1,2} \;=\; \pm\, i\,\omega_0, \qquad \lambda_{3,4} \;=\; \pm\, i\,\sqrt{\omega_0^2 + 2k_0}$$

Eigenvektoren:

$$v^1 \;=\; \begin{pmatrix} 1 \\ 1 \\ i\,\omega_0 \\ i\,\omega_0 \end{pmatrix}, \quad v^2 \;=\; \begin{pmatrix} 1 \\ 1 \\ -i\,\omega_0 \\ -i\,\omega_0 \end{pmatrix}, \quad v^3 \;=\; \begin{pmatrix} 1 \\ -1 \\ i\,\omega \\ -i\,\omega \end{pmatrix}, \quad v^4 \;=\; \begin{pmatrix} 1 \\ -1 \\ -i\,\omega \\ i\,\omega 0 \end{pmatrix}$$

mit $\quad \omega := \sqrt{\omega_0^2 + 2k_0}.$

Hieraus erhält man nun das folgende reelle Fundamentalsystem:

$$y^1(t) \;=\; \mathrm{Re}\ \left(\mathrm{e}^{i\,\omega_0\,t}\,v^1\right) \;=\; \begin{pmatrix} \cos(\omega_0 t) \\ \cos(\omega_0 t) \\ -\omega_0 \sin(\omega_0 t) \\ -\omega_0 \sin(\omega_0 t) \end{pmatrix}$$

$$y^2(t) \;=\; \mathrm{Im}\ \left(\mathrm{e}^{i\,\omega_0\,t}\,v^1\right) \;=\; \begin{pmatrix} \sin(\omega_0\,t) \\ \sin(\omega_0\,t) \\ \omega_0 \cos(\omega_0\,t) \\ \omega_0 \cos(\omega_0\,t) \end{pmatrix}$$

$$y^3(t) \;=\; \mathrm{Re}\ \left(\mathrm{e}^{i\,\omega\,t}\,v^3\right) \;=\; \begin{pmatrix} \cos(\omega\,t) \\ -\cos(\omega\,t) \\ -\omega \sin(\omega\,t) \\ \omega \sin(\omega\,t) \end{pmatrix}$$

$$y^4(t) \;=\; \mathrm{Im}\ \left(\mathrm{e}^{i\,\omega\,t}\,v^3\right) \;=\; \begin{pmatrix} \sin(\omega\,t) \\ -\sin(\omega\,t) \\ \omega \cos(\omega\,t) \\ -\omega \cos(\omega\,t) \end{pmatrix}.$$

Die ersten beiden Fundamentallösungen beschreiben parallele Schwingungszustände der Pendel, die letzten beiden Lösungen beschreiben genau entgegengesetzt schwingende Pendel.

C. Lineare Differentialgleichungen höherer Ordnung.

Wir betrachten eine skalare lineare DGL n–ter Ordnung:

$$L\,[y] \;:=\; y^{(n)}(t) \;+\; a_{n-1}(t)\,y^{(n-1)}(t) \;+\; \ldots \;+\; a_0(t)\,y(t) \;=\; b(t)\,. \tag{2.28}$$

Wir sagen auch:

$$L \quad := \quad \sum_{k=0}^{n} a_k(t)\,\frac{d^k}{dt^k}\,, \qquad a_n \;\equiv\; 1\,, \tag{2.29}$$

ist ein *linearer Differentialoperator der Ordnung n.*

Die $a_k(t),\ k = 0, 1, \ldots, n-1$, seien stetige Funktionen auf $\mathbb{R}$.

Vermöge der Definition $y_k(t) := y^{(k-1)}(t),\ k = 1, \ldots, n,$ lässt sich die DGL (2.28) in ein äquivalentes DGLsystem erster Ordnung transformieren. Man erhält:

$$\frac{d}{dt}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & 0 & 1 \\ -a_0 & -a_1 & \ldots & \ldots & -a_{n-1} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}. \tag{2.30}$$

Die Ergebnisse aus den Abschnitten A. und B. lassen sich daher unmittelbar auf den Fall einer skalaren linearen DGL n–ter Ordnung übertragen.

Nachfolgend geben wir die wesentlichen Resultate an, verzichten jedoch weitgehend auf eigene Beweise.

Die homogene Differentialgleichung.

Ein Funktionensystem $(y_1, \ldots, y_n),\ y_k \in C^1(\mathbb{R}),$ heißt ein *Fundamentalsystem* der DGL $L[y] = h,$ falls

a) y_k löst die homogene DGL, $L\,[y_k] \;=\; 0,\ k = 1, \ldots, n.$

b) Die *Wronski–Determinante* verschwindet nicht

$$W(t) \quad := \quad \det \begin{pmatrix} y_1 & \ldots & y_n \\ y_1' & \ldots & y_n' \\ \vdots & & \vdots \\ y_1^{(n-1)} & \ldots & y_n^{(n-1)} \end{pmatrix} \;\neq\; 0. \tag{2.31}$$

W genügt nach (2.10) der DGL $W'(t) = -a_{n-1}(t)\cdot W(t)$ und besitzt damit die Darstellung

$$W(t) \;=\; W(t_0)\cdot \exp\left(-\int_{t_0}^{t} a_{n-1}(\tau)\,d\tau\right). \tag{2.32}$$

Ist W also an einer Stelle t_0 von Null verschieden, so verschwindet W nirgends.

Ein Fundamentalsystem $(y_1, \ldots, y_n)$ lässt sich durch Lösung der folgenden n AWAen $(k = 1, \ldots, n)$ gewinnen:

$$\begin{aligned} L[y_k] &= 0 \\ y_k^{(i)}(t_0) &= \begin{cases} 0, & i \neq k-1 \\ 1, & i = k-1 \end{cases} \qquad (i = 0, 1, \ldots, n-1)\,. \end{aligned} \tag{2.33}$$

Ist $(y_1, \ldots, y_n)$ ein Fundamentalsystem, so lautet die allgemeine Lösung der inhomogenen linearen DGL (2.28):

$$y(t) = y_p(t) + \sum_{k=1}^{n} C_k\, y_k(t), \qquad C_k \in \mathbb{R}\,; \tag{2.34}$$

dabei ist $y_p(t)$ eine partikuläre Lösng der inhomogenen DGL.

Die inhomogene Differentialgleichung.

Sei $(y_1, \ldots, y_n)$ ein Fundamentalsystem. Analog zu (2.12) verwenden wir den Ansatz der *Variation der Konstanten*

$$y_p(t) = \sum_{i=1}^{n} C_i(t)\, y_i(t)\,. \tag{2.35}$$

Für die unbekannten Funktionen $C_i(t)$ fordern wir:

$$\begin{array}{lcccl} C_1'(t)\, y_1(t) & + \ldots + & C_n'(t)\, y_n(t) & = & 0 \\ C_1'(t)\, y_1'(t) & + \ldots + & C_n'(t)\, y_n'(t) & = & 0 \\ \vdots & & \vdots & \vdots & \\ C_1'(t)\, y_1^{(n-2)}(t) & + \ldots + & C_n'(t)\, y_n^{(n-2)}(t) & = & 0\,. \end{array} \tag{2.36}$$

Damit folgt:

$$\begin{aligned} y^{(k)}(t) &= \sum_{i=1}^{n} C_i(t)\, y_i^{(k)}(t), \qquad k = 0, 1, \ldots, n-1 \\ y^{(n)}(t) &= \sum_{i=1}^{n} C_i'(t)\, y_i^{(n-1)}(t) + \sum_{i=1}^{n} C_i(t)\, y_i^{(n)}(t)\,, \end{aligned}$$

und somit

$$\begin{aligned} L[y] &= \sum_{k=0}^{n} a_k(t)\, y^{(k)}(t) \\ &= \sum_{i=1}^{n} C_i(t) \underbrace{\left(\sum_{k=0}^{n} a_k(t)\, y_i^{(k)}(t) \right)}_{=\,0} + \sum_{i=1}^{n} C_i'(t)\, y^{(n-1)}(t) = b(t)\,. \end{aligned}$$

Zusammen mit (2.35) ergibt sich das folgende lineare Gleichungssystem für die $C_i' = C_i'(t)$, $i = 1, \ldots, n$:

$$\begin{pmatrix} y_1^{(0)} & \ldots & y_n^{(0)} \\ y_1^{(1)} & \ldots & y_n^{(1)} \\ \vdots & & \vdots \\ y_1^{(n-1)} & \ldots & y_n^{(n-1)} \end{pmatrix} \begin{pmatrix} C_1' \\ C_2' \\ \vdots \\ C_n' \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix} . \tag{2.37}$$

Die Koeffizientenmatrix ist regulär, vgl. (2.31). Somit ist das obige Gleichungssystem eindeutig lösbar.

Durch Integration der Lösung erhält man die $C_1, \ldots, C_n$. Hierbei genügt es, eine *beliebige* Stammfunktion der C_i' zu bestimmen. Mit (2.34) hat man dann eine partikuläre Lösung gefunden.

Lineare DGL mit konstanten Koeffizienten.

Gegeben sei eine homogene lineare DGL

$$L\,[y] \;=\; \sum_{k=0}^{n} a_k\, y^{(k)}(t) \;=\; 0 \tag{2.38}$$

mit konstanten Koeffizienten $a_k \in \mathbb{R}$, $k = 0, 1, \ldots, n-1$ und $a_n = 1$.

Wie im eindimensionalen Fall verwenden wir den *Ansatz* $y(t) := \mathrm{e}^{\lambda t}$. Es folgt:

$$L[y] \;=\; \left(\sum_{k=0}^{n} a_k\, \lambda^k \right) \mathrm{e}^{\lambda t} \;=\; 0.$$

Damit ist y genau dann Lösung der homogenen DGL, wenn λ eine Nullstelle der *charakteristischen Gleichung* ist

$$p(\lambda) \quad := \quad \sum_{k=0}^{n} a_k\, \lambda^k \quad = \quad 0\,. \tag{2.39}$$

Sind $\lambda_1, \ldots, \lambda_m$ die (paarweise verschiedenen) Nullstellen der charakteristischen Gleichung, so gelten folgende Eigenschaften.

Satz (2.40)

a) Ist λ_k eine r_k–fache reelle Wurzel, so hat man die folgenden Lösungen der homogenen Gleichung:

$$\begin{aligned} y_{k1}(t) &:= \mathrm{e}^{\lambda_k t} \\ y_{k2}(t) &:= t \cdot \mathrm{e}^{\lambda_k t} \\ &\vdots \\ y_{k,r_k}(t) &:= t^{r_k - 1} \cdot \mathrm{e}^{\lambda_k t}\,. \end{aligned}$$

b) Ist λ_k eine r_k–fache komplexe Wurzel, $\lambda_k \notin \mathbb{R}$, so ist auch $\overline{\lambda}_k = \lambda_\ell$ eine weitere r_k–fache Wurzel. Reelle Lösungen sind dann:

$$\begin{aligned} y_{kj}(t) &= t^{j-1}\,\mathrm{e}^{\alpha_k t}\,\cos(\beta_k t) \\ y_{\ell j}(t) &= t^{j-1}\,\mathrm{e}^{\alpha_k t}\,\sin(\beta_k t) \\ \text{für} \quad & j = 1, \ldots, r_k, \quad \lambda_k = \alpha_k + i\beta_k\,. \end{aligned}$$

c) Die gemäß a) und b) konstruierten Lösungen bilden ein Fundamentalsystem von $L\,[y] = 0$.

Beispiel (2.41)

$$y^{(4)} + 2\,y'' + y = 0.$$

Die charakteristische Gleichung $p(\lambda) = \lambda^4 + 2\,\lambda^2 + 1 = 0$ hat die Nullstellen $\lambda_{1,2} = i,\ \lambda_{3,4} = -i$.

Ein Fundamentalsystem lautet damit:

$$\begin{aligned} y_1(t) &= \cos t, \quad & y_3(t) &= t \cdot \cos t \\ y_2(t) &= \sin t, \quad & y_4(t) &= t \cdot \sin t\,. \end{aligned}$$

Beispiel (2.42)

$$y'' - 2\,y' + y = \frac{\mathrm{e}^t}{t^2}$$

a) Für die *homogene* DGL hat man die charakteristische Gleichung

$$p(\lambda) = \lambda^2 - 2\,\lambda + 1 = 0$$

mit der doppelten Wurzel $\lambda_{1,2} = 1$.

Die allgemeine Lösung des homogenen Systems lautet damit:

$$y_h(t) = C_1\,\mathrm{e}^t + C_2\,t\,\mathrm{e}^t\,.$$

b) Für die *inhomogene* DGL findet man mittels Variation der Konstanten

$$\begin{aligned} C_1'\,\mathrm{e}^t &+ C_2'\,t\,\mathrm{e}^t &&= 0 \\ C_1'\,\mathrm{e}^t &+ C_2'\,(1+t)\,\mathrm{e}^t &&= \mathrm{e}^t/t^2 \end{aligned}$$

mit der Lösung: $C_1' = -\dfrac{1}{t},\ C_2' = \dfrac{1}{t^2}$, also: $C_1 = -\ln|t|,\ C_2 = -\dfrac{1}{t}$.

Eine partikuläre Lösung lautet damit

$$y_p(t) = -(\ln|t| + 1)\,\mathrm{e}^t\,.$$

Spezielle Ansätze (2.43).

Hat die Inhomogenität die spezielle Form $b(t) = \mathrm{e}^{\mu t} \sum_{j=0}^{m} \beta_j t^j$ so lässt sich anstelle der Variation der Konstanten der folgende Ansatz verwenden.

a) Falls μ keine Nullstelle des charakteristischen Polynoms $p(\lambda)$ ist, setze man $y_p(t) = \mathrm{e}^{\mu t} \sum_{j=0}^{m} \gamma_j t^j$, γ_j : Parameter,

b) falls μ eine r–fache Nullstelle von $p(\lambda)$ ist, $y_p(t) = \mathrm{e}^{\mu t} t^r \sum_{j=0}^{m} \gamma_j t^j$.

Beispiel (2.44) $y'' - y = t\,\mathrm{e}^t$.

Hier ist $\mu = 1$ eine einfache Nullstelle des charakteristischen Polynoms $p(\lambda) = \lambda^2 - 1$. Wir verwenden daher den Ansatz:

$$y_p(t) = \mathrm{e}^t (\gamma_0 t + \gamma_1 t^2) .$$

Setzt man diesen in die DGL ein, so folgt mittels Koeffizientenvergleich

$$\gamma_1 = -\gamma_0 = \frac{1}{4}, \quad \text{also:} \quad y_p(t) = \frac{t}{4}(t-1)\,\mathrm{e}^t .$$

Das Superpositionsprinzip (2.45).

Ist die Inhomogenität einer lineare DGL von der Form

$$L[y] = b(t) = b_1(t) + b_2(t)$$

und sind y_1 und y_2 partikuläre Lösungen der DGL $L[y] = b_k$, $k = 1, 2$, so ist $y_p(t) := y_1(t) + y_2(t)$ eine partikuläre Lösung von $L[y] = b$.

Komplexe Differentialgleichungen (2.46).

Ist die Inhomogenität b Real– oder Imaginärteil einer komplexwertigen Funktion, also $b(t) = \mathrm{Re}\,(c(t))$ bzw. $b(t) = \mathrm{Im}\,(c(t))$, und ist z (komplexe) Lösung der DGL $L[y] = c$, so ist $y := \mathrm{Re}\, z$ bzw. $y := \mathrm{Im}\, z$ eine (reelle) Lösung der DGL $L[y] = b$.

Beispiel (2.47). $y'' + 2y' + 5y = \mathrm{e}^{-t}(\cos t + \sin(2t))$.

Wir wenden das Superpositionsprinzip an und lösen zunächst:

a) $$y'' + 2y' + 5y \;=\; \mathrm{e}^{-t}\cos t \;=\; \mathrm{Re}\,\left\{\mathrm{e}^{(-1+i)t}\right\}.$$

$\mu = -1+i$ löst nicht die charakteristische Gleichung $p(\lambda) = \lambda^2+2\lambda+5 = 0$; daher verwenden wir den Ansatz $z_p(t) = C_0\mathrm{e}^{(-1+i)t}$. Diesen in die Differentialgleichung $z'' + 2z' + 5z = \mathrm{e}^{(-1+i)t}$ eingesetzt, liefert $C_0 = 1/3$.

Man hat also

$$z_p(t) \;=\; \frac{1}{3}\,\mathrm{e}^{(-1+i)t}\,, \qquad y_{p1}(t) \;=\; \frac{1}{3}\,\mathrm{e}^{-t}\cos t\,.$$

b) $$y'' + 2y' + 5y \;=\; \mathrm{e}^{-t}\sin(2t) \;=\; \mathrm{Im}\,\left\{\mathrm{e}^{(-1+2i)t}\right\}.$$

$\mu = -1+2i$ ist hierbei eine einfache Nullstelle von $p(\lambda)$; wir verwenden also den Ansatz $z_p = C_0\,t\,\mathrm{e}^{(-1+2i)t}$.

Diesen in die komplexe Differentialgleichung $z'' + 2z' + 5z = \mathrm{e}^{(-1+2i)t}$ eingesetzt, liefert: $C_0 = -i/4$, also $z_p = -i/4\;t\,\mathrm{e}^{(-1+2i)t}$. Damit lautet eine partikuläre Lösung der zweiten Differentialgleichung:

$$y_{p2}(t) \quad = \quad -\frac{t}{4}\,\mathrm{e}^{-t}\cos(2t)\,.$$

Das Superpositionsprinzip liefert damit die folgende partikuläre Lösung für die Ausgangsgleichung

$$y_p(t) \quad = \quad \mathrm{e}^{-t}\left(\frac{1}{3}\cos t - \frac{1}{4}\,t\,\cos(2t)\right)\,.$$

3. Existenz, Eindeutigkeit und Stabilität bei Anfangswertaufgaben

Wir beschäftigen uns in diesem Abschnitt mit der Lösungstheorie, d.h. den Fragen der Existenz und Eindeutigkeit einer Lösung und deren Abhängigkeit von Parametern für AWA der Form

$$y'(t) \;=\; f(t, y(t)), \quad y(t_0) \;=\; y_0. \tag{3.40}$$

Hierbei ist die rechte Seite des DGLsystems (3.1) eine Funktion $f : I \times D \to \mathbb{R}^n$, wobei $I \subset \mathbb{R}$ ein offenes Intervall und $D \subset \mathbb{R}^n$ eine offene Menge ist. Ferner sei natürlich $(t_0, y_0) \in I \times D$.

Schon in sehr einfachen Fällen lässt sich eine Lösung von (3.1) nicht explizit durch elementare Funktionen beschreiben. Daher sind Aussagen von Interesse, die unter möglichst allgemeinen Voraussetzungen an die rechte Seite f die Existenz, Eindeutigkeit und Stabilität einer Lösung garantieren.

Im Einzelnen interessieren uns hierbei die folgenden Fragen:

- Existiert eine Lösung y in einer Umgebung $|t - t_0| < \varepsilon$ der Anfangszeit? (Lokale Existenz?)
- Ist diese eindeutig bestimmt?
- Wie weit lässt sich eine solche Lösung fortsetzen? (Globale Existenz?)
- Wie verändert sich die Lösung bei Störung der Anfangsdaten (t_0, y_0) oder der rechten Seite f?

Zunächst sei ein kurzer historischer Rückblick gegeben:

Auf Augustin Louis Cauchy (1789 – 1857) geht ein Satz zurück, der die lokale Existenz und Eindeutigkeit garantiert unter der Voraussetzung, dass die rechte Seite f in einem Gebiet $I \times D$ stetig und beschränkt ist und sämtliche partiellen Ableitungen $\partial f / \partial y_i$, $i = 1, \ldots, n$, dort existieren und beschränkt sind (1826).

Rudolf Lipschitz (1832 – 1903) ersetzte 1876 die Voraussetzung an die partiellen Ableitungen durch eine schwächere Bedingung, die so genannte *Lipschitz–Bedingung*) :

$$\|f(t, \tilde{y}) - f(t, y)\| \quad \leq \quad L \;\; \|\tilde{y} - y\| \; . \tag{3.41}$$

Émile Picard (1856–1941) und Ernst Lindelöf (1870–1946) gaben um 1890 einen konstruktiven Beweis des Satzes von Lipschitz an, bei dem sie das *Verfahren der sukzessiven Approximation* verwendeten.

Im gleichen Jahr 1890 konnte GIUSEPPE PEANO (1858 – 1932) zeigen, dass die Existenz einer Lösung von (3.1) bereits dann garantiert ist, wenn die rechte Seite lediglich stetig und beschränkt ist. Die Eindeutigkeit der Lösung ist allerdings unter diesen schwachen Voraussetzungen nicht mehr gesichert.

Beispiel (3.3) Wir betrachten die AWA

$$y'(t) \quad = \quad \sqrt{|y(t)|}\,, \qquad y(0) \;=\; 0\,.$$

Die rechte Seite $f(t,y) = \sqrt{|y|}$ ist stetig und beschränkt auf $\mathbb{R} \times [-a,a]$, $a > 0$, erfüllt jedoch dort *keine* Lipschitz–Bedingung.

In der Tat ist für beliebige $\alpha \le 0 \le \beta$

$$y(t) \quad = \quad \begin{cases} -\frac{1}{4}(t-\alpha)^2\,, & -\infty < t \le \alpha \\ 0 \quad\;\; , & \alpha \le t \le \beta \\ \frac{1}{4}(t-\beta)^2\;\, , & \beta \le t < \infty \end{cases}$$

eine Lösung der AWA.

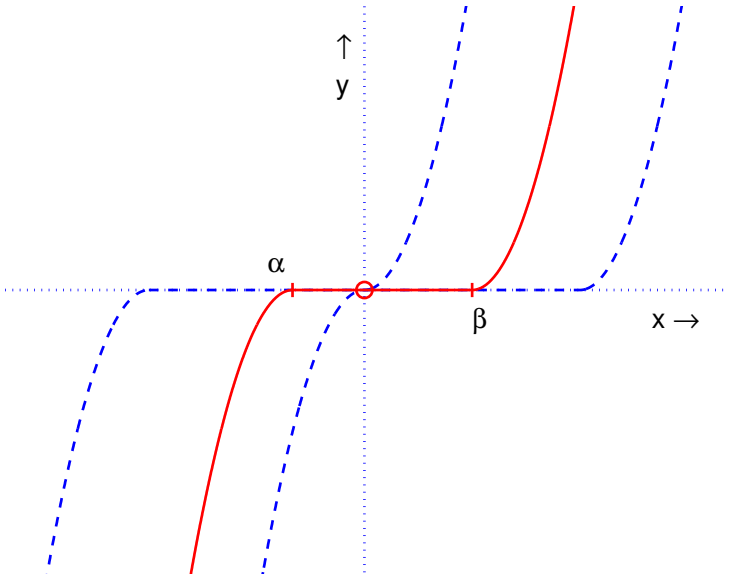

Abb. 3.1. Mehrdeutigkeit der Lösungen einer Anfangswertaufgabe

A. Der Existenzsatz von Peano.

Der Kernpunkt in unserem Beweis des Existenzsatzes von Peano ist eine Konvergenzaussage für das so genannte *Eulersche Polygonzugverfahren*, auch *Euler-Cauchy-Verfahren* genannt nach Leonard Euler (1707 – 1783) und August Louis Cauchy (1789 – 1857).

Wir wollen zeigen, dass unter Stetigkeitsvoraussetzungen an die rechte Seite f der Differentialgleichung das Euler-Verfahren zu einer gegen Null konvergenten Schrittweitenfolge

Näherungslösungen liefert, die eine konvergente Teilfolge besitzen. Deren Grenzwert ist dann notwendigerweise eine Lösung der Anfangswertaufgabe.

Ein beweistechnisches Hilfsmittel ist der

Satz (3.4) (Satz von Arzela und Ascoli[3])

Eine Folge gleichmäßig beschränkter und gleichgradig stetiger Funktionen $z_m : [a,b] \to \mathbb{R}^n$, $m \in \mathbb{N}$, besitzt eine gleichmäßig konvergente Teilfolge.

Dabei heißt $(z_m)_{m\in\mathbb{N}} \in \mathrm{C}[a,b]^{\mathbb{N}}$ *gleichmäßig beschränkt*, falls $\|z_m(t)\| \le C$ für eine geeignete positive Konstanten C und alle $m \in \mathbb{N}$ und $t \in [a,b]$ gilt. Die Folge $(z_m)_{m\in\mathbb{N}}$ heißt *gleichgradig stetig*, falls gilt

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, t,\, \tilde{t} \in [a,b],\, m \in \mathbb{N} :\ |\, t - \tilde{t}| < \delta \implies \|z_m(t) - z_m(\tilde{t})\| < \varepsilon. \qquad (3.5)$$

Beweis: Es sei $A = \{t_k :\ k \in \mathbb{N}\}$ eine abzählbar dichte Teilmenge von $[a,b]$, etwa eine Abzählung der rationalen Zahlen in diesem Intervall. Wir konstruieren nun iterativ Teilfolgen vom (z_m), die jeweils an einer der Stellen t_k konvergieren:

$k = 0$: $\quad z_m^{(0)}(t) \ := \ z_m(t)$,

$k \Rightarrow k+1$: $\quad (z_m^{(k)}(t_{k+1}))_{m\in\mathbb{N}}$ ist eine beschränkte Folge im $\mathbb{R}^n$. Sie besitzt daher eine konvergente Teilfolge $(z_{m_j}^{(k)}(t_{k+1}))_{j\in\mathbb{N}}$. Wir wählen die entsprechende Teilfolge $(z_{m_j}^{(k)})$ und bezeichnen diese mit $(z_m^{(k+1)})$. $(z_m^{(k+1)})$ ist also eine Teilfolge von $(z_m^{(k)})$, die an der Stelle t_{k+1} konvergiert (für $m \to \infty$). Nach Konstruktion konvergiert $(z_m^{(k)})$ damit aber auch an allen früheren Stellen $t_j, j \le k$.

	t_1	t_2	t_3	$\dots$
z_1	$\mathbf{z_1}$	z_1	z_5	$\dots$
z_2	z_3	$\mathbf{z_5}$	z_9	$\dots$
z_3	z_5	z_9	$\mathbf{z_{17}}$	$\dots$
z_4	z_7	z_{13}	z_{21}	$\dots$
z_5	z_9	z_{17}	z_{29}	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Wir bilden nun die Diagonalfolge $w_m := z_m^{(m)}$, $m \in \mathbb{N}$. Für jedes k ist $(w_m)_{m\ge k}$ eine Teilfolge von $(z_m^{(k)})$ und damit in t_k konvergent. Somit ist (w_m) auch eine Teilfolge von (z_m), die an *allen* Stellen t_k konvergiert.

Wir zeigen nun, dass (w_m) im Raum $(\mathrm{C}[a,b], \|\cdot\|_\infty)$ eine Cauchy-Folge bildet und daher gleichmäßig konvergiert. Dazu sei $\varepsilon > 0$ beliebig vorgegeben und $\delta > 0$ gemäß (3.5) gewählt. Da (w_m) eine Teilfolge von (z_m) ist, folgt mit (3.5):

$$\forall\, t,\, \tilde{t} \in [a,b],\, m \in \mathbb{N} :\ |\, t - \tilde{t}| < \delta \implies \|w_m(t) - w_m(\tilde{t})\| < \varepsilon. \qquad (3.6)$$

[3]Nach Cesare Arzela (1847 – 1912) und Giulio Ascoli (1843 – 1896)

Nun sei $\ell \in \mathbb{N}$ so groß gewählt, dass zu jedem $t \in [a,b]$ ein $k \in \{1,\ldots,\ell\}$ existiert mit $|\,t-t_k| < \delta$ (beachte, dass A dicht in $[a,b]$ ist). Da nun die $w_m(t_k)$ für $m \to \infty$ konvergieren, existiert ein (von $k \in \{1,\ldots,\ell\}$ unabhängiges) $N = N(\varepsilon) \in \mathbb{N}$ mit:

$$\forall m,\, \widetilde{m} \geq N \;\; \forall k \in \{1,\ldots,\ell\}: \;\; \|w_m(t_k) - w_{\widetilde{m}}(t_k)\| \;<\; \varepsilon.$$

Für ein beliebiges $t \in [a,b]$ und $m, \widetilde{m} \geq N$ erhält man dann mittels (3.6) die Abschätzung:

$$\begin{aligned} \|w_m(t) - w_{\widetilde{m}}(t)\| \;&\leq\; \|w_m(t) - w_m(t_k)\| \;+\; \|w_m(t_k) - w_{\widetilde{m}}(t_k)\| \\ &+\; \|w_{\widetilde{m}}(t_k) - w_{\widetilde{m}}(t)\| \;\leq\; 3\,\varepsilon. \end{aligned}$$

□

Satz (3.7) (Existenzsatz von Peano)

Ist f auf dem Gebiet $G = I \times D \subset \mathbb{R}^{n+1}$ stetig und ist $(t_0, y_0) \in G$, so existiert ein $\delta > 0$, so dass die AWA (3.1) im Intervall $|\,t - t_0| < \delta$ eine Lösung besitzt.

Beweis: Da G offen ist, gibt es einen Quader Q mit Mittelpunkt (t_0, y_0), der ganz in G liegt:

$$Q \;=\; \{(t,y): \, |\,t-t_0| \leq a \;\wedge\; \|y - y_0\|_\infty \leq b\} \;\text{ mit }\; a,\; b > 0. \tag{3.8}$$

Auf dem Kompaktum Q ist f beschränkt, es gibt also $M > 0$ mit

$$\forall (t,y) \in Q: \quad \|f(t,y)\|_\infty \;\leq\; M. \tag{3.9}$$

Schließlich sei $$\delta \;:=\; \min(a, b/M) \;>\; 0. \tag{3.10}$$

Der Doppelkegel $K := \{(t,y): \, |\,t-t_0| \leq \delta \wedge \|y-y_0\|_\infty \leq M\,|\,t-t_0|\}$ liegt damit ganz in Q.

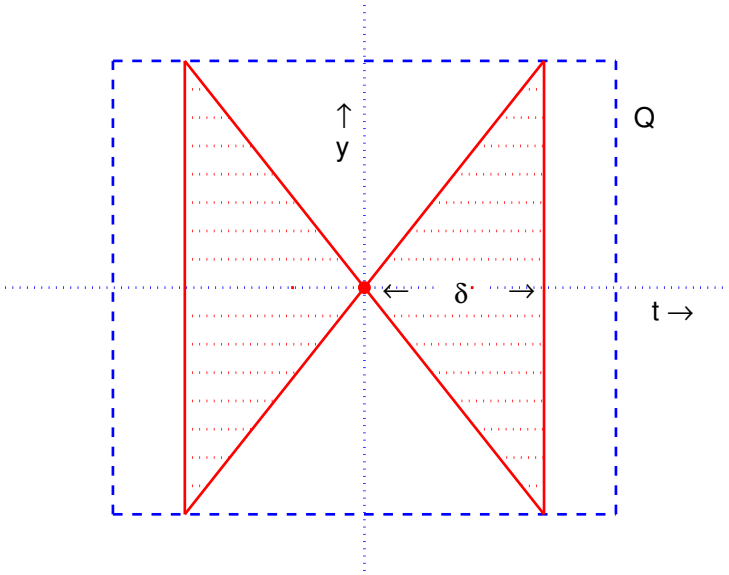

Abb. 3.2. Doppelkegel

Wir konstruieren nun Eulersche Polygonzüge mit den Ecken

$$t_{j+1}^{(m)} := t_j^{(m)} + h_j^{(m)}, \quad y_{j+1}^{(m)} := y_j^{(m)} + h_j^{(m)} f(t_j^{(m)}, y_j^{(m)}), \quad j = 0, 1, \ldots, \ell_m^+ - 1,$$
$$t_{j-1}^{(m)} := t_j^{(m)} - h_{j-1}^{(m)}, \quad y_{j-1}^{(m)} := y_j^{(m)} - h_{j-1}^{(m)} f(t_j^{(m)}, y_j^{(m)}), \quad j = 0, -1, \ldots, 1 - \ell_m^-.$$

Startpunkt ist jeweils $t_0^{(m)} := t_0$ und $y_0^{(m)} := y_0$. Alle Schrittweiten $h_j^{(m)}$ seien positiv und die maximalen Schrittweiten $\Delta_m := \max\{h_j^{(m)} : -\ell_m^- \le j \le \ell_m^+ - 1\}$ mögen eine Nullfolge bilden, $\Delta_m \to 0$, für $m \to \infty$.

Die *Diskretisierungspunkte* oder *Gitterpunkte* $t_j^{(m)}$ mögen bis an den Rand des zulässigen Intervalls $|t - t_0| \le \delta$ führen, genauer seien $t_{-\ell_m^-}^{(m)} - \Delta_m < t_0 - \delta \le t_{-\ell_m^-}^{(m)}$ und $t_{\ell_m^+}^{(m)} \le t_0 + \delta < t_{\ell_m^+}^{(m)} + \Delta_m$.

Die Punkte $(t_j^{(m)}, y_j^{(m)})$ werden nun durch Geradenstücke verbunden, wobei an den Rändern die letzten Geradenstücke nur bis zum Rand fortgesetzt werden.Die entstehenden Polygonzüge $y^{(m)}$ sind damit auf dem gesamten Intervall $[t_0 - \delta, t_0 + \delta]$ wohldefinierte, stetige Funktionen.

Ferner sieht man aufgrund der Konstruktion: Alle Polygonzüge verlaufen im Kegel K und für beliebige t, $\widetilde{t} \in [t_0 - \delta, t_0 + \delta]$ gilt

$$\|y^{(m)}(\widetilde{t}) - y^{(m)}(t)\|_\infty \le M\,|\widetilde{t} - t|. \tag{3.11}$$

Für benachbarte Gitterpunkte folgt (3.11) aus dem Mittelwertsatz und (3.9). Für beliebige Abszissen $\widetilde{t} > t$ schiebt man die inneren Gitterpunkte ein und schätzt mit der Dreiecksungleichung ab:

$$\begin{aligned}\|y^{(m)}(\widetilde{t}) - y^{(m)}(t)\| &\le \|y^{(m)}(\widetilde{t}) - y_k^{(m)}\| + \|y_k^{(m)} - y_{k-1}^{(m)}\| + \ldots \\ &+ \|y_{j+1}^{(m)} - y_j^{(m)}\| + \|y_j^{(m)} - y^{(m)}(t)\|.\end{aligned}$$

Wegen (3.11) sind die $(y^{(m)})_{m\in\mathbb{N}}$ gleichmäßig beschränkt und gleichgradig stetig, also existiert nach dem Satz von Arzela und Ascoli eine gleichmäßig konvergente Teilfolge, die der Einfachheit halber wieder mit $(y^{(m)})$ bezeichnet werde. Die Grenzfunktion $y := \lim y^{(m)}$ ist somit (als gleichmäßiger Limes stetiger Funktionen) stetig und erfüllt die Anfangsbedingung $y(t_0) = y_0$.

Es bleibt nun noch zu zeigen, dass y auch differenzierbar ist und die DGL (3.1) erfüllt.

Dazu sei o.E.d.A. $t_0 \le t_1 < t_0 + \delta$ und $y_1 := y(t_1)$.

Zu $\varepsilon > 0$ existiert aufgrund der Stetigkeit von f ein $\eta = \eta(\varepsilon) > 0$, so dass

$$Q_\eta := \{(t, y) : |t - t_1| \le 2\,\eta \wedge \|y - y_1\|_\infty < 4\,M\eta\} \subset Q$$

und
$$\forall (t, y) \in Q_\eta : \|f(t, y) - f(t_1, y_1)\| < \varepsilon. \tag{3.12}$$

Ferner sei $N = N(\varepsilon) \in \mathbb{N}$ so groß gewählt, dass $\Delta_m < \eta$ und $\|y(t) - y^{(m)}(t)\| < M\,\eta$ für alle t mit $|t - t_1| \le 2\,\eta$ und $m \ge N$ gilt.

Damit liegt der Polygonzug $(t, y^{(m)}(t))$ für $|t - t_1| \leq 2\eta$ ganz in Q_η, denn:

$$\begin{aligned} \|y^{(m)}(t) - y_1\| &\leq \|y^{(m)}(t) - y^{(m)}(t_1)\| + \|y^{(m)}(t_1) - y(t_1)\| \\ &\leq M\,|t - t_1| + M\,\eta \\ &\leq 3\,M\,\eta < 4\,M\,\eta. \end{aligned}$$

Sei nun $|t_2 - t_1| < \eta$, $m \geq N(\varepsilon)$ und o.E.d.A. $t_2 > t_1$. Wie zuvor schieben wir zwischen t_2 und t_1 die dazwischen liegenden Gitterpunkte $t_i^{(m)}$, $j \leq i \leq k$, ein. $t_{j-1}^{(m)}$ ist dann der erste Gitterpunkt links von t_1. Wegen $\Delta_m < \eta$ gibt es einen solchen Gitterpunkt und dieser liegt dann auch im Bereich $|t - t_1| < 2\,\eta$:

$$\begin{aligned} y^{(m)}(t_2) - y^{(m)}(t_1) &= \left(y^{(m)}(t_2) - y_k^{(m)}\right) + \left(y_k^{(m)} - y_{k-1}^{(m)}\right) + \ldots \\ &+ \left(y_{j+1}^{(m)} - y_j^{(m)}\right) + \left(y_j^{(m)} - y^{(m)}(t_1)\right) \\ &= f(t_k^{(m)}, y_k^{(m)})\,(t_2 - t_j^{(m)}) + f(t_{k-1}^{(m)}, y_{k-1}^{(m)})\; h_{k-1}^{(m)} + \ldots \\ &+ f(t_j^{(m)}, y_j^{(m)})\; h_j^{(m)} + f(t_{j-1}^{(m)}, y_{j-1}^{(m)})\,(t_j^{(m)} - t_1). \end{aligned}$$

Die Dreiecksungleichung und (3.12) liefert die Abschätzung:

$$\begin{aligned} \|y^{(m)}(t_2) - y^{(m)}(t_1) - f(t_1, y_1)\,(t_2 - t_1)\| &\leq \|\,f(t_k^{(m)}, y_k^{(m)}) - f(t_1, y_1)\|\;(t_2 - t_k^{(m)}) \\ &+ \sum_{i=j}^{k-1} \|\,f(t_i^{(m)}, y_i^{(m)}) - f(t_1, y_1)\|\; h_i^{(m)} \\ &+ \|\,f(t_{j-1}^{(m)}, y_{j-1}^{(m)}) - f(t_1, y_1)\|\;(t_j^{(m)} - t_1) \\ &\leq \varepsilon\,(t_2 - t_1). \end{aligned}$$

Für alle $m \geq N$ und $|t_2 - t_1| < \eta$, $t_2 \neq t_1$, folgt demnach

$$\left\| \frac{y^{(m)}(t_2) - y^{(m)}(t_1)}{t_2 - t_1} - f(t_1, y_1) \right\| \leq \varepsilon$$

und hiermit für $m \to \infty$:

$$\left\| \frac{y(t_2) - y(t_1)}{t_2 - t_1} - f(t_1, y_1) \right\| \leq \varepsilon.$$

Damit ist gezeigt, dass y in t_1 differenzierbar ist und dort die DGL $y' = f(t, y)$ erfüllt. $\square$

Bemerkungen (3.13)

a) Ist die rechte Seite f der DGL (3.1) auf einem durch (3.8) definierten Quader Q stetig, so existiert eine Lösung der AWA wenigstens auf dem Intervall $|t - t_0| \leq \delta$, wobei δ gemäß (3.10) erklärt ist.

b) Jeder Häufungspunkt der oben konstruierten Folge $y^{(m)}$ von Polygonzügen liefert eine Lösung der AWA. Im Allgemeinen lässt sich aber umgekehrt nicht jede Lösung mit dem Euler–Verfahren gewinnen.

c) Ist f stetig auf dem Gebiet $G := I \times D$, so lässt sich jede Lösung y der AWA (3.1) auf ein *maximales* Existenzintervall $t_{\min} < t < t_{\max}$ fortsetzen. Dabei kommt $(t, y(t))$ für $t \to t_{\min}$ bzw. $t \to t_{\max}$ dem Rand von G beliebig nahe, d.h. jeder (endliche) Häufungspunkt einer Folge $(t_k, y(t_k))_{k\in\mathbb{N}}$ mit $t_k \to t_{\min}$ bzw. $t_k \to t_{\max}$ $(k \to \infty)$ liegt auf dem Rand von G.

Beispiel (3.14) $\quad y' = y, \quad G = \mathbb{R} \times \mathbb{R}.$

Die allgemeine Lösung ist $y(t) = C\mathrm{e}^t$. Jede Lösung lässt sich auf $\mathbb{R}$ fortsetzen. Wegen $t_{\min} = -\infty$ und $t_{\max} = \infty$ existieren keine (endlichen) Häufungspunkte.

Beispiel (3.15) $\quad y' = -t/y, \quad G = \mathbb{R} \times \mathbb{R}^+.$

Man beachte, dass G ein Gebiet, also insbesondere zusammenhängend sein muss. Die allgemeine Lösung der DGL lautet $y(t) = +\sqrt{r^2 - t^2}, \quad r > 0$, vgl. auch Beispiel (1.24). Damit ist $t_{\min} = -r$ und $t_{\max} = r$ und die beiden Häufungspunkte $(t_{\min}, 0)$ und $(t_{\max}, 0)$ liegen auf dem Rand von G.

Beispiel (3.16) $\quad y' = y^2, \quad y(0) = 1, \quad G = \mathbb{R} \times \mathbb{R}.$

Mittels Variablentrennung erhält man die Lösung $y(t) = 1/(1-t)$ und damit $t_{\min} = -\infty$, $t_{\max} = 1$. Wiederum existieren keine (endlichen) Häufungspunkte für $t \to t_{\min}$ oder $t \to t_{\max}$.

Beispiel (3.17) $\quad y' = -\dfrac{\sqrt{1-y^2}}{t^2}, \quad G = \mathbb{R}^+ \times [-1, 1].$

Die Lösung kann wieder mittels Variablentrennung ermittelt werden. Man findet $y(t) = \sin(1/t + C)$ und $t_{\min} = 0$, sowie $t_{\max} = \infty$. Häufungspunkte existieren für $t \to 0$ mit den Werten $(0, \lambda), \quad \lambda \in [-1, 1]$.

Man beachte, dass es neben den obigen Lösungen die beiden singulären Lösungen $y(t) = \pm 1$ gibt, und in den Punkten $(t, \pm 1)$ die lokale Eindeutigkeit verletzt ist.

Wir können nun auch schon eine Aussage über die Stabilität von AWA beweisen. Dabei gehen wir von konvergenten Folgen von rechten Seiten und Anfangswerten aus und zeigen, dass die Lösungsfolge der zugehörigen AWAen gegen die Lösung der Grenz-AWA

konvergiert. Grenzwertbildung und die Lösung von AWA sind also in diesen Sinn vertauschbare Prozesse.

Satz (3.18) (Stabilität)

Sei $f^{(m)} : G \to \mathbb{R}^n$ eine Folge stetiger Funktionen auf einem Gebiet $G = I \times D \subset \mathbb{R}^{n+1}$ und es konvergiere $f^{(m)} \to f\ (m \to \infty)$ lokal gleichmäßig auf G.
Ferner seien $(t_m, y_m) \in G$ Anfangswerte mit $(t_m, y_m) \to (t_0, y_0) \in G\ (m \to \infty)$ und es bezeichne $y^{(m)}$ bzw. y Lösungen der zugehörigen AWAen

$$y' = f^{(m)}(t, y),\ y^{(m)}(t_m) = y_m \quad \text{bzw.} \quad y' = f(t, y),\ y(t_0) = y_0.$$

Ist die Lösung y der Grenz-AWA dann eindeutig bestimmt und auf einem kompakten Intervall $I_0 \subset I$ definiert, so sind auch die $y^{(m)}$ für hinreichend großes m auf I_0 erklärt (bzw. fortsetzbar) und konvergieren gleichmäßig gegen y.

Beweis: Wir wählen wie im Existenzsatz von Peano einen kompakten Quader $Q = \{(t, y) : |t - t_0| \le a \wedge \|y - y_0\|_\infty \le b\} \subset G$ und $M > 0$ mit $\|f(t, y)\| < M$ für alle $(t, y) \in Q$. Wegen der gleichmäßigen Konvergenz gilt dann auch $\|f^{(m)}(t, y)\| < M$ für hinreichend große $m \ge m_1$. Sei weiter $\delta := \min(a, b/M)$. Wegen $t_m \to t_0$ und $y_m \to y_0$ existieren dann aufgrund des Satzes von Peano sowohl die $y^{(m)}$ (für hinreichend großes $m \ge m_2 \ge m_1$) wie auch y im gesamten Intervall $|t - t_0| \le \delta/2$.

Die Folge $(y^{(m)})_{m \ge m_2}$ ist dann auf $|t - t_0| \le \delta/2$ gleichmäßig beschränkt und gleichgradig stetig, besitzt also nach dem Satz von Arzela und Ascoli eine gleichmäßig konvergente Teilfolge $(y^{(m_k)})$. Für die Grenzfunktion $\overline{y}$ gilt dann die Integralbeziehung (Integration der Anfangswertaufgabe):

$$\begin{aligned} \overline{y}(t) &= \lim\ y^{(m_k)}(t) = \lim \left(y_{m_k} + \int\limits_{t_{m_k}}^{t} f^{(m_k)}(\tau, y^{(m_k)}(\tau))\, d\tau \right) \\ &= y_0 + \int\limits_{t_0}^{t} f(\tau, \overline{y}(\tau))\, d\tau. \end{aligned}$$

Damit ist $\overline{y}$ zugleich Lösung der Grenz-AWA, also wegen der vorausgesetzten Eindeutigkeit: $\overline{y} = y$.

Die obige Überlegung gilt für jeden Häufungspunkt der Folge $(y^{(m)})$, d.h. die Folge besitzt überhaupt nur einen Häufungspunkt, nämlich y. Hieraus folgt mit dem Satz von Arzela und Ascoli, dass die Folge selbst gleichmäßig gegen y konvergieren muss. (Gäbe es unendlich viele Folgenglieder außerhalb eine ε-Streifens um y, so hätten diese Folgenglieder einen Häufungspunkt, im Widerspruch zur obigen Aussage).

Damit ist die Behauptung für das Teilintervall $|t - t_0| \le \delta/2$ gezeigt. Für das gesamte Intervall I_0 folgt sie mittels Kompaktheitsschluss. □

B. Der Satz von Picard und Lindelöf.

Die bisherigen Beispiele für nicht eindeutig lösbare AWAen (Beispiele 3.3 und 3.17) waren dadurch gekennzeichnet, dass sich die rechte Seite f in der Nähe einer kritischen Stelle mit y stark änderte. Dies legt nahe, die Eindeutigkeit dadurch zu erzwingen, dass man die Variation von f bei Änderung von y beschränkt. Dies kann beispielsweise durch die *Lipschitz-Bedingung* (3.2) mit einer festen Lipschitz-Konstanten L erfolgen.

Wir beschreiben wieder einen Zugang, der die Existenz und Eindeutigkeit einer Lösung mit Hilfe eines Näherungsverfahrens, dem *Verfahren der sukzessiven Approximation*, zeigt. Allerdings ist dieses Verfahren, anders als das Euler-Verfahren, für die tatsächliche numerische Rechnung wenig geeignet.

Satz (3.19) (Satz von Picard und Lindelöf)

Sei $f : Q \to \mathbb{R}^n$ eine stetige Funktion auf dem Quader $(a,\, b > 0)$

$$Q \;=\; \big\{(t,y) \in \mathbb{R}^{n+1} \;:\; |\,t-t_0| \le a \;\wedge\; \|y-y_0\|_\infty \le b\big\}\;.$$

Ferner gebe es Konstante M, $L > 0$, so dass für alle $(t,\widetilde{y})$, $(t,y) \in Q$ gilt: $\|f(t,y)\|_\infty \le M$, sowie die Lipschitz-Bedingung: $\|f(t,\widetilde{y}) - f(t,y)\|_\infty \le L\,\|\widetilde{y}-y\|_\infty$.

Dann besitzt die AWA (3.1) eine eindeutig bestimmte Lösung y, die mindestens im Intervall $[\,t_0-\delta, t_0+\delta]$, $\delta := \min(a, b/M)$ definiert ist. Diese lässt sich als gleichmäßiger Limes der folgenden Funktionenfolge (oberer Index = Folgenindex!!) erhalten:

$$y^{(0)}(t) \;:=\; y_0, \quad y^{(k+1)}(t) \;:=\; y_0 \;+\; \int_{t_0}^{t} f(\tau, y^{(k)}(\tau))\,d\tau, \quad k = 0,1,\ldots. \tag{3.20}$$

Beweis: Durch komponentenweise Integration lässt sich die AWA (3.1) in eine äquivalente Integralgleichung umwandeln

$$y(t) \;=\; y_0 \;+\; \int_{t_0}^{t} f(\tau, y(\tau))\,d\tau \;=:\; \Phi(y)(t). \tag{3.21}$$

Dies ist eine Fixpunktgleichung für eine Funktion $y : [\,t_0-\delta, t_0+\delta] \to \mathbb{R}$, und es ist daher naheliegend, zur Lösung dieser Gleichung die Fixpunktiteration (3.20) zu verwenden. Der obere Index ist hierbei der Iterationsindex des Verfahrens. Die Iteration heißt *Verfahren der sukzessiven Approximation*.

Wir führen den Konvergenzbeweis nun analog zum Beweis des Fixpunktsatzes (vgl. z.B. Königsberger, Analysis 2, Seite 107).

Zunächst sieht man, dass alle Iterierten $y^{(k)}$ auf dem Intervall $|\,t-t_0| \le \delta$ erklärt sind und ganz im Quader Q verlaufen (die Norm ist stets die Maximumsnorm im $\mathbb{R}^n$):

$$\big\|\,y^{(k+1)}(t) - y_0\big\| \;=\; \|\int_{t_0}^{t} f(\tau, y^{(k)}(\tau))\,d\tau\| \;\le\; \int_{t_0}^{t} \|f(\tau, y^{(k)}(\tau))\|\,d\tau \;\le\; M\,|\,t-t_0| \;\le\; b.$$

Die Lipschitz–Bedingung liefert nun die Abschätzung:

$$\begin{aligned} \| y^{(k+1)}(t) - y^{(k)}(t)\| &= \| \int_{t_0}^{t} f(\tau, y^{(k)}(\tau)) - f(\tau, y^{(k-1)}(\tau))\, d\tau \| \\ &\leq L \int_{t_0}^{t} \| y^{(k)}(\tau) - y^{(k-1)}(\tau) \|\, d\tau, \end{aligned}$$

woraus sich wegen $\| y^{(1)}(t) - y_0\| \leq M\,|t - t_0|$ mittels vollständiger Induktion ergibt:

$$\forall\, k \in \mathbb{N},\ |t - t_0| \leq \delta: \quad \left\| y^{(k)}(t) - y^{(k-1)}(t) \right\| \leq M \cdot L^{k-1}\, \frac{|t - t_0|^k}{k!}\,,$$

Dies zeigt nun die gleichmäßige Konvergenz der Reihe $\sum_{j=1}^{\infty} \left(y^{(j)}(t) - y^{(j-1)}(t) \right)$ und damit auch die gleichmäßige Konvergenz von $y^{(k)}$ (für $k \to \infty$) gegen eine stetige Lösung y der Fixpunktgleichung (3.21) auf dem Intervall $|t - t_0| \leq \delta$.

Zur Eindeutigkeit: Sind $\widetilde{y}$, y stetige Lösungen der Fixpunktgleichung, so folgt:

$$\begin{aligned} \|\widetilde{y}(t) - y(t)\| &= \| \int_{t_0}^{t} f(\tau, \widetilde{y}(\tau)) - f(\tau, y(\tau))\, d\tau \| \\ &\leq L \int_{t_0}^{t} \|\widetilde{y}(\tau) - y(\tau)\|\, d\tau \\ &\leq L\, C\, |t - t_0|, \quad C := \max_{|t-t_0|\leq\delta} \|\widetilde{y}(t) - y(t)\|\,. \end{aligned}$$

Setzt man diese Abschätzung nun wieder in das obige Integral ein und iteriert diesen Prozess, so folgt schließlich

$$\|\widetilde{y}(t) - y(t)\| \leq L^k\, C\, \frac{|t - t_0|^k}{k!} \to 0 \qquad (k \to \infty),$$

und somit $\widetilde{y}(t) = y(t), \quad \forall t : |t - t_0| \leq \delta.$ □

Bemerkungen (3.22)

a) Erfüllt f auf dem Streifen $[a, b] \times \mathbb{R}^n$ eine (globale) Lipschitz–Bedingung (3.2), so besitzt die AWA (3.1) mit $t_0 \in [a, b]$ eine eindeutig bestimmte Lösung y, die auf ganz $[a, b]$ erklärt ist (globale Existenz).

Beweis: Man hat in dem obigen Beweis des Satzes von Picard und Lindelöf lediglich die Anfangsabschätzung zu ersetzen durch:

$$\| y^{(1)}(t) - y_0\| = \| \int_{t_0}^{t} f(\tau, y_0)\, d\tau \| \leq \int_{t_0}^{t} \|f(\tau, y_0)\|\, d\tau \leq M\,|t - t_0|.$$

mit $M := \max\{\|f(t, y_0)\| :\ t \in [a, b]\}$. □

b) Eine *lineare* AWA

$$y'(t) \;=\; A(t)\,y(t) + b(t)\,, \quad y(t_0) \;=\; y_0 \tag{3.23}$$

mit stetigen Funktionen $A : \mathbb{R} \to \mathbb{R}^{(n,n)}$, $b : \mathbb{R} \to \mathbb{R}^n$ besitzt eine eindeutig bestimmte Lösung y, die auf ganz $\mathbb{R}$ definiert ist.

Beweis: Die Aussage folgt aus a) wegen

$$\|f(t,\widetilde{y}) - f(t,y)\| \quad\leq\quad \|A(t)\|\;\|\widetilde{y} - y\|\;.$$

Die globale Lipschitz–Bedingung ist also auf *jedem* Streifen $[a,b] \times \mathbb{R}^n$ erfüllt. □

c) Zum Nachweis der Lipschitz-Bedingung und zur Berechnung der Lipschitz-Konstanten ist der folgende Sachverhalt hilfreich: Ist f stetig und bzgl. y differenzierbar auf Q und sind die partiellen Ableitungen beschränkt:

$$L_i \;:=\; \sup\left\{ \sum_{j=1}^{n} \left|\frac{\partial f_i}{\partial y_j}(t,y)\right| :\; (t,y) \in Q \right\} \;<\; \infty, \tag{3.24}$$

so ist f Lipschitz-stetig mit der Lipschitz-Konstanten $L := \max(L_1\,,\ldots,L_n)$.

Beweis: Nach dem Mittelwertsatz gilt für $i = 1,\ldots,n$:

$$f_i(t,\widetilde{y}) - f_i(t,y) \;=\; \nabla_y f_i(t, y + \Theta_i(\widetilde{y} - y))^{\mathrm{T}}\,(\widetilde{y} - y), \quad \Theta_i \in\,]0,1[,$$

und damit $|f_i(t,\widetilde{y}) - f_i(t,y)| \leq L_i\,\|\widetilde{y} - y\|_\infty$. Bildet man hier das Maximum über $i = 1,\ldots,n$, so folgt die Behauptung. □

Beispiel (3.25) $\qquad y' \;=\; y, \quad y(0) \;=\; 1.$

Das Verfahren der sukzessiven Approximation liefert die Näherungen (Beweis per Induktion):

$$y^{(k)}(t) \quad=\quad \sum_{j=0}^{k} \frac{1}{j!}\,t^j\;.$$

Für $k \to \infty$ ergibt sich daher: $y(t) \;=\; \lim_{k\to\infty}\, y^{(k)}(t) \;=\; \exp(t)$.

C. Abhängigkeit von Parametern, Stabilität

Wir betrachten wieder die AWA (3.1) und setzen nun voraus, dass f auf einem Gebiet $G = I \times D \subset \mathbb{R}^{n+1}$ stetig differenzierbar ist.

Die Lösung der AWA ist damit für $(t_0, y_0) \in G$ nach (3.19) lokal eindeutig bestimmt. Wir denken uns die Lösung in G maximal fortgesetzt und bezeichnen diese Fortsetzung mit $y(t; t_0, y_0)$.

In Anwendungen sind die Anfangsdaten (t_0, y_0) häufig nur mit einer gewissen Genauigkeit gegeben. Wir fragen, wie sich Fehler in diesen Daten auf die Lösung auswirken.

Als technisches Hilfsmittel verwenden wir das folgende Gronwall–Lemma:

Satz (3.26) (Lemma von Gronwall[4])

Gilt für eine auf einem Intervall $I := \{t : |t - t_0| \le \delta\}$ stetige Funktion $r : I \to \mathbb{R}$ eine Abschätzung der Form

$$r(t) \;\le\; \alpha + \beta \int_{t_0}^{t} r(\tau)\, d\tau, \qquad \alpha \ge 0,\ \beta > 0\,, \tag{3.27}$$

so folgt für alle $t \in I$: $r(t) \le \alpha\, \mathrm{e}^{\beta\,|t-t_0|}$.

Beweis: Wir multiplizieren (3.27) mit $\mathrm{e}^{-\beta t}$ und setzen dann $u(t) := \mathrm{e}^{-\beta t} \int_{t_0}^{t} r(\tau)\, d\tau$.

Es folgt: $$u'(t) \;=\; -\beta\, u(t) + \mathrm{e}^{-\beta t}\, r(t) \;\le\; \alpha\, \mathrm{e}^{-\beta t}\,,$$

also $\alpha\, \mathrm{e}^{-\beta t} - u'(t) \ge 0$. Integration über $[t_0, t]$ ergibt:

$$-\frac{\alpha}{\beta}\, \mathrm{e}^{-\beta t} - u(t) \;\ge\; -\frac{\alpha}{\beta}\, \mathrm{e}^{-\beta t_0} \quad (\,\forall\, t :\ t_0 \le t \le t_0 + \delta)\,.$$

Wir lösen nach u auf:

$$u(t) \;\le\; \frac{\alpha}{\beta}\, \left(\mathrm{e}^{-\beta t_0} - \mathrm{e}^{-\beta t}\right)$$

und erhalten mit Hilfe der Ausgangsungleichung:

$$r(t) \;\le\; \alpha + \beta\, \mathrm{e}^{\beta t}\, u(t) \;\le\; \alpha\, \mathrm{e}^{\beta (t-t_0)},$$

was zu zeigen war. Für $t \le t_0$ kann die Behauptung vermöge der Transformation

$$\tilde{r}(t) \;:=\; r(2t_0 - t), \qquad t \ge t_0$$

auf den obigen Fall ($t \ge t_0$) zurückgeführt werden. □

Mit dem Gronwall–Lemma lässt sich nun die folgende Fehlerabschätzung beweisen.

Satz (3.28) (Stabilität, Fehlerabschätzung)

Für Anfangswerte $y_0, z_0 \in \mathbb{R}^n$ seien die Lösungen $y(t; t_0, y_0)$ und $y(t; t_0, z_0)$ auf dem Intervall $|t - t_0| \le \delta$ definiert.

$L > 0$ bezeichne eine Lipschitz–Konstante von f auf einem (kompakten) Quader $Q = [t_0 - \delta, t_0 + \delta] \times \tilde{Q}$, welcher beide Lösungen enthält.
Dann gilt für $|t - t_0| \le \delta$:

$$\|y(t; t_0, y_0) - y(t; t_0, z_0)\| \;\le\; \mathrm{e}^{L|t-t_0|}\, \|y_0 - z_0\|\,. \tag{3.29}$$

[4]Nach dem schwedischen Mathematiker Thomas Hakon Gronwall (1877-1932)

Beweis: Die integrale Form der Anfangswertaufgabe

$$y(t;t_0,y_0) \;=\; y_0 + \int\limits_{t_0}^{t} f\left(\tau, y(\tau;t_0,y_0)\right)\, d\tau$$

liefert mittels Dreiecksungleichung die Abschätzung:

$$\begin{aligned}\|y(t;t_0,y_0) - y(t;t_0,z_0)\| \;&\leq\; \|y_0 - z_0\| + \int\limits_{t_0}^{t} \|f(\tau, y(\tau;t_0,y_0)) - f(\tau, y(\tau;t_0,z_0))\| d\tau \\ &\leq\; \|y_0 - z_0\| + L \cdot \int\limits_{t_0}^{t} \|y(\tau;t_0,y_0)) - y(\tau;t_0,z_0))\|\; d\tau\;.\end{aligned}$$

Dies ist aber gerade eine Abschätzung der Form, wie sie im Lemma von Gronwall auftritt mit $r(t) := \|y(t;t_0,y_0) - y(t;t_0,z_0)\|$ und $\alpha := \|y_0 - z_0\| \geq 0, \quad \beta := L > 0.$

Das Lemma von Gronwall liefert somit die behauptete Abschätzung. □

Bemerkungen (3.30)

a) Die in obigem Satz bewiesene Abschätzung bedeutet gerade die Lipschitz–stetige Abhängigkeit der Lösung einer AWA von den Anfangswerten.

b) Die obige Abschätzung ist auch (in gewissem Sinne) nicht zu verbessern, da beispielsweise für die lineare Anfangswertaufgabe $y' = L\,y,\;\; y(t_0) = y_0$ mit $L > 0$ und $t \geq t_0$ die Abschätzung mit Gleichheit gilt. Für $t < t_0$ wird jedoch der tatsächliche Fehler erheblich überschätzt.

c) In Verallgemeinerung des Satzes (3.28) lassen sich auch Fehler in der rechten Seite f und in der Anfangszeit t_0 berücksichtigen. Ohne Beweis bemerken wir hierzu:

   Sind f, g stetig differenzierbare Funktionen auf einem Quader Q, und gelten dort die Abschätzungen

$$\|f(t,y) - g(t,y)\| \;\leq\; \delta,\;\; \|g(t,y)\| \;\leq\; M,\;\; \|f(t,\tilde{y}) - f(t,y)\| \;\leq\; L\|\tilde{y} - y\|,$$

   so folgt für die Lösungen y und z der AWA

$$\begin{aligned} y' \;&=\; f(t,y),\;\; y(t_0) \;=\; y_0 \\ z' \;&=\; g(t,z),\;\; z(t_1) \;=\; z_0 \end{aligned}$$

   mit $(t_0,y_0),\ (t_0,z_0) \in Q^0$, die Abschätzung:

$$\|y(t) - z(t)\| \;\leq\; \|y_0 - z_0\|\, \mathrm{e}^{L\,|t-t_0|} + M\,|\,t_1 - t_0|\, \mathrm{e}^{L\,|t-t_0|} + \frac{\delta}{L}\left(\mathrm{e}^{L\,|t-t_0|} - 1\right). \quad (3.31)$$

Der erste Summand beschreibt hierbei den Fehler, der in $y(t)$ aufgrund der Änderung der Anfangswerte auftritt, der zweite Summand den Fehler, der durch die Veränderung der Anfangszeit auftritt, und der dritte Summand beschreibt schließlich den Fehler, der

durch die veränderte rechte Seite des DGLsytems hervorgerufen wird.

Weiterhin ist von Interesse, wie sich die Lösung einer *parameter–abhängigen* AWA

$$y'(t) \;=\; f(t,y,\lambda), \qquad y(t_0) \;=\; y_0\,, \tag{3.32}$$

in Abhängigkeit von den Parametern $\lambda \in \mathbb{R}^m$ verhält.

Vermöge der Transformation von (3.32) in die äquivalente AWA

$$\begin{array}{rclrcl} y' &=& f(t,y,z), & y(t_0) &=& y_0 \\ z' &=& 0, & z(t_0) &=& \lambda \end{array} \tag{3.33}$$

lässt sich dieses Problem jedoch auf den zuvor betrachteten Fall der Variation der Anfangswerte zurückführen.

Mitunter interessiert man sich über die recht groben Abschätzungen (3.29) bzw. (3.31) hinaus für die konkrete Auswertung der Größen $\dfrac{\partial}{\partial t_0}\, y(t;t_0,y_0)$ und $\dfrac{\partial}{\partial y_0}\, y(t;t_0,y_0)$.

Diese Daten lassen sich als die (absoluten) *Konditionszahlen* für die Abbildung

$$(t_0,y_0) \mapsto y(t;t_0,y_0)$$

interpretieren, vgl. Vorlesung über Numerik.

Die Existenz der hierbei auftretenden partiellen Ableitungen ist unter den folgenden Voraussetzungen sichergestellt.

Satz (3.34) (Variationsgleichungen)

Die rechte Seite f sei eine C^1–Funktion auf einem Gebiet $G = I \times D \subset \mathbb{R}^{n+1}$. $\widetilde{y}$ sei eine auf einem kompakten Intervall $I_0 \subset I$ erklärte Lösung der DGL $y' = f(t,y)$.

a) Es gibt einen Streifen um $\widetilde{y}$

$$S_\varepsilon \;:=\; \big\{(t,y)^T :\; t \in I_0 \;\wedge\; \|y - \widetilde{y}(t)\| \le \varepsilon\big\} \;\subset\; G\,,\;\; \varepsilon > 0\,,$$

so dass die Lösung $y(t;t_0,y_0)$ der AWA (3.1) für alle Anfangswerte $(t_0,y_0) \in S_\varepsilon$ auf ganz I_0 erklärt ist.

b) Die Lösung $y(t;t_0,y_0)$ ist eine C^1–Funktion (bezüglich aller Variablen) auf dem Innern $I_0^0 \times S_\varepsilon^0$.

c) Die so genannten Variationen (auch Propagationsmatrizen)

$$W(t;t_0) \;:=\; \frac{\partial}{\partial y_0}\, y(t;t_0,y_0) \;\in\; \mathbb{R}^{(n,n)}, \quad w(t;t_0) \;:=\; \frac{\partial}{\partial t_0}\, y(t;t_0,y_0) \;\in\; \mathbb{R}^n \tag{3.35}$$

lassen sich als Lösungen der folgenden linearen AWAen (Variationsgleichungen) erhalten

$$\begin{aligned} W'(t;t_0) &= f_y(t,y(t;t_0,y_0))\,W(t;t_0), \qquad W(t_0,t_0) = I_n \\ w'(t,t_0) &= f_y(t,y(t;t_0,y_0))\,w(t;t_0), \qquad w(t_0;t_0) = -f(t_0,y_0)\,. \end{aligned} \tag{3.36}$$

Auf den recht technischen Beweis dieser Aussagen wird hier verzichtet, er kann im Wesentlichen mit Hilfe einer *parameterabhängigen* Variante des Fixpunktsatzes geführt werden. Dieser Satz besagt, dass ein parameterabhängiges Fixpunktverfahren unter der Voraussetzung einer *gleichmäßigen* Kontraktionsbedingung gegen einen Fixpunkt konvergiert, der stetig und bei entsprechenden Voraussetzungen auch differenzierbar von den Parametern abhängt. Für die Details sei auf die Literatur (z.B. J. Hale, Abschnitt I.3) verwiesen.

Eine einfache Herleitung der Variationsgleichungen erhält man dagegen, wenn man voraussetzt, dass $y(t;t_0,y_0)$ sogar eine C^2–Funktion auf $I_0^0 \times S_\varepsilon^0$ ist.

Aus der Differentialgleichung

$$\frac{\partial}{\partial t}\, y(t;t_0,y_0) \;=\; f(t,y(t;t_0,y_0))$$

folgt dann nämlich durch partielle Differentiation nach y_0 mit Hilfe der Kettenregel:

$$\frac{\partial}{\partial y_0}\,\frac{\partial}{\partial t}\, y(t;t_0,y_0) \;=\; f_y(t;t_0,y_0)\,\frac{\partial}{\partial y_0}\, y(t;t_0,y_0)\,.$$

Vertauscht man nach dem Satz von Schwarz die partiellen Ableitungen auf der linken Seite, so erhält man gerade die erste Variationsgleichung (20) für W.

Die zugehörige Anfangsbedingung ergibt sich ebenso durch Differentiation der Identität (in y_0): $y(t_0;t_0,y_0) = y_0$. □

D. Differentialungleichungen.

In diesem Abschnitt betrachten wir Modifikationen der Lipschitz-Bedingung in der folgenden Form:

$$\|f(t,\widetilde{y}) - f(t,y)\| \;\le\; \omega(t,\|\widetilde{y}-y\|). \tag{3.37}$$

Dabei ist ω einer hinreichend glatte Funktion mit $\omega(t,0)=0$. Speziell für $\omega(t,u) := L\,u$ erhält man aus (3.37) die ursprüngliche Lipschitz–Bedingung (3.2).

Um zu sehen, wie sich aus (3.37) die Eindeutigkeit für die Lösung einer AWA ergibt, betrachten wir *Differentialungleichungen*. Dazu bezeichne $D^+g(t) = g'(t^+)$ die rechtsseitige Ableitung einer Funktion $g:[a,b]\to\mathbb{R}$, also

$$D^+g(t_0) \;:=\; \lim_{t\downarrow t_0}\,\frac{g(t)-g(t_0)}{t-t_0}. \tag{3.38}$$

Ein wichtiges Beispiel für einseitig differenzierbare Funktionen sind Normen.

Satz (3.39) (Über die Ableitung von Normen)

Sei $y : [a,b] \to \mathbb{R}^n$ eine C^1-Funktion und $\|\cdot\|$ eine Norm auf dem $\mathbb{R}^n$. Dann ist die Funktion $g(t) := \|y(t)\|$ auf $[a,b[$ rechtsseitig differenzierbar und es gilt

$$D^+\|y(t)\| \;\leq\; \|y'(t)\|. \tag{3.40}$$

Beweis: Für $y,\, u \in \mathbb{R}^n$, $h > 0$ und $0 < \mu \leq 1$ gilt aufgrund der Dreiecksungleichung

$$\begin{aligned}
& \|y + \mu h u\| \;-\; \|\mu y + \mu h u\| \;\leq\; \|y - \mu y\| \;=\; (1-\mu)\|y\| \\
\Longrightarrow\;\; & \|y + \mu h u\| \;-\; \|y\| \;\leq\; \|\mu y + \mu h u\| \;-\; \mu\|y\| \\
\Longrightarrow\;\; & \frac{\|y + \mu h u\| \;-\; \|y\|}{\mu h} \;\leq\; \frac{\|y + h u\| \;-\; \|y\|}{h}.
\end{aligned}$$

Damit sieht man, dass die Funktion $(\|y + hu\| - \|y\|)/h$ bzgl. $h > 0$ monoton wächst. Sie ist auch nach unten beschränkt, denn

$$\frac{\|y + h u\| \;-\; \|y\|}{h} \;\geq\; \frac{\|y\| - h\|u\| - \|y\|}{h} \;=\; -\|u\|.$$

Damit existiert der Grenzwert $\lim_{h \downarrow 0} \left(\|y + hu\| - \|y\|\right)/h$ und damit für $t \in [a,b[$ und $0 < h < b - t$ auch der Grenzwert

$$\lim_{h \downarrow 0} \frac{\|y(t) + h y'(t)\| - \|y(t)\|}{h} \;\leq\; \|y'(t)\|. \tag{3.41}$$

Schließlich folgt nun mit

$$\begin{aligned}
& \left| \left(\frac{\|y(t+h)\| - \|y(t)\|}{h} \right) \;-\; \left(\frac{\|y(t) + h y'(t)\| - \|y(t)\|}{h} \right) \right| \\
= \;\; & (1/h)\, |\, \|y(t+h)\| \;-\; \|y(t) + h y'(t)\|\, | \\
\leq \;\; & (1/h)\, \|y(t+h) \;-\; y(t) \;-\; h y'(t)\| \;\longrightarrow\; 0 \;\; (h \downarrow 0),
\end{aligned}$$

dass auch der Grenzwert $\lim_{h \downarrow 0} \left(\|y(t+h\|) - \|y(t)\|\right)/h$ existiert und die in (3.41) angegebenen Abschätzung genügt. □

Satz (3.42) (Vergleichssatz)

Sei $\omega : [t_0, t_0 + a] \times \mathbb{R} \to \mathbb{R}$ stetig. Die AWA $u' = \omega(t,u)$, $u(t_0) = u_0$ habe eine eindeutige Lösung u, die auf $[t_0, t_0 + a]$ definiert sei.

Gilt dann für eine stetige, rechtsseitig differenzierbare Funktion $v : [t_0, t_0 + a] \to \mathbb{R}$ die Differentialungleichung $D^+ v(t) \leq \omega(t, v(t))$, $t_0 \leq t < t_0 + a$, und $v(t_0) \leq u_0$, so folgt hieraus für alle $t \in [t_0, t_0 + a]$: $v(t) \leq u(t)$.

Beweis : Wir betrachten die folgende Familie von AWAen (der obere Index $m \in \mathbb{N}$ bezeichnet den Folgenindex!)

$$u' \;=\; \omega^{(m)}(t,u) \;:=\; \omega(t,u) + 1/m, \quad u^{(m)}(t_0) \;=\; u_0.$$

Die Folge $(\omega^{(m)})$ konvergiert auf $[\,t_0, t_0+a] \times \mathbb{R}$ gleichmäßig gegen ω. Mit dem Stabilitätssatz (3.18) folgt, dass $u^{(m)}$ für hinreichend große m auf $[\,t_0, t_0+a]$ definiert ist und gleichmäßig gegen u konvergiert.

Wir zeigen, dass $v(t) \le u^{(m)}(t)$ für alle $t \in [\,t_0, t_0+a]$ und hinreichend große m gilt. Wegen der gleichmäßigen Konvergenz folgt hieraus dann die Behauptung für u. Die Ungleichung gilt für $t = t_0$. Wäre sie nicht für alle t gültig, so gäbe es $t_0 < t_1 < t_2$ mit $v(t_1) = u^{(m)}(t_1)$ und $v(t) > u^{(m)}(t)$ für alle $t \in]t_1, t_2]$. Für diese t folgt damit $v(t) - v(t_1) > u^{(m)}(t) - u^{(m)}(t_1)$. Damit erhält man mittels Grenzübergang

$$D^+v(t_1) \ge \omega(t_1, u^{(m)}(t_1)) + 1/m = \omega(t_1, v(t_1)) + 1/m > \omega(t_1, v(t_1)),$$

im Widerspruch zur Voraussetzung. □

Beispiel (3.43) Wir betrachten die folgende AWA für eine Riccatische DGL

$$y' \;=\; t^2 + y^2, \quad y(0) \;=\; 1.$$

Eine untere Schranke für $y(t)$ erhält man durch Verkleinerung der rechten Seite, etwa zu $u' = u^2$, $u(0) = 1$. Der Vergleichssatz besagt dann, dass auf dem gemeinsamen Existenzintervall $y(t) \ge 1/(1-t)$ gelten muss. Damit ist auch klar, dass y *spätestens* in $t = 1$ eine Singularität besitzen muss, genauer: das maximale Existenzintervall (nach rechts) ist $[0, t_{\max}[$ mit $t_{\max} \le 1$.

Eine obere Schranke für $y(t)$ erhält man durch Vergrößerung der rechten Seite auf $[0,1]$, etwa zu $v' = 1 + v^2$, $v(0) = 1$. Der Vergleichssatz besagt dann wiederum, dass auf dem gemeinsamen Existenzintervall $y(t) \le \tan(t + \pi/4)$ gelten muss. Damit ist aber auch klar, dass y *frühestens* in $t = \pi/4$ eine Singularität besitzen kann, d.h. $\pi/4 \le t_{\max} \le 1$.

Die numerisch berechnete Lösung der AWA ist zusammen mit oberer und unterer Schranke in Abbildung 3.3 dargestellt. Die (numerisch bestimmte) Singularität der Lösung liegt bei $t_{\max} \approx 0.9698106539$.

Aus den Sätzen (3.39) und (3.42) lässt sich die folgende Eindeutigkeitsaussage ableiten.

Satz (3.44) (Eindeutigkeit)

Seien $f : [t_0, t_0+a] \times \mathbb{R}^n \to \mathbb{R}^n$ und $\omega : [t_0, t_0+a] \times [0, \infty[\to [0, \infty[$ stetige Funktionen mit $\omega(t,0) = 0$ und

$$\forall\, \widetilde{y},\ y \;\in \mathbb{R}^n : \quad \|f(t,\widetilde{y}) - f(t,y)\| \;\le\; \omega(t, \|\widetilde{y} - y\|). \tag{3.45}$$

Ferner habe die AWA $u' = \omega(t,u)$, $u(t_0) = 0$ auf $[t_0, t_0+a]$ die eindeutige Lösung $u = 0$.

Sind dann $\widetilde{y}$ und y auf $[t_0, t_0+a]$ definierte Lösungen der AWA (3.1), so folgt $\widetilde{y} = y$.

Beweis: Aufgrund der Voraussetzung ist $\widetilde{y}' - y' = f(t,\widetilde{y}) - f(t,y)$. Mit (3.39) folgt hieraus

$$D^+\|\widetilde{y} - y\| \leq \|\widetilde{y}' - y'\| = \|f(t,\widetilde{y}) - f(t,y)\| \leq \omega(t, \|\widetilde{y} - y\|).$$

Der Vergleichssatz (3.42) ergibt dann $v(t) := \|\widetilde{y}(t) - y(t)\| \leq u(t) = 0$ und somit $\forall t \in [t_0, t_0 + a] : \widetilde{y}(t) = y(t)$. □

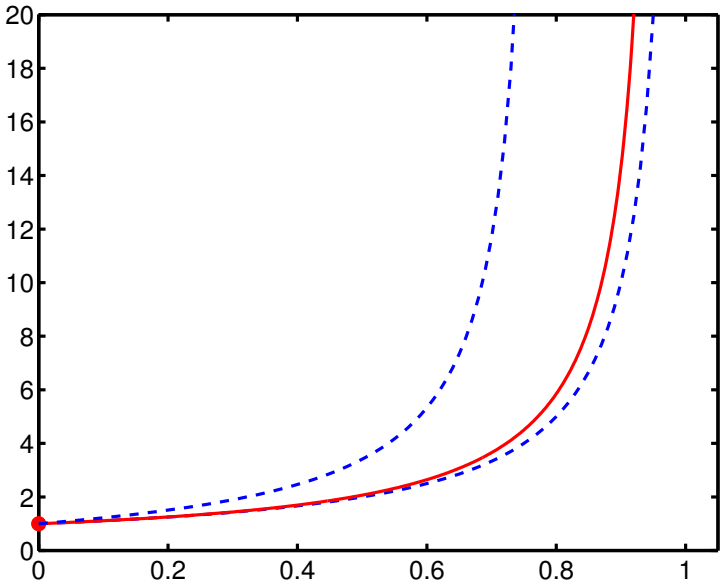

Abb. 3.3. Einschließung der Lösung von Beispiel (3.43)

Bemerkung (3.46) Die Eindeutigkeitsaussage (3.44) gilt analog für ein Intervall links von t_0. Man kann dies etwa mit Hilfe der folgenden Transformation zeigen:

$$z(t) := y(2\,t_0 - t), \quad g(t,z) := -f(2\,t_0 - t, z).$$

Der Vergleichssatz lässt sich zusammen mit der Fortsetzungseigenschaft (vgl. Anmerkungen zum Satz von Peano) zu einer Aussage über die *globale Existenz* kombinieren.

Satz (3.47) (Globale Existenz)

Sei $f : [t_0, t_0+a]\times\mathbb{R}^n \to \mathbb{R}^n$ stetig . Es gebe eine stetige Funktion $\omega : [t_0, t_0+a]\times[0,\infty[\to [0,\infty[$ so dass die AWA $u' = \omega(t,u),\ u(t_0) = u_0 > 0$ eine eindeutig bestimmte positive Lösung auf $[t_0, t_0 + a]$ besitzt.

Ist weiter für alle (t,y): $\|f(t,y)\| \leq \omega(t, \|y\|)$, so lässt sich jede Lösung y der AWA (3.1) mit $\|y_0\| \leq u_0$ auf $[t_0, t_0 + a[$ fortsetzen und es gilt $\|y(t)\| \leq u(t)$.

Beweis Hat eine Lösung y ein maximales Existenzintervall $[t_0, t_0 + \varepsilon[$ mit $\varepsilon < a$, so kann aufgrund des Fortsetzungssatzes $(t, y(t))$ für $t \uparrow t_0 + \varepsilon$ keinen endlichen Häufungspunkt

besitzen, insbesondere muss daher $\|y(t)\|$ für $t \to t_0 + \varepsilon$ unbeschränkt sein. Andererseits folgt aus (3.39)

$$D^+\|y(t)\| \;\leq\; \|y'(t)\| \;=\; \|f(t, y(t)\| \;\leq\; \omega(t, \|y(t)\|)$$

und damit nach dem Vergleichssatz (3.42) $\|y(t)\| \leq u(t)$. Widerspruch! □

4. Einschrittverfahren, insbesondere Runge–Kutta–Verfahren

A. Allgemeines.

Es geht in diesem Abschnitt um die numerische Lösung einer AWA

$$y'(t) \;=\; f(t, y(t)), \qquad y(t_0) \;=\; y_0\,. \tag{4.46}$$

Aufgabe ist es, zu vorgegebenem $t_b \neq t_0$ eine numerische Approximation für die Lösung $y(t_b)$ zu berechnen. O.B.d.A. sei hierbei $t_b > t_0$ und wir setzen voraus, dass f auf einem Gebiet $I \times \mathbb{R}^n$ mit $[t_0, t_b] \subset I$ hinreichend oft stetig differenzierbar ist und die AWA (4.1) auch eine (eindeutig bestimmte) Lösung y besitzt, die im gesamten Intervall $[t_0, t_b]$ erklärt ist. Mitunter wird eine Lipschitz-Konstante der rechten Seite f benötigt. Damit ist stets eine (lokale) Lipschitz-Konstante gemeint, die zu einem kompakten Quader $Q = [t_0, t_b] \times \widetilde{Q}$ gehört, der die Lösung umfasst, d.h. $y(t) \in \widetilde{Q}^0$, für alle $t \in [t_0, t_b]$.

An dieser Stelle ist eine besondere *Warnung* angebracht. In vielen Anwendungen tauchen DGLn auf, deren rechte Seite sich in gewissen Zeitpunkten nichtdifferenzierbar oder sogar unstetig ändert. Sie können beispielsweise von folgender Form sein

$$y'(t) \;=\; \begin{cases} f_1(t, y), & \text{falls } S(y(t)) \leq 0 \\ f_2(t, y), & \text{falls } S(y(t)) > 0. \end{cases} \tag{4.47}$$

Hierbei ist S eine so genannte *Schaltfunktion*. Ein klassisches Beispiel in der Mechanik ist das Phänomen der *trockenen Reibung*, bei der die Richtung der Reibungskraft von der Geschwindigkeitsrichtung abhängt. Die rechte Seite der zugehörigen DGL hängt damit vom Vorzeichen einer Zustandsgröße (abhängige Variable der DGL) ab. Eine solche DGL erfüllt jedenfalls die genannten Voraussetzungen *nicht*, und man hat besondere Vorkehrungen zu treffen, um diese numerisch zu lösen.

Numerische Integratoren arbeiten mit einer *Diskretisierung*, d.h., anstelle einer kontinuierlichen Lösung $y(t)$, $t_0 \leq t \leq t_b$ betrachtet man eine Zerlegung des Integrationsintervalls

$$t_0 \;<\; t_1 \;<\; \ldots \;<\; t_m \;=\; t_b \tag{4.48}$$

und Näherungen $Y_j \approx y(t_j),\ \ j = 0, 1, \ldots, m$.

Die t_j heißen *Integrationsknoten*, $I_h = \{t_0, \ldots, t_m\}$ heißt das *Integrationsgitter*, die $h_j := t_{j+1} - t_j$, $j = 0, \ldots, m-1$ heißen *Schrittweiten*, $\delta_h := \max h_j$ heißt die *Feinheit* des Gitters I_h.

Schließlich lassen sich die Näherungen auch als Funktionswerte einer so genannten *Gitterfunktion* $Y_h : I_h \to \mathbb{R}^n$ interpretieren. Unter einem *Diskretisierungsverfahren* versteht man dann eine Vorschrift, die jeden Gitter I_h eine Gitterfunktion Y_h zuordnet.

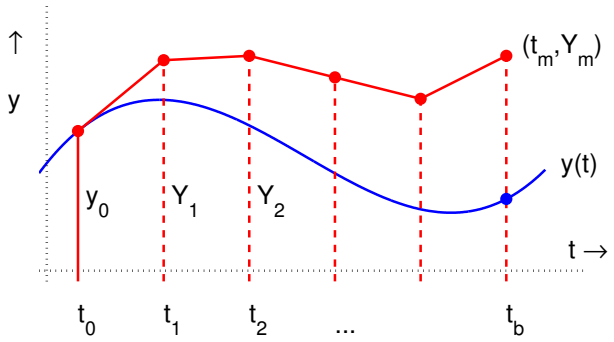

Abb. 4.1. Diskrete Lösung einer DGL

Die numerischen Verfahren zur Integration gewöhnlicher DGLn werden üblicherweise in die Klassen *Einschritt-, Mehrschritt- und Extrapolationsverfahren* unterteilt (wobei nicht immer eine scharfe Abgrenzung dieser Klassen möglich ist).

Einschrittverfahren (ESV) verwenden jeweils nur die zuletzt berechnete Näherung (t_j, Y_j), um hieraus eine nächste Näherung (t_{j+1}, Y_{j+1}) zu bestimmen. Sie haben die allgemeine Form

$$Y_{j+1} \;=\; Y_j + h_j\,\Phi\,(t_j, Y_j, Y_{j+1}; h_j)\,. \tag{4.49}$$

Die Funktion Φ heißt *Verfahrensfunktion* oder *Inkrementfunktion* des konkreten Verfahrens. Sie gibt die „Fortschreiterichtung“ eines Integrationsschrittes wieder. Sie hängt natürlich von der rechten Seite f der DGL ab und wird im Allgemeinen mit Hilfe mehrerer f–Auswertungen berechnet. Ist Φ unabhängig von Y_{j+1}, so definiert (4.4) ein *explizites ESV* – Y_{j+1} kann dann direkt mittels (4.4) ausgewertet werden, andernfalls ist (4.4) ein *implizites Verfahren*.

Mehrschrittverfahren (MSV) verwenden dagegen *mehrere* zuvor berechnete Näherungen $(t_i, Y_i),\ j < i < j+s$, um hieraus eine neue Näherung Y_{j+s} zu berechnen. Zumeist schränkt man sich dabei auf lineare Ansätze in den f_i–Daten (Quadraturformeln) und den Y_i–Daten (Differentiationsformeln) ein. Im Fall äquidistanter Schrittweiten erhält man somit den folgenden allgemeinen Ansatz für ein *lineares Mehrschrittverfahren*:

$$\sum_{i=0}^{s} \alpha_i\, Y_{j+i} \;=\; h \sum_{i=0}^{s} \beta_i\, f_{j+i}\,. \tag{4.50}$$

Dabei sind $\alpha_i,\ \beta_i \in \mathbb{R},\ i = 0,\ldots,s,\ \alpha_s \neq 0,\ t_i := a + ih,\ h := (b-a)/m$ und $f_i := f(t_i, Y_i),\ i = 0, 1, \ldots$.

Ist $\beta_s = 0$, so lässt sich (4.5) nach Y_{j+s} auflösen; man hat dann ein *explizites Verfahren*. Ist dagegen $\beta_s \neq 0$, so ist (4.5) eine implizite Gleichung zur (numerischen) Berechnung von Y_{j+s}, man spricht dann von einem *impliziten Verfahren*.

Extrapolationsverfahren beruhen auf einem, im Allgemeinen einfachen Ein– oder Mehrschrittverfahren. Es werden Näherungen für $y(t_b)$ zu *verschiedenen Schrittweiten* berechnet. Diese Näherungen werden durch Extrapolation bzgl. der Schrittweite verbessert.

In diesem Kapitel wollen wir uns zunächst mit expliziten ESV vom Typ (4.4) beschäftigen. Das einfachste ESV ist das bereits erwähnte *Eulersche Polygonzugverfahren*.

$$Y_{j+1} \;=\; Y_j + h_j\, f(t_j, Y_j)\,. \tag{4.51}$$

Man erhält das Verfahren, wenn man in der DGL $y'(t_j) = f(t_j, Y_j)$ die Ableitung durch den Vorwärts–Differenzenquotienten $(Y_{j+1} - Y_j)/h_j$ ersetzt.

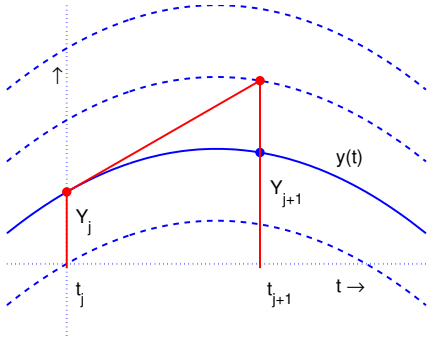

Abb. 4.2. Explizites Euler-Verfahren

Nimmt statt dessen den Rückwärts–Differenzenquotienten $(Y_j - Y_{j-1})/h_{j-1}$, so ergibt sich nach Indexverschiebung das so genannte *implizite Euler-Verfahren*

$$Y_{j+1} \;=\; Y_j + h_j\, f(t_{j+1}, Y_{j+1})\,. \tag{4.52}$$

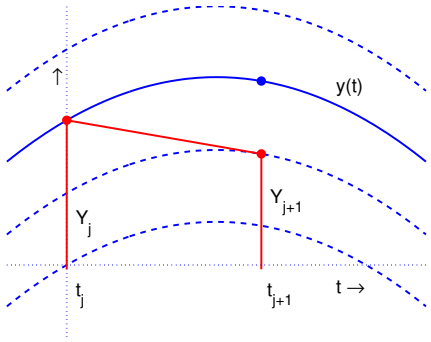

Abb. 4.3. Implizites Euler-Verfahren

Man beachte, dass man auch dieses Verfahren formal als ESV in der Form (4.4) schreiben kann, allerdings ist Y_{j+1} und damit auch $\Phi(t_j, Y_j; h_j)$ implizit durch die Beziehung (4.7) festgelegt. Diese ist im Übrigen eine Fixpunktgleichung, die für hinreichend kleine Integrationsschrittweiten kontrahiert.

Aus der geometrischen Bedeutung der beiden Euler-Verfahrens (die Fortschreiterichtung ist gleich der Tangentenrichtung im linken bzw. rechten Punkt) lassen sich sofort „Verbesserungen“ des Euler-Verfahrens finden:

Wählt man als Fortschreiterichtung etwa den Mittelwert zweier Steigungen, so erhält man das *Verfahren von Heun*:

$$\begin{aligned} k_1 &:= f(t_j, Y_j)\,, \\ k_2 &:= f(t_j + h_j, Y_j + h_j\, k_1)\,, \\ Y_{j+1} &:= Y_j + h_j \left(\frac{1}{2}\, k_1 + \frac{1}{2}\, k_2\right)\,. \end{aligned} \tag{4.53}$$

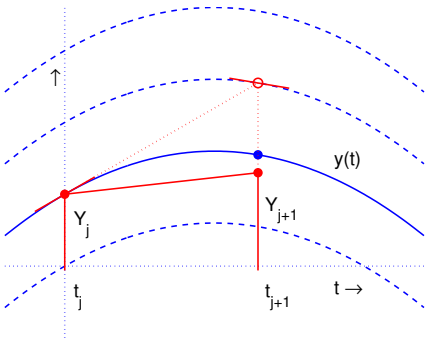

Abb. 4.4. Verfahren von Heun

Alternativ könnte man als Fortschreiterichtung auch eine mittlere Steigung wählen. Ein zugehöriges ESV ist beispielsweise das *modifizierte Euler-Verfahren* (Runge, 1895):

$$\begin{aligned} k_1 &:= f(t_j, Y_j)\,, \\ k_2 &:= f\left(t_j + \tfrac{1}{2}\, h_j, Y_j + h_j \left(\tfrac{1}{2}\, k_1\right)\right)\,, \\ Y_{j+1} &:= Y_j + h_j\, k_2\,. \end{aligned} \tag{4.54}$$

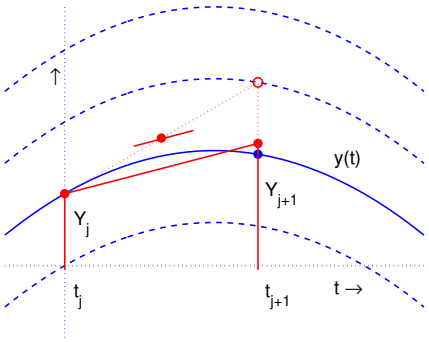

Abb. 4.5. Modifiziertes Euler-Verfahren

Beispiel (4.10) Wir greifen nochmal das Beispiel (3.43) auf und bestimmen numerisch die Lösung der AWA

$$y' \;=\; t^2 + y^2, \quad y(0) \;=\; 1.$$

im Punkt $t_b = 0.95$ mit Hilfe des expliziten Euler-Verfahrens, des Heunschen Verfahrens und des modifizierten Euler-Verfahrens. Wir verwenden zur Integration jeweils eine konstante Schrittweite und bestimmen den *relativen Fehler* der Näherungslösung in t_b. Der Referenzwert für die Lösung ist

$$y(t_b) \approx 0.50471867247946 \times 10^2.$$

Tabelle 4.1: Relative Fehler für Beispiel (4.10).

m	Schrittweite	Euler-Verf.	Heun-Verf.	Modif. Euler-V.
19	0.500D–01	0.82984D+00	0.46801D+00	0.51635D+00
95	0.100D–01	0.59076D+00	0.82046D–01	0.10688D+00
190	0.500D–02	0.44575D+00	0.25811D–01	0.35798D–01
950	0.100D–02	0.15551D+00	0.12034D–02	0.17809D–02
1900	0.500D–03	0.86164D–01	0.30536D–03	0.45585D–03
9500	0.100D–03	0.18896D–01	0.12350D–04	0.18564D–04
19000	0.500D–04	0.95643D–02	0.30915D–05	0.46510D–05
95000	0.100D–04	0.19319D–02	0.12379D–06	0.18636D–06
190000	0.500D–05	0.96718D–03	0.30951D–07	0.46600D–07

B. Konsistenz, Ordnung und Konvergenz.

Die Güte eines ESVs wird durch den so genannten *lokalen Diskretisierungsfehler* gemessen. Dieser gibt an, wie sich für einen einzelnen Integrationsschritt die Fortschreiterichtung $\Phi(t_j, Y_j; h_j)$ von der theoretisch exakten Fortschreiterichtung unterscheidet.

Definition (4.11)

Zu einer aktuellen Näherung $(t_j, Y_j) \in Q$ bezeichne $z(t)$, genauer $z(t; t_j, Y_j)$ die Lösung der *lokalen AWA*

$$z' \;=\; f(t, z(t)), \qquad z(t_j) \;=\; Y_j\,.$$

a) Zu hinreichend kleinem $h > 0$ heißt dann

$$\Delta(t_j, Y_j; h) \quad := \quad \frac{z(t_j + h) - Y_j}{h} \tag{4.12}$$

das *exakte Inkrement* oder *die exakte Fortschreiterichtung.*

b) Die Differenz zwischen exakter und numerischer Fortschreiterichtung

$$\tau(t_j, Y_j; h) \quad := \quad \Delta(t_j, Y_j; h) \;-\; \Phi(t_j, Y_j; h) \tag{4.13}$$

heißt der *lokale Diskretisierungsfehler* des ESVs.

c) Ein ESV heißt *konsistent*, falls für alle hinreichend oft stetig differenzierbaren rechten Seiten f und Näherungen $(t_j, Y_j) \in Q$ eine Abschätzung der folgenden Form gilt:

$$\| \tau(t_j, Y_j; h) \| \quad \leq \quad \sigma(h), \quad \text{mit } \sigma(h) \to 0 \ (h \to 0). \tag{4.14}$$

Hier und im Folgenden sei mit $\| \cdot \|$ die Maximumsnorm im $\mathbb{R}^n$ bezeichnet. Die Konsistenzbedingung fordert also die *gleichmäßige* Konvergenz des lokalen Diskretisierungsfehlers für alle Schrittweitenfolgen $h \to 0$.

d) Wir sagen, ein ESV besitzt die *Konsistenzordnung* $p \in \mathbb{N}$, falls gilt:

$$\| \tau(t_j, Y_j, h) \| \quad \leq \quad \sigma(h), \quad \text{mit } \sigma(h) \ = \ O(h^p), \tag{4.15}$$

d.h., es gibt nur von f und Q abhängige Konstante C, $h_0 > 0$, so dass gilt:
$\forall\, h \in]0, h_0] : \ \| \tau(t_j, Y_j; h) \| \ \leq \ C\, h^p$.

Bemerkungen (4.16)

a) Der lokale Diskretisierungsfehler wird mitunter in der Literatur etwas anders definiert, nämlich durch $\varepsilon := z(t_{j+1}) - Y_{j+1}$; dies entspricht gerade dem lokalen Fehler in den Funktionswerten. Dieser hängt aber auch direkt mit dem Fehler in den Steigungen (unsere Definition) zusammen. Man erhält nämlich durch einfache Umformung mittels (4.4) und (4.12)

$$z(t_{j+1}) - Y_{j+1} \quad = \quad h_j \, \tau(t_j, Y_j; h_j) \, ,$$

d.h., τ ist der (absolute) *Integrationsfehler pro Schrittweite (local error per unit step).*

b) Eine andere Interpretation des lokalen Diskretisierungsfehlers ist die folgende: τ ist das Residuum, welches man erhält, wenn man in der Relation (4.4)

$$\frac{Y_{j+1} - Y_j}{h} \quad = \quad \Phi(t_j, Y_j, h)$$

Y_{j+1} durch die exakte (lokale) Lösung $z(t_{j+1})$ ersetzt.

Zur Bestimmung der Ordnung eines vorgegebenen Einschrittverfahrens vergleicht man die Taylor–Entwicklungen von $\Phi(t, Y; h)$ bezüglich h (Entwicklungspunkt: $h = 0$) mit der entsprechenden Entwicklung von $\Delta(t, Y; h)$.

Für $\Delta(t, Y; h)$ finden wir mittels Taylor–Entwicklung von $z(t + h)$ um $h = 0$

$$\begin{aligned} \Delta(t, Y; h) \ &= \ \frac{z(t+h) - Y}{h} \ = \ \frac{z(t+h; t, Y) - Y}{h} \\ &= \ z'(t) \ + \ \frac{h}{2!}\, z''(t) \ + \ \frac{h^2}{3!}\, z'''(t) \ + \ \ldots \end{aligned}$$

Hierin verwenden wir nun die vorgegebene DGL $z' = f(t,z)$, die wir, so oft wie benötigt, mittels Kettenregel weiter differenzieren, also

$$\begin{aligned} z''(t) &= f_t(t,z) + f_y(t,z)\cdot f(t,z)\,, \quad z(t) = Y\,, \\ z'''(t) &= f_{tt} + 2\,f_{ty}\,f + f_{yy}\,(f,f) + f_y\,f_t + f_y\,f_y\,f \end{aligned}$$

und so fort. Damit finden wir:

$$\begin{aligned} \Delta &= f + \frac{h}{2}\,(f_t + f_y\,f) \\ &+ \frac{h^2}{6}\,(f_{tt} + 2f_{ty}\,f + f_{yy}\,(f,f) + f_y\,f_t + f_y\,f_y\,f) + O(h^3)\,. \end{aligned} \tag{4.17}$$

Hierbei sind f und sämtliche partiellen Ableitungen von f (außer denen im hier nicht angegebenem Restglied) jeweils im aktuellen Bezugspunkt (t,Y) auszuwerten.

Ferner ist zu beachten, dass es sich sowohl bei f wie bei y um vektorwertige Funktionen handeln kann. Der Term $f_{yy}\,(f,f)$ beispielsweise ist in Koordinaten folgendermaßen zu lesen: $\sum\limits_{k,\ell} \frac{\partial^2 f}{\partial y_k \partial y_\ell} f_k f_\ell$. Analog ist $f_y\,f_y\,f = \sum\limits_{k,\ell} \frac{\partial f}{\partial y_k}\frac{\partial f_k}{\partial y_\ell}\,f_\ell\,.$

Beispiele (4.18)

a) Für das Euler–Verfahren ist $\Phi = f(t,Y) = f$, also:

$$\tau \;=\; \Delta - \Phi \;=\; \frac{h}{2}\,(f_t + f_y\,f) + O(h^2)\,.$$

Das Euler–Verfahren ist also konsistent und hat die Ordnung $p = 1$.

b) Für das Heun–Verfahren erhält man durch Taylor–Entwicklung

$$\begin{aligned} \Phi &= \frac{1}{2}\,f(t,Y) + \frac{1}{2}\,f\,(t+h, Y + h\,f(t,Y)) \\ &= f + h\left\{\frac{1}{2}\,(f_t + f_y f)\right\} + \frac{h^2}{2}\left\{\frac{1}{2}\,(f_{tt} + 2f_{ty}f + f_{yy}(f,f))\right\} + O(h^3)\,. \end{aligned}$$

Zusammen mit (14) ergibt sich

$$\tau \;=\; \Delta - \Phi \;=\; h^2\left\{\frac{1}{6}\,(f_y f_t + f_y f_y f) - \frac{1}{12}\,(f_{tt} + 2f_{ty}f + f_{yy}(f,f))\right\} + O(h^3)\,.$$

Das Heun–Verfahren ist also konsistent und besitzt die Ordnung $p = 2$.

Aufgabe: Berechnen Sie genauso den führenden Term des lokalen Diskretisierungsfehlers für das modifizierte Euler–Verfahren.

Satz (4.19) (Konvergenzsatz)

Die Lösung y der AWA (4.1) existiere im Intervall $t_0 \le t \le t_b$. Ein ESV sei konsistent und besitze die Ordnung p, es gelte also $\| \tau(t, Y; h)\| \le C\, h^p$.
Ferner sei die Verfahrensfunktion Φ des ESVs auf dem Quader Q Lipschitz–stetig bezüglich der Variablen Y:

$$\|\Phi(t, \widetilde{Y}; h) - \Phi(t, Y; h)\| \quad \le \quad L_\Phi \, \|\widetilde{Y} - Y\| \, .$$

für alle (t, Y), $(t, \widetilde{Y}) \in Q$ und hinreichend kleinen Schrittweiten $h > 0$.

Dann liegen alle auf einem hinreichend feinen Gitter I_h berechneten Näherungen (t_j, Y_j) im Quader Q und für die Näherungen $Y_m = Y(t_b; I_h)$ im Endpunkt t_b gelten:

$$\|Y(t_b; I_h) - y(t_b)\| \;\le\; \frac{C}{L_\Phi} \left(\mathrm{e}^{L_\Phi (t_b - t_0)} - 1\right) \cdot h_{\max}^p \, . \tag{4.20}$$

Beweis:

Wir schätzen für einen Integrationsschritt $t_j \to t_{j+1}$ mit Schrittweite $h_j > 0$ ab

$$\begin{aligned}
\|Y_{j+1} - y(t_{j+1})\| &= \| \left(Y_j + h_j \, \Phi(t_j, Y_j; h_j)\right) \, - \, \left(y(t_j) + h_j \, \Delta(t_j, Y_j; h_j)\right) \| \\
&= \|(Y_j - y(t_j)) \, + \, h_j \, \left(\Phi(t_j, Y_j; h_j) - \Phi(t_j, y(t_j); h_j)\right) \, + \\
&\qquad h_j \, \left(\Phi(t_j, y(t_j); h_j) - \Delta(t_j, y(t_j); h_j)\right) \| \\
&\le (1 + h_j \, L_\Phi) \, \|Y_j - y(t_j)\| \, + \, h_j \, \| \tau(t_j, y(t_j); h_j)\| \\
&\le \mathrm{e}^{L_\Phi (t_{j+1} - t_j)} \, \|Y_j - y(t_j)\| \, + \, h_j \, \| \tau(t_j, y(t_j); h_j)\| \, .
\end{aligned}$$

Setzt man diese Abschätzung nun iterativ ineinander ein und beachtet $Y_0 = y(t_0) = y_0$, so erhält man

$$\|Y_{j+1} - y(t_{j+1})\| \;\le\; \sum_{k=0}^{j} \mathrm{e}^{L_\Phi (t_{j+1} - t_{k+1})} \, h_k \, \| \tau(t_k, y(t_k); h_k)\| \, .$$

Mittels vollständiger Induktion zeigt man nun die Abschätzung (Übungsaufgabe!)

$$\sum_{k=0}^{j} \mathrm{e}^{L_\Phi (t_{j+1} - t_{k+1})} \, h_k \;\le\; \frac{1}{L_\Phi} \left(\mathrm{e}^{L_\Phi (t_{j+1} - t_0)} - 1\right) ,$$

woraus sich zusammen mit der vorausgesetzten Ordungseigenschaft schließlich ergibt:

$$\|Y_{j+1} - y(t_{j+1})\| \;\le\; \frac{C}{L_\Phi} \left(\mathrm{e}^{L_\Phi (t_{j+1} - t_0)} - 1\right) \cdot h_{\max}^p \, .$$

Insbesondere lässt sich also - in Abhängigkeit der Konstanten C, L_Φ und der Integrationslänge $(t_b - t_0)$ - eine Schrittweite $h > 0$ angeben, so dass die diskrete Lösung zu jedem Gitter I_h mit Feinheit $h_{\max} \le h$ ganz in einem ε-Streifen verläuft, der selbst im vorgebenen Quader Q liegt.

Für $j = m - 1$ ergibt sich ferner die gewünschte Abschätzung (4.20). □

Bemerkungen (4.21)

a) Der Satz (4.19) zeigt, dass die (lokale) Konsistenzordung mit der globalen Konvergenzordnung übereinstimmt (*Konsistenzordnung = Konvergenzordnung*). Im Übrigen ist jedes konsistente ESV auch konvergent. Diese Aussage lässt sich, wie wir sehen werden, auf Mehrschrittverfahren nicht übertragen!

b) Aus (4.20) folgt unmittelbar die mitunter nützliche, jedoch etwas schwächere Abschätzung

$$\|Y(t_b; I_h) - y(t_b)\| \;\leq\; C\,(t_b - t_0)\,\mathrm{e}^{L_\Phi (t_b - t_0)} \cdot h_{\max}^p\,. \tag{4.22}$$

c) Für viele ESV impliziert die lokale Lipschitz-Eigenschaft der rechten Seite f auch unmittelbar die (lokale) Lipschitz-Stetigkeit der Verfahrensfunktion.

So folgt beispielsweise für das Verfahren von Heun (4.8) mit $\Phi(t, Y; h) = 0.5\,(f(t, Y) + f(t + h, Y + hf(t, Y)))$ die Abschätzung

$$\|\Phi(t, \widetilde{Y}; h) - \Phi(t, Y; h)\| \;\leq\; 0.5\,L\,\|\widetilde{Y} - Y\| \;+\; 0.5\,L\,\|\widetilde{Y}^* - Y^*\|,$$

wobei $\widetilde{Y}^* := \widetilde{Y} + hf(t, \widetilde{Y})$, $Y^* := Y + hf(t, Y)$ und L eine lokale Lipschitzkonstante der rechten Seite f bezeichnen. Damit ergibt sich weiter

$$\|\widetilde{Y}^* - Y^*\| \;\leq\; (1 + h\,L)\,\|\widetilde{Y} - Y\|$$

und somit insgesamt

$$\|\Phi(t, \widetilde{Y}; h) - \Phi(t, Y; h)\| \;\leq\; (L \;+\; 0.5\,h\,L^2)\;\|\widetilde{Y} - Y\|\,.$$

$L_\Phi := L + 0.5\,(t_b - t_0)\,L^2$ ist also eine (von h unabhängige) lokale Lipschitz-Konstante der Verfahrensfunktion Φ.

Auf die gleiche Art lässt sich auch für die Verfahrensfunktionen der Runge-Kutta Methoden (vgl. Abschnitt D) die lokale Lipschitz-Stetigkeit zeigen.

d) Man könnte versuchen, aus der Abschätzung (4.20) des Konvergenzsatzes eine optimale, äquidistante Schrittweite zu berechnen. Hierzu würde man fordern

$$\frac{C}{L_\Phi}\,\left(\mathrm{e}^{L_\Phi (t_b - t_0)} - 1\right) \cdot h^p \;=\; \|Y_h\|_\infty \cdot \mathrm{tol},$$

wobei tol (von Toleranz) eine (vom Benutzer) vorzugebende Schranke für den relativen Fehler bezeichnet und $\|Y_h\|_\infty$ die Maximumnorm der Gitterfunktion ist, also $\|Y_h\|_\infty := \max\{\|Y_j\| :\; j = 0, \ldots, m\}$.

Aus dem obigen Ansatz würde man also die folgende Formel für die Schrittweite erhalten:

$$h \;=\; \left[\,\frac{L_\Phi\,\|Y_h\|_\infty\,\mathrm{tol}}{C\,(\mathrm{e}^{L_\Phi (t_b - t_0)} - 1)}\,\right]^{1/p}\,. \tag{4.23}$$

Natürlich ist diese Beziehung für die praktische Wahl der Schrittweite wenig hilfreich, da die Größen C und L_Φ im Allgemeinen kaum abgeschätzt werden können. Sie zeigt jedoch ein Phänomen, dem wir bereits im Beispiel (4.10) begegnet sind: Je größer die Ordnung des Verfahrens ist, desto größere Schrittweiten werden wir im Allgemeinen verwenden können, um eine vorgegebene Genauigkeit zu erreichen. Insbesondere scheint es also numerisch sinnvoll zu sein, Verfahren höherer Ordung zu konstruieren.

C. Rundungsfehler.

Nach der Abschätzung (4.20) des Konvergenzsatzes konvergiert der absolute (aber auch der relative) Fehler der Näherungslösungen $Y(t_b, I_h)$ eines ESVs der Ordnung p für eine Gitterfolge mit $h_{\max} \to 0$ wie $h_{\max}^p$ gegen Null. Dabei sind wir von exakter Rechnung ausgegangen, haben also die bei der Durchführung auf einem Computer auftretenden Rundungsfehler vernachlässigt. Diese können natürlich insbesondere bei kleiner Schrittweite die erreichte Genauigkeit erheblich beeinflussen.

Wir wollen in diesem Abschnitt den Einfluss der Rundungsfehler überschlagsmäßig erfassen, wobei wir von exakter Gleitpunkt-Arithmetik ausgehen, d.h. alle elementaren Operationen $+, -, *, /$ werden mit einer relativen Genauigkeit durchgeführt, die durch eine universelle Konstante, der Maschinengenauigkeit eps, beschränkt ist. Diese liegt bei einfacher Genauigkeit bei eps $\approx 10^{-7}$, bei doppelter Genauigkeit etwa bei eps $\approx 10^{-14}$. Sind a, b Maschinenzahlen und ist $\circ \in \{+, -, *, /\}$, so gilt also für die auf einem Computer berechnete Verknüpfung (fl bezeichnet die in Gleitpunktrechnung ausgeführte Operation)

$$fl(a \circ b) \;=\; (a \circ b)\,(1+\varepsilon), \qquad |\varepsilon| \le \text{eps}.$$

Desweiteren nehmen wir an, dass die Verfahrensfunktion numerisch stabil ausgewertet werden kann, so dass für alle (t, Y, h) gilt

$$fl(\Phi(t, Y; h)) \;=\; \Phi(t, Y; h)\,(1+\alpha), \qquad \|\alpha\| \le K\,\text{eps}$$

mit einer nicht zu großen und von (t, Y, h) unabhängigen Konstanten $K > 0$.

Berücksicht man nun bei der Auswertung eines ESVs (4.4) die Rundungsfehler (einschließlich der Eingangsfehler!), so ergibt sich für die numerisch berechneten Näherungen $\widetilde{Y}_j,\ j = 0, \ldots, m,$ die folgende Rekursion:

$$\begin{aligned}
\widetilde{Y}_0 \quad &= \quad y_0 \,+\, \Delta y_0, \\[2ex]
\widetilde{Y}_{j+1} \quad &= \quad \Big[\widetilde{Y}_j + h_j\,\Phi(t_j, \widetilde{Y}_j; h_j)\,(1+\alpha_j)\,(1+\mu_j)\Big]\,(1+\sigma_j), \\[2ex]
&\text{mit} \qquad \|\,\alpha_j\| \;\le\; K\,\text{eps},\ \|\,\mu_j\|,\ \|\,\sigma_j\| \;\le\; \text{eps}.
\end{aligned}$$

Die Linearisierung des Fehlers, d.h. die Vernachlässigung aller Terme der Größenordnung eps^2 ergibt

$$\begin{aligned}
\widetilde{Y}_{j+1} \quad &= \quad \widetilde{Y}_j + h_j\,\Phi(t_j, \widetilde{Y}_j; h_j) \,+\, \varepsilon_{j+1} \\[2ex]
\varepsilon_{j+1} \quad &:= \quad \widetilde{Y}_j\,\sigma_j \,+\, h_j\,\Phi(t_j, \widetilde{Y}_j; h_j)\,[(1+\alpha_j)\,(1+\mu_j)\,(1+\sigma_j) \,-\, 1] \qquad (4.24) \\[2ex]
&\approx \quad \widetilde{Y}_j\,\sigma_j \,+\, h_j\,\Phi(t_j, \widetilde{Y}_j; h_j)\,(\alpha_j + \mu_j + \sigma_j).
\end{aligned}$$

Ist, wie wir bereits angenommen hatten, K nicht zu groß und sind die Schrittweiten h_j so klein, dass die Terme $h_j\Phi$ gegenüber $\widetilde{Y}_j$ vernachlässigt werden können, so kann man in erster Näherung abschätzen:

$$\|\,\varepsilon_{j+1}\| \;\le\; \|\widetilde{Y}_h\|_\infty\ \text{eps}. \qquad (4.25)$$

Von der ersten Gleichung in (4.24) wird nun die exakte Rekursion (4.4) subtrahiert. Wir erhalten:

$$\widetilde{Y}_{j+1} - Y_{j+1} \;=\; (\widetilde{Y}_j - Y_j) \;+\; h_j\,(\Phi(t_j, \widetilde{Y}_j; h_j) \;-\; \Phi(t_j, Y_j; h_j)) \;+\; \varepsilon_{j+1}$$

und damit – mit Hilfe der Lipschitz-Bedingung für Φ:

$$\begin{aligned} \|\widetilde{Y}_{j+1} - Y_{j+1}\| \;&\leq\; (1 + h_j\,L_\Phi)\,\|\widetilde{Y}_j - Y_j\| \;+\; \|\varepsilon_{j+1}\| \\ &\leq\; \mathrm{e}^{L_\Phi(t_{j+1}-t_j)}\,\|\widetilde{Y}_j - Y_j\| \;+\; \|\varepsilon_{j+1}\|. \end{aligned}$$

Diese Abschätzung ist genau von der Art, wie wir sie im Beweis des Konvergenzsatzes (4.19) kennengelernt haben. Mit gleicher Technik (ineinander einsetzen!) folgt daher

$$\begin{aligned} \|\widetilde{Y}_{j+1} - Y_{j+1}\| \;&\leq\; \mathrm{e}^{L_\Phi(t_{j+1}-t_0)}\,\|\Delta y_0\| \;+\; \sum_{k=0}^{j} \mathrm{e}^{L_\Phi(t_{j+1}-t_{k+1})}\,\|\,\varepsilon_{k+1}\| \\ &\leq\; \mathrm{e}^{L_\Phi(t_{j+1}-t_0)}\,\|\Delta y_0\| \;+\; \frac{1}{L_\Phi}\,\left(\mathrm{e}^{L_\Phi(t_{j+1}-t_0)} - 1\right) \cdot \max_k \frac{\|\varepsilon_{k+1}\|}{h_k}\,. \end{aligned}$$

Zur Abschätzung des tatsächlichen Fehlers $\|\widetilde{Y}_{j+1} - y(t_{j+1})\|$ verwenden wir nun noch die Beziehung (4.20) sowie die Dreieckungleichung. Wir fassen das Ergebnis im folgenden Satz zusammen:

Satz (4.26) (Rundungsfehler bei ESV)

Die Lösung y der AWA (4.1) existiere im Intervall $t_0 \leq t \leq t_b$. Ein ESV (4.4) sei konsistent und besitze die Ordnung p, es gelte also $\|\,\tau(t, Y; h)\| \leq C\,h^p$. Ferner sei die Verfahrensfunktion Φ auf dem Quader Q Lipschitz–stetig bezüglich der Variablen Y mit der Lipschitz-Konstanten L_Φ .

Liegen dann die auf einem hinreichend feinen Gitter I_h numerisch berechneten Näherungen $(t_j, \widetilde{Y}_j)$ im Quader Q, so gilt für die Näherung $\widetilde{Y}_m = \widetilde{Y}(t_b; I_h)$ im Endpunkt t_b die Abschätzung:

$$\begin{aligned} \|\widetilde{Y}(t_b; I_h) - y(t_b)\| \;&\leq\; \mathrm{e}^{L_\Phi(t_b-t_0)}\,\|\Delta y_0\| \;+ \\ &+\; \frac{1}{L_\Phi}\,\left(\mathrm{e}^{L_\Phi(t_b-t_0)} - 1\right) \cdot \left(C\,h_{\max}^p \;+\; \max_j \frac{\|\varepsilon_{j+1}\|}{h_j}\right)\,. \end{aligned} \tag{4.27}$$

In der Beziehung (4.27) beschreibt Δy_0 den absoluten Fehler im Anfangswert (*Einlesefehler*), der Term mit $C\,h_{\max}^p$ beschreibt den *Diskretisierungsfehler* und schließlich der Term mit $\|\varepsilon_{j+1}\|/h_j$ den *Rundungsfehlereinfluß*.

Unter den bei (4.25) genannten Voraussetzungen lässt sich der Ausdruck in (4.27) nochmals vereinfachen zu

$$\begin{aligned} \|\widetilde{Y}(t_b; I_h) - y(t_b)\| \;&\leq\; \mathrm{e}^{L_\Phi(t_b-t_0)}\,\|\Delta y_0\| \;+ \\ &+\; \frac{1}{L_\Phi}\,\left(\mathrm{e}^{L_\Phi(t_b-t_0)} - 1\right) \cdot \left(C\,h_{\max}^p \;+\; \frac{\|\widetilde{Y}_h\|_\infty\,\mathrm{eps}}{h_{\min}}\right)\,. \end{aligned} \tag{4.28}$$

Für den Fall äquidistanter Schrittweite (also $h_{\min} = h_{\max}$) sind die Ausdrücke in der rechten Klammer von (4.28) in der Abbildung 4.6 qualitativ wiedergegeben.

Man erkennt, dass es bei vorgegebener Maschinengenauigkeit eine Grenzgenauigkeit und eine zugehörige optimale Schrittweite gibt, bis zu der der Gesamtfehler des numerischen Ergebnisses fällt. Bei weiterer Verkleinerung der Schrittweite wächst jedoch der Fehler dann aufgrund der Rundungsfehler wieder an.

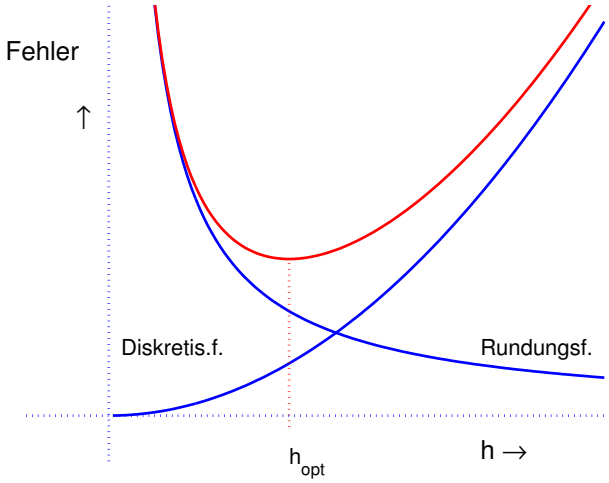

Abb. 4.6. Gesamtfehler bei ESV

D. Runge–Kutta–Verfahren.

Die meistgebräuchlichen Einschrittverfahren sind die so genannten Runge–Kutta–Verfahren (kurz: RK–Verfahren) benannt nach Carl Runge (1856–1927) und Martin Wilhelm Kutta (1867–1927). Es handelt sich dabei um ESV, die die Verfahrensfunktion als Linearkombination von Auswertungen der rechten Seite f ansetzen. Insoweit sind dies direkte Verallgemeinerungen des Heunschen Verfahrens (4.8) bzw. des modifizierten Euler-Verfahrens (4.9). Erste Verfahren dieser Art wurden von Runge (1895), Heun (1900) und Kutta (1901) angegeben. Letzterer gab auch *das klassische Runge-Kutta Verfahren* vierter Ordnung an. Erst fünfzig Jahre später bemühte man sich um die Konstruktion von RK–Verfahren höherer Ordnung.

Die allgemeine Form eines (expliziten) RK–Verfahrens lautet:

$$\begin{aligned} Y_{j+1} &= Y_j + h_j \sum_{i=1}^{s} b_i\, k_i(t_j, Y_j; h_j) \\ k_1(t, Y; h) &= f(t, Y) \\ k_i(t, Y; h) &= f(t + c_i h, Y + h \sum_{\ell=1}^{i-1} a_{i\ell}\, k_\ell\,(t, Y; h))\,, \end{aligned} \tag{4.29}$$

Dabei heißt s die *Stufenzahl* des RK–Verfahrens, die c_i heißen *Knoten*, die b_i *Gewichte* und die (a_{ij}) werden zu einer *Verfahrensmatrix* zusammengefasst. Alle Koefizienten b_i, c_i und $a_{i\ell}$, welche ja das konkrete Verfahren festlegen, werden üblicherweise in einem Tableau, dem so genannten *Butcher–Schema* angeordnet:

Tabelle 4.2: Allgemeines Butcher–Schema.

0					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
$\vdots$	$\vdots$	$\vdots$	$\ddots$		
c_s	a_{s1}	a_{s2}	$\dots$	$a_{s,s-1}$	
	b_1	b_2	$\dots$	b_{s-1}	b_s

Beispiele hatten wir schon kennengelernt. So sind die Schemata für das Heun-Verfahren (4.8) bzw. für das modifizierte Euler-Verfahren (4.9) (für beide ist $s = 2$) in der Tabelle 4.3 angegeben.

Tabelle 4.3: Heun–Verfahren (p=2) und Modifiziertes Euler–Verfahren (p=2)

0		
1	1	
	$1/2$	$1/2$

0		
$1/2$	$1/2$	
	0	1

Für die so genannte *Kutta–Regel* und ein weiteres auf Heun zurückgehendes dreistufiges Verfahren (beide mit $s = p = 3$) sind die Schemata der Tabelle 4.4 zu entnehmen.

Tabelle 4.4: Kutta–Regel (p=3) und Heun–Verfahren (p=3)

0			
$1/2$	$1/2$		
1	-1	2	
	$1/6$	$2/3$	$1/6$

0			
$1/3$	$1/3$		
$2/3$	0	$2/3$	
	$1/4$	0	$3/4$

Zwei Beispiele vierter Ordnung gehen auf Kutta zurück. Dies ist zum Einen *das klassische Runge-Kutta Verfahren* RK4 und zum Andern die so genannte 3/8–Regel. Die Schemata sind in Tabelle 4.5 angegeben.

In allen bisher angegebenen Beispielen stimmt jeweils die Stufenzahl s mit der Ordnung p des Verfahrens überein. Dies ist allerding für Verfahren höherer Ordnung nicht mehr der Fall. So werden für ein RK–Verfahren der Ordung sieben bereits neun, für ein Verfahren der Ordnung acht sogar elf Stufen benötigt (Butcher, 1987).

Tabelle 4.5: Klassisches RK4–Verfahren und 3/8–Regel (p=4)

0				
1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	1/3	1/3	1/6

0				
1/3	1/3			
2/3	−1/3	1		
1	1	−1	1	
	1/8	3/8	3/8	1/8

Zur Konstruktion von RK–Verfahren hat man für die allgemeine Verfahrensfunktion nach (4.29):

$$\Phi(t, Y; h) \;:=\; \sum_{i=1}^{s} b_i \, k_i(t, Y; h)$$

einen Taylor–Abgleich mit dem exakten Inkrement $\Delta(t, Y; h)$ gemäß (4.12) und (4.17) durchzuführen.

Durch Abgleich des Absolutterms ($h = 0$) findet man beispielsweise sofort:

Satz (4.30)

Ein RK–Verfahren ist genau dann konsistent, falls $\sum\limits_{i=1}^{s} b_i \;=\; 1$.

Zumeist schränkt man sich bei der Aufstellung der Ordnungsbedingungen auf den Fall *autonomer* DGLn ein. Dies ist, wie der folgende Satz zeigt, gerechtfertigt, wenn die so genannte *Knotenbedingung* erfüllt ist:

$$c_i \;=\; \sum_{j=1}^{i-1} a_{ij} \,, \quad i = 2, \ldots s \,. \tag{4.31}$$

Satz (4.32)

Hat ein RK–Verfahren die Ordnung p für alle autonomen DGLn und ist die Knotenbedingung (4.31) erfüllt, so ist das Verfahren auch von gleicher Ordnung für nichtautonome DGLn.

Beweis: Eine nichtautonome AWA

$$z'(t) \;=\; g(t, z(t)), \quad z(t_0) \;=\; z_0$$

lässt sich wie folgt in ein autonomes Problem $y' = f(y),\ y(t_0) = y_0$ transformieren:

$$y(t) := \begin{pmatrix} t \\ z(t) \end{pmatrix}, \quad f(y) := \begin{pmatrix} 1 \\ g(y_1, y_2) \end{pmatrix}, \quad y_0 := \begin{pmatrix} t_0 \\ z_0 \end{pmatrix}.$$

Das RK–Verfahren für dieses Problem lautet

$$Y_{j+1} \;=\; Y_j \;+\; h_j \sum_{i=1}^{s} b_i\, k_i\,, \qquad k_i \;=\; f(Y_j + h_j \sum_{\ell=1}^{i-1} a_{i\ell}\, k_\ell) \tag{4.33}$$

Wir schreiben dies wieder in Koordinaten mit $Y = (t, Z)^{\mathrm{T}}$ und $k_i = (1, \widetilde{k}_i)^{\mathrm{T}}$ und erhalten

$$\begin{aligned} t_{j+1} &= t_j \;+\; h_j \sum_{i=1}^{s} b_i\,, \\ Z_{j+1} &= Z_j \;+\; h_j \sum_{i=1}^{s} b_i\, \widetilde{k}_i\,, \\ \widetilde{k}_i &= g(t_j + h_j \sum_{\ell=1}^{i-1} a_{i\ell}\,,\; Z_j + h_j \sum_{\ell=1}^{i-1} a_{i\ell}\, \widetilde{k}_\ell)\,. \end{aligned}$$

Wegen der Konsistenz (4.30) des Verfahrens und der vorausgesetzten Knotenbedingung (4.31) ist dies aber genau das RK–Verfahren für das nichtautonome Anfangswertproblem und dieses hat demnach die gleiche Konsistenzordnung wie (4.33). □

Die in den Tabellen 4.3–4.5 angegebenen RK-Verfahren erfüllen alle die Knotenbedingung, es genügt dort also zur Ordnungsbestimmung, sich auf autonome Differentialgleichungen einzuschränken.

Die Einschränkung auf autonome Differentialgleichungen bedeutet eine erhebliche Vereinfachung für das Aufstellen der Ordnungsbedingungen.

Beispiel (4.34) Will man beispielsweise die Ordnungsbedingungen für ein dreistufiges RK-Verfahren der Ordnung $p = 3$ aufstellen, so hat man die Funktionen

$$\begin{aligned} \Phi &= b_1\, k_1 + b_2\, k_3 + b_3\, k_3, \\ k_1 &= f(Y), \quad k_2 = f(Y + h\, a_{21}\, k_1), \\ k_3 &= f(Y + h\,(a_{31}\, k_1 + a_{32}\, k_2)) \end{aligned}$$

bzgl. der Schrittweite h in eine Taylor-Reihe zu entwickeln und diese bis zur Potenz h^2 mit der der exakten Inkrementfunktion

$$\Delta \;=\; f + \frac{h}{2}\,(f'f) + \frac{h^2}{6}\,(f''(f,f) + f'f'f) + O(h^3)\,,$$

abzugleichen. Wir schreiben hier f' anstelle von f_y, vgl. auch (4.17).

Die Terme $f,\ f'f,\ f''(f,f)$ und $f'f'f$ heißen *elementare Differentiale*. Sie treten ganz analog bei der Taylor–Entwicklung der Verfahrensfunktion auf. Man erhält

$$\begin{aligned} \Phi \;=\; & (\textstyle\sum b_i)\; f \;+\; h\,[b_2\, a_{21} + b_3(a_{31} + a_{32})]\; f'\, f \\ & + \frac{h^2}{2}\,[(b_2\, a_{21}^2 \;+\; b_3\,(a_{31} + a_{32})^2)\; f''(f,f) \;+\; 2\, b_3\, a_{21}\, a_{32}\, f'\, f'\, f] \;+\; O(h^3). \end{aligned}$$

Nun sind die elementaren Differentiale (bei hinreichend großer Dimension n) linear unabhängig (vgl. Lemma 4.25 in Deuflhard, Bornemann (2002)), so dass die Entwicklungen von Δ und Φ nicht nur bzgl. der h-Potenzen, sondern auch bzgl. der elementaren Differentiale übereinstimmen müssen.

Mittels Koeffizientenvergleichs erhalten wir damit die folgenden vier Ordnungsgleichungen für die sechs Unbekannten b_i, a_{ik}:

$$\begin{aligned} b_1 + b_2 + b_3 &= 1 \\ b_2\, a_{21} + b_3\,(a_{31} + a_{32}) &= 1/2 \\ b_2\, a_{21}^2 + b_3\,(a_{31} + a_{32})^2 &= 1/3 \\ b_3\, a_{21}\, a_{32} &= 1/6. \end{aligned} \tag{4.35}$$

Ein dreistufiges RK-Verfahren besitzt also genau dann die Konsistenzordnung $p = 3$, wenn das obige Gleichungssystem erfüllt ist. Man überzeugt sich unmittelbar, dass die in Tabelle 4.4 angegebenen dreistufigen RK-Verfahren dieses Gleichungsystem lösen, und damit die Konsistenzordnung $p = 3$ besitzen. Die Parameter c_i sind jeweils durch die Knotenbedingung (4.31) festgelegt.

E. Ordnungsbedingungen nach Butcher.

J.C. Butcher (1963) hat zur Aufstellung der Ordnungsgleichungen ein relativ einfaches graphentheoretisches Verfahren angegeben. Will man die Gleichungen dafür aufstellen, dass ein s-stufiges RK–Verfahren die Ordnung $\geq$ p besitzt, so hat man sämtliche, paarweise nicht isomorphen Wurzelbäume mit höchstens p Knoten aufzustellen. Diese Wurzelbäume entsprechen genau den elementaren Differentialen in den Taylor-Entwicklungen von Δ und Φ.

Wir benötigen einige Grundbegriffe aus der Graphentheorie, die wir zunächst hier zusammenstellen wollen. Ein *Graph* $\mathbf{g} = (P, K, v)$ besteht aus einer endlichen Menge $P = \{x_1, \ldots, x_q\}$ von Knoten (Punkte), einer endlichen Menge K von Kanten und einer Abbildung v mit

(i) für *ungerichtete Graphen*: $v : K \to \wp(P)$, $v(k) = \{a, b\}$ ($a = b$ ist zugelassen).

(ii) für *gerichtete Graphen*: $v : K \to P \times P$, $v(k) = (a, b)$. In diesem Fall heißt $v_A(k) := a$ der *Anfangspunkt* und $v_E(k) := b$ der *Endpunkt* der Kante k.

Aus jedem gerichteten Graphen lässt sich natürlich durch Vergessen der Richtungen ein ungerichteter Graph machen.

Mit $\sharp\,\mathbf{g}$ wird die Knotenzahl des Graphen $\mathbf{g}$ bezeichnet.

Zwei Graphen $\mathbf{g}_1 = (P_1, K_1, v_1)$ und $\mathbf{g}_2 = (P_2, K_2, v_2)$ heißen *isomorph*, falls es Bijektionen $\phi : P_1 \to P_2$ und $\psi : K_1 \to K_2$ mit der folgenden Eigenschaft gibt: Für jede Kante $k \in K_1$ mit $v_1(k) = \{a, b\}$ bzw. $v_1(k) = (a, b)$ ist $v_2(\psi(k)) = \{\phi(a), \phi(b)\}$ bzw. $v_2(\psi(k)) = (\phi(a), \phi(b))$.

Für einen gerichteten Graphen $\mathbf{g}$ und $x \in P$ heißt $g^-(x) := \sharp\{k : v_E(k) = x\}$ der *negative*

Grad (Zahl der einlaufenden Kanten) und $g^+(x) := \sharp\{k : v_A(k) = x\}$ der *positive Grad* des Knotens x (Zahl der auslaufenden Kanten). $g(x) := g^-(x) + g^+(x)$ heißt der *Grad* von x.

Eine endliche Folge $\omega = \langle x_1, k_1, x_2, \ldots, x_{m-1}, k_{m-1}, x_m \rangle$ von Knoten und Kanten heißt ein *ungerichteter* bzw. *gerichteter Kantenzug*, falls $v(k_i) = \{x_i, x_{i+1}\}$ bzw. $v(k_i) = (x_i, x_{i+1})$ für alle $i = 1, \ldots, m-1$. Wir sagen auch, der Kantenzug ω verbindet x_1 und x_m. Im Fall $x_1 = x_m$ heißt der Kantenzug ein *Kreis*.

Ein Graph **g** heißt *zusammenhängend*, falls je zwei (verschiedene) Knoten durch einen Kantenzug verbunden werden können.

Ein ungerichteter, zusammenhängender Graph ohne Kreise heißt ein *Baum*. Schließlich heißt ein gerichteter Graph ein *Wurzelbaum*, falls er als ungerichteter Graph ein Baum ist, und es einen ausgezeichneten Knoten $x_1 \in P$ gibt, die *Wurzel*, mit der Eigenschaft:

$$g^-(x_1) \;=\; 0 \quad \text{und} \quad \forall\, x \neq x_1 : \;\; g^-(x) \;=\; 1\,.$$

Mit dieser Eigenschaft ist klar, dass ein Wurzelbaum mit q Knoten genau $q-1$ Kanten besitzt.

Ist **g** ein Wurzelbaum und $x \in P$ ein Knoten von **g**, so erzeugt x einen Teil-Wurzelbaum von **g** mit der Wurzel x, der aus allen Knoten von **g** besteht, die durch einen gerichteten Kantenzug – von x ausgehend – erreicht werden können, zusammen mit den zugehörigen Kanten. Dieser von x erzeugte Teil-Wurzelbaum werde mit $[x]$ bezeichnet.

In der Abbildung 4.7 sind sämtliche paarweise nicht isomorphe Wurzelbäume mit maximal vier Knoten aufgezeichnet. Die Richtung der Kanten weist dabei stets von unten nach obe

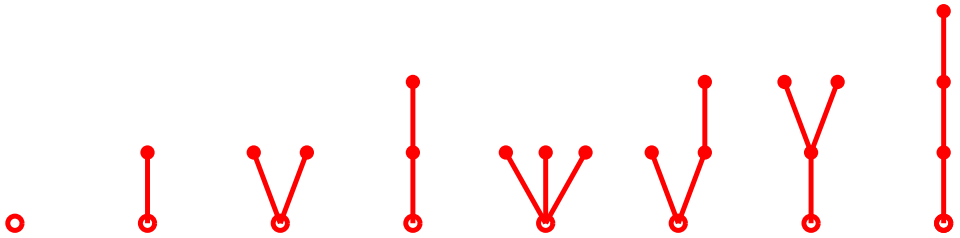

Abb. 4.7. Alle Wurzelbäume mit bis zu vier Knoten

Den Wurzelbäumen entsprechen in eineindeutiger Weise die elementaren Differentiale.

So gehören die in Abbildung 4.7 dargestellten Wurzelbäume (in dieser Reihenfolge) zu den folgenden elementaren Differentialen f, $f'\,f$, $f''(f,f)$, $f'\,f'\,f$, $f'''\,(f,f,f)$, $f''\,(f,f'\,f)$, $f'\,f''\,(f,f)$ und $f'\,f'\,f'\,f$.

Es sei nun **g** ein Wurzelbaum mit der Knotenmenge $P = \{x_1, \ldots, x_q\}$. x_1 sei die Wurzel von **g**. Zu einem Knoten $x_j \in P$ bezeichne J_j die Indizes der Nachfolgerknoten, also

$$J_j \;:=\; \{\ell : \;\; \exists\, k \in K : v(k) = (x_j, x_\ell)\,\}$$

Wir ordnen dem Wurzelbaum $\mathbf{g}$ nun den folgenden polynomialen Ausdruck in den RK–Koeffizienten $b_i\, a_{jk}$ zu:

$$p(\mathbf{g}) \;:=\; \sum_{i_1,\ldots,i_q=1}^{s} b_{i_1} \prod_{j=1}^{q} \left(\prod_{\ell \in J_j} a_{i_j\, i_\ell} \right) . \tag{4.36}$$

Hierbei ist zu beachten, dass, wie üblich, leere Produkte ($J_j = \emptyset$) Eins gesetzt werden. Ferner sind alle Koeffizienten a_{jk} mit $k \geq j$, die also im expliziten Butcher-Schema nicht auftreten, Null zu setzen.

Sodann wird dem Wurzelbaum $\mathbf{g}$ noch eine natürliche Zahl $\gamma(\mathbf{g})$ zugeordnet, nämlich:

$$\gamma(\mathbf{g}) \;:=\; \prod_{i=1}^{q} \sharp\,[x_i]\,. \tag{4.37}$$

Mit diesen Größen lässt sich nun der folgende Satz von Butcher (1963) formulieren:

Satz (4.38) (Ordungsbedingungen für RK–Verfahren)

Ein s-stufiges RK-Verfahren mit Koeffizienten ($b_i,\ a_{jk}$) und der Knotenbedingung (4.31) besitzt genau dann die Konsistenzordnung p, wenn für alle Wurzelbäume $\mathbf{g}$ mit maximal p Knoten gilt: $\quad p(\mathbf{g}) \;=\; 1/\gamma(\mathbf{g})$.

Beweise dieses Satzes findet man u.a. in den Lehrbüchern von Hairer, Norsett und Wanner, von Strehmel und Weiner, sowie von Deuflhard und Bornemann.

Wir wollen uns die Aussage des Butcherschen Satzes für den Fall eines RK–Verfahren der Ordnung vier ansehen. Hierzu sind genau die Wurzelbäume aus Abbildung 4.7 zu betrachten. Diese werden im Folgenden mit $\mathbf{g}_1, \ldots \mathbf{g}_8$ bezeichnet.

Es ergeben sich damit die folgenden acht Ordnungsbedingungen:

$$\begin{aligned}
p(\mathbf{g}_1) &= \sum_{i_1=1}^{s} b_{i_1} \;=\; \sum_i b_i \;= && \gamma(\mathbf{g}_1)^{-1} \;=\; 1,\\
p(\mathbf{g}_2) &= \sum_{i_1,i_2=1}^{s} b_{i_1} a_{i_1 i_2} \;=\; \sum_i b_i\, c_i \;= && \gamma(\mathbf{g}_2)^{-1} \;=\; 1/2,\\
p(\mathbf{g}_3) &= \sum_{i_1,i_2,i_3=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_1 i_3} \;=\; \sum_i b_i\, c_i^2 \;= && \gamma(\mathbf{g}_3)^{-1} \;=\; 1/3,\\
p(\mathbf{g}_4) &= \sum_{i_1,i_2,i_3=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_2 i_3} \;=\; \sum_{i,j} b_i\, a_{ij}\, c_j \;= && \gamma(\mathbf{g}_4)^{-1} \;=\; 1/6,\\
p(\mathbf{g}_5) &= \sum_{i_1,i_2,i_3,i_4=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_1 i_3} a_{i_1 i_4} \;=\; \sum_i b_i\, c_i^3 \;= && \gamma(\mathbf{g}_5)^{-1} \;=\; 1/4,\\
p(\mathbf{g}_6) &= \sum_{i_1,i_2,i_3,i_4=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_1 i_3} a_{i_3 i_4} \;=\; \sum_{i,j} b_i\, c_i\, a_{ij}\, c_j \;= && \gamma(\mathbf{g}_6)^{-1} \;=\; 1/8,
\end{aligned}$$

$$p(\mathbf{g}_7) \;=\; \sum_{i_1,i_2,i_3,i_4=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_2 i_3} a_{i_2 i_4} \;=\; \sum_{i,j} b_i\, a_{ij}\, c_j^2 \;=\; \gamma(\mathbf{g}_7)^{-1} \;=\; 1/12,$$

$$p(\mathbf{g}_8) \;=\; \sum_{i_1,i_2,i_3,i_4=1}^{s} b_{i_1} a_{i_1 i_2} a_{i_2 i_3} a_{i_3 i_4} \;=\; \sum_{i,j,k} b_i\, a_{ij}\, a_{jk}\, c_k \;=\; \gamma(\mathbf{g}_8)^{-1} \;=\; 1/24.$$

Man kann sich nun wiederum davon überzeugen, dass die in Tabelle 4.5 angegebenen vierstufigen RK–Verfahren tatsächlich diese Ordnungsgleichungen erfüllen.

Die Anzahl der Ordnungsgleichungen nimmt mit wachsender Ordnung p stark zu. So hat man für ein Verfahren der Ordnung sieben schon 85 nichtlineare Gleichungen zu lösen, wozu übrigens ein wenigstens neunstufiges RK–Verfahren benötigt wird. Für ein Verfahren der Ordnung zehn sind es sogar 1205 nichtlineare Gleichungen und man benötigt wenigstens 13 Stufen.

F. Schrittweitensteuerung.

Hierbei geht es um die automatische Generierung eines Integrationsgittes I_h, das einerseits fein genug sein soll, um eine vorgegebene Genauigkeit der numerischen Lösung zu garantieren, andererseits aber auch nicht feiner, um den numerischen Aufwand (dieser schließt auch die Gittererzeugung selbst ein) und den Einfluss von Rundungsfehlern möglichst gering zu halten.

Die Schrittweitensteuerung wird dabei ein *lokaler* Prozess sein, also im Allgemeinen eine *nicht äquidistante* Schrittweite generieren. Dass die Wahl einer äquidistanten Schrittweite über größere Integrationsdistanzen häufig zu einem unvertretbaren numerischen Aufwand führt, zeigt sehr eindrucksvoll das folgende Beispiel des restringierten Dreikörperproblems, dass als numerisches Testproblem für Anfangswertproblemlöser vielfach in der Literatur verwendet worden ist.

Beispiel (4.39) (Das restringierte Dreikörperproblem)

Wir betrachten die bereits in (1.18) vorgestellte AWA zur Beschreibung einer ebenen periodischen Satellitenbahn im Gravitationsfeld von Erde und Mond. Mit den dort genannten Bezeichnungen haben wir die AWA

$$\begin{aligned}
\ddot{x} &= x + 2\,\dot{y} - \hat{\mu}\,\frac{x+\mu}{[(x+\mu)^2+y^2]^{3/2}} - \mu\,\frac{x-\hat{\mu}}{[(x-\hat{\mu})^2+y^2]^{3/2}} \\
\ddot{y} &= y - 2\,\dot{x} - \hat{\mu}\,\frac{y}{[(x+\mu)^2+y^2]^{3/2}} - \mu\,\frac{y}{[(x-\hat{\mu})^2+y^2]^{3/2}}, \qquad (4.40)\\
x(0) &= 1.2, \quad y(0) = 0, \quad \dot{x}(0) = 0, \quad \dot{y}(0) = -1.049357510
\end{aligned}$$

Eine Lösung dieser AWA soll im Periodenintervall $[0, t_b]$ mit $t_b \doteq 6.1921\,69331$ berechnet werden.

Wir lösen die AWA (transformiert in ein System erster Ordnung) auf zwei Arten, wobei wir jeweils ein fünfstufiges RK–Verfahren der Ordnung vier anwenden, dass auf Fehlberg (1969) zurückgeht.

Zum Einen arbeiten wir mit der konstanten Schrittweite $h := t_b/1000$, führen also 1000 Integrationsschritte mit je fünf Auswertungen der rechten Seite aus. Die sich ergebenden Integrationspunkte sind in Abbildung 4.8 blau eingezeichnet. Man erkennt, dass jeweils bei den erdnahen Bereichen die Schrittweite noch zu groß ist und daher Integrationsfehler auftreten, die schließlich die Bahn völlig verfälschen.

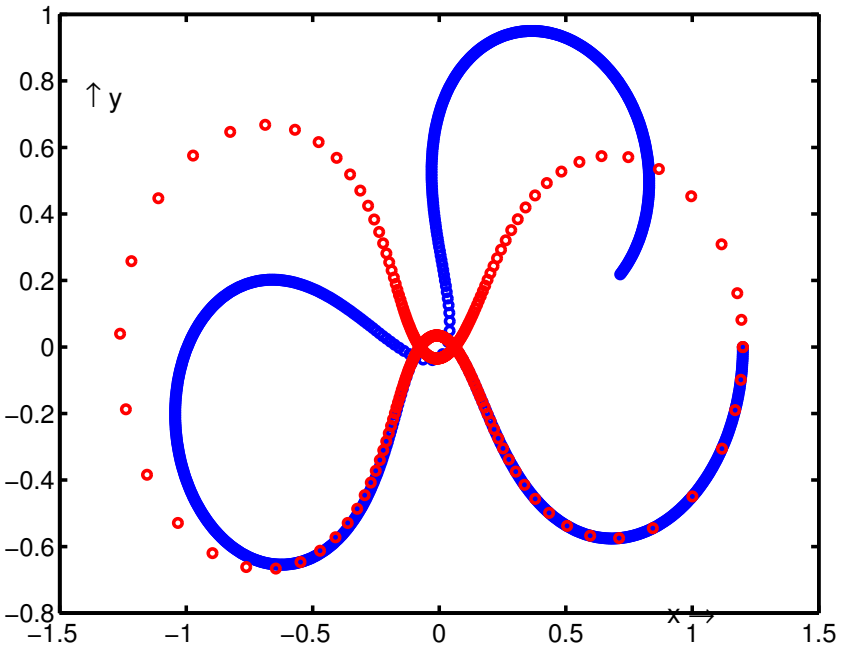

Abb. 4.8. Dreikörperproblem.

Zum Anderen arbeiten wir mit dem gleichen Integrationsverfahren, benutzen jedoch eine automatische Schrittweitensteuerung nach Fehlberg. Die sich ergebenden Integrationsknoten bei vorgegebener Toleranzanforderung tol $= 10^{-5}$ sind ebenfalls in Abbildung 3.8 eingezeichnet (rot). Man erkennt, dass sich die Bahn hier tatsächlich (im Rahmen der Zeichengenauigkeit) schließt. Im Endpunkt ergibt sich sogar ein relativer/absoluter Fehler $\leq 1.4 \times 10^{-4}$.

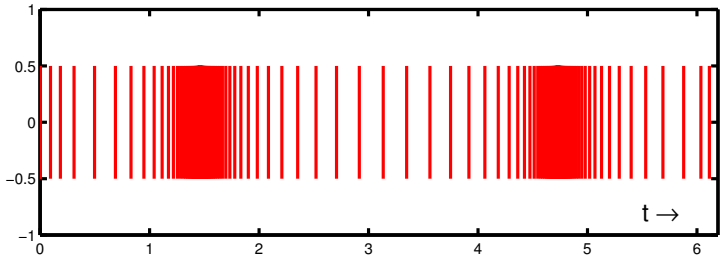

Abb. 4.9. Integrationsgitter.

In der Abbildung 4.9 ist das von der Schrittweitensteuerung erzeugte Gitter dargestellt. Die von der Schrittweitensteuerung erzeugten Schrittweiten sind fern der Erde relativ groß (Größenordnung ≈ 0.3) und werden erdnahe auf etwa 2×10^{-4} reduziert. Das schrittweitengesteuerte Verfahren kommt insgesamt mit nur 2196 Auswertungen der rechten Seite aus, ist also weniger als halb so teuer wie unsere Rechnung mit konstanter Schrittweite!

Die Algorithmen zur Schrittweitensteuerung arbeiten mit einer *Schätzung des lokalen Diskretisierungsfehlers*, d.h. zu jedem Integrationsschritt

$$(t_j, Y_j) \;\rightarrow\; (t_j + h, Y_j + h\Phi(t_j, Y_j; h))$$

mit einer aktuellen Schrittweite $h > 0$ wird zugleich ein (numerisch berechenbarer!) Schätzwert τ_{est} (est von Estimation) ermittelt:

$$\tau_{\text{est}}(t_j, Y_j; h) \;\approx\; \tau(t_j, Y_j; h) \;=\; (1/h)\;[y(t_{j+1}; t_j, Y_j) - Y_j - h\Phi(t_j, Y_j; h)]\;.$$

Von einer optimalen lokalen Schrittweite h_j^* wird nun mit einer vom Benutzer vorzugebenden Genauigkeitsanforderung tol (von Toleranz) gefordert:

$$\| \tau(t_j, Y_j; h_j^*) \| \;=\; \text{tol}.$$

Zusammen mit der Ordnungseigenschaft: $\tau(t_j, Y_j; h) = C(t_j)\, h^p + O(h^{p+1})$ folgt hiermit die folgende Heuristik:

$$\begin{aligned} \text{tol} \;&=\; \| \tau(t_j, Y_j; h_j^*) \| \;\approx\; \| C(t_j) \| \, (h_j^*)^p \;=\; \| C(t_j)\, h^p \| \, (h_j^*/h)^p \\ &\approx\; \| \tau(t_j, Y_j; h) \| \, (h_j^*/h)^p \;\approx\; \| \tau_{\text{est}} \| \, (h_j^*/h)^p \end{aligned}$$

und somit

$$h_j^* \;\approx\; \left(\frac{\text{tol}}{\| \tau_{\text{est}} \|} \right)^{(1/p)} \; h\;. \tag{4.41}$$

Diese Beziehung muss für die praktische Anwendung noch modifiziert werden. Die Schrittweite wird dazu etwas kleiner gewählt (Sicherheitsfaktor $q \in]0, 1[$) als optimal, um die Zahl der nicht erfolgreichen Integrationsschritte ($\|\tau_{\text{est}}\| > \text{tol}$) klein zu halten. Ferner werden Schranken $0 < \nu < 1 < \mu$ eingeführt, mit denen ein starkes Oszillieren der Schrittweite vermieden werden soll. Insgesamt erhält man folgenden Grundalgorithmus zur adaptiven Schrittweitenwahl.

Algorithmus (4.42) (Schrittweitensteuerung)

Start: Toleranzschranke: $\text{tol} > 0$, Parameter: $q,\ \nu \in]0, 1[,\ \mu > 1$,
$j := 0$, $Y_0 := y_0$, Startschrittweite: h_0 mit $0 < h_0 \le (t_b - t_0)$,
Minimale Schrittweite: $h_{\min} > 0$;

Iteration: $Y_{j+1} := Y_j + h_j\, \Phi(t_j, Y_j; h_j)$, $\tau_{\text{est}}(t_j, Y_j; h_j)$,
$h := q\, (\text{tol}/\|\tau_{\text{est}}\|)^{(1/p)}\, h_j$,
$h := \max[\, \min(h,\ \mu\, h_j),\ \nu\, h_j\,]$, Falls: $h < h_{\min}$, Stop!

Falls: $\|\tau_{\text{est}}\| > \text{tol}$ (Integrationsschritt wird verworfen)
$h_j := h$, gehe zu Iteration;

Sonst: (Integrationsschritt wird akzeptiert)
$t_{j+1} := t_j + h_j, \quad h_{j+1} := \min(h, t_b - t_{j+1}), \quad j := j+1;$
Falls: $t_j = t_b$, Stop! Sonst: Gehe zu Iteration.

Natürlich gibt es in den professionellen Realisierungen der Schrittweitensteuerung verschiedene Varianten und Verfeinungen des obigen Grundalgorithmus. So werden häufig Skalierungen verwendet und es wird an Stelle einer universellen Toleranzanforderung tol (hier absoluter Fehler pro Schrittweite) mit relativen und absoluten Genauigkeitsforderungen gearbeitet, die auch komponentenweise unterschiedlich vorgegeben werden können.

G. Eingebettete Runge-Kutta Verfahren.

Eine effiziente Methode, den lokalen Diskretisierungsfehler zu schätzen, besteht in der Verwendung so genannter eingebetteter RK-Verfahren. Hierunter versteht man ein Paar von RK-Verfahren mit gemeinsamen Knoten c_i, gemeinsamer Verfahrensmatrix (a_{ij}), aber unterschiedlichen Gewichten.

Tabelle 4.6: Eingebettete RK-Verfahren.

0					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
$\vdots$	$\vdots$	$\vdots$	$\ddots$		
c_s	a_{s1}	a_{s2}	$\dots$	$a_{s,s-1}$	
	b_1	b_2	$\dots$	b_{s-1}	b_s
	$\widehat{b}_1$	$\widehat{b}_2$	$\dots$	$\widehat{b}_{s-1}$	$\widehat{b}_s$

Das Verfahren $\Phi(t, Y; h) = \sum b_i\, k_i$ habe hierbei die Konsistenzordnung p, das zweite Verfahren $\widehat{\Phi}(t, Y; h) = \sum \widehat{b}_i k_i$ habe die Konsistenzordung $\widehat{p}$, wobei üblicherweise $\widehat{p} = p-1$ oder $\widehat{p} = p+1$ ist. Man bezeichnet solche Verfahren auch kurz mit RK $p\,(\widehat{p})$. Das Verfahren Φ ist das eigentliche Integrationsverfahren, das Verfahren $\widehat{\Phi}$ dient zur Schätzung des lokalen Diskretisierungsverfahrens:

$$\tau_{\text{est}}(t_j, Y_j; h_j) \;:=\; \sum_{i=1}^{s} (\widehat{b}_i - b_i)\, k_i(t_j, Y_j; h_j)\,. \tag{4.42}$$

Die ersten eingebetteten RK–Verfahren sind von Merson (1957), Ceschino (1962) und Zonneveld (1963) konstruiert worden. Viele Verfahren dieser Klasse , die in den Anwendungen

besonders erfolgreich warnen und sind, gehen auf Fehlberg zurück. In Tabelle 4.7 sind die Koeffizienten des Fehlbergschen RKF4(5) Verfahrens angegeben. Dieses Verfahren haben wir zur Lösung des Beispiels (4.39) verwendet.

Tabelle 4.7: RK4(5)–Verfahren von Fehlberg (1969)

0						
$\frac{1}{4}$	$\frac{1}{4}$					
$\frac{3}{8}$	$\frac{3}{32}$	$\frac{9}{32}$				
$\frac{12}{13}$	$\frac{1932}{2197}$	$-\frac{7200}{2197}$	$\frac{7296}{2197}$			
1	$\frac{439}{216}$	-8	$\frac{3680}{513}$	$-\frac{845}{4104}$		
$\frac{1}{2}$	$-\frac{8}{27}$	2	$-\frac{3544}{2565}$	$\frac{1859}{4104}$	$-\frac{11}{40}$	
	$\frac{25}{216}$	0	$\frac{1408}{2565}$	$\frac{2197}{4104}$	$-\frac{1}{5}$	0
	$\frac{16}{135}$	0	$\frac{6656}{12825}$	$\frac{28561}{56430}$	$-\frac{9}{50}$	$\frac{2}{55}$

Ein weiteres außerordentlich erfolgreiches Verfahren ist ein RK7(8) Verfahren von Fehlberg, dessen Koeffizienten in Tabelle 4.8 angegeben sind.

Alle von Fehlberg angegebenen Verfahren sind vom Typ RK$p(q)$ mit $q > p$. Dabei ist das Verfahren der Konsistenzordnung p das eigentliche Integrationsverfahren, das Verfahren höherer Ordnung, in der Regel $q = p + 1$ dient lediglich zur Schrittweitensteuerung. Die Verfahren sind daher auch so konzipiert worden, dass die Abbrechfehler der Verfahren niedrigerer Ordnung möglichst kleine Koeffizienten (Faktoren bei den elementaren Differentialen) besitzen.

Dormand und Price (1980) bemühten sich statt dessen, eingebettete RK-Verfahren zu konstruieren, bei denen die Fehlerkoeffizienten des Verfahrens *höherer* Ordnung minimiert werden und verwenden natürlich dann auch dieses Verfahren als eigentliches Integrationsverfahren. Sie konstruierten so die vielfach verwendeten Verfahren vom Typ RK5(4) und RK8(7), genannt DOPRI5 und DOPRI8. Die Koeffizienten beider Verfahren (respektive eine genaue rationale Approximation dieser) findet man bei Deuflhard, Bornemann. Anwendungen dieser Verfahren auf das restringierte Dreikörperproblem sind in Hairer, Norsett und Wanner beschrieben.

Tabelle 4.8: RK7(8)–Verfahren von Fehlberg (1968)

0													
$\frac{2}{27}$	$\frac{2}{27}$												
$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{12}$											
$\frac{1}{6}$	$\frac{1}{24}$	0	$\frac{1}{8}$										
$\frac{5}{12}$	$\frac{5}{12}$	0	$-\frac{25}{16}$	$\frac{25}{16}$									
$\frac{1}{2}$	$\frac{1}{20}$	0	0	$\frac{1}{4}$	$\frac{1}{5}$								
$\frac{5}{6}$	$-\frac{25}{108}$	0	0	$\frac{125}{108}$	$-\frac{65}{27}$	$\frac{125}{54}$							
$\frac{1}{6}$	$\frac{31}{300}$	0	0	0	$\frac{61}{225}$	$-\frac{2}{9}$	$\frac{13}{900}$						
$\frac{2}{3}$	2	0	0	$-\frac{53}{6}$	$\frac{704}{45}$	$-\frac{107}{9}$	$\frac{67}{90}$	3					
$\frac{1}{3}$	$-\frac{91}{108}$	0	0	$\frac{23}{108}$	$-\frac{976}{135}$	$\frac{311}{54}$	$-\frac{19}{60}$	$\frac{17}{6}$	$-\frac{1}{12}$				
1	$\frac{2383}{4100}$	0	0	$-\frac{341}{164}$	$\frac{4496}{1025}$	$-\frac{301}{82}$	$\frac{2133}{4100}$	$\frac{45}{82}$	$\frac{45}{164}$	$\frac{18}{41}$			
0	$\frac{3}{205}$	0	0	0	0	$-\frac{6}{41}$	$-\frac{3}{205}$	$-\frac{3}{41}$	$\frac{3}{41}$	$\frac{6}{41}$	0		
1	$-\frac{1777}{4100}$	0	0	$-\frac{341}{164}$	$\frac{4496}{1025}$	$-\frac{289}{82}$	$\frac{2193}{4100}$	$\frac{51}{82}$	$\frac{33}{164}$	$\frac{12}{41}$	0	1	
	$\frac{41}{840}$	0	0	0	0	$\frac{34}{105}$	$\frac{9}{35}$	$\frac{9}{35}$	$\frac{9}{280}$	$\frac{9}{280}$	$\frac{41}{840}$	0	0
	0	0	0	0	0	$\frac{34}{105}$	$\frac{9}{35}$	$\frac{9}{35}$	$\frac{9}{280}$	$\frac{9}{280}$	0	$\frac{41}{840}$	$\frac{41}{840}$

5. Mehrschrittverfahren

A. Konstruktion von Mehrschrittverfahren.

Die Grundidee von Mehrschrittverfahren besteht darin, mehrere bisher berechnete Näherungen (t_k, Y_k), $k = j,\ j-1, \ldots, j-s$ zur Berechnung einer neuen Näherung Y_{j+1} zu verwenden. Man hat damit die allgemeine Form eine Mehrschrittverfahrens

$$Y_{j+1} \ = \ \Psi(Y_{j+1}, Y_j, \ldots, Y_{j-s}; h_j)\ , \ \ j = s,\ s+1,\ \ldots \tag{5.43}$$

Man unterscheidet hierbei *explizite Verfahren*, auch *Prediktor–Verfahren* genannt und *implizite Verfahren*, welche auch *Korrektor–Verfahren* genannt werden. Bei expliziten Verfahren hängt Ψ nicht von Y_{j+1} ab, so dass (1) direkt ausgewertet werden kann. Bei impliziten Verfahren stellt die Gleichung (5.1) dagegen ein im Allg. nichtlineares Gleichungssystem zu Bestimmung von Y_{j+1} dar.

Adams–Verfahren.

Viele spezielle Verfahren ergeben sich dabei durch formale Integration der Differentialgleichung $y' = f(t,y)$ über dem Knotenbereich $[\, t_{j-k}, t_{j+\ell}\,]$, $k,\ \ell \geq 0$:

$$y(t_{j+\ell}) \ - \ y(t_{j-k}) \ = \ \int\limits_{t_{j-k}}^{t_{j+\ell}} f(t, y(t))\, dt\ . \tag{5.44}$$

Hierin ersetzt man nun den Integranden durch ein Interpolationspolynom $P_s \in \Pi_s$ vom Maximalgrad s zu den Stützstellen $(t_i, f(t_i, Y_i)) \ =: \ (t_i, f_i),\ i = j, j-1, \ldots, j-s$.

Wir verwenden die Lagrange–Darstellung des Interpolationspolynoms, die sich allerdings nur im äquidistanten Fall bewähren wird. Hiernach ist

$$P_s(t) \ = \ \sum_{i=0}^{s} f_{j-i}\, L_i(t), \quad L_i(t) \ = \ \prod_{\substack{p=0\\ p\neq i}}^{s} \left(\frac{t - t_{j-p}}{t_{j-i} - t_{j-p}} \right)$$

und damit

$$Y_{j+\ell} - Y_{j-k} \ = \ \sum_{i=0}^{s} f_{j-i} \int\limits_{t_{j-k}}^{t_{j+\ell}} \prod_{\substack{p=0\\ p\neq i}}^{s} \left(\frac{t - t_{j-p}}{t_{j-i} - t_{j-p}} \right)\ dt\ . \tag{5.45}$$

Wir vereinfachen diesen Ausdruck nun für den Fall *äquidistanter Integrationsschritte* $t_i = t_0 + i\, h,\ i = 1, 2, \ldots$.

Mit der Transformation $t = t_j + \tau\, h, \quad dt = h\, d\tau$ erhält man

$$\frac{t - t_{j-p}}{t_{j-i} - t_{j-p}} \ = \ \frac{(t_j + \tau\, h) - (t_j - p\, h)}{(t_j - i\, h) - (t_j - p\, h)} \ = \ \frac{\tau + p}{-i + p}\ .$$

Damit lautet das Verfahren

$$\begin{aligned} Y_{j+\ell} &= Y_{j-k} + h \sum_{i=0}^{s} b_{i,s}\, f_{j-i}, \quad f_{j-i} := f(t_{j-i}, Y_{j-i}) \\ b_{i,s} &= \int_{-k}^{\ell} \prod_{\substack{p=0 \\ p\neq i}}^{s} \left(\frac{\tau + p}{-i + p} \right) d\tau, \qquad i = 0,\, 1,\, \ldots\,,\, s\,. \end{aligned} \tag{5.46}$$

Wir notieren einige spezielle Verfahren:

I. Verfahren nach Adams–Bashforth (1883) $(\ell = 1,\ k = 0)$

$$\begin{aligned} Y_{j+1} &= Y_j + h \sum_{i=0}^{s} b_{i,s}\, f_{j-i} \qquad \text{(explizites Verfahren)} \\ b_{i,s} &= \int_{0}^{1} \prod_{\substack{p=0 \\ p\neq i}}^{s} \left(\frac{p + \tau}{p - i} \right) d\tau \;=:\; \beta_{i,s}/\gamma_s \end{aligned} \tag{5.47}$$

s	γ_s	$\beta_{i,s}$			
0	1	1			
1	2	3	-1		
2	12	23	-16	5	
3	24	55	-59	37	-9

II. Verfahren nach Adams–Moulton (1926) $(\ell = 0,\ k = 1)$

$$\begin{aligned} Y_j &= Y_{j-1} + h \sum_{i=0}^{s} b^*_{i,s}\, f_{j-i} \qquad \text{(implizites Verfahren)} \\ b^*_{i,s} &= \int_{-1}^{0} \prod_{\substack{p=0 \\ p\neq i}}^{s} \left(\frac{p + \tau}{p - i} \right) d\tau \;=:\; \beta^*_{i,s}/\gamma_s \end{aligned} \tag{5.48}$$

s	γ_s	$\beta^*_{i,s}$			
0	1	1			
1	2	1	1		
2	12	5	8	-1	
3	24	9	19	-5	1

III. Verfahren nach Nyström (1925) $(\ell = 1,\ k = 1)$

$$\begin{aligned} Y_{j+1} &= Y_{j-1} + h \sum_{i=0}^{s} a_{i,s}\, f_{j-i} \qquad \text{(explizites Verfahren)} \\ a_{i,s} &= \int_{-1}^{1} \prod_{\substack{p=0 \\ p\neq i}}^{s} \left(\frac{p+\tau}{p-i}\right) d\tau \;=:\; \alpha_{i,s}/\delta_s \end{aligned} \tag{5.49}$$

s	δ_s	$\alpha_{i,s}$			
0	1	2			
1	1	2	0		
2	3	7	−2	1	
3	3	8	−5	4	−1

IV. Verfahren nach Milne–Simpson $(\ell = 0,\ k = 2)$

$$\begin{aligned} Y_{j} &= Y_{j-2} + h \sum_{i=0}^{s} a^*_{i,s}\, f_{j-i} \qquad \text{(implizites Verfahren)} \\ a^*_{i,s} &= \int_{-2}^{0} \prod_{\substack{p=0 \\ p\neq i}}^{s} \left(\frac{p+\tau}{p-i}\right) d\tau \;=:\; \alpha^*_{i,s}/\delta_s \end{aligned} \tag{5.50}$$

s	δ_s	$\alpha^*_{i,s}$				
0	1	2				
1	1	0	2			
2	3	1	4	1		
3	3	1	4	1	0	
4	90	29	124	24	4	−1

Einige bekannte Verfahren finden wir in den Tabellen wieder. So liefert der Adams-Bashforth Ansatz für $s = 0$ gerade das explizite Euler–Verfahren, Adams-Moulton liefert für $s = 0$ das (zugehörige) implizite Euler-Verfahren. Die Nyström–Verfahren für $s = 0$ und $s = 1$ ergeben ebenso wie das Milne–Simpson–Verfahren für $s = 1$ die so genannte *Mittelpunktsregel* $Y_{j+1} = Y_{j-1} + 2\,h\,f_j$. Das Adams–Moulton–Verfahren für $s = 1$ entspricht bei Quadraturen gerade der *Trapezregel*, das Milne–Simpson–Verfahren für $s = 2$ und $s = 3$ entspricht bei Quadraturen der *Simpson–Regel* oder *Keplerschen Fassregel*.

Mehrschrittverfahren verlangen im Unterschied zu den Einschrittverfahren beim Start eine Anlaufrechnung, mit der zunächst einmal der Beginn der Datenkette $Y_0,\ Y_1, \ldots, Y_s$

berechnet werden muss, vgl. (5.1). Hierzu lässt sich ein Einschrittverfahren entsprechender Ordnung oder ein Mehrschrittverfahren kleinerer Schrittzahl verwenden.

Schließlich werden aus Stabilitätsgründen häufig explizite und implizite Mehrschrittverfahren gekoppelt. Dabei wird das implizite Verfahren zumeist mittel *Fixpunkt-Iteration* gelöst, also etwa für die Adams–Moulton–Verfahren:

$$Y_j^{k+1} \;=\; Y_{j-1} \;+\; h\,\left\{ b_{0,s}^*\, f(t_j, Y_j^k) \;+\; \sum_{i=1}^{s} b_{i,s}^*\, f_{j-i} \right\}, \qquad k = 0,\, 1, \ldots \tag{5.51}$$

Für hinreichend kleine Schrittweiten ist die rechte Seite von (5.9) bzgl. Y_j^k kontrahierend, die Fixpunktiteration also konvergent. Dennoch werden häufig nur wenige Iterationen von (5.9) durchgeführt, mitunter nur eine einzige. Der Startwert wird dabei mit dem zugehörigen expliziten Verfahren berechnet.

Für die Kopplung der Adams–Verfahren erhält man beispielsweise das folgene Verfahren für einen Integrationsschritt, wenn man nur einen Schritt der Fixpunktiteration ausführt:

$$\begin{aligned}
Y_{j+1}^0 \;&:=\; Y_j \;+\; h \sum_{i=0}^{s} b_{i,s}\, f_{j-i} && \text{(Predictor)}\\
f_{j+1}^0 \;&:=\; f(t_{j+1}, Y_{j+1}^0) && \text{(Evaluation)}\\
Y_{j+1} \;&:=\; Y_j \;+\; h\,\left\{ b_{0,s}^*\, f_{j+1}^0 \;+\; \sum_{i=1}^{s+1} b_{i,s}^*\, f_{j+1-i} \right\}, && \text{(Corrector)}\\
f_{j+1} \;&:=\; f(t_{j+1}, Y_{j+1}) && \text{(Evaluation)}
\end{aligned}$$

Man spricht dann von einem PECE-Verfahren und auch allgemeiner bei mehreren Fixpunktschritten von einem P(EC)^{m}E–Verfahren.

BDF–Verfahren.

Eine andere Möglichkeit, Mehrschrittverfahren zu gewinnen, besteht darin, anstelle der Approximation des Integrals in (5.2) eine Approximation der Ableitung $y'(t_{j+1})$ durch numerische Differentiation aus y_{j+1} und den bisher berechneten y_k–Werten zu verwenden. Man spricht daher von *Rückwärts-Differentiationsformeln (backward differentiation formulas, kurz BDF-Verfahren).*

Sei also $p_s \in \Pi_s$ das Interpolationspolynom zu den Stützstellen (t_{j+1-k}, Y_{j+1-k}), $k = 0, 1, \ldots, s$. Nach der Lagrange-Darstellung erhält man (äquidistanter Fall!)

$$\begin{aligned}
p_s(t) \;&=\; \sum_{k=0}^{s} Y_{j+1-k}\; \ell_k(\frac{t_{j+1} - t}{h}),\\
\ell_k(\tau) \;&=\; \prod_{m \neq k}^{s} \left(\frac{\tau - m}{k - m} \right)
\end{aligned}$$

Man approximiert nun die Ableitung $y'(t_{j+1})$ durch

$$p_s'(t_{j+1}) \;=\; -\sum_{k=0}^{s} \frac{1}{h}\, Y_{j+1-k}\, \ell_k'(0)$$

und erhält damit das folgende BDF-Verfahren

$$\begin{aligned} \sum_{k=0}^{s} c_{k,s}\, Y_{j+1-k} &= h\, f(t_{j+1}, Y_{j+1}), \\ c_{k,s} &= -\,\ell_k'(0). \end{aligned} \tag{5.52}$$

Die Koeffizienten $c_{k,s},\ k = 0, \ldots, s$ der BDF-Verfahren bis zur Schrittzahl $s = 6$ sind in der folgenden Tabelle angegeben

s	$c_{k,s}$						
1	1	-1					
2	3/2	-2	1/2				
3	11/6	-3	3/2	$-1/3$			
4	25/12	-4	3	$-4/3$	1/4		
5	137/60	-5	5	$-10/3$	5/4	$-1/5$	
6	49/20	-6	15/2	$-20/3$	15/4	$-6/5$	1/6

Für Schrittzahlen $s > 6$ werden die Verfahren jedoch instabil.

Die BDF-Verfahren sind bereits 1952 von C.F. Curtiss und J.O. Hirschfelder untersucht worden und haben später als Grundlage eines Integrationsprogramms von C.W. Gear (1971) zur Lösung so genannter steifer Differentialgleichungen große Popularität erhalten.

B. Theorie allgemeiner Mehrschrittverfahren.

Wir betrachten allgemeine lineare Mehrschrittverfahren der Schrittzahl s von der folgenden Form

$$\sum_{k=0}^{s} \alpha_k\, Y_{j+k} \;=\; h\, \sum_{k=0}^{s} \beta_k\, f_{j+k}\,. \tag{5.53}$$

Hierbei sind die $\alpha_k,\ \beta_k \in \mathbb{R},\ \alpha_s \neq 0,\ f_k := f(t_k, Y_k)$. Wir setzen weiterhin voraus, dass das Mehrschrittverfahren mit konstanter Schrittweite $h = (t_b - t_0)/N$ durchgeführt wird.

Die *Anfangswerte* seien $Y_j = y(t_j) + \varepsilon_j,\ j = 0,\, 1,\, \ldots,\, s-1$ mit den (absoluten) Fehlern ε_k der Anlaufrechnung.

Definition (5.12):

Zu einer aktuellen Näherung $(t_j, Y_j) \in Q$ bezeichne wieder $z(t)$, genauer $z(t; t_j, Y_j)$ die Lösung der *lokalen* Anfangswertaufgabe

$$z' \;=\; f(t, z(t)), \qquad z(t_j) \;=\; Y_j\,.$$

a) Zu hinreichend kleinem $h > 0$ heißt dann

$$\tau(t_j, Y_j; h) \quad := \quad \frac{1}{h}\left[\sum_{k=0}^{s} \alpha_k\, z(t_j + kh) \;-\; h\,\sum_{k=0}^{s} \beta_k\, z'(t_j + kh)\right] \tag{5.13}$$

der *lokale Diskretisierungsfehler* des Mehrschrittverfahrens (5.11).

b) Das Mehrschrittverfahren heißt *konsistent*, falls für alle hinreichend oft stetig differenzierbaren rechten Seiten f und Näherungen $(t_j, Y_j) \in Q$ eine Abschätzung der folgenden Form gleichmäßig in Q gilt:

$$\|\,\tau(t_j, Y_j; h)\,\| \quad \leq \quad \phi(h), \quad \text{mit } \phi(h) \to 0 \;\; (h \to 0). \tag{5.14}$$

c) Wir sagen, das Mehrschrittverfahren besitzt die *Konsistenzordnung* $p \in \mathbb{N}$, falls gilt:

$$\|\,\tau(t_j, Y_j, h)\,\| \quad \leq \quad \phi(h), \quad \text{mit } \phi(h) \;=\; O(h^p), \tag{5.15}$$

d.h., es gibt nur von f und Q abhängige Konstante C, $h_0 > 0$, so dass gilt: $\forall\, h \in]0, h_0] : \;\|\,\tau(t_j, Y_j; h)\,\| \;\leq\; C\,h^p$.

Beispiel (5.16)

Für die *Mittelpunktsregel* $Y_{j+2} = Y_j + 2h\, f_{j+1}$ erhält man den lokalen Diskretisierungsfehler:

$$\tau \quad = \quad \frac{1}{h}\,[z(t + 2h) - z(t) - 2h\, z'(t + h)]$$

und hieraus mittels Taylor–Entwicklung von $z(t + 2h)$ und $z'(t + h)$ um $h = 0$:

$$\tau \quad = \quad \frac{1}{3}\,h^2\, z'''(t) \;+\; O(h^3)\,.$$

Damit ist die Mittelpunktsregel also konsistent und von der Ordnung $p = 2$.

Bemerkung (5.17)

Durch Taylor–Entwicklung von $\tau(t_j, Y_j; h)$ um $h = 0$ erhält man die folgende Aussage zur Konsistenz:

Ein MSV ist genau dann konsistent, falls die folgenden beiden Gleichungen erfüllt sind:

$$\sum_{k=0}^{s} \alpha_k \;=\; 0, \qquad \sum_{k=0}^{s} k \cdot \alpha_k \;-\; \sum_{k=0}^{s} \beta_k \;=\; 0\,. \tag{5.18}$$

Definiert man zum MSV (5.11) die zugehörigen *charakteristischen Polynome* durch

$$\rho(z) \;:=\; \sum_{k=0}^{s} \alpha_k \, z^k, \qquad \sigma(z) \;:=\; \sum_{k=0}^{s} \beta_k \, z^k \,, \tag{5.19}$$

so lässt sich die Konsistenzbedingung (5.17) auch folgendermaßen formulieren:

$$\rho(1) \;=\; 0, \qquad \rho'(1) \;=\; \sigma(1). \tag{5.20}$$

Auch die Ordnungseigenschaft eines MSVs lässt sich direkt anhand der charakteristischen Polynome überprüfen. Es lässt sich nämlich zeigen, dass das MSV (5.11) genau dann (wenigstens) die Konsistenzordnung $p \geq 2$ besitzt, wenn neben (5.20) gelten

$$\sum_{k=0}^{s} \alpha_k \, k^\ell \;=\; \ell \sum_{k=0}^{s} \beta_k \, k^{\ell-1} \qquad \text{für alle } \ell = 2, \ldots, p. \tag{5.21}$$

Einen Beweis dieser Aussage findet man beispielsweise in Deuflhard, Bornemann, Lemma 7.8, oder in Strehmel, Weiner, Satz 4.1.1. Die obige Bedingung (5.21) besagt - anders ausgedrückt, dass der lokale Diskretisierungsfehler τ für alle Polynome $z \in \Pi_p$ verschwindet.

Nun sind die Ordnungseigenschaften eines Mehrschrittverfahrens nicht alleine für die numerische Güte des Verfahrens verantwortlich. Der Grund liegt darin, dass gewisse (instabile) Mehrschrittverfahren die Fehler in der Anlaufrechnung, aber damit auch die Rundungsfehler, erheblich verstärken und dabei zu völlig unbrauchbaren Ergebnissen führen können. Diese Situation ist also für Mehrschrittverfahren prinzipiell anders, als wir dies von den Einschrittverfahren kennen. Um dies an einem einfachen Beispiel sehen zu können, betrachten wir zunächst den Lösungsraum *linearer, homogener Differenzengleichungen.*

Satz (5.22)

Gegeben sei eine lineare, homogene Differenzengleichung mit reellen (oder komplexen) Koeffizienten α_k

$$\sum_{k=0}^{s} \alpha_k \, w_{j+k} \;=\; 0, \qquad j = 0, 1, \ldots, \qquad \alpha_s \neq 0.$$

a) Ist z_1 eine Nullstelle des charakteristischen Polynoms $\rho(z) := \sum_{k=0}^{s} \alpha_k \, z^k$, so ist durch $w_j := z_1^j$ eine Lösung der Differenzengleichung gegeben.

b) Ist z_1 eine doppelte Nullstelle von ρ, so ist neben $w_j := z_1^j$ auch $\widetilde{w}_j := j \, z_1^j$ eine Lösung der Differenzengleichung.

c) Sind $z_1, \ldots, z_s$ die paarweise verschiedenen (einfachen) Nullstellen von ρ, so lautet die allgemeine Lösung der Differenzengleichung $w_j \;=\; \sum_{k=1}^{s} c_k \, z_k^j$.

Beweis:

zu a): $\displaystyle\sum_{k=0}^{s} \alpha_k \, w_{j+k} \;=\; \sum_{k=0}^{s} \alpha_k \, z_1^{j+k} \;=\; z_1^j \, \rho(z_1) \;=\; 0.$

zu b): $\displaystyle\sum_{k=0}^{s} \alpha_k \, \widetilde{w}_{j+k} \;=\; \sum_{k=0}^{s} \alpha_k \, (j+k) \, z_1^{j+k} \;=\; z_1^j \, [j \, \rho(z_1) \,+\, z_1 \, \rho'(z_1)] \;=\; 0.$

zu c): Der Lösungsraum ist ein s–dimensionaler linearer Raum, die z_k^j sind linear unabhängig. □

Beispiel (5.23)

Durch

$$Y_{j+2} \;+\; \frac{1}{5} \, Y_{j+1} \;-\; \frac{6}{5} \, Y_j \;=\; h \, \left(\frac{21}{10} \, f_{j+1} \;+\; \frac{1}{10} \, f_j \right)$$

ist ein explizites Zweischrittverfahren gegeben. Die zugehörigen charakteristischen Polynome lauten

$$\rho(z) \;=\; z^2 \;+\; \frac{1}{5} \, z \;-\; \frac{6}{5}, \qquad \sigma(z) \;=\; \frac{21}{10} \, z \;+\; \frac{1}{10}.$$

Man zeigt nun entweder mittels Taylor-Entwicklung des lokalen Diskretisierungsfehlers (5.13), oder direkt mittels (5.20) und (5.21), dass das Verfahren die Konsistenzordnung $p = 2$ besitzt.

Für die triviale Anfangswertaufgabe

$$y' \;=\; 0, \qquad y(0) \;=\; 1$$

liefert das Mehrschrittverfahren die lineare, homogene Differenzengleichung

$$Y_{j+2} \;+\; \frac{1}{5} \, Y_{j+1} \;-\; \frac{6}{5} \, Y_j \;=\; 0, \qquad Y_0 \;=\; 1, \qquad Y_1 = 1 \;+\; \varepsilon \, h.$$

Die Lösung lässt sich daher mittels Satz 4.3 unmittelbar angeben. Man erhält

$$Y_j \;=\; 1 \;+\; \frac{5 \, h \, \varepsilon}{11} \, \left(1 \;-\; (-\frac{6}{5})^j \right), \qquad j = 0, 1, \ldots$$

Durch die *parasitäre* Nullstelle $z_2 = -6/5$ des charakteristischen Polynoms wird also die eigentliche Lösung des Differenzenverfahrens (d.h. $\varepsilon = 0$, $Y_j = 1$) selbst bei sehr kleinen Werten von $h \, \varepsilon$ für entsprechend große Werte von j qualitativ völlig zerstört. Der Fehler überwuchert die Lösung und verhält sich zudem oszillatorisch. Insbesondere beobachten wir auch, dass die Näherungslösung $Y(t; h)$ an einer festen Stelle $t > 0$ für $h \to 0$ nicht gegen die exakte Lösung der Anfangswertaufgabe $y(t) = 1$ konvergiert.

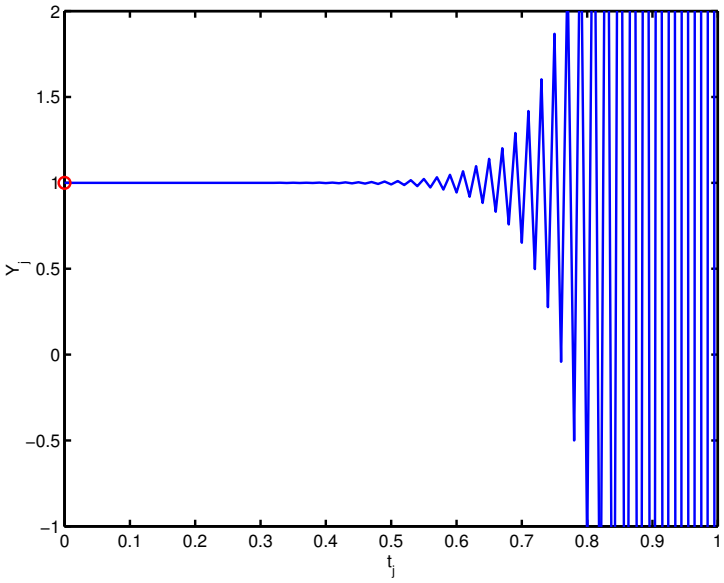

Abb. 5.1. Instabiles Verhalten eines Mehrschrittverfahrens

Definition (5.24)

Ein lineares MSV (5.11) heißt *stabil* (auch *nullstabil*), wenn alle Nullstellen z_j des charakteristischen Polynoms ρ in der Einheitskreisscheibe liegen, $|z_j| \leq 1$, und auf dem Rand, $|z_j| = 1$, nur einfache Nullstellen auftreten. Man sagt dann auch, die *Dahlquistsche Wurzelbedingung*[5] sei erfüllt.

Da bei einem konsistenten Verfahren stets $z_1 = 1$ eine Nullstelle von ρ ist, vgl. (5.20), muss diese Nullstelle bei einem stabilen Verfahren einfach sein. Liegen nun alle anderen Nullstellen im Innern des Einheitskreises, so heißt das Verfahren auch *stark stabil* oder *strikt stabil*, andernfalls *schwach stabil*.

Bemerkung (5.25)

Das Verfahren im obigen Beispiel (5.23) hat die Wurzeln $z_1 = 1$ und $z_2 = -6/5$ und ist daher instabil.

Für die Adams-Bashforth und Adams-Moulton Verfahren gilt $\rho(z) = z^s - z^{s-1}$. Damit ist $z_1 = 1$ (einfach) und $z_2 = 0$ ($(s-1)$–fach). Diese Verfahren sind also alle stark stabil.

Für die Verfahren von Nyström und Milne-Simpson ist dagegen $\rho(z) = z^s - z^{s-2}$. Diese Verfahren sind also nur schwach stabil.

Definition (5.26)

Wir wollen die *Konvergenz* eines linearen MSVs (5.11) definieren. Dazu sei eine beliebige AWA $y' = f(t, y),\ y(t_0) = y_0,$ mit einer (lokal) Lipschitz-stetigen rechten Seite f und

[5] Nach dem schwedischen Mathematiker Germund Dahlquist (1925-2005)

der auf einem Intervall $t_0 \le t \le t_b$ erklärten (exakten) Lösung y vorgegeben.

Zur Gitterfolge $t_j^{(m)} := t_0 + j\, h_m, \quad h_m := (t_b - t_0)/m, \quad m \ge s,$ und Startnäherungen $Y(t_j^{(m)}; h_m),\ j = 0, \ldots, s-1,$ mit $\lim\limits_{m\to\infty} |Y(t_j; h_m) - y(t_j)| \ = \ 0$ berechnen wir hieraus mit einem linearen MSV die Näherungen $Y(t_b; h_m)$. Gilt dann stets

$$\lim_{m\to\infty} |Y(t_b; h_m) - y(t_b)| \ = \ 0,$$

so heißt das MSV *konvergent*.

Gilt für alle hinreichend glatten rechten Seiten f und Startdaten mit der Approximationsgüte $|Y(t_j; h_m) - y(t_j)| \ = \ O(h_m^p),$ dass auch im Endpunkt

$$|Y(t_b; h_m) - y(t_b)| \ = \ O(h_m^p)$$

ist, so heißt das MSV *konvergent von der Ordnung p*.

Satz (5.27)

Ein konvergentes lineares Mehrschrittverfahren ist notwendig konsistent und stabil.

Beweis: Wir betrachten die triviale AWA $y' = 0,\ y(0) = 0,\ t_b = 1,$ und berechnen die Näherungen zur Schrittweitenfolge $h_m = 1/m$ und den Anfangsdaten $Y(t_j; h_m) = h_m w_j,\ j = 0, \ldots, s-1,$ mit beliebig (fest) vorgegebenen $w_0, \ldots, w_{s-1}$. Die Anfangsdaten erfüllen dann die in Definition (5.26) geforderte Approximationseigenschaft, so dass das Mehrschrittverfahren im Punkt $t = 1$ konvergiert. Die Näherungen sind gegeben durch:

$$\sum_{k=0}^{s} \alpha_k\, Y(t_{j+k}; h_m) \ = \ 0, \ \ j \ = \ 0,\ 1,\ \ldots,\ m - s\,.$$

Mit den obigen Anfangsdaten folgt $Y(t_j; h_m) = h_m\, w_j,$ wobei w_j die Lösung der Differenzengleichung $\sum_{k=0}^{s} \alpha_k\, w_{j+k} \ = \ 0$ zu den (beliebig) vorgegebenen Anfangsdaten ist. Die Konvergenz impliziert damit

$$\lim_{m\to\infty} \ Y(t_m; h_m) \ = \ \lim_{m\to\infty} \ \frac{w_m}{m} \ = \ 0,$$

woraus sich mit Satz (5.22) gerade die Stabilität des Mehrschrittverfahrens ergibt.

Analog betrachten wir die AWA $y' = 0,\ y(0) = 1,\ t_b = 1$. Die vorausgesetzte Konvergenz des MSVs impliziert, dass die Lösung der Differenzengleichung

$$\sum_{k=0}^{s} \alpha_k\, Y_{j+k} \ = \ 0, \ \ j \ = \ 0,\ 1,\ \ldots \quad \text{mit} \quad Y_0 \ := \ \ldots \ := \ Y_{s-1} \ := \ 1$$

gegen 1 konvergieren muss. Damit folgt aber $\sum_{k=0}^{s} \alpha_k \ = \ \rho(1) \ = \ 0.$

Schließlich betrachten wir die AWA $y' \ = \ 1,\ y(0) \ = \ 0,\ t_b \ = \ 1.$ Das MSV ergibt die Differenzengleichung

$$\sum_{k=0}^{s} \alpha_k\, Y(t_{j+k}; h_m) \ = \ h_m \ \sum_{k=0}^{s} \beta_k, \quad j \ = \ 0,\ 1,\ \ldots,\ m - s\,.$$

Hierin verwenden wir den Ansatz $Y(t_j; h_m) := j\, h_m\, K$. Für $j = 0, \ldots, s-1$ sind dies jedenfalls zulässige Startwerte. Wir erhalten

$$\sum_{k=0}^{s} \alpha_k\, (j+k)\, h_m\, K \;=\; h_m \sum_{k=0}^{s} \beta_k.$$

Wegen $\sum_{k=0}^{s} \alpha_k = 0$ ist diese Relation für K eindeutig lösbar mit

$$K \;=\; \left(\sum_{k=0}^{s} \beta_k \right) / \left(\sum_{k=0}^{s} k\, \alpha_k \right).$$

Man beachte, dass hierbei der Nenner aufgrund der bereits gezeigten Stabilität nicht verschwindet. Mit diesem Wert für K haben wir also die Lösung der Differenzengleichung (mit den entsprechenden Anfangswerten) gefunden. Damit wird $Y(t_m; h_m) = m h_m K = K$ und die vorausgesetzte Konvergenz des Verfahrens liefert $K = 1$, also $\rho'(1) = \sigma(1)$. Damit ist die Konsistenz des MSV gezeigt. □

Die soeben gezeigte notwendigen Bedingungen für die Konvergenz eines linearen MSV, nämlich Konsistenz uns Stabilität, sind tatsächlich auch hinreichend für die Konvergenz. Dieses wichtige Ergebnis konnte G. Dahlquist 1956 zeigen. Wir verzichten hier auf einen Beweis (einen solchen findet man in den schon mehrfach zitierten Lehrbüchern von Hairer, Norsett und Wanner, Strehmel und Weiner sowie Deuflhard und Bornemann) und referieren nur das Ergebnis zusammen mit den Dahlquistschen Schranken für die Konsistenzordung.

Satz (5.28) (G. Dahlquist, 1956/58)

a) Ein lineares MSV ist genau dann konvergent, wenn es konsistent und stabil ist.

b) Hat das MSV darüber hinaus die Konsistenzordnung p und verwendet man Startnäherungen der gleichen Approximationsgüte, so hat das Verfahren auch die Konvergenzordnung p; vgl. die Definition (5.26).

c) Ein stabiles linearen MSV der Schrittzahl s hat eine maximale Konsistenzordnung

$$p \;\leq\; \begin{cases} s+2 & \text{falls} \quad s \text{ gerade} \\ s+1 & \text{falls} \quad s \text{ ungerade.} \end{cases}$$

Für stabile explizite Mehrschrittverfahren gilt sogar $p \leq s$.

d) Für die Konsistenzordnung eines stark stabilen Mehrschrittverfahrens der Schrittzahl s gilt die Schranke $p \leq s + 1$.

C. Adams Verfahren für variable Schrittweiten.

Wie in Abschnitt 4 ausführlich dargelegt worden ist, benötigen effiziente Verfahren zur Lösung von AWA eine adaptive (lokale) Bestimmung der Integrationsschrittweite. Bei MSV stellt die lokale Änderung der Schrittweite ein nicht unerhebliches Problem dar, da die bisherigen Verfahren alle unter der Annahme eines äquidistanten Gitters konstruiert worden sind.

Im Wesentlichen gibt es zwei Ansätze diese Voraussetzung zu überwinden. Der erste Ansatz, der auf F.T. Krogh 1969/74 zurückgeht, verwendet anstelle der Lagrange–Darstellung des Interpolationspolynoms, vgl. (5.3), die Newton-Darstellung und kann somit mit völlig beliebigen Gittern arbeiten. Der wesentliche Trick besteht darin, dass man die benötigten Integrale der Interpolationpolynome ebenso wie die Newtonschen dividierten Differenzen rekursiv berechnen kann.

Eine anderer, auf A.Nordsiek 1962 zurückgehender Vorschlag arbeitet mit einem *virtuell* äquidistanten Gitter, speichert jedoch die Informationen über das Interpolationspolynom als Ableitungsinformation am letzten aktuellen Integrationsknoten. So lässt sich die Schrittweite bei Beibehaltung der Ableitungsinformation lokal variieren.

Wir sehen uns im Folgenden den Kroghschen Ansatz etwas genauer an. Ausgangspunkt ist wieder der allgemeine Ansatz für Adams Verfahren

$$y(t_j) \;-\; y(t_{j-1}) \;=\; \int\limits_{t_{j-1}}^{t_j} f(t, y(t))\, dt\,. \tag{5.29}$$

Wir konstruieren nun ein PECE-Verfahren, indem wir den Integranden durch ein Interpolationspolynom ersetzen.

Im *Prädiktorschritt* wird $f(t, y(t))$ ersetzt durch das Interpolationspolynom $P_s^{(0)} \in \Pi_{s-1}$ zu den schon berechneten Stützstellen (t_{j-i}, f_{j-i}), $i = 1, \ldots, s$. Damit wird

$$\begin{aligned} Y_j^{(0)} \quad &:= \quad Y_{j-1} \;+\; \int\limits_{t_{j-1}}^{t_j} P_s^{(0)}(t)\, dt\,, \\ P_s^{(0)}(t) \quad &:= \quad \sum_{i=0}^{s-1} p_i(t)\, f[t_{j-1}, \ldots, t_{j-1-i}]\,, \\ p_i(t) \quad &:= \quad \prod_{k=1}^{i} (t - t_{j-k})\,,\;\; i \;=\; 0,\, 1,\, \ldots\,. \end{aligned} \tag{5.30}$$

Hierbei bezeichnen wie üblich $f[t_{j-1}, \ldots, t_{j-1-i}]$ die Newtonschen dividierten Differenzen, die sich rekursiv in einem Dreiecks-Tableau berechnen lassen gemäß

$$f[t_k] \;:=\; f(t_k); \quad f[t_k, \ldots, t_\ell] \;:=\; \frac{f[t_k, \ldots, t_{\ell-1}] \;-\; f[t_{k-1}, \ldots, t_\ell]}{t_k \;-\; t_\ell}\,. \tag{5.31}$$

Nach Berechnung von (5.30) wird $f_j^{(0)} := f(t_j, Y_j^{(0)})$ ausgewertet (*Evaluation*) und sodann wird wiederum in (5.29) $f(t, y(t))$ ersetzt durch das Interpolationspolynom $P_s \in \Pi_{s-1}$

zu den nun verschobenen Stützstellen

$$(t_j, f_j^{(0)}), \quad (t_{j-i}, f_{j-i}) \,, \quad i = 1, \ldots, s-1 \,.$$

Zusammen mit (5.30) lässt sich der *Korrektorschritt* dann folgendermaßen formulieren

$$\begin{aligned} Y_j &:= Y_j^{(0)} + \int_{t_{j-1}}^{t_j} (P_s(t) - P_s^{(0)}(t))\, dt \,, \\ P_s(t) - P_s^{(0)}(t) &:= (f_j^{(0)} - f_j^{(1)}) \, \frac{p_{s-1}(t)}{p_{s-1}(t_j)} \,, \\ f_j^{(1)} &:= P_s^{(0)}(t_j) = \sum_{i=0}^{s-1} p_i(t_j)\, f[t_{j-1}, \ldots, t_{j-1-i}] \,. \end{aligned} \tag{5.32}$$

Nun wird neben Prädiktor– und Korrektorschritt noch eine Schätzung $\tau_{\text{est},s}$ für den *lokalen Diskretisierungsfehler* benötigt. Dieser wird analog zu den Einschrittverfahren definiert durch

$$\tau := \frac{1}{h_{j-1}} \, (z(t_j; t_{j-1}, Y_{j-1}) - Y_j) \,,$$

wobei mit z wieder die Lösung der entsprechenden lokalen Anfangswertaufgabe bezeichnet wird. Setzt man Y_j und z in Integralform ein, so erhält man

$$\tau = \frac{1}{h_{j-1}} \int_{t_{j-1}}^{t_j} (f(t, z(t)) - P_s(t))\, dt \,.$$

Zur Schätzung von τ wird hierin $f(t, z(t))$ ersetzt durch das Interpolationspolynom $P_{s+1} \in \Pi_s$ zu den Stützstellen

$$(t_j, f_j^{(0)}), \quad (t_{j-i}, f_{j-i}) \,, \quad i = 1, \ldots, s$$

Damit ist also nach der Newton-Datstellung des Interpolationspolynoms

$$\begin{aligned} \tau_{\text{est},s} &:= \frac{1}{h_{j-1}} \int_{t_{j-1}}^{t_j} (P_{s+1}(t) - P_s(t))\, dt \,, \\ P_{s+1}(t) - P_s(t) &:= f^{(0)}[t_j, t_{j-1}, \ldots, t_{j-s}] \, (t - t_j) \, p_{s-1}(t) \,, \end{aligned} \tag{5.33}$$

wobei die dividierte Differenz $f^{(0)}[t_j, t_{j-1}, \ldots, t_{j-s}]$ mit den Daten $f_j^{(0)}, f_{j-1}, \ldots, f_{j-s}$ auszuwerten ist.

Wir fassen die Terme eines Integrationsschrittes der Schrittzahl s nochmals zusammen:

$$\begin{aligned} Y_j^{(0)} &= Y_{j-1} + \sum_{i=0}^{s-1} \left(\int_{t_{j-1}}^{t_j} p_i(t)\, dt \right) f[t_{j-1}, \ldots, t_{j-1-i}] \,, \\ f_j^{(0)} &= f(t_j, Y_j^{(0)}) \,, \\ f_j^{(1)} &= \sum_{i=0}^{s-1} p_i(t_j)\, f[t_{j-1}, \ldots, t_{j-1-i}] \,, \end{aligned} \tag{5.34}$$

$$
\begin{aligned}
Y_j &= Y_j^{(0)} + \frac{f_j^{(0)} - f_j^{(1)}}{p_{s-1}(t_j)} \left(\int\limits_{t_{j-1}}^{t_j} p_{s-1}(t)\, dt \right) , \\
\tau_{\mathrm{est},s} &= \frac{1}{h_{j-1}}\, f^{(0)}[t_j, t_{j-1}, \ldots, t_{j-s}] \left(\int\limits_{t_{j-1}}^{t_j} (t - t_j)\; p_{s-1}(t)\, dt \right) .
\end{aligned}
\tag{5.34}
$$

Die hierin auftretenden Integrale lassen sich nun rekursiv berechnen. Wir definieren dazu

$$
g_{i\,k} := \int\limits_{t_{j-1}}^{t_j} p_i(t)\, (t - t_j)^{k-1}\, dt \tag{5.35}
$$

und finden die Rekursion

$$
\begin{aligned}
g_{0\,k} &= (-1)^{k+1}\, \frac{1}{k}\, h_{j-1}^k , \\
g_{i\,k} &= (t_j - t_{j-i})\, g_{i-1,\,k} + g_{i-1,\,k+1} .
\end{aligned}
\tag{5.36}
$$

Tableau:

$$
\begin{array}{llll}
g_{0,\,1} & & & \\
g_{0,\,2} & g_{1,\,1} & & \\
\vdots & \vdots & \ddots & \\
g_{0,\,s} & g_{1,\,s-1} & \cdots & g_{s-1,\,1} \\
g_{0,\,s+1} & g_{1,\,s} & \cdots & g_{s-1,\,2} \quad g_{s,\,1}
\end{array}
$$

Die Strategie zur Wahl der Schrittweite und der Ordnung kann nun in Anlehnung an die Strategie bei Einschrittverfahren erfolgen.

Wir akzeptieren einen aktuellen Integrationschritt, falls für eine vorgegebene Genauigkeitsschranke TOL:

$$
|\tau_{\mathrm{est},s}| \;\leq\; \mathrm{TOL}
$$

erfüllt ist und verwenden sodann eine neue Schrittweite

$$
h_{\mathrm{neu}} := q\; (r\,\mathrm{tol}/\|\tau_{\mathrm{est},s}\|)^{(1/s)}\quad h_{j-1}\,; \tag{5.37}
$$

q und r sind dabei geeignet zu wählende Sicherheitsfaktoren, etwa $q = 0.95,\;\; r = 0.5$.

Mit dieser Schrittweitenstrategie lässt sich nun aber zugleich auch die Ordnung des Verfahrens, genauer die aktuelle Schrittzahl s steuern. Dazu rechnet man in einem Integrationsschritt mit der Schrittzahl s nicht nur den Verstärkungsfaktor aus (5.37) aus, sondern auch die bzgl. der Schrittzahl benachbarten Faktoren

$$
(r\,\mathrm{tol}/\|\tau_{\mathrm{est},s-1}\|)^{(1/(s-1))}\;,\;\; (r\,\mathrm{tol}/\|\tau_{\mathrm{est},s}\|)^{(1/s)}\;,\;\; (r\,\mathrm{tol}/\|\tau_{\mathrm{est},s+1}\|)^{(1/(s+1))}\;.
$$

Gewählt wird dann im nächsten Integrationsschritt die Schrittzahl, die die größte neue Schrittweite geliefert hat. Das Verfahren kann damit als Einschrittverfahren mit $s = 1$ und sehr kleiner Schrittweite gestartet werden. Danach wächst s bis zu einer Arbeitsschrittzahl von etwa $s = 10$ an.

6. Extrapolationsverfahren

A. Allgemeines.

Extrapolation ist ein Verfahren zur *Konvergenzbeschleunigung*, das sich in der folgenden Situation anwenden lässt:

Zu bestimmen sei eine Größe τ_0, für die sich Näherungen $T(h)$ numerisch berechnen lassen, die von einer *Schrittweite* h abhängen. Dabei kann h aus einer diskreten Menge

$$h \in H = \{h: \ h = H_0/n, \ n \in \mathbb{N}\}$$

gewählt werden. Es wird vorausgesetzt, dass sich für die Näherungen eine so genannte *asymptotische Entwicklung*

$$T(h) = \tau_0 + \tau_1 h^\gamma + \tau_2 h^{2\gamma} + \ldots + \tau_m h^{m\gamma} + \alpha_{m+1}(h) h^{(m+1)\gamma} \tag{6.38}$$

mit einem bezüglich h beschränkten Restglied $|\alpha_{m+1}(h)| \le E_{m+1}, \ \forall h \in H$, zeigen lässt.

In der Praxis ist hierbei zumeist $\gamma \in \{1, 2\}$, und man spricht von einer *linearen* bzw. *quadratischen* Entwicklung. Es ist zu beachten, dass in (6.1) keine Konvergenz für $m \to \infty$ vorausgesetzt wird. Man ist vielmehr am Grenzwert $h \to 0$ interessiert, insbesondere, um mittels (6.1) die Konvergenzordnung zu erhöhen.

Beispiele für das Vorliegen einer solchen Situation sind aus der Numerik bekannt. So gestattet im Zusammenhang mit der numerischen Berechnung eines bestimmten Integrals die Trapezsumme eine quadratische asymptotische Entwicklung, die so genannte Euler-Maclaurin-Entwicklung. Das zugehörige Extrapolationsverfahren geht auf Romberg, 1955, zurück. Eine andere Anwendung liefert die Approximation der Ableitung einer hinreichend glatten Funktion durch den symmetrischen Differenzenquotienten. Die Taylor-Entwicklung zeigt, dass auch hierbei eine quadratische asymptotische Entwicklung vorliegt, die Extrapolation ermöglicht.

Die Grundidee des Extrapolationsverfahrens ist nun die Folgende:

Vernachlässigt man in der Entwicklung (6.1) das Restglied, so ist $T(h)$ näherungsweise ein Polynom in h^γ, dessen Absolutterm wir ermitteln wollen. Dazu bestimmen wir zu einer Schrittweitenfolge

$$h_0 > h_1 > \ldots > h_m$$

die Werte $T(h_k), k = 0, \ldots, m$ und berechnen hieraus das Interpolationspolynom $P_m \in \Pi_m$ zu den Stützstellen $(h_k^\gamma, T(h_k)), \ k = 0, \ldots, m$.

Es ist nun zu erwarten, dass dieses Interpolationspolynom bis auf einen Fehler der Größenordnung $O(h_0^{(m+1)\gamma})$ mit der Entwicklung (6.1) übereinstimmt und daher auch für den Absolutterm $P_m(0)$ des Interpolationspolynoms gilt

$$P_m(0) - \tau_0 = O(h_0^{(m+1)\gamma}). \tag{6.39}$$

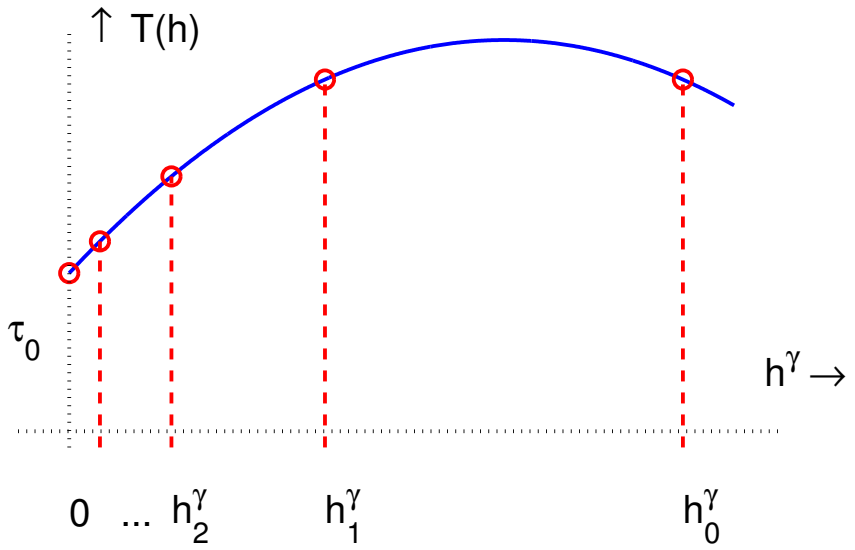

Abb. 6.1. Extrapolationsverfahren

Beispiel (6.3)

Zur numerischen Differentiation einer C^∞–Funktion $f : \mathbb{R} \to \mathbb{R}$ an einer Stelle x_0 verwenden wir den symmetrischen Differenzenquotienten

$$D(h; x_0) \;:=\; \frac{f(x_0 + h) - f(x_0 - h)}{2\,h}\,. \tag{6.4}$$

Setzen wir die Taylor–Entwicklung von f

$$f(x_0 + h) \;=\; a_0 \,+\, a_1\,h \,+\, \ldots \,+\, a_\ell\,h^\ell \,+\, O(h^{\ell+1}); \quad a_k \;:=\; f^{(k)}(x_0)/k!$$

hierin ein, so erhalten wir die quadratische asymptotische Entwicklung

$$D(h; x_0) \;=\; f'(x_0) \,+\, a_3\,h^2 \,+\, a_5\,h^4 \,+\, \ldots \,+\, a_{2m+1}h^{2m} \,+\, O(h^{2(m+1)}). \tag{6.5}$$

Wir werten nun $D(h; x_0)$ für zwei Schrittweiten $h_0 > h_1 > 0$ aus, etwa $h_1 = h_0/2$, und bestimmen die interpolierende Gerade in h^2 durch die Stützstellen $(h_k^2, D(h_k; x_0)), k = 0, 1$:

$$P_1(h^2; h_0, h_1) \;=\; D(h_0; x_0) \,+\, \frac{h^2 - h_0^2}{h_1^2 - h_0^2}\,(D(h_1; x_0) - D(h_0; x_0)).$$

Die Auswertung für $h = 0$ ergibt damit die Näherung

$$P_1(0; h_0, h_1) \;=\; \frac{h_1^2}{h_1^2 - h_0^2}\,D(h_0; x_0) \;-\; \frac{h_0^2}{h_1^2 - h_0^2}\,D(h_1; x_0)\,.$$

Um die Abhängigkeit von der Schrittweite zu untersuchen, setzen wir hierin die asymptotische Entwicklung (6.5) ein und erhalten

$$\begin{aligned} P_1(0; h_0, h_1) \;&=\; \frac{h_1^2}{h_1^2 - h_0^2}\,(f'(0) + a_3 h_0^2 + \ldots) \;-\; \frac{h_0^2}{h_1^2 - h_0^2}\,(f'(0) + a_3 h_1^2 + \ldots) \\ &=\; f'(0) \;-\; a_5 h_0^2\,h_1^2 \;-\; a_7(h_0^2 + h_1^2)\,h_0^2\,h_1^2 \,+\, \ldots \end{aligned}$$

Wir sehen also, dass wir mit $P_1(0; h_0, h_1)$ tatsächlich eine Approximation von $f'(0)$ konstruiert haben, deren Fehler von der Ordnung h_0^4 ist. Ferner besitzt auch diese Näherungswert eine quadratische asymptotische Entwicklung, allerdings mit eliminierten h^2–Term, so dass eine iterierte Anwendung des Verfahrens, d.h. die Elimination des h^4–, des h^6–Terms und so fort, mittels weiterer Auswertungen von $D(h; x_0)$ möglich ist.

Praktische Auswertung:

Die algorithmische Auswertung eines Interpolationspolynoms an einer vorgegebenen Stelle (in unserem Fall in $h = 0$) erfolgt numerisch effizient mit Hilfe des *Algorithmus von Aitken und Neville* (vgl. Vorlesung über Numerische Mathematik). Dazu setzen wir vorübergehend $z := h^\gamma$.

Bezeichnet nun $P_{i,k}(z) \in \Pi_k$ das Interpolationspolynom höchstens k-ten Grades zu den Stützstellen $(z_j, T(h_j)), j = i, \ldots i+k,$ so gilt nach dem *Lemma von Aitken* die Rekursion

$$\begin{aligned} P_{i,0}(z) &= T(h_i), \qquad i = 0, \ldots, m \\ P_{i,k}(z) &= P_{i+1,k-1}(z) + \frac{z - z_{i+k}}{z_i - z_{i+k}} \left(P_{i,k-1}(z) - P_{i+1,k-1}(z)\right), \\ &\qquad k = 1, \ldots, m, \qquad i = 0, \ldots, m-k. \end{aligned} \tag{6.6}$$

Speziell für $z = 0$ (wir lassen einfach das Argument z weg) und $z_i = h_i^\gamma$ ergibt sich dann die Rekursion

$$\begin{aligned} P_{i,0} &= T(h_i), \qquad i = 0, \ldots, m \\ P_{i,k} &= P_{i+1,k-1} + \frac{P_{i+1,k-1} - P_{i,k-1}}{(h_i/h_{i+k})^\gamma - 1}, \quad 1 \le i + k \le m. \end{aligned} \tag{6.7}$$

Üblicherweise ordnet man die $P_{i,k}$–Werte in einem *Extrapolationstableau* (Neville–Tableau) an:

$$\begin{array}{lllll} P_{0,0} & & & & \\ P_{1,0} & P_{0,1} & & & \\ P_{2,0} & P_{1,1} & P_{0,2} & & \\ \vdots & \vdots & \vdots & \ddots & \\ P_{m,0} & P_{m-1,1} & P_{m-2,2} & \ldots & P_{0,m} \end{array} \tag{6.8}$$

Ist (h_i) eine streng monoton fallende Nullfolge, so konvergiert die erste Spalte dieses Tableaus wie h_i^γ gegen τ_0, die zweite Spalte wie $h_i^{2\gamma}$ und so fort. Extrapoliert man also bis zur Spalte m, so hat man aus dem Ausgangsverfahren der Ordnung γ ein Verfahren der Ordnung $m\gamma$ gewonnen.

Erwähnt werden soll auch, dass das obige Tableau natürlich unter der Verwendung eine *eindimensionalen* Feldes berechnet werden kann, wobei stets nur die letzte Zeile in der

Form

$$P_m \quad P_{m-1} \quad \dots \quad P_1 \quad P_0$$

aktuell aufgehoben wird (mitunter zusätzlich die Zeile darüber). Ein Algorithmus zur Berechnung des Extrapolationstableaus sieht dann etwa folgendermaßen aus:

Algorithmus (6.9):

$$\begin{aligned}
&P_0 \;:=\; T(h_0)\\
&\text{für } k \;=\; 1, 2, \dots, m\\
&\qquad P_k \;:=\; T(h_k)\\
&\qquad \text{für } i \;=\; k-1, \dots, 0\\
&\qquad\qquad P_i \;:=\; P_{i+1} \;+\; \frac{P_{i+1} \;-\; P_i}{(h_i/h_k)^\gamma \;-\; 1}
\end{aligned}$$

Für die Wahl der Schrittweitenfolge $(h_i)_{i\in\mathbb{N}_0}$ gibt es mehrerer Vorschläge. Gebräuchlich ist die sukzessive Halbierung der Schrittweite, die so genannte *Romberg-Folge*

$$h_0,\;\; h_0/2,\;\; h_0/4,\;\; h_0/8,\;\; h_0/16, h_0/32,\;\; h_0/64,\;\; \dots$$

Diese Folge hat bei Quadraturverfahren den Vorteil, dass bei Verfeinerung der Schrittweite die alten Funktionsauswertungen weiter verwenden kann, allerdings auch den Nachteil, dass die Schrittweiten schnell klein werden. Günstiger ist hier das Verhalten der so genannten *Bulirsch-Folge*

$$h_0,\;\; h_0/2,\;\; h_0/3,\;\; h_0/4,\;\; h_0/6,\;\; h_0/8,\;\; h_0/12,\;\; \dots$$

bei der auch die alte Information wiederverwendet werden kann, ohne dass die Folge zu schnell klein wird.

Da bei der Anwendung auf Differentialgleichungsprobleme jedoch in der Regel auf die Wiederverwendung alter Information verzichtet werden muss, benutzt man auch die harmonische Folge

$$h_0,\;\; h_0/2,\;\; h_0/3,\;\; h_0/4,\;\; h_0/5,\;\; h_0/6,\;\; h_0/7,\;\; \dots$$

Fehlerdarstellung:

Für die extrapolierten Werte gilt die Fehlerdarstellung

$$P_{i\,k} \;=\; \tau_0 \;+\; h_i^\gamma \dots h_{i+k}^\gamma \, \sigma_{k+1}(h_i, \dots, h_{i+k}) \tag{6.10}$$

wobei $|\sigma_{k+1}(h_i, \dots, h_{i+k})|$ beschränkt ist für festes k und $i \to \infty$.

B. Extrapolation von Einschrittverfahren.

Wir betrachten ein allgemeines ESV zur Integration einer AWA $y' = f(t,y),\ y(t_0) = y_0$:

$$Y_{j+1} \ = \ Y_j \ + \ h_j \, \Phi \, (t_j, Y_j; h_j) \, , \ t_{j+1} \ = \ t_j \ + \ h_j, \ Y_0 \ = \ y_0, \tag{6.11}$$

und setzen im Folgenden voraus, dass die rechte Seite f des Differentialgleichungssystems wie auch die Verfahrensfunktion Φ bezüglich aller Variablen hinreichend oft stetig differenzierbar ist.

Zur Anwendung des Extrapolationsverfahrens betrachten wir *einen* Integrationsschritt, wobei (t, Y) die aktuelle Näherung bezeichne. Den Index für die Integrationsschritte (bisher j) werden wir der Einfachheit halber hier weglassen.

Zunächst wird mit einer *Makroschrittweite* $H > 0$ ein Integrationsschritt ausgeführt

$$Y(t + H; H) = Y \ + \ H \, \Phi \, (t, Y; H).$$

Sodann werden zu einer streng monoton fallenden Nullfolge von Schrittweiten

$$h_0 \ := \ H; \quad h_i \ := \ H/n_i, \quad n_i \in \mathbb{N}, \quad i = 0, \ 1, \ 2, \ \ldots$$

mit jeweils n_i äquidistanten Integrationsschritten (*Mikroschritten*) mit Schrittweite h_i Näherungen $Y(t + H; h_i)$ berechnet, $i = 1, \ 2, \ \ldots$.

Um auf diese Näherungen $Y(t + H; h)$ nun erfolgreich ein Extrapolationsverfahren anwenden zu können, benötigt man die Existenz einer asymptotischen Entwicklung dieser Näherung bezüglich der Schrittweite h. Dies bedeutet gerade, dass wir eine asymptotische Entwicklung für den *globalen* Diskretisierungsfehler suchen.

Den *lokalen* Diskretisierungsfehler

$$\tau(t,h) \ = \ \frac{1}{h} \, [\, z(t+h) - Y - h \, \Phi(t, Y; h) \,]$$

hatten wir bereits in Abschnitt 4 B in eine Taylor-Summe mit Restglied entwickelt, wobei wir allerdings die Entwicklung nur bis zur Konsistenzordnung p des Verfahrens vorgenommen haben. Damit ergab sich $\tau(t,h) \ = \ \sigma_p(t,h) \, h^p$ mit $|\sigma_p(t,h)| \ \leq \ C_p \, , \ \forall \, (t,h)$. Entwickelt man (bei der vorausgesetzten Glattheit von f und Φ) weiter, so ergibt sich

$$\tau(t,h) \ = \ \tau_p(t) \, h^p \ + \ \tau_{p+1}(t) \, h^{p+1} \ + \ldots + \ \tau_m(t) \, h^m \ + \ \sigma_{m+1}(t,h) \, h^{m+1},$$

mit $|\sigma_{m+1}(t,h)| \leq C_{m+1}, \ \forall \, (t,h)$.

Aus dieser Entwicklung konnte W.B. Gragg 1965 tatsächlich eine asymptotische des globalen Fehlers (unter entsprechenden Differenzierbarkeitsannahmen) ableiten. Er erweiterte damit zugleich auch die Aussage des (globalen) Konvergenzsatzes (4.19) auf höhere h–Potenzen.

Satz (6.12) (Gragg, 1965)

Der lokale Diskretisierungsfehler des Einschrittverfahrens (6.10) besitze eine asymptotische Entwicklung der Form

$$\tau(t,h) \;=\; \tau_p(t)\,h^p \;+\; \tau_{p+1}(t)\,h^{p+1} \;+\ldots+\; \tau_m(t)\,h^m \;+\; O(h^{m+2}).$$

Das Verfahren sei also insbesondere von der Konsistenzordnung p, also $\tau_p \neq 0$.

Dann besitzt auch der globale Diskretisierungsfehler eine asymptotische Entwicklung der Form

$$Y(t;h) \;-\; y(t) \;=\; e_p(t)\,h^p \;+\; \ldots\; e_m(t)\,h^m \;+\; \alpha_{m+1}(t,h)\,h^{m+1}$$

mit einem bzgl. h beschränktem Restglied $\|\alpha_{m+1}(t,h)\| \leq M_{m+1}$.

Ferner genügen die Koeffizientenfunktionen e_k einem gestaffelten, linearen, inhomogenen Differentialgleichungssystem

$$e_k'(t) \;=\; f_y(t,y(t))\,e_k(t) \;+\; \psi_k(t;y,e_p,\ldots,e_{k-1}); \quad e_k(t_0)=0.$$

Für eine Beweis des Graggschen Satzes sei auf die Literatur verwiesen, z.B. auf das Lehrbuch von Strehmel und Weiner, Satz 3.1.2.

Der Graggsche Satz besagt, dass sich beispielsweise die Ergebnisse des Eulerschen Polygonzugverfahrens, aber auch die von Runge-Kutta Verfahren durch Extrapolation mit $\gamma = 1$ verbessern lassen.

Nun ist Extrapolation bei einer nur linearen asymptotischen Entwicklung weniger effizient, als dies für eine quadratische asymptotische Entwicklung der Fall ist, wie sie etwa bei der Trapezsummenextrapolation vorliegt. Es ist daher naheliegend, nach Integrationsverfahren Ausschau zu halten, die eine *quadratische* asymptotische Entwicklung gestatten. Eine hinreichende Bedingung hierfür ist, dass das ESV bezüglich der Schrittweite h eine Symmetrie aufweist.

Mathematisch lässt sich dies wie folgt feststellen.

Spiegelung von Einschrittverfahren (6.13)

Gegeben ist ein ESV der Form $Y(t+h,h) = Y(t,h) + h\,\Phi(t,Y(t,h);h)$. Zur Spiegelung dieses Verfahrens geht man folgendermaßen vor.

(i) Ersetze h durch $-h$:

$$Y(t-h,-h) \;=\; Y(t,-h) \;-\; h\,\Phi(t,Y(t,-h);-h)$$

(ii) Ersetze t durch $t+h$:

$$Y(t,-h) \;=\; Y(t+h,-h) \;-\; h\,\Phi(t,Y(t+h,-h);-h)$$

oder – umgeschrieben

$$Y(t+h,-h) \;=\; Y(t,-h) \;+\; h\,\Phi(t,Y(t+h,-h);-h).$$

Die letzte Gleichung wird aufgefasst als eine implizite Gleichung zur Bestimmung von $Y(t+h,-h)$. Für hinreichend kleine Schrittweiten h ist diese Gleichung nach dem Satz über implizite Funktionen auch (lokal eindeutig) lösbar. Wir schreiben dann für die Lösung

$$Y(t+h,-h) \;=\; Y(t,-h) \;+\; h\,\Phi^*(t,Y(t,-h);-h)$$

und bezeichnen mit Φ^* das *gespiegelte ESV* von Φ.

Schießlich heißt das ESV Φ *symmetrisch*, falls $\Phi = \Phi^*$ gilt.

Beispiel (6.14)

Die Spiegelung des Euler-Verfahrens $Y(t+h,h) = Y(t,h) + h\,f(t,Y(t,Y(t,h))$ ergibt mittels (i), (ii) und Umstellung

$$Y(t+h,-h) \;=\; Y(t,-h) + h\,f(t+h,Y(t+h,-h).$$

Das gespiegelte Euler Verfahren ist also gerade das implizite Euler Verfahren, vgl. (4.7). Insbesondere ist das Euler-Verfahren daher nicht symmetrisch!

Anmerkungen (6.15)

Ohne Beweise zitieren wir die folgenden Eigenschaften.

a) Die Spiegelung eines RK-Verfahrens ergibt wieder ein (ev. implizites) RK-Verfahren. Hat das Ausgangsverfahren das Butcher-Tableau

$$\Phi:\quad \begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1s} \\ c_2 & a_{21} & \dots & a_{2s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \dots & a_{ss} \\ \hline & b_1 & \dots & b_s \end{array}$$

so ergibt sich für das gespiegelte Verfahren das Tableau

$$\Phi^*:\quad \begin{array}{c|ccc} (1-c_s) & (b_s-a_{ss}) & \dots & (b_1-a_{s1}) \\ \vdots & \vdots & & \vdots \\ (1-c_1) & (b_s-a_{1s}) & \dots & (b_1-a_{11}) \\ \hline & b_s & \dots & b_1 \end{array}$$

b) Zweimalige Spiegelung ergibt wieder das Ausgangsverfahren, $\Phi^{**} = \Phi$.

c) Die Konsistenzordnung des ESV bleibt bei Spiegelung unverändert, genauer gilt für den lokalen Diskretisierungsfehler $\tau^* = (-1)^p\,\tau$.

d) Besitzt das ESV Φ eine asymptotische Entwicklung, wie im Satz von Gragg (6.11) angegeben, so besitzt der globale Fehler des gespiegelten Verfahrens die asymptotische Entwicklung

$$Y^*(t;-h) \;-\; y(t) \;=\; e_p(t)\,(-h)^p \;+\; \dots\; e_m(t)\,(-h)^m \;+\; \alpha_{m+1}(t,-h)\,(-h)^{m+1}$$

mit einem bzgl. h beschränktem Restglied $\|\alpha_{m+1}(t,-h)\| \le M^*_{m+1}$.

e) Aus d) ergibt sich die Folgerung: Ist das ESV Φ symmetrisch, so besitzt es unter den Voraussetzungen des Graggschen Satzes eine quadratische asymptotische Entwicklung.!!

Beispiele (6.16)

Sowohl die **implizite Trapezregel**

$$Y(t+h,h) \;=\; Y(t,h) + \frac{h}{2}\,[\,f(t,Y(t,h)) \;+\; f(t+h,Y(t+h,h)\,],$$

wie auch die **implizite Mittelpunktsregel**

$$Y(t+h,h) \;=\; Y(t,h) + h\,f(t+h/2,\,(Y(t,h)+f(t+h,Y(t+h,h))/2)$$

sind symmetrisch und besitzen daher eine *quadratische* asymptotische Entwicklung.

C. Extrapolation der Mittelpunktsregel.

Eines der frühesten und erfolgreichsten Verfahren, die auf Extrapolation beruhen und die mit einer automatischen Schrittweitensteuerung ausgestattet sind, ist das auf R. Bulirsch, W.B Gragg und J. Stoer (1966) zurückgehende Programm DIFSYS. Es wird hierin rationale Extrapolation der Mittelpunktsregel verwendet.

Gragg konnte in seiner Dissertation 1964 zeigen, dass der globale Diskretisierungsfehler für die Mittelpunktsregel, gestartet mit einem Euler-Schritt, eine quadratische asymptotische Entwicklung besitzt.

Die Mittelpunktsregel wird dabei für äquidistante Schrittweiten folgendermaßen ausgewertet: Für $t > t_0$, $h := (t-t_0)/N$ und $t_i = t_0 + i\,h$ $(i = 1,\ldots,N)$ wird berechnet:

$$\begin{aligned}
Y_0 \;&:=\; y_0 \\
Y_1 \;&:=\; Y_0 \;+\; h\,f(t_0,Y_0) \\
\text{für}\quad i \;&=\; 1,\,\ldots\,,N-1 \\
\qquad Y_{i+1} \;&:=\; Y_{i-1} \;+\; 2\,h\,f(t_i,Y_i) \\
Y(t;h) \;&:=\; Y_N
\end{aligned} \tag{6.13}$$

Beachten Sie, dass in der Anwendung nur die Mikroschritte äquidistant sind, während für die Makroschritte (Schrittweite H) eine Schrittweitensteuerung entwickelt werden wird.

Satz (6.14) (Gragg, 1965)

Der globale Diskretisierungsfehler der Mittelpunktsregel (6.12) besitzt eine quadratische asymptotische Entwicklung der Form

$$Y(t;h) \; - \; y(t) \;\; = \;\; \sum_{k=1}^{m} \; \left[e_k(t) \; + \; (-1)^N \, f_k(t) \right] \; h^{2\,k} \; + \; \alpha_{m+1}(t,h) \, h^{2(m+1)}$$

mit einem bzgl. h beschränktem Restglied $\|\alpha_{m+1}(t,h)\| \leq M_{m+1}$.

Ferner genügen die Koeffizientenfunktionen e_k und f_k einem gestaffelten, linearen, inhomogenen Differentialgleichungssystem

$$\begin{aligned} e_k'(t) \;\; &= \;\; f_y(t,y(t)) \, e_k(t) \; + \; \psi_k(t; y, e_1, \ldots, e_{k-1}); \quad e_k(t_0) = 0, \\ f_k'(t) \;\; &= \;\; -f_y(t,y(t)) \, f_k(t) \; + \; \phi_k(t; y, f_1, \ldots, f_{k-1}); \quad f_k(t_0) = 0. \end{aligned}$$

Bemerkungen (6.15)

De facto liegen mit Satz (6.14) *zwei* asymptotische Entwicklungen vor, je nachdem ob N gerade oder ungerade ist. Für die numerische Anwendung schränkt man sich daher auf den Fall N gerade ein.

Die Funktionen f_k repräsentieren die instabilen Fehlerterme, in ihnen spiegelt sich die Tatsache wieder, dass die Mittelpunktsregel nur ein schwach stabiles Verfahren ist.

Beispiel (6.16)

Wendet man das Verfahren (6.13) auf die AWA

$$y' \;\; = \;\; -y, \qquad y(0) \; = \; 1,$$

an, so findet man für die Lösung der Rekursion

$$Y_0 \; = \; 1, \quad Y_1 \; = \; 1 \; - \; h, \quad Y_{j+1} \; = \; Y_{j-1} \; - \; 2\,h\,Y_j, \quad j = 1, \ldots, N-1; \quad N \; = \; t/h$$

die folgende Darstellung (Übungsaufgabe)

$$\begin{aligned} y(t;h) \;\; &= \;\; \Phi_1(t;h) \;\; + \;\; (-1)^N \, \Phi_2(t;h) \\ \Phi_1(t;h) \;\; &= \;\; \frac{\sqrt{1+h^2} \; + \; 1}{2\,\sqrt{1+h^2}} \; \left[\sqrt{1+h^2} \; - \; h \right]^N \\ \Phi_2(t;h) \;\; &= \;\; \frac{\sqrt{1+h^2} \; - \; 1}{2\,\sqrt{1+h^2}} \; \left[\sqrt{1+h^2} \; + \; h \right]^N \end{aligned}$$

Die Funktionen Φ_1 und Φ_2 besitzen jeweils eine quadratische asymtotische Entwicklung in h mit

$$\Phi_1(t;h) \; = \; \mathrm{e}^{-t} \; + \; O(h^2), \quad \Phi_2(t;h) \; = \; \frac{1}{4} \, h^2 \, \mathrm{e}^{t} \; + \; O(h^4)$$

Es ist also hier zu erkennen, dass Φ_2 einen instabilen Anteil der diskreten Lösung beschreibt, der allerdings für kleine Schrittweiten mit h^2 gedämpft wird.

Der Graggsche Schlussschritt

Gragg gelang es nun weiterhin, den führenden instabilen Fehlerterm $f_1(t)$ in der Entwicklung (6.14) mittels eines speziellen *Schlussschrittes* zu eliminieren. Dieser Schlussschritt symmetrisiert gewissermaßen wieder das Verfahren durch die Verwendung eines weiteren Euler-Schrittes zum Abschluss. Das Graggsche Verfahren sieht wie folgt aus:

$$\begin{aligned} Y_0 &:= y_0 \\ Y_1 &:= Y_0 + h\, f(t_0, Y_0) \\ \text{für} \quad i &= 1, \ldots, N-1 \\ Y_{i+1} &:= Y_{i-1} + 2\, h\, f(t_i, Y_i) \\ Y(t;h) &:= 0.5\, [Y_{N-1} + (Y_N + h\, f(t_N, Y_N))]\,. \end{aligned} \tag{6.17}$$

Lässt man die Schleife in (6.17) bis zum Index N laufen, so lässt sich der Schlussschritt auch folgendermaßen formulieren

$$Y(t;h) \;=\; \frac{1}{4}\,[Y_{N-1} + 2\,Y_N + Y_{N+1}]\;.$$

Für die Terme in der rechten Seite gilt jeweils die asymptotische Entwicklung von Satz (6.13), also

$$\begin{aligned} Y_{N-1} &= y(t-h) + h^2\,\left[e_1(t-h) + (-1)^{N-1} f_1(t-h)\right] + \ldots \\ Y_N &= y(t) + h^2\,\left[e_1(t) + (-1)^{N} f_1(t)\right] + \ldots \\ Y_{N+1} &= y(t+h) + h^2\,\left[e_1(t+h) + (-1)^{N+1} f_1(t+h)\right] + \ldots \end{aligned}$$

Setzt man diese Entwicklungen oben ein und entwickelt nach h-Potenzen, so folgt in der Tat

$$Y(t;h) \;=\; y(t) + h^2\left[e_1(t) + \frac{1}{2}y''(t)\right] + \ldots,$$

d.h. der instabile h^2–Anteil wurde eliminiert.

Zur Extrapolation wird nun ausgehend von den nach (6.17) berechneten Näherungen $Y(t;h)$ das Extrapolationstableau (6.8) aufgestellt. Hierbei wird aber lediglich bis zu einem festen maximalen Polynomgrad m (Praxiswert $m = 6 - 8$) extrapoliert, d.h. man berechnet nur $m+1$ Spalten des Tableaus.

Das Hinzufügen weiterer Tableauzeilen wird dann solange fortgesetzt, bis der folgende *Konvergenztest* erfüllt ist:

$$|P_{k,\,m} - P_{k+1,\,m}| \;\le\; \text{tol}\;|P_{k,\,m}|. \tag{6.18}$$

Hierbei bezeichnet tol eine vom Benutzer vorzugegebende relative Genauigkeitsschranke.

$$
\begin{array}{lllll}
P_{0,0} & & & & \\
P_{1,0} & P_{0,1} & & & \\
P_{2,0} & P_{1,1} & P_{2,2} & & \\
\vdots & \vdots & \vdots & \ddots & \\
P_{m,0} & P_{m-1,1} & P_{m-2,2} & \dots & P_{0,m} \\
P_{m+1,0} & P_{m,1} & P_{m-1,2} & \dots & P_{1,m} \\
\vdots & \vdots & \vdots & \vdots & \vdots
\end{array}
$$

Schrittweitensteuerung:

Die Folge in der ℓ-ten Spalte des obigen Extrapolationstableaus ($\ell = 0, 1, \dots, m$) konvergiert wie $H^{2(\ell+1)}$ gegen die Lösung $y(t+H)$ der Differentialgleichung. Hierbei bezeichnet H wieder die Grundschrittweite $H = t_{j+1} - t_j$ mit der das Extrapolationstableau begonnen wird.

Genauer gilt aufgrund der asymptotischen Entwicklung

$$
\begin{aligned}
P_{k,\ell} &= y(t+H) + H^{2\ell+2} g_\ell(t+H) + O(H^{2\ell+4}) \\
P_{k+1,\ell} &= y(t+H) + \delta_k H^{2\ell+2} g_\ell(t+H) + O(H^{2\ell+4}),
\end{aligned}
$$

wobei $\delta_k \in]0,1[$ der Verkleinerungsfaktor der Grundschrittweite ist. Damit wird

$$
|P_{k,\ell} - P_{k+1,\ell}| = (1-\delta_k) H^{2\ell+2} |g_\ell(t+H)| + O(H^{2\ell+4})
$$

und per Taylor-Entwicklung von g_ℓ mit $g_\ell(t_j) = 0$

$$
|P_{k,\ell} - P_{k+1,\ell}| \approx (1-\delta_k) H^{2\ell+3} |g'_\ell(t)|. \tag{6.19}
$$

Für eine *optimale* Schrittweite H_{neu} fordert man nun

$$
|P_{k,\ell} - P_{k+1,\ell}| \equiv \alpha_k |P_{k+1,\ell}|
$$

mit

$$
\alpha_k = \alpha_0^k + 1, \quad \alpha_0 \approx 0.04.
$$

Zusammen mit (6.19) hat man dann

$$
\alpha_k |P_{k+1,\ell}| \approx (1-\delta_k) H_{\text{neu}}^{2\ell+3} |g'_\ell(t)|
$$

und schließlich durch Elimination von g'_ℓ aus dieser Gleichung und (6.19):

$$
H_{\text{neu}} = \left[\frac{\alpha_k |P_{k+1,\ell}|}{|P_{k,\ell} - P_{k+1,\ell}|} \right]^{1/(2\ell+3)} H_{\text{alt}}. \tag{6.20}
$$

Zur Schrittweitensteuerung geht man nun folgendermaßen vor:

Mit einer geschätzten Schrittweite H wird das Extrapolationstableau aufgebaut. In den Spalten $k = 0, 1, 2, 3$ wird jeweils H_{neu} gemäß (6.20) berechnet und getestet, ob $H_{\text{neu}} < 0.7\,H$ gilt. Ist dies der Fall, so verwirft man das berechnete Tableau und beginnt mit H_{neu} ein neues Extrapolationstableau aufzubauen. Ist die Bedingung dagegen nicht erfüllt, so vervollständigt man das Tableau bis der Konvergenztest (6.18) erfüllt ist. Die Integration wird dann mit H_{neu} fortgesetzt.

Natürlich gibt es in den verwendeten Algorithmen Varianten der Schrittweitensteuerung, die sich in den Details unterscheiden. Erwähnt werden sollte auch, dass sich mit der Information des Extrapolationstableus auch entscheiden lässt, ob es sich lohnt, den Grad des Interpolationspolynoms weiter zu erhöhen. Man erhält damit nicht nur eine adaptive Steuerung der Schrittweite sondern auch der Ordnung des Verfahrens, ganz analog zu der Ordnungssteuerung für Mehrschrittverfahren, die wir in Abschnitt 5.C. kennengelernt haben. Dies findet man ausführlicher in Strehmel, Weiner.

7. Steife Differentialgleichungen

A. Allgemeines.

Bei der numerischen Lösung von AWAen tritt mitunter ein Phänomen auf, dass als *Steifheit* der DGL (besser der AWA) bezeichnet wird. Besitzt ein DGLsystem etwa einen Lösungsanteil, dser schnell klein wird und dann gegenüber den anderen langsam veränderlichen Lösungsanteilen nicht mehr sichtbar ist, so kann denoch ein numerisches Verfahren gezwungen sein, aus Stabilitätsgründen die Schrittweite nach diesem schnell verschwindenden Lösungsanteil auszurichten und daher unverhältnismäßig viele Integrationsschritte zu benötigen. Wir verdeutlichen dies anhand des so genannten Van der Pol-Oszillators.

Beispiel (7.1) (Van der Pol-Oszillator)

Wir betrachten die folgende AWA

$$\begin{aligned} y_1' &= -y_2\,, & y_1(0) &= 1,\\ y_2' &= \frac{1}{\varepsilon}\,(y_1 - (1/3)\,y_2^3 + y_2)\,, & y_2(0) &= 2. \end{aligned}$$

Hierbei sei $\varepsilon = 10^{-4}$ und die Integrationslänge $t_b = 2$.

Wir lösen diese AWA mittels der in MATLAB implementierten Integratoren und der Genauigkeitsvorgabe $\mathrm{TOL} = 10^{-5}$. Neben den Standardintegratoren **ode45** und **ode113** verwenden wir auch solche Integratoren aus MATLAB, die insbesondere für steife DGLn geeignet sind:

ode15s: Mehrschrittverfahren basierend auf BDF

ode23s: Rosenbrock-RK-Verfahren der Ordnung 2

ode23t: Implizite Trapezregel

ode23tb: Kombination von impliziten RK-Verfahren und BDF

In der folgenden Tabelle 7.1 sind die Ergebnisse der numerischen Integration angegeben, Alle Integratoren liefern vergleichbare Genauigkeiten, allerdings ist der Aufwand (gemessen durch die Zahl der Integrationsschritte und die Anzahl NFC der Auswertungen der rechten Seite der DGL) sehr unterschiedlich. Wodurch lässt sich der verhebliche Mehraufwand der Verfahren **ode45** und **ode113** erklären? Am Lösungsverlauf, siehe Abb. 7.1, sind zwei unterschiedliche Zeitskalen zu erkennen, es gibt sehr kurze, schnelle Phasen und lange Phasen, in denen sich die Lösung nur langsam ändert. Die schnellen Phasen heißen *Grenzschichten*, die langsamen Phasen heißen *asymptotische oder transiente Phasen*.

Tabelle 7.1: Ergebnisse der numerischen Integration von (7.1)

	erfolgr. Schritte	nicht erfolgr. Schritte	NFC
ode45	12.112	740	77.113
ode113	24.951	3.378	53.281
ode15s	361	90	926
ode23s	888	5	4.452
ode23t	539	16	1303
ode23tb	393	28	1678

In der Abbildung 7.1. sind neben den beiden Lösungskomponenten y_1 (gestrichelt) und y_2 noch die Integrationsknoten für das Verfahren **ode23tb** eingezeichnet. Man erkennt, dass der Integrator nur in unmittelbarer Nähe der Grenzschichten sehr kleine Schrittweiten wählt. Für die Integratoren **ode45** und **ode113** sind die Schrittweiten jedoch überall extrem klein.

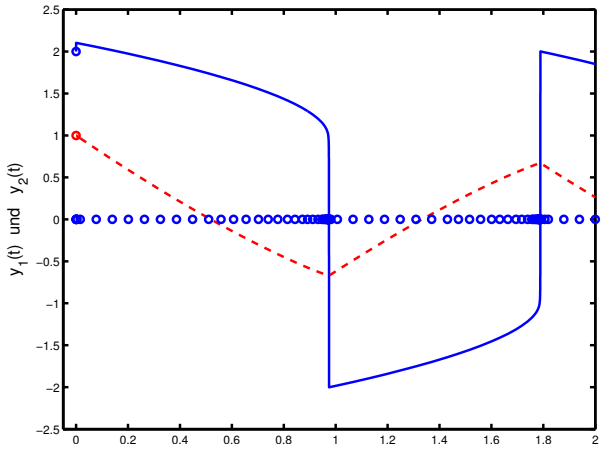

Abb. 7.1 Lösung zu Beispiel (7.1).

Beispiel (7.2) (Feder)

Wir betrachten eine Feder mit einer zeitlich veränderlichen, bewegten Aufhängung (beschrieben durch eine Funktion $f(t)$). Für den Ort $y(t)$ der an der Feder aufgehängten Masse m gilt dann nach dem Hookschen Gesetz die DGL

$$m\,y''(t) \;+\; \rho\,y'(t) \;\;=\;\; k\,(f(t) - y(t)).$$

Dabei bezeichnet ρ einen Reibungskoeffizient und k die Federkonstante. Wir gehen nun von den folgenden Größenverhältnissen aus: $0 < m \ll \rho \ll k$.

Vereinfachend können wir deshalb den Term zweiter Ordnung vernachlässigen und betrachten das *Ersatzproblem*

$$y'(t) \;=\; \lambda\,(f(t) \,-\, y(t)), \qquad \lambda \,:=\, k/\rho \,\gg\, 0.$$

Die Lösung der zugehörigen AWA mit $y(0) = y_0$ lässt sich hier explizit angeben

$$y(t) \;=\; y_0\,\mathrm{e}^{-\lambda\,t} \;+\; \int_0^t f(\tau)\,\lambda\,\mathrm{e}^{-\lambda\,(\tau-t)}\,d\tau\,.$$

Der erste Summand beschreibt eine schnelle Einschwingphase, der zweite Summand eine langsam veränderliche Mittelung von f-Daten.

Die Zeitkonstanten sind dann $\tau_1 \approx 1/\lambda$ (nach unserer Annahme sehr klein) und $\tau_2 \approx f(t)/f'(t)$ (relative Änderung von f, moderat).

Wir untersuchen nun das numerische Lösungverhalten für das Ersatzproblem bei Anwendung des expliziten und des impliziten Euler-Verfahrens (bei konstanter Schrittweite).

Explizites Euler-Verfahren:

$$\begin{aligned} & Y_{j+1} \;=\; Y_j \;+\; h\,\lambda\,(f_j - Y_j) \\ \Longrightarrow\quad & (Y_{j+1} - f_{j+1}) \;\approx\; (Y_{j+1} - f_j) \;=\; (1 - \lambda\,h)\,(Y_j - f_j) \\ \Longrightarrow\quad & (Y_{j+1} - f_{j+1}) \;\approx\; (1 - \lambda\,h)^{j+1}\,(Y_0 - f_0) \end{aligned}$$

Der Fehler $|Y_j - f_j|$ fällt also nur dann, wenn $|1 - \lambda\,h| < 1$ gilt, also $0 < \lambda\,h < 2$ ist. Für große Werte von λ muss die Schrittweite daher sehr klein gewählt werden.

Implizites Euler-Verfahren:

$$\begin{aligned} & Y_{j+1} \;=\; Y_j \;+\; h\,\lambda\,(f_{j+1} - Y_{j+1}) \\ \Longrightarrow\quad & (Y_{j+1} - f_{j+1}) \;=\; \frac{1}{1+\lambda\,h}\,(Y_j - f_{j+1}) \;\approx\; \frac{1}{1+\lambda\,h}\,(Y_j - f_j) \\ \Longrightarrow\quad & (Y_{j+1} - f_{j+1}) \;\approx\; \left(\frac{1}{1+\lambda\,h}\right)^{j+1}\,(Y_0 - f_0) \end{aligned}$$

Da $\lambda > 0$ fällt der Fehler in jedem Integrationsschritt und zwar unabhängig von der Schrittweitenwahl.

Es ist zu vermerken, dass es keine mathematisch präzise Definition dafür gibt, wann eine DGL (besser eine AWA) als *steif* bezeichnet wird. Kennzeichned sind stark unterschiedliche Zeitskalen bei den Lösungsanteilen.

B. A–Stabilität.

Wir hatten für Mehrschrittverfahren den Begriff der *Nullstabilität* kennengelernt (vgl. Abschnitt 5). Ein MSV ist nullstabil, wenn es - angewendet auf die triviale DGL $y' = 0$ - nur nichtwachsende (numerische) Lösungen liefert.

In Verallgemeinerung hiervon betrachtet man nach Dahlquist nun das nur wenig kompliziertere **Modellproblem**

$$y'(t) \;=\; \lambda\, y(t), \quad y(t_0) = y_0, \quad \lambda \in \mathbb{C}, \quad \mathrm{Re}\,\lambda < 0. \tag{7.3}$$

Definition (7.4)

Ein numerisches Integrationsverfahren heißt *absolut stabil* (auch *A-stabil*), falls für das Modellproblem (7.3) (bei beliebigem λ und Anfangswerten (t_0, y_0) und allen konstanten Schrittweiten $h > 0$) gilt, dass die Gitterabbildung $t \mapsto |Y(t;h)|$ monoton fällt.

In der Regel schränkt man sich bei der Definition (7.4) auf solche Integrationsverfahren ein, die angewendet auf das Modellproblem eine Rekursion der Form

$$Y_{j+1} \;=\; R(\lambda\, h) \,\cdot\, Y_j \tag{7.5}$$

ergeben, wobei $R : \mathbb{C} \supset D \to \mathbb{C}$ eine (komplexwertige) und zumindest in einer Umgebung von $z = 0$ analytische Funktion ist. Zu diesen Integrationsverfahren (manchmal auch Verfahren der Klasse (D) genannt) gehören alle bisher betrachteten Ein- und Mehrschrittverfahren.

Die Funktion R heißt dann die *Stabilitätsfunktion* des Integrationsverfahrens. Ferner heißt dann die Menge

$$S \;:=\; \{\, z \in \mathbb{C} :\, |R(z)| \,\le\, 1\} \tag{7.6}$$

der *Stabilitätsbereich* des Integrationsverfahrens.

Mit der Definition (7.4) und der Relation (7.5) ergibt sich damit, dass ein Integrationsverfahren der Klasse (D) genau dann A-stabil ist, wenn gilt

$$\mathbb{C}^- \;:=\; \{\, z \in \mathbb{C} :\; \mathrm{Re}\; z \,<\, 0\,\} \;\subset\; S. \tag{7.7}$$

Das Euler-Verfahren.

Das Euler-Verfahren liefert für das Modellproblem (7.3) die Rekursion

$$Y_{j+1} \;=\; Y_j \,+\, h\,(\lambda\, Y_j) \;=\; (1 \,+\, \lambda\, h)\, Y_j\,.$$

Damit lautet die Stabilitätsfunktion $R(z) = 1 + z$. Der Stabilitätsbereich $S = \{\, z : |1 + z| < 1\,\}$ ist also das Innere der Kreisscheibe um $z = -1$ mit Radius $r = 1$.

Insbesondere ist das Euler-Verfahren also *nicht* A-stabil.

Das implizite Euler-Verfahren.

Das implizite Euler-Verfahren ergibt für (7.3) die Rekursion

$$Y_{j+1} \;=\; Y_j \;+\; h\,(\lambda\, Y_{j+1}) \quad \Rightarrow \quad Y_{j+1} \;=\; \frac{1}{1-\lambda\, h}\, Y_j\,.$$

Damit ist $R(z) := 1/(1-z)$ die Stabilitätsfunktion des impliziten Euler-Verfahrens und $S := \{\, z \,:\, |1-z| \,>\, 1\,\}$ der Stabilitätsbereich. Dieser beschreibt also das Äußere des Kreises um $z = 1$ mit Radius $r = 1$ und umfasst damit die Menge $\mathbb{C}^-$. Insbesondere ist das implizite Euler-Verfahren also A-stabil.

Das klassische RK4-Verfahren.

Das klassische Runge-Kutta Verfahren vierter Ordnung

$$\begin{aligned} Y_{j+1} \;&=\; Y_j \;+\; \frac{h}{6}\,(k_1 \;+\; 2\,k_2 \;+\; 2\,k_3 \;+\; k_4), \\ k_1 \;&=\; f(Y_j), \quad k_2 \;=\; f(Y_j \;+\; \frac{h}{2}\,k_1), \quad k_3 \;=\; f(Y_j \;+\; \frac{h}{2}\,k_2), \quad k_4 \;=\; f(Y_j \;+\; h\,k_3) \end{aligned}$$

liefert für das Modellproblem die Rekursion

$$Y_{j+1} \;=\; \left[\, 1 \;+\; (\lambda\, h) \;+\; \frac{1}{2}\,(\lambda\, h)^2 \;+\; \frac{1}{6}\,(\lambda\, h)^3 \;+\; \frac{1}{24}\,(\lambda\, h)^4 \,\right]\, Y_j.$$

Die Stabilitätsfunktion lautet also $R(z) \;:=\; 1 \;+\; z \;+\; \frac{1}{2}\,z^2 \;+\; \frac{1}{6}\,z^3 \;+\; \frac{1}{24}\,z^4$. Der Stabilitätsbereich ist das Innere der in Abb. 7.2 dargestellten Kurve.

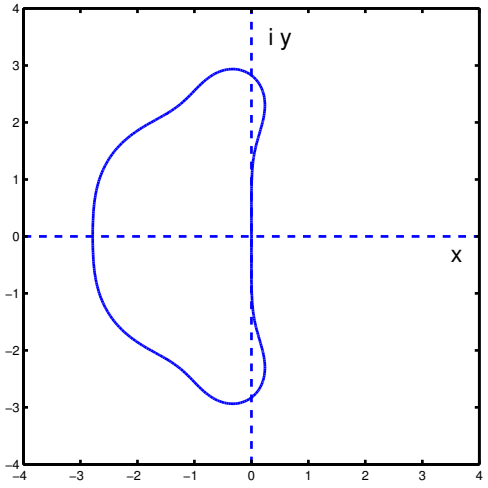

Abb. 7.2 Stabilitätsbereich des RK4-Verfahrens.

Das RK4-Verfahren ist also nicht A-stabil.

Man erkennt anhand des obigen Beispiels, dass *alle expliziten RK-Verfahren* auf eine polynomiale Stabilitätsfunktion führen. Damit gilt aber $\lim_{z\to\infty} R(z) \;=\; \infty$, so dass diese Verfahren nicht A-stabil sein können.

Implizite RK-Verfahren.

Die Stabilitätsfunktion eines impliziten RK-Verfahrens

$$\begin{aligned} Y_{j+1} &= Y_j + h \sum_{i=1}^{s} b_i\, k_i \\ k_i &= f(t_j + c_i\, h, Y_j + h \sum_{\ell=1}^{s} a_{i\,\ell}\, k_\ell) \end{aligned} \tag{7.8}$$

ist eine rationale Funktion mit den Darstellungen

$$R(z) = 1 + z\, b^{\mathrm{T}}(I_s - z\, A)^{-1}\mathbf{1}, \quad \mathbf{1} := (1, \ldots, 1)^{\mathrm{T}}, \tag{7.9}$$

$$R(z) = \frac{\det(I_s - zA + z\mathbf{1}b^{\mathrm{T}})}{\det(I_s - zA)}\,. \tag{7.10}$$

Beweis: Für $f(t, y) := \lambda\, y$ folgt aus (7.8)

$$\begin{pmatrix} k_1 \\ \vdots \\ k_s \end{pmatrix} = \lambda\, Y_J\, \mathbf{1} + \lambda\, h\, A \begin{pmatrix} k_1 \\ \vdots \\ k_s \end{pmatrix}$$

$$\Rightarrow \qquad (I_s - \lambda\, h\, A) \begin{pmatrix} k_1 \\ \vdots \\ k_s \end{pmatrix} = \lambda\, Y_j\, \mathbf{1}$$

$$\Rightarrow \qquad Y_{j+1} = Y_j + h\, b^{\mathrm{T}} k = Y_j + h\, \lambda\, Y_j\; b^{\mathrm{T}}(I_s - \lambda\, h\, A)^{-1}\, \mathbf{1}$$

Damit ist (7.9) gezeigt. Zu (7.10) schreibt man

$$\begin{pmatrix} I_s - \lambda\, h\, A & 0 \\ -\, h\, b^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} k \\ Y_{j+1} \end{pmatrix} = \begin{pmatrix} \lambda\, Y_j\, \mathbf{1} \\ Y_j \end{pmatrix}$$

und wendet hierauf die Cramersche Regel an:

$$Y_{j+1} = \det \begin{pmatrix} I_s - \lambda\, h\, A & \lambda\, Y_j\, \mathbf{1} \\ -\, h\, b^{\mathrm{T}} & Y_j \end{pmatrix} \Big/ \det \begin{pmatrix} I_s - \lambda\, h\, A & 0 \\ -\, h\, b^{\mathrm{T}} & 1 \end{pmatrix}.$$

Für den Zähler folgt durch Subtraktion vom λ–fachen der letzten Zeile von allen anderen Zeilen:

$$\begin{aligned} \text{Zähler} &= \det \begin{pmatrix} I_s - \lambda\, h\, A + \lambda\, h\, \mathbf{1}\, b^{\mathrm{T}} & 0 \\ -\, h\, b^{\mathrm{T}} & Y_j \end{pmatrix} \\ &= Y_j\, \det(I_s - \lambda\, h\, A + \lambda\, h\, \mathbf{1}\, b^{\mathrm{T}}) \end{aligned}$$

□

Kollokationsverfahren[1]**.**

Implizite RK–Verfahren lassen sich durch sogenannte Kollokationsverfahren erzeugen. Dazu betrachtet man einen Integrationsschritt im Intervall $[t_j, t_j + h]$ und bestimmt zu vorgegebenen Kollokationspunkten

$$0 \leq c_1 < \ldots < c_s \leq 1$$

ein Polynom $w \in \Pi_s$ mit den Eigenschaften

$$\begin{aligned} w(t_j) &= Y_j \\ w'(t_j + c_i\, h) &= f(t_j + c_i\, h, w(t_j + c_i\, h)), \quad i = 1, \ldots, s \end{aligned} \tag{7.11}$$

In den Kollokationspunkten soll w (transformiert auf $[t_j, t_j + h]$) also die vorgegebene DGL erfüllen. Damit setzt man nun $Y_{j+1} := w(t_j + h)$.

Es gelten dann die folgenden Eigenschaften

a) Das Kollokationsverfahren ist äquivalent zu einem s–stufigen impliziten RK–Verfahren mit den Koeffizienten

$$a_{i\,\ell} = \int_0^{c_i} L_\ell(t)\, dt, \quad b_i = \int_0^1 L_i(t)\, dt\,. \tag{7.12}$$

Hierbei bezeichnet $L_i(t) := \Pi_{\nu \neq i}(t - c_\nu)/(c_i - c_\nu)$ die Lagrange-Polynome zu den Knoten c_i. Wegen $\sum_{i=1}^{s} L_i = 1$ ist hiermit die Knotenbedingung $c_i = \sum_{\ell=1}^{s} a_{i\,\ell}$ erfüllt.

b) Ein durch Kollokation erzeugtes RK–Verfahren hat genau dann die Konsistenzordnung p, wenn die durch die Stützstellen c_i und die Gewichte b_i gegebene Quadraturformel die Ordnung p besitzt, d.h. wenn Polynome vom Grad kleiner oder gleich p durch die Quadraturformel exakt integriert werden.

Beispiele (7.13)

a) Gauß–Verfahren.
Hierbei sind die c_i als Nullstellen der verschobenen Legendre-Polynome $\frac{d^s}{dt^s}[t^s\,(t-1)^s]$ gewählt. Die Gauß–RK–Verfahren sind A-stabil und haben die Konsistenzordnung $p = 2\,s$.

b) Radau IA bzw. Radau IIA.
Hierbei werden die c_i als Nullstellen der Polynome $\frac{d^{s-1}}{dt^{s-1}}[t^s\,(t-1)^{s-1}]$ bzw. $\frac{d^{s-1}}{dt^{s-1}}[t^{s-1}\,(t-1)^s]$ gewählt. Die Konsistenzordnung dieser Verfahren ist $p = 2\,s - 1$. Die Verfahren sind A-stabil und erfüllen zudem die Bedingung $\lim_{z\to\infty} R(z) = 0$. Man nennt solche Verfahren *L-stabil*.

[1] aus dem Lateinischen: con = zusammen, locus = Ort

Mehrschrittverfahren.

Wendet man ein lineares Mehrschrittverfahren

$$\sum_{k=0}^{s} \alpha_k \, Y_{j+k} \;=\; h \sum_{k=0}^{s} \beta_k \, f_{j+k} \tag{7.14}$$

auf das Modellproblem (7.3) an, so ergibt sich

$$\sum_{k=0}^{s} (\alpha_k - (\lambda\, h)\, \beta_k)\, Y_{j+k} \;=\; 0.$$

Damit folgt unmittelbar:

Satz (7.15)

Ein lineares MSV (7.14) ist genau dann A–stabil, wenn für alle $z \in \mathbb{C}$ mit $\mathrm{Re}\, z < 0$ gilt: Alle Nullstellen ζ des Polynoms $p(\zeta) := \rho(\zeta) - z\,\sigma(\zeta)$ liegen im Einheitskreis, $|\zeta| \le 1$, uns sind einfach, falls sie auf dem Rand des Einheitskreises, $|\zeta| = 1$, liegen. Dabei bezeichnen ρ und σ die beiden charakteristischen Polynome des MSV, vgl. (5.19).

Wieder heißt die Menge

$$S \;:=\; \{z \in \mathbb{C} : \quad \mathrm{Re}\, z < 0,\;\; p := \rho - z\,\sigma \text{ erfüllt (7.15)}\} \tag{7.16}$$

der *Stabilitätsbereich* des MSV.

Bemerkungen

Die (expliziten) Adams-Bashforth Verfahren sind für keien Schrittzahl s A–stabil, die zugehörigen impliziten Adams–Moulton Verfahren sind nur für $s = 1$ A–stabil.

Dagegen haben die BDF–Verfahren sehr gute Stabilitätseigenschaften. Sie sind für $s = 1$ und $s = 2$ A–stabil und für $3 \le s \le 6$ zumindest noch A(α)–stabil, d.h. es gibt ein $\alpha > 0$, so dass der Sektor $\{z : \; |\arg(-z)| < \alpha\}$ ganz zum Stabilitätsbereich gehört.

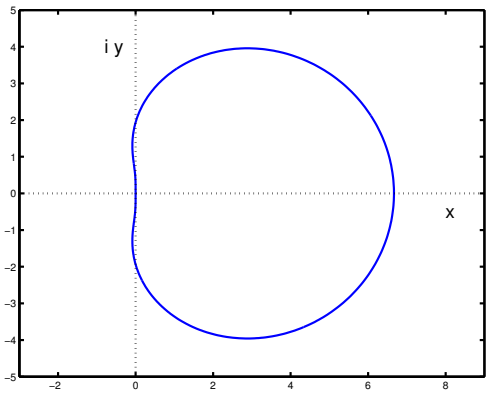

Abb. 7.3 Stabilitätsbereich des BDF-Verfahrens der Stufenzahl 3.

Zur Skizzierung des Stabilitätsbereiches bestimmt man diejenigen Werte $z \in \mathbb{C}$ für die $p(\zeta) = \rho(\zeta) - z\,\sigma(\zeta)$ eine Nullstelle vom Betrag $|\zeta| = 1$ besitzt. Dazu setzt man $\zeta = \mathrm{e}^{i\,\phi}$, $\phi \in [0, 2\,\pi]$, löst die Gleichung $p(\zeta) = 0$ nach z auf uns stellt z in Abhängigkeit von ϕ dar.

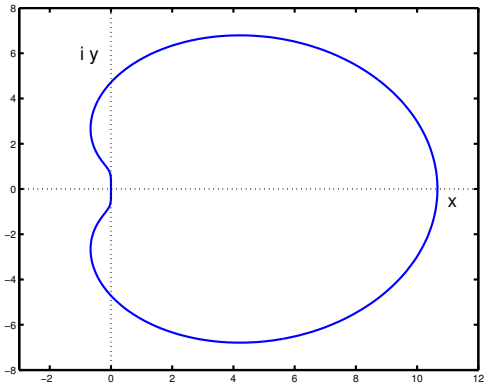

Abb. 7.4 Stabilitätsbereich des BDF-Verfahrens der Stufenzahl 4.

In den Abbildungen 7.3 und 7.4. sind die Stabilitätsbereiche der BDF–Verfahren mit Stufenzahlen $s = 3$ und $s = 4$ dargestellt. Die Stabilitätsbereiche sind dabei jeweils das *Äußere* der dargestellten Kurven, wobei allerdings nur der Bereich $\mathrm{Re}\,z \leq 0$ relevant ist.

8. Randwertaufgaben

A. Allgemeines.

Wir betrachten wieder ein DGLsystem erster Ordnung:

$$y'(t) \quad = \quad f(t, y(t)) \tag{8.17}$$

mit $y(t) \in \mathbb{R}^n$ und hinreichend glatter rechter Seite $f : \mathbb{R} \times D \to \mathbb{R}^n$, wobei D ein Gebiet in $\mathbb{R}^n$ sei.

Sind zur Festlegung einer speziellen Lösung y nicht alle Koordinaten $y_i,\ i = 1, \ldots, n,$ an *einer* Stelle a vorgegeben, sondern jeweils nur gewisse Komponenten y_i an verschiedenen Stellen $t_j = a, b, c, \ldots,$ so spricht man von einer *Randwertaufgabe (RWA)*, je nachdem auch von einem *Zweipunkt–Randwertproblem* oder einem *Mehrpunkt–Randwertproblem.*

Beispiele (8.2)

a) *Sturmsche RWA*

$$\begin{aligned} y''(t) + a_1(t)\, y'(t) + a_0(t)\, y(t) &= h(t) \\ \alpha_1\, y(a) \quad + \alpha_2\, y'(a) &= d_1 \\ \beta_1\, y(b) \quad + \beta_2\, y'(b) &= d_2\,, \end{aligned} \tag{8.3}$$

b) *Lineare RWA*

$$\begin{aligned} & y'(t) \quad = \quad A(t)\, y(t) \ + \ h(t) \\ & B_a\, y(a) \ + \ B_b\, y(b) \quad = \quad d \\ & \text{mit} \quad A(t), B_a, B_b \in \mathbb{R}^{(n,n)}, \quad h(t), d \in \mathbb{R}^n\,, \end{aligned} \tag{8.4}$$

c) *Allgemeine Zweipunkt–RWA*

$$\begin{aligned} & y'(t) \quad = \quad f(t, y(t)) \\ & r\,(y(a), y(b)) \quad = \quad 0 \\ & \text{mit} \quad y(t) \in \mathbb{R}^n, \quad r : \mathbb{R}^n \times \mathbb{R}^n \ \to \ \mathbb{R}^n\,. \end{aligned} \tag{8.5}$$

Im Unterschied zu den AWA lassen sich Existenz und Eindeutigkeit einer Lösung der Randwertaufgabe – selbst bei glatten Daten – nicht garantieren. Dies belegt etwa das folgende Beispiel.

Beispiel (8.6)

Die RWA

$$\begin{aligned} y'' &= -y \\ y(0) &= 0, \qquad y(\pi/2) = 1 \end{aligned}$$

besitzt die eindeutig bestimmte Lösung $y(t) = \sin t$.

Ändert man die Randbedingungen ab zu:

$$y(0) = 0, \qquad y(\pi) = 1\,,$$

so besitzt die Randwertaufgabe keine Lösung, bei den Randbedingungen

$$y(0) = 0, \qquad y(\pi) = 0$$

besitzt sie dagegen unendlich viele Lösungen, nämlich $y(t) = C\,\sin t,\ C \in \mathbb{R}$ beliebig.

Für lineare Randwertaufgaben (8.4) erhält man mit Hilfe eines Fundamentalystems ein Kriterium für die eindeutige Lösbarkeit:

Satz (8.7)

Gegeben sei eine lineare RWA (8.4) mit stetigen, auf $\mathbb{R}$ definierten Funktionen A und h. Y bezeichne ein Fundamentalsystem der homogenen Gleichung $y' = A\,y$. Dann sind die folgenden Aussagen äquivalent:

a) Die RWA (8.4) ist für *alle* (stetigen) Inhomogenitäten h und d stets eindeutig lösbar.

b) Die zugehörige homogene RWA

$$y' = A\,y, \qquad B_a\,y(a) + B_b\,y(b) = 0$$

hat nur die triviale Lösung $y = 0$.

c) Die so genannte *shooting Matrix*

$$E := B_a\,Y(a) + B_b\,Y(b) \in \mathbb{R}^{(n,n)} \tag{8.8}$$

ist regulär.

Beweis:

Die allgemeine Lösung der DGL lautet

$$y(t) = y_p(t) + Y(t)\,c\,, \qquad c \in \mathbb{R}^n\,.$$

Dabei ist y_p eine partikuläre Lösung.

Dies in die Randbedingungen eingesetzt liefert:

$$\begin{aligned} & B_a\ (y_p(a)\ +\ Y(a)\,c)\ +\ B_b\ (y_p(b)\ +\ Y(b)\,c) \quad = \quad d \\ \Longleftrightarrow \qquad & E\,c \quad = \quad d\ -\ B_a\,y_p(a)\ -\ B_b\,y_p(b)\,. \end{aligned}$$

Somit ist die eindeutige Lösbarkeit der RWA äquivalent zur Regularität der Matrix E. Dies gilt unabhängig von den speziellen Inhomogenitäten h und d. $\square$

Beispiel (8.9)

Für das obige Beispiel (8.6) – umgeschrieben in ein System erster Ordnung – erhält man eine Fundamentalmatrix

$$Y(t) \quad = \quad \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Damit folgt:

$$\begin{aligned} E \ &= \ B_a\,Y(0)\ +\ B_b\,Y(b) \\ &= \ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ + \ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \cos b & \sin b \\ -\sin b & \cos b \end{pmatrix} \\ &= \ \begin{pmatrix} 1 & 0 \\ \cos b & \sin b \end{pmatrix}. \end{aligned}$$

Die Matrix E ist also für $b = \pi/2$ regulär, für $b = \pi$ jedoch singulär.

Eine Reihe anderer Aufgaben lässt sich auf die *Normalform* (8.5) einer Zweipunkt-RWA transformieren. Zu diesen gehören u.a. *Eigenwertaufgaben bei gewöhnlichen DGLen* sowie *RWA mit freier Endzeit*.

Beispiel (8.10) (Eigenwertaufgabe)

Zu einer vorgegebenen Funktion $q : [a, b] \to \mathbb{R}$ sind diejenigen $\lambda \in \mathbb{R}$ (*Eigenwerte*) gesucht, für die die RWA

$$\begin{aligned} & z'' \ + \ (\lambda \ - \ q(t))\ z \ = \ 0, \qquad a \le t \le b \\ & z(a) \ = \ 0, \qquad z(b) \ = \ 0 \end{aligned} \tag{8.11}$$

eine nicht identisch verschwindende Lösung (*Eigenfunktion*) besitzt.

Setzt man

$$y_1 \ := \ z, \qquad y_2 \ := \ z', \qquad y_3 \ := \ \lambda$$

so erhält man die folgende Randwertaufgabe in Normalform

$$\begin{aligned} y_1' \ &= \ y_2 \\ y_2' \ &= \ (q(t) - y_3)\ y_1 \\ y_3' \ &= \ 0 \\ y_1(a) \ &= \ 0, \quad y_1(b) \ = \ 0, \quad y_2(a) \ = \ 1. \end{aligned} \tag{8.12}$$

Die dritte Randbedingung $y_2(a) = 1$ ist eine *Normierungsbedingung*. Man beachte, dass die obige RWA durchaus unendlich viele Lösungen besitzen kann. Dennoch lassen sich die Lösungen numerisch bestimmen, sofern sie separiert sind (Matrix E regulär).

Beispiel (8.13) (RWA mit freier Endzeit)

Eine RWA mit freier Endzeit hat die allgemeine Form

$$\begin{aligned} y'(t) &= f(t, y(t)), \qquad 0 \le t \le t_b \\ r(y(0), y(t_b)) &= 0\,. \end{aligned} \tag{8.13}$$

Hierbei ist die Endzeit t_b nicht vorgegeben (frei) und $r : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n+1}$ ($n+1$ Randbedingungen!).

Setzt man nun

$$z(\tau) := y(\tau \cdot t_b), \qquad z_{n+1}(\tau) := t_b, \qquad 0 \le \tau \le 1,$$

so ist das Problem (8.13) zu der folgenden RWA in Normalform äquivalent

$$\begin{aligned} z'(\tau) &= z_{n+1} \cdot f(z_{n+1}\,\tau, z(\tau))\,, \qquad 0 \le \tau \le 1 \\ z'_{n+1}(\tau) &= 0 \\ r(z(0), z(1)) &= 0\,. \end{aligned} \tag{8.14}$$

RWA mit freier Endzeit treten häufig als notwendige Bedingungen von Optimalsteuerungsaufgaben auf, bei denen die Endzeit minimiert werden soll.

Beispiel (8.15) (Navigationsproblem von Zermelo[1])

Ein Fährmann habe die Aufgabe, mit seinem Boot einen Fluss zu überqueren. Dieser sei in der (x, y)-Ebene gegeben durch den Streifen $-1 \le x \le 1$. Ausgangspunkt sei $A = (-1, 0)$; das Ziel liege am gegenüberliegenden Ufer in $B = (1, 0)$. Die Flussströmung sei gegeben durch ein quadratische Profil $\mathbf{v}(x, y) = (0, v_0\,(1 - x^2))^{\mathrm{T}}$.

Ferner habe das Boot habe eine Eigengeschwindigkeit $\mathbf{w}$ mit konstantem Betrag $w = \|\mathbf{w}\|$, deren Richtung (Winkel $u(t)$) jedoch zeitabhängig beliebig vorgegeben werden kann.

Das Problem lässt sich nun als eine Aufgabe der optimalen Steuerung formulieren: Wie muss der Fährmann bei vorgegebenen Daten v_0 und w den Winkelverlauf $u(t)$ wählen, damit der Fluss in kürzester Zeit überqert wird?

Mit den angegebenen Bezeichnung liefert die Theorie der optimalen Steuerung, dass eine zeitoptimale Lösung der folgenden RWA mit freier Endzeit genügen muss

[1] Ernst Zermelo, 1871–1953; Berlin, Göttingen, Zürich, Freiburg.

$$
\begin{aligned}
x'(t) &= w\,\cos u(t), && x(0) = -1,\ x(t_b) = 1,\\
y'(t) &= w\,\sin u(t) + v(x(t)), && y(0) = 0,\quad y(t_b) = 0,\\
\lambda_x'(t) &= -\lambda_y(t)\,v'(x(t)), && \lambda_x^2(t_b) + \lambda_y^2(t_b) = 1,\\
\lambda_y'(t) &= 0.
\end{aligned}
\tag{8.16}
$$

Hierbei bezeichnen λ_x und λ_y die so genannten *adjungierten Variablen*. Die Steuerung u wird durch eine weitere notwendige Optimalitätsbedingung festgelegt. Im vorliegenden Problem lautet diese

$$
\cos u(t) = -\frac{\lambda_x(t)}{\sqrt{\lambda_x(t)^2+\lambda_y(t)^2}}, \qquad \sin u(t) = -\frac{\lambda_y(t)}{\sqrt{\lambda_x(t)^2+\lambda_y(t)^2}}. \tag{8.17}
$$

Mit (8.17) liesse sich das DGLsystem (8.16) nun numerisch integrieren, allerdings sind fünf Randbedingungen für die vier DGL vorgegeben und die Endzeit t_b ist frei. Die Transformation (8.14) auf das (feste) Intervall $[0,1]$ ergibt nun das DGLsystem:

$$
\begin{aligned}
x'(t) &= -t_b(t)\,w\,\frac{\lambda_x(t)}{\sqrt{\lambda_x(t)^2+\lambda_y(t)^2}},\\
y'(t) &= -t_b(t)\left(w\,\frac{\lambda_x(t)}{\sqrt{\lambda_x(t)^2+\lambda_y(t)^2}} + v(x(t))\right),\\
\lambda_x'(t) &= t_b(t)\,\lambda_y(t)\,v'(x(t)),\\
\lambda_y'(t) &= 0,\\
t_b'(t) &= 0,
\end{aligned}
\tag{8.18}
$$

welches zusammen mit den obigen Randbedingungen eine RWA in Normalform bildet.

B. Lineare RWA zweiter Ordnung.

Wir betrachten eine lineare RWA zweiter Ordnung

$$
\begin{aligned}
L[y] &:= y''(t) + a_1(t)\,y'(t) + a_0(t)\,y(t) &&= h(t)\\
R_1[y] &:= \alpha_1\,y(a) + \beta_1\,y'(a) + \gamma_1\,y(b) + \delta_1\,y'(b) &&= d_1\\
R_2[y] &:= \alpha_2\,y(a) + \beta_2\,y'(a) + \gamma_2\,y(b) + \delta_2\,y'(b) &&= d_2\,,
\end{aligned}
\tag{8.19}
$$

und nehmen an, dass die zugehörige homogene RWA

$$
L[y] = 0, \qquad R_1[y] = R_2[y] = 0 \tag{8.20}
$$

nur die triviale Lösung besitzt. Nach Satz (8.7) besitzt dann auch (8.19) eine eindeutig bestimmte Lösung y.

Wir wollen zunächst die Aufgabe auf den *Fall homogener Randbedingungen*, d.h. $d_1 = d_2 = 0$, reduzieren.

Dazu sei y_0 eine beliebige C^2–Funktion, die die Randbedingungen $R_1[y_0] = d_1$ und $R_2[y_0] = d_2$ erfüllt.

Wir setzen: $y(t) := y_0(t) + z(t)$. Damit folgt

$$\begin{aligned} L[y] &= L[y_0] + L[z] &= h(t) \\ \iff \quad L[z] &= h(t) - L[y_0] &=: \tilde{h}(t) \end{aligned}$$

und $\quad R_k[y] = d_k \iff R_k[z] = 0, \ k = 1, 2$.

Anstelle der ursprünglichen RWA genügt es also, eine RWA (für $z(t)$) mit homogenen Randbedingungen zu lösen.

Im Folgenden nehmen wir o.B.d.A. an, dass in (8.19) bereits $d_1 = d_2 = 0$ vorliegt.

Wir versuchen nun eine Lösungsdarstellung der folgenden Form zu finden

$$y(t) \quad = \quad \int_a^b G(t,\tau)\, h(\tau)\, d\tau \,. \tag{8.21}$$

Die Funktion $G(t,\tau)$, $a \le t$, $\tau \le b$ heißt dann eine *Greensche Funktion* der RWA (8.19).

Diese hängt nur vom Differentialoperator $L[y]$ und den Randbedingungen $R_1[y]$, $R_2[y]$ ab, nicht aber von der Inhomogenität h.

Konstruktion der Greenschen Funktion G :

Wir nehmen an, dass G auf jedem der Bereiche

$$\begin{aligned} D_1 &:= \{(t,\tau) \ : \ a \le \tau \le t \le b\} \\ D_2 &:= \{(t,\tau) \ : \ a \le t \le \tau \le b\} \end{aligned}$$

glatt ist, d.h. sich als eine C^2–Funktion auf den Rand fortsetzen lässt, dass jedoch für $t = \tau$ Sprünge auftreten können.

Wir bilden nun die Ableitungen von (8.21):

$$
\begin{aligned}
y(t) &= \int_a^b G(t,\tau)\, h(\tau)\, d\tau \\
y'(t) &= \frac{d}{dt}\left\{ \int_a^t G(t,\tau)\, h(\tau)\, d\tau \;+\; \int_t^b G(t,\tau)\, h(\tau)\, d\tau \right\} \qquad (8.22) \\
&= \int_a^b G_t(t,\tau)\, h(\tau)\, d\tau \;+\; \left[G(t,t^-) - G(t,t^+)\right]\, h(t)
\end{aligned}
$$

und fordern, dass $G(t,t^-) - G(t,t^+)$ verschwindet, $G(t,\tau)$ also in $t = \tau$ stetig ist.

Genauso folgt dann für die zweite Ableitung

$$
y''(t) \quad = \quad \int_a^b G_{tt}(t,\tau)\, h(\tau)\, d\tau \;+\; \left[G_t(t,t^-) - G_t(t,t^+)\right]\, h(t) \qquad (8.23)
$$

und damit

$$
L[y] \quad = \quad \int_a^b L\left[G(\cdot,\tau)\right]\, h(\tau)\, d\tau \;+\; \left[G_t(t,t^-) - G_t(t,t^+)\right]\, h(t)\,.
$$

Für die Randbedingungen erhält man ferner:

$$
R_k[y] \quad = \quad \int_a^b R_k\left[G(\cdot,\tau)\right]\, h(\tau)\, d\tau \quad = \quad 0\,, \quad k = 1,2\,. \qquad (8.24)
$$

Insgesamt haben wir damit den folgenden Satz bewiesen:

Satz (8.25)

Erfüllt die Greensche Funktion $G : [a,b]^2 \to \mathbb{R}$ die Eigenschaften:

a) G ist stetig auf $[a,b]^2$ und lässt sich auf D_1 und auf D_2 als C^2–Funktion fortsetzen,

b) $G(t,\tau)$ erfüllt bei festem τ die homogene Differentialgleichung $L\left[G(\cdot,\tau)\right] = 0$, für $t \in [a,\tau]$ und $t \in [\tau,b]$ und die Randbedingungen

$$
R_k\left[G(\cdot,\tau)\right] \;=\; 0, \quad k = 1,\, 2\,,
$$

c) $G_t(t,t^-) \;-\; G_t(t,t^+) \;=\; 1,$

so ist die Lösung y der RWA gegeben durch $\quad y(t) \;=\; \int_a^b G(t,\tau)\, h(\tau)\, d\tau\,.$

Zur *Konstruktion einer Greenschen Funktion* kann man folgendermaßen vorgehen:

1.) Man bestimme ein Fundamentalsystem $y_1,\ y_2$ der homogenen Differentialgleichung und setze (einseitige Grenzwerte für $\tau = t$):

$$G(t,\tau) \quad := \quad \begin{cases} \sum\limits_{i=1}^{2} \left(a_i(\tau) + b_i(\tau)\right)\ y_i(t) \ : \ \tau \ \leq \ t \\ \sum\limits_{i=1}^{2} \left(a_i(\tau) - b_i(\tau)\right)\ y_i(t) \ : \ \tau \ \geq \ t \end{cases} \tag{8.26}$$

2.) Aufgrund der geforderten Stetigkeit und der Sprungbedingung (8.25) c) hat man

$$\begin{aligned} \sum_{i=1}^{2} b_i(t)\, y_i(t) &= 0 \\ \sum_{i=1}^{2} b_i(t)\, y_i'(t) &= \frac{1}{2}\,. \end{aligned} \tag{8.27}$$

Dies ist ein lineares Gleichungssystem zur Bestimmung von $b_1(t), b_2(t)$. Die Koeffizientenmatrix ist die zugehörige Fundamentalmatrix, sie ist also insbesondere regulär.

3.) Man setze (8.26) in die Randbedingungen ein. Dies ergibt ein lineares Gleichungssystem für $a_1(t), a_2(t)$, welches ebenfalls eindeutig lösbar ist, da die homogene RWA nach Voraussetzung nur die triviale Lösung besitzt.

Beispiel (8.28)

$$\begin{aligned} y''(t) + y(t) &= h(t) \\ y(0) \ - y(\pi) &= 0 \\ y'(0) - y'(\pi) &= 0\,. \end{aligned}$$

Ein Fundamentalsystem ist gegeben durch $y_1(t) = \cos\, t$ und $y_2(t) = \sin\, t$. Man hat also den Ansatz:

$$G(t,\tau) \quad = \quad \begin{cases} [a_1(\tau) + b_1(\tau)]\, \cos\, t + [a_2(\tau) + b_2(\tau)]\, \sin\, t \ : \ \tau \ \leq \ t \\ [a_1(\tau) - b_1(\tau)]\, \cos\, t + [a_2(\tau) - b_2(\tau)]\, \sin\, t \ : \ \tau \ \geq \ t \end{cases}.$$

Die Stetigkeit und die Sprungbedingung liefern das Gleichungssystem

$$\begin{aligned} b_1(t)\, \cos\, t + b_2(t)\, \sin\, t &= 0 \\ -\, b_1(t)\, \sin\, t + b_2(t)\, \cos\, t &= \frac{1}{2} \end{aligned}$$

mit der Lösung: $b_1(t) = -\frac{1}{2}\, \sin\, t,\ b_2(t) = \frac{1}{2}\, \cos\, t.$

Schließlich erhält man durch Auswerten der Randbedingungen:

$$\begin{aligned} G(0,\tau) - G(\pi,\tau) &= [a_1(\tau) - b_1(\tau)] + [a_1(\tau) + b_1(\tau)] &= 0 \\ G_t(0,\tau) - G_t(\pi,\tau) &= [a_2(\tau) - b_2(\tau)] + [a_2(\tau) + b_2(\tau)] &= 0 \end{aligned}$$

und damit $a_1(\tau) = a_2(\tau) = 0$.

Die Greensche Funktion lautet also

$$G(t,\tau) \quad = \quad \begin{cases} \dfrac{1}{2}\,\sin(t-\tau)\,, & \tau \le t \\ -\dfrac{1}{2}\,\sin(t-\tau)\,, & \tau \ge t \end{cases}$$

und die Lösung der Randwertaufgabe ergibt sich zu

$$y(t) \quad = \quad \frac{1}{2}\int\limits_0^t \sin(t-\tau)\,h(\tau)\,d\tau \;-\; \frac{1}{2}\int\limits_t^\pi \sin(t-\tau)\,h(\tau)\,d\tau\,.$$

Bemerkung (8.29)

Es sei abschließend angemerkt, dass sich die Methode der Greenschen Funktion ohne Mühe auf den Fall linearer Randwertaufgaben höherer Ordnung übertragen lässt.

C. Eigenwertaufgaben.

Wir betrachten eine homogene lineare RWA n–ter Ordnung

$$\begin{aligned} L\,[y] \quad &= \quad y^{(n)}(t) \;+\; a_{n-1}(t,\lambda)\,y^{(n-1)}(t) \;+\; \ldots \;+\; a_0(t,\lambda)\,y(t) \quad = \quad 0 \\ R_k[y,\lambda] \quad &= \quad \sum_{\ell=0}^{n-1} \left[\alpha_{k,\ell}(\lambda)\,y^{(\ell)}(a) \;+\; \beta_{k,\ell}(\lambda)\,y^{(\ell)}(b)\right] \quad = \quad 0\,, \qquad k = 1,2,\ldots,n\,. \end{aligned} \tag{8.30}$$

Die Koeffizienten der DGL und/oder der Randbedingungen mögen dabei Funktionen eines Parameters $\lambda \in \mathbb{R}$ bzw. $\in \mathbb{C}$ sein. Wir fragen nach *nichttrivialen Lösungen* dieser RWA.

Dazu sei $(y_1,\ldots,y_n)$ ein Fundamentalsystem von $L\,[y] = 0$. Die $y_k = y_k(t,\lambda)$ können dabei ebenfalls vom Paramter λ abhängen.

Eine Linearkombination $y(t) = \sum\limits_{k=j}^{n} c_j\,y_j(t,\lambda)$ löst nun genau dann die RWA, falls

$$\forall\, k \;:\; R_k[y] \quad = \quad \sum_{j=1}^{n} c_j\,R_k[y_j] \quad = \quad 0 \tag{8.31}$$

gilt. Dies ist ein homogenes lineares Gleichungssystem für die $c_1,\ldots,c_n$ mit der Koeffizientenmatrix

$$E(\lambda) \quad := \quad \begin{pmatrix} R_1[y_1] & \ldots & R_1[y_n] \\ \vdots & & \vdots \\ R_n[y_n] & \ldots & R_n[y_n] \end{pmatrix}\,. \tag{8.32}$$

Die RWA hat also genau dann nichttriviale Lösungen $y \neq 0$, falls gilt

$$D(\lambda) \quad := \quad \det\left(E(\lambda)\right) \quad = \quad 0\,. \tag{8.33}$$

Definition (8.34)

Zahlen $\lambda \in \mathbb{R}$ bzw. $\in \mathbb{C}$ mit $D(\lambda) = 0$ heißen *Eigenwerte* der Randwertaufgabe (8.30). Die zugehörigen nichttrivialen Lösungen (diese sind höchstens bis auf skalare Vielfache eindeutig) heißen die zugehörigen *Eigenfunktionen*.

Die Relation (8.33) ist im Allgemeinen ein nichtlineares Nullstellenproblem mit unendlich vielen Lösungen.

Bemerkung (8.35)

Zur numerischen Berechnung von Eigenwerten und Eigenfunktionen lässt sich das Eigenwertproblem in eine – allerdings nichtlineare – Randwertaufgabe in Normalform transformieren. Dazu setzt man $y_{n+1}(t) := \lambda$ und findet aus (8.30):

$$\begin{aligned}
y^{(n)}(t) \;&=\; -a_{n-1}(t, y_{n+1}(t))\, y^{(n-1)}(t) \;-\; \ldots \;-\; a_0(t, y_{n+1}(t))\, y(t) \\
y'_{n+1}(t) \;&=\; 0 \\
R_k[y, y_{n+1}] \;&=\; 0, \qquad k = 1, 2, \ldots, n \\
y'(a) \;&=\; 1 \qquad \text{(Normierung)}\,.
\end{aligned} \tag{8.36}$$

Wendet man hierauf ein numerisches Verfahren zur Lösung von RWA an, so wird man (abhängig von einer vorzugebenden Anfangsnäherung) natürlich nur jeweils *eine* spezielle Lösung erhalten.

Beispiel (8.37)

a) $\quad y'' + \lambda^2\, y \;=\; 0, \qquad y(0) \;=\; y(1) \;=\; 0\,.$

Die allgemeine Lösung der Differentialgleichung lautet $y(t) \;=\; C_1 \cos(\lambda\, t) + C_2 \sin(\lambda\, t)$. Die Randbedingungen ergeben:

$$\begin{aligned}
y(0) \;&=\; 0 \qquad \Rightarrow \qquad C_1 \;=\; 0 \\
y(1) \;&=\; 0 \qquad \Rightarrow \qquad C_2 \cdot \sin(\lambda) \;=\; 0\,.
\end{aligned}$$

Für die Eigenwerte ergibt sich also $\lambda_k = k\pi$, $k \in \mathbb{Z} \setminus \{0\}$ mit zugehörigen Eigenfunktionen $y_k(t) = \sin(\lambda_k t)$.

b) $\quad y'' + \lambda^2\, y \;=\; 0, \qquad y(0) \;=\; 0, \quad y(1) - y'(1) \;=\; 0.$

Wie oben ergibt sich

$$\begin{aligned}
y(0) \;&=\; 0 \qquad \Rightarrow \qquad C_1 \;=\; 0 \\
y(1) \;&=\; y'(1) \qquad \Rightarrow \qquad C_2(\sin\lambda - \lambda\cos\lambda) \;=\; 0\,.
\end{aligned}$$

Die Eigenwerte λ_k sind also die Lösungen der nichtlinearen Gleichung

$$\lambda \quad = \quad \tan \lambda \,,$$

die zugehörigen Eigenfunktionen sind

$$y_k(t) \quad = \quad \begin{cases} \sin(\lambda_k t) \quad , & \text{für} \quad \lambda_k \neq 0 \\ t \qquad\quad , & \text{für} \quad \lambda_k = 0 \,. \end{cases}$$

Beispiel (8.38)

Die Biegelinie (neutrale Faser) eines Balkens genügt bei kleiner Auslenkung $y(t)$ der Differentialgleichung

$$y''(t) \quad = \quad \frac{M(t,y)}{E \cdot I} \,,$$

$M(t,y)$: Biegemoment, E: Elastizitätsmodul, I: axiales Flächenträgheitsmoment.

a) *Kragbalken:* in $t = 0$ eingespannt, in $t = \ell$ frei, $M = P(\ell - t)$.

Man erhält das *Anfangswertproblem*:

$$y''(t) \quad = \quad \frac{P}{E\,I}\,(\ell - t), \qquad y(0) \;=\; y'(0) \;=\; 0\,.$$

b) *Gestützter Balken:* $M = -\dfrac{P}{2}\,\left(\frac{\ell}{2} - |\,t|\right)$.

Man erhält das *Randwertproblem*:

$$y''(t) \quad = \quad \frac{P}{2E\,I}\,\left(|\,t| - \frac{\ell}{2}\right) , \qquad y(-\ell/2) \;=\; y(\ell/2) \;=\; 0\,.$$

c) *Balkenknickung:*

Man erhält das *Eigenwertproblem*:

$$y'' \,+\, \frac{P}{E\,I}\,y \;=\; 0, \qquad y(0) \;=\; y(\ell) \;=\; 0\,.$$

Der kleinste positive Eigenwert ist

$$\lambda_1^2 \;=\; \frac{P_1}{E\,I} \;=\; \left(\frac{\pi}{\ell}\right)^2$$

$$\Rightarrow \quad P \;=\; P_1 \quad = \quad E\,I\,\frac{\pi^2}{\ell^2} \quad (\textit{Eulersche Knicklast})\,.$$

D. Das einfache Schießverfahren (single shooting).

Die Grundidee der so genannten Schießverfahren ist die iterative Verwendung von numerischen Integrationsverfahren für AWA. Die bei einer RWA fehlenden Anfangsdaten werden dabei geschätzt und dann so iterativ verbessert, so dass im Grenzwert die Randbedingungen erfüllt werden.

Wir beschreiben das Verfahren der Einfachheit halber zunächst für eine RWA zweiter Ordnung der Form

$$\begin{aligned} y'' &= f(t,y,y'), \qquad y(t) \in \mathbb{R} \\ y(a) &= y_a, \qquad y(b) = y_b\,. \end{aligned} \tag{8.39}$$

Die zugehörige AWA lautet dann mit geschätztem zweiten Anfangswert z:

$$\begin{aligned} y'' &= f(t,y,y') \\ y(a) &= y_a, \qquad y'(a) = z\,. \end{aligned} \tag{8.40}$$

Wir bezeichnen die Lösung dieser Anfangswertaufgabe mit $y(t;z)$ und nehmen an, dass diese Lösungen (zumindest für alle interessierenden z–Werte) auf dem gesamten Intervall $a \le t \le b$ existieren. Aufgabe ist es somit, eine Nullstelle der folgenden Funktion zu bestimmen

$$F(z) \quad := \quad y(b;z) - y_b\,. \tag{8.41}$$

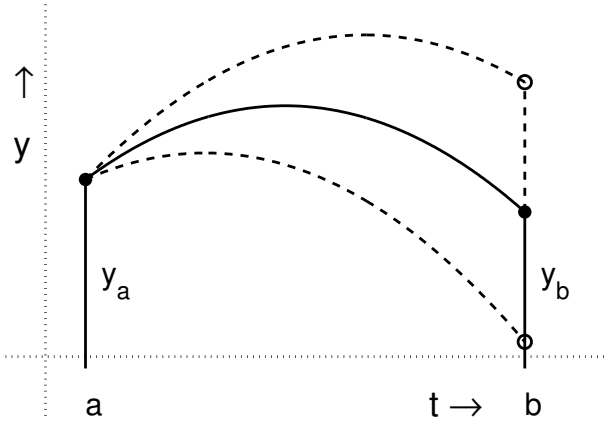

Abb. 8.1. Einfaches Schießverfahren

Beispiel (8.42)

Für die RWA

$$y'' = \frac{3}{2}\,y^2, \quad y(0) = 4, \quad y(1) = 1$$

ergibt sich der folgenden Graph für die Funktion F gemäß (8.41).

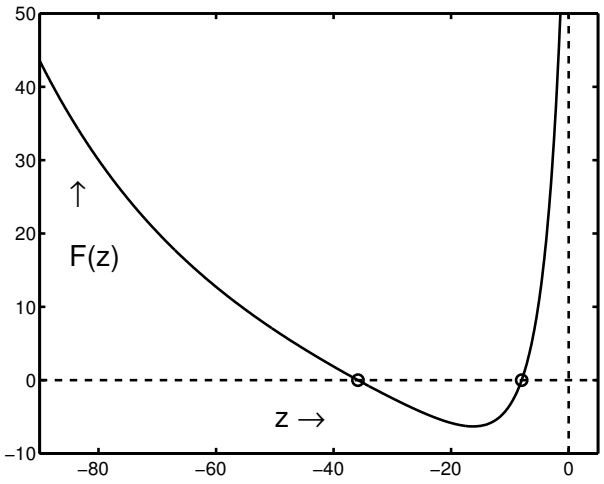

Abb. 8.2. $F(z)$ für das Beispiel (8.42)

$F(z)$ hat demnach zwei Nullstellen $z_1^* = -8$, und $z_2^* = -35.8585\ldots$, die RWA hat also auch genau zwei Lösungen!

Zur numerische Berechnung einer Nullstelle z^* von $F(z)$ lassen sich im Prinzip alle Verfahren zur numerischen Nullstellenberechnung verwenden.

Für das Verfahren der **Bisektion** geht man dabei von zwei Punkten $z_1 < z_2$ mit $F(z_1) \cdot F(z_2) < 0$ aus. Die Iteration lautet:

$$\begin{aligned} z &:= z_1 + 0.5 \cdot (z_2 - z_1)\,; \\ &\text{Falls} \quad F(z) \cdot F(z_2) < 0 \quad : \quad z_1 := z\,, \\ &\text{anderenfalls} \qquad\qquad\quad : \quad z_2 := z\,. \end{aligned} \tag{8.43}$$

Die Iteration wird abgebrochen, falls $|z_2 - z_1|$ hinreichend klein ist.

Das Bisektionsverfahren hat den Vorteil, nur Funktionsauswertungen (und keine Ableitungen) zu benötigen. Allerdings ist das Verfahren relativ langsam.

Das **Newton–Verfahren**

$$z_{k+1} = z_k - \frac{F(z_k)}{F'(z_k)}\,, \qquad k = 1, 2, \ldots \tag{8.44}$$

hat dagegen den Vorteil der schnellen (quadratischen) Konvergenz. Allerdings ist in jedem Iterationsschritt die Ableitung $F'(z)$ zu berechnen. Hierfür gibt es im Prinzip zwei Zugänge.

Eine naheliegende Methode ist es, $F'(x)$ durch *numerische Differentiation* zu approximieren. Man setzt also z.B.

$$\begin{aligned} F'(z) &\approx \frac{F(z + \Delta z) - F(z)}{\Delta z}\,, \\ |\Delta z| &\approx |z| \sqrt{\text{eps}} \quad \text{(Faustregel)}\,. \end{aligned} \tag{8.45}$$

Hierbei bezeichnet eps die relative Maschinengenauigkeit.

Eine andere Möglichkeit zur Berechnung von $F'(x)$ besteht in der *numerischen Integration der Variationsgleichung*; vgl. Satz (3.34).

Demnach ist $F'(z) = \dfrac{\partial}{\partial z}\, y(b;z)$. Setzt man also $w(t) := \dfrac{\partial}{\partial z}\, y(t;z)$, so lässt sich w als Lösung der folgenden AWA (Variationsgleichung) erhalten:

$$\begin{aligned} w''(t) &= f_y(t,y,y')\, w(t) + f_{y'}(t,y,y')\, w'(t)\,, \\ w(a) &= 0, \qquad w'(a) = 1\,. \end{aligned} \tag{8.46}$$

Die AWAen (8.40) und (8.46) lassen sich simultan (d.h. als ein DGLsystem) im Intervall $[a,b]$ lösen. Man hat dann:

$$F(z) = y(b;z) - y_b, \qquad F'(z) = w(b)\,. \tag{8.47}$$

Wir übertragen das obige Verfahren nun auf **allgemeine Zweipunkt–RWA** der Form (8.5)

$$\begin{aligned} y'(t) &= f(t,y(t)), \qquad y(t) \in \mathbb{R}^n \\ r\,(y(a),y(b)) &= 0 \in \mathbb{R}^n \end{aligned}$$

Die zugehörige AWA lautet:

$$\begin{aligned} y'(t) &= f(t,y(t)), \qquad a \le t \le b\,, \\ y(a) &= z \in \mathbb{R}^n, \end{aligned} \tag{8.48}$$

wobei der Vektor z wieder die geschätzten Anfangsdaten bezeichnet.

Die Lösung von (8.48) werde wiederum mit $y(t;z)$ bezeichnet. Sie sei im gesamten Intervall $a \le t \le b$ definiert.

Damit ist die RWA überführt worden in ein äquivalentes Nullstellenproblem für die Funktion

$$F(z) := r(z, y(b;z)); \quad F:\ \mathbb{R}^n \supset D \to \mathbb{R}^n \tag{8.49}$$

F ist eine „glatte“ Funktion, sofern die Funktionen $r(u,v)$ und $f(t,y)$ hinreichend oft stetig differenzierbar sind. D ist eine offene Menge und enthält die Anfangswerte $z^* = y^*(a)$ der Lösungen y^* der RWA (Existenz wird vorausgesetzt).

Zur numerischen Berechnung einer Nullstelle z^* lässt sich das gedämpfte (globalisierte) Newton–Verfahren verwenden. Der Iterationsschritt lautet folgendermaßen:

$$\begin{aligned} F'(z^k)\,\Delta z^k &= -F(z^k) \\ z^{k+1} &:= z^k + \lambda_k\, \Delta z^k\,, \qquad 0 < \lambda_k \le 1\,. \end{aligned} \tag{8.50}$$

Hierbei beschreibt $F'(z)$ die Jacobi–Matrix von F und λ den Dämpfungsparameter (Schrittweitendämpfung).

Für die Jacobi–Matrix $F'(z)$ erhält man aus (8.49)

$$\begin{aligned} F'(z) &= B_a + B_b \cdot Y(b), \\ B_a &:= \frac{\partial}{\partial u} r(u,v)\Big|_{(z,y(b;z))}, \quad B_b := \frac{\partial}{\partial v} r(u,v)\Big|_{(z,y(b;z))}, \\ Y(b) &:= \frac{\partial}{\partial z} y(b;z)\,. \end{aligned} \tag{8.51}$$

Bemerkung (8.52)

Die Matrizen B_a und B_b beschreiben die Ableitungen der Randbedingungen. Sie lassen sich i. Allg. ohne Schwierigkeiten analytisch berechnen.

Im Fall linearer Randbedingungen stimmen sie mit den in (8.4) vorgegebenen Matrizen überein, zudem bildet die Matrix $Y(t) := \frac{\partial}{\partial z} y(t;z)$ in diesem Fall ein Fundamentalsystem der homogenen DGL und die Jacobi–Matrix $F'(z)$ stimmt mit der in Satz (8.7) angegebenen Matrix E überein. Insbesondere gelten die in Satz (8.7) genannten Kriterien für die Regularität der Jacobi–Matrix.

Zur Berechnung der Matrix $Y(b)$ in (8.51) bieten sich wiederum zwei Wege an. Zum Einen lässt sich $Y(b)$ durch *numerische Differentiation* etwa mit Vorwärtsdifferenzen analog zu (8.45) approximieren:

$$\begin{aligned} Y(b) &= (y_{ij}) \\ y_{ij} &\approx \frac{y_i(b;\, z_1, \ldots, z_j + \Delta z_j, \ldots, z_n) - y_i(b;\, z_1, \ldots, z_n)}{\Delta z_j} \\ |\Delta z_j| &\approx |z_j| \sqrt{\text{eps}}. \end{aligned} \tag{8.53}$$

Zum anderen lässt sich $Y(b)$ durch numerische Integration der *Variationsgleichung* berechnen. Hierzu hat man die folgende Matrix–AWA zu lösen, vgl. (3.34)

$$\begin{aligned} Y'(t) &= f_y(t, y(t;z))\, Y(t), \qquad a \le t \le b \\ Y(a) &= I_n\,. \end{aligned} \tag{8.54}$$

Da die rechte Seite von (8.54) von der Lösung $y(t;z)$ der AWA abhängt, ist wiederum die simultane Integration des Matrix–DGLsytems (8.54) mit der Ausgangs-DGL erforderlich.

Wählt man zur Approximation der Jacobi–Matrix das Verfahren der numerischen Differentiation, so erhält man den folgenden (groben) Algorithmus für das einfache Schießverfahren.

Algorithmus (8.55) (Einfaches Schießverfahren)

(a) Wähle einen Startvektor $z^0 \in \mathbb{R}^n$; $\quad k := 0, \quad \lambda := 1.$

(b) Löse die AWA (8.48) zur Berechnung von $y(b; z^{(k)})$ und setze

$$\begin{aligned} F(z^k) &:= r(z^k, y(b; z^k)); \\ T^{(k)} &:= \left\| F(z^k) \right\|^2 . \end{aligned}$$

Falls $k = 0$: gehe nach (d).

(c) Falls $T^{(k)} > T^{(k-1)}$: $\quad \ell := 1, \quad \lambda := \lambda/2;$

$$z^k := z^{k-1} + \lambda \, \Delta \, z^{k-1};$$

gehe nach (b).

Falls $\ell = 0$: $\quad \lambda := \min(1, 2\lambda)$.

Falls $T^{(k)}$ hinreichend klein: *Abbruch!*

(d) *Newton–Korrektur:*
Für $j = 1, 2, \ldots, n$:

$$\Delta z_j := \sqrt{\text{eps}} \ \max\left(1, \left| z_j^k \right| \right);$$

Berechnung von $y\left(b; z_1^k, \ldots, z_j^k + \Delta z_j, \ldots, z_n^k\right)$ nach (8.48) .

Auswertung der Jacobi–Matrix $F'(z^k)$ nach (8.53);
Lösung des linearen Gleichungssystems $F'(z^k) \, \Delta z^k = -F(z^k)$.
Falls $\left\| \Delta z^k \right\|$ hinreichend klein: *Abbruch*;
Setze $k := k + 1; \quad \ell := 0; \quad z^k := z^{k-1} + \lambda \, \Delta z^k$.
Gehe nach (b).

Das einfache Schießverfahren besitzt **zwei wesentliche Nachteile**, die seine Anwendbarkeit bei komplizierteren Randwertaufgaben häufig stark einschränken. Beide Schwierigkeiten hängen damit zusammen, dass zur Auswertung von $F(z)$ und $F'(z)$ AWA über den gesamten Zeitbereich $a \le t \le b$ gelöst werden müssen.

1. Bewegliche Singularitäten Nichtlineare DGLen können so genannte *bewegliche Singularitäten* besitzen, d.h., die Lösung $y(t; z)$ einer AWA existiert i. Allg. nur in einer von den Anfangswerten z abhängigen (maximalen) Umgebung $]t_{min}, t_{max}[$ der Anfangszeit $t = a$. In $t = t_{\min}$ bzw. $t = t_{\max}$ besitzt die Lösung eine Singlarität (die Grenzwerte $\lim_{t \downarrow t_{\min}} y(t)$ bzw. $\lim_{t \uparrow t_{\max}} y(t)$ existieren nicht oder liegen am Rand des Definitionsbereichs von f) . Liegt die „Singularität“ t_{max} nun im inneren Intervall $]a, b[$, so ist die Auswertung von $F(z) = r(z, y(b; z))$ nicht möglich und das Verfahren bricht ab.

Beispiel (8.56) (nach Troesch [2])

Gegeben sei folgende RWA

$$\begin{aligned} y'' &= \lambda \sinh(\lambda y) \\ y(0) &= 0, \qquad y(1) = 1\,. \end{aligned}$$

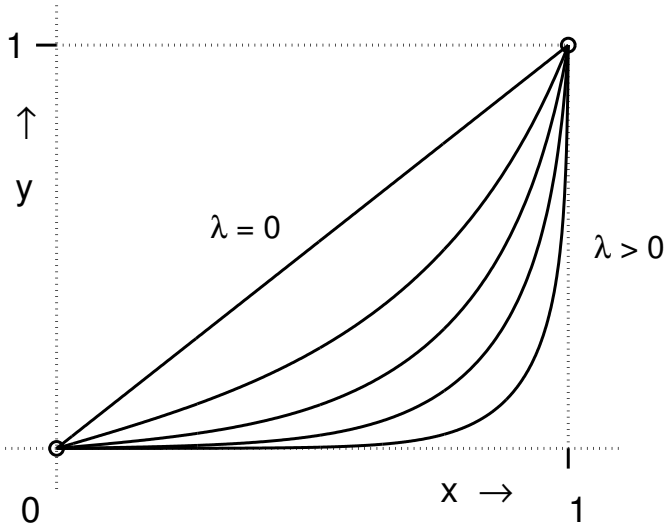

Abb. 8.3. Beispiel von Troesch

Für $\lambda = 5$ besitzt die zugehörige AWA

$$y'' = \lambda \sinh(\lambda y), \quad y(0) = 0, \quad y'(0) = z$$

nur für $|z| \leq 0.05$ eine Lösung, die im ganzen Intervall $[0,1]$ existiert; vgl. auch Stoer, Bulirsch, Abschnitt 7.3.
Für die tatsächliche Lösung der Randwertaufgabe erhält man: $z^* = 0.0457504\ldots$.

2. Potentielle Instabilität Der zweite Nachteil des einfachen Schießverfahrens hängt mit der möglicherweise empfindlichen Abhängigkeit der Lösung $y(t;z)$ von den Anfangsdaten z zusammen. Die entsprechende Fehlerabschätzung lautet, vgl. Satz (3.28):

$$\|y(b;z) - y(b,\tilde{z})\| \quad \leq \quad \mathrm{e}^{L|b-a|}\ \|z - \tilde{z}\|\ . \tag{8.57}$$

Hierbei bezeichnet L eine Lipschitz–Konstante von $f(t,y)$ bezüglich y. Vernachlässigt man die Abhängigkeit der Lipschitz–Konstanten von dem jeweils betrachteten (t,y)–Bereich, so hängt die Empfindlichkeit der Anfangswertaufgabe also im wesentlichen von dem Produkt aus Lipschitz-Konstanter und Integrationslänge $(b-a)$ ab.

[2] B.A. Troesch: Intrinsic difficulties in the numerical solution of a boundary value problem. Internal Report NN–142, TRW, Inc. Redondo Beach, California, 1960.

Beispiel (8.58)

Gegeben sei die RWA

$$y'' \;=\; 12y \,+\, y', \qquad y(0) \;=\; y(10) \;=\; 1\,.$$

Die Lösung der zugehörigen AWA mit $y(0) = z_1$ und $y'(0) = z_2$ lautet:

$$y(t; z_1, z_2) \quad = \quad \frac{4z_1 - z_2}{7}\,\mathrm{e}^{-3t} \,+\, \frac{3z_1 + z_2}{7}\,\mathrm{e}^{4t}\,.$$

Setzt man hierin die Randbedingungen ein, so erhält man die eindeutig bestimmten Anfangsdaten

$$\begin{aligned} z_1^* \;&=\; y(0) \;=\; 1 \\ z_2^* \;&=\; y'(0) \;=\; -3 \,+\, 2.9\ldots 10^{-17}\,. \end{aligned}$$

Bei numerischer Lösung kann man hierfür jedoch nur Approximationen im Rahmen der Maschinengenauigkeit erwarten, also

$$\tilde{z}_1 \;=\; 1, \quad \tilde{z}_2 \;=\; -3 + \varepsilon, \quad |\varepsilon| \;\leq\; \mathrm{eps}$$

erwarten. Andererseits hängen die Lösungen $y(t; \tilde{z}_1, \tilde{z}_2)$ der AWA sehr empfindlich von der Approximation für z_2^* ab.

So findet man beispielsweise:

$$\begin{aligned} y(10; 1, -3) = \mathrm{e}^{-30} \quad &\approx \quad 9.36E - 14\,, \\ y(10; 1, -3 + 10^{-10}) \;&\approx\; \frac{1}{7}\,\mathrm{e}^{30} \quad \approx \quad 1.53E + 12\,. \end{aligned}$$

Der gesuchte Anfangswert z_2^* lässt sich also numerisch gar nicht so genau berechnen, dass der Lösungsverlauf auch nur in etwa qualitativ richtig ist (für Maschinengenauigkeiten in der Größenordung 10^{-10}).

Schätzt man für dieses Beispiel die in (8.57) auftretenden Größen ab, so findet man:

$$\begin{aligned} L \;&\approx\; 12, \qquad b - a \;=\; 10, \\ \mathrm{e}^{L(b-a)} \;&\approx\; 10^{52}, \\ \|z - \tilde{z}\| \;&\approx\; 10^{-10} \quad \text{(bei 10stelliger Rechnung)}\,. \end{aligned}$$

E. Lineare Randwertaufgaben, Superposition.

In diesem Abschnitt beschäftigen wir uns mit den Vereinfachungen, die sich bei Anwendung des einfachen Schießverfahrens auf *lineare* RWAen ergeben. Es sei ausdrücklich angemerkt, dass der zweite Nachteil des Schießverfahren, nämlich das instabile Verhalten der AWAen, auch bei linearen RWAen auftritt und man daher die folgenden Verfahren nur auf gutmütige RWAen (solche mit gut konditionierten AWAen) anwenden sollte.

Wir betrachten zunächst wieder den Fall einer skalaren, linearen DGL zweiter Ordnung

$$\begin{aligned} L[y] \;&:=\; y'' \,-\, p(t)\,y' \,-\, q(t)\,y \;=\; r(t), \\ y(a) \;&=\; y_a, \quad y(b) \;=\; y_b, \end{aligned} \tag{8.59}$$

mit stetigen auf $[a,b]$ erklärten Funktionen p, q und r.

Wir verwenden das einfache Schießverfahren und berechnen die Lösungen $u(t) := y(t;0)$ und $v(t) := y(t;1)$ der folgenden beiden AWAen

$$L[y] \;=\; r(t); \quad y(a) \;=\; y_a, \quad y'(a) \;=\; z \tag{8.60}$$

für $z = 0$ uns $z = 1$. Durch Einsetzen überzeugt man sich, dass dann durch

$$y(t;z) \;:=\; u(t) \;+\; z\,(v(t) \,-u(t)) \tag{8.61}$$

die Lösung der AWA (8.60) für allgemeines z gegeben ist.

Die Lösung der RWA erhält man nun durch Einsetzen von (8.61) in die Randbedingung $y(b) = y_b$. Es ergibt sich die eindeutige Lösung

$$y(t) \;=\; \left(\frac{v(b)-y_b}{v(b)-u(b)}\right)\,u(t) \;+\; \left(\frac{y_b-u(b)}{v(b)-u(b)}\right)\,v(t)\,, \tag{8.62}$$

wobei wir voraussetzen, dass der Nenner nicht verschwindet.

Dass diese Annahme unter einer Zusatzvoraussetzung tatsächlich erfüllt ist, zeigt der folgende Satz, der damit zugleich die Existenz und Eindeutigkeit einer Lösung von (8.59) belegt.

Satz (8.63) (Existenz)

Sind $p, q, r : [a,b] \to \mathbb{R}$ stetig und ist $q(t) \;>\; 0$ für alle $t \in [a,b]$, so gilt für die Lösungen von (8.60): $y(b;1) \neq y(b;0)$.

Beweis:

Sei wie oben $u(t) := y(t;0)$, $v(t) := y(t;1)$ und $w(t) := v(t) - u(t)$. Dann gelten

$$L[w] \;=\; 0, \quad w(a) \;=\; 0, \quad w'(a) \;=\; 1, \quad w(b) = y(b;1) - y(b,0).$$

Wir zeigen, dass $w(t) > 0$ für alle $t \in]a,b]$. Wegen $w(a) = 0$ und $w'(a) > 0$ ist dies in einem Intervall $]a, a+\varepsilon[$, $\varepsilon > 0$ hinreichend klein, richtig.

Wenn w in $]a,b]$ eine Nullstelle besäße, so gäbe es daher auch eine kleinste Nullstelle τ_0 von w, wobei $\tau_0 \in [a+\varepsilon, b]$.

Mit $P(t) \;:=\; -\int\limits_a^t p(\tau)d\tau$ und $L[w] = 0$ folgt dann für $t \in]a, \tau_0[$

$$\frac{d}{dt}\,\left[\mathrm{e}^{P(t)}w'(t)\right] \;=\; \mathrm{e}^{P(t)}\,[w'' \,-\, p(t)w'(t)] \;=\; \mathrm{e}^{P(t)}\,q(t)\,w(t) \;>\; 0.$$

Damit ist die Funktion $e^{P(t)}w'(t)$ also in $[a, \tau_0]$ streng monoton wachsend, genauer:

$$\forall\, t \in [a, \tau_0]: \quad e^{P(t)}w'(t) \;\geq\; e^{P(a)}w'(a) \;=\; 1.$$

Anderseits ist $w(a) = w(\tau_0) = 0$. w' besitzt also nach dem Satz von Rolle eine Nullstelle in $]a, \tau_0[$, im Widerspruch zu der obigen Abschätzung. □

Im Fall einer linearen RWA für ein System erster Ordnung

$$\begin{aligned} y'(t) \quad &= \quad A(t)\, y(t) \;+\; h(t) \\ B_a\, y(a) \;+\; B_b\, y(b) \quad &= \quad d \\ \text{mit} \quad A(t), B_a, B_b \in \mathbb{R}^{(n,n)}, &\quad h(t), d \in \mathbb{R}^n\,, \end{aligned} \tag{8.64}$$

mit stetigen Funktionen A und h kann man wie folgt vorgehen. Man löst die AWAen

$$\begin{aligned} Y'(t) \;&=\; A(t)\, Y(t), \quad Y(a) \;=\; I_n \\ w'(t) \;&=\; A(t)\, w(t) + h(t), \quad w(a) \;=\; 0. \end{aligned}$$

für $Y(t) \in \mathbb{R}^{(n,n)}$ (Hauptfundamentalsystem) und $w(t) \in \mathbb{R}^n$ (partikuläre Lösung).

Die allgemeine Lösung der inhomogenen DGL lautet dann

$$y(t; z) \;=\; Y(t)\, z \;+\; w(t); \quad z \in \mathbb{R}^n.$$

Setzt man diese nun in die Randbedingungen ein, so erhält man ein lineares Gleichungssystem zur Bestimmung von z:

$$E\, z \;=\; d \;- B_b\, w(b), \qquad \text{mit} \quad E \;:=\; B_a \;+\; B_b\, Y(b). \tag{8.65}$$

Besitzt die Randwertaufgabe (für alle Inhomogenitäten) eine eindeutige Lösung, so ist die Matrix E nach Satz (8.7) regulär, und damit das lineare Gleichungssystem (8.65) eindeutig lösbar.

F. Die Mehrzielmethode.

Die Mehrzielmethode ist eine Modifikation des einfachen Schießverfahrens, welche die numerische Integration über „große“ Bereiche vermeidet und damit (in gewissem Umfang) eine Abhilfe für die genannten Probleme des einfachen Schießverfahrens bietet. Das Verfahren geht auf Keller [3], Osborne [4] und Bulirsch [5] zurück.

Die Grundidee des Verfahrens ist es, mit einer festen Intervallunterteilung zu arbeiten und AWAen jeweils nur über diesen Teilintervallen zu lösen. Die Anfangswerte (für alle Teilintervalle) sind dann so zu bestimmen, dass im Grenzwert nicht nur die Randbedingungen

[3] H.B. Keller: Numerical Methods for Two-Point Boundary–Value Problems. Blaisdell, London 1968.

[4] M.R. Osborne: On shooting methods for boundary value problems. J. Math. Anal. Appl. **27** (1969)

[5] R. Bulirsch: Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. Report der Carl–Cranz–Gesellschaft, 1971.

erfüllt werden, sondern auch die aus den Teiltrajektorien zusammengesetzte Lösung glatt ist.

Wir betrachten wieder eine allgemeine Zweipunkt-RWA der Form (8.5). Ferner wird eine *Intervallunterteilung* vorgegeben:

$$a \;=\; t_1 \;<\; t_2 \;<\; \ldots \;<\; t_m \;=\; b\,. \tag{8.66}$$

Die t_j heißen *Mehrzielknoten*.

Für jedes Teilintervall werden Anfangswerte

$$z_j \;\approx\; y(t_j), \qquad j = 1, \ldots, m-1$$

geschätzt und die jeweiligen AWAen auf den Teilintervallen

$$\begin{aligned} y' \;&=\; f(t,y), \qquad t_j \;\leq\; t \;\leq\; t_{j+1} \\ y(t_j) \;&=\; z_j \end{aligned} \tag{8.67}$$

(numerisch) gelöst. Die Lösungen werden wie üblich mit $y(t; t_j, z_j)$ bezeichnet.

Die zusammengesetzte Trajektorie

$$y(t; z_1, \ldots, z_{m-1}) \quad := \quad \begin{cases} y(t; t_1, z_1) & : \; t_1 \;\leq\; t \;<\; t_2 \\ y(t; t_2, z_2) & : \; t_2 \;\leq\; t \;<\; t_3 \\ \quad\vdots & \quad\vdots \\ y(t; t_{m-1}, z_{m-1}) & : \; t_{m-1} \;\leq\; t \;\leq\; t_m \end{cases} \tag{8.68}$$

löst nun die Randwertaufgabe genau dann, falls die folgenden Bedingungen erfüllt sind:

$$\begin{aligned} F_j(z_j, z_{j+1}) \quad &:= \quad y(t_{j+1}; t_j, z_j) - z_{j+1} \;=\; 0\,, \qquad j = 1, 2, \ldots, m-2 \\ F_{m-1}(z_1, z_{m-1}) \quad &:= \quad r(z_1, y(t_m; t_{m-1}, z_{m-1})) \;=\; 0\,. \end{aligned} \tag{8.69}$$

Die ersten Bedingungen sind die *Sprungbedingungen*, die letzte Bedingung beschreibt die vorgegebenen Randbedingungen. Damit ist die Randwertaufgabe wieder in ein äquivalentes Nullstellenproblem für die zusammengesetzte Funktion

$$F \quad = \quad (F_1, \ldots, F_{m-1})^{\mathrm{T}} \;:\; \mathbb{R}^{(m-1)\,n} \;\supset\; D \;\to\; \mathbb{R}^{(m-1)\,n}$$

überführt worden.

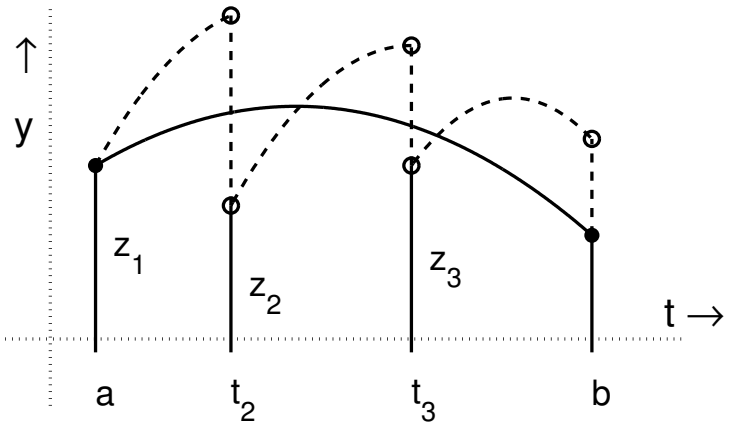

Abb. 8.4. Mehrzielmethode

Zur Nullstellenbestimmung wird das globalisierte (gedämpfte) Newton–Verfahrens verwendet. Für die hierzu benötigte Jacobi–Matrix (auch *Mehrzielmatrix* genannt) $F'(z)$, $z = (z_1, \ldots, z_{m-1})^{\mathrm{T}}$ erhält man die folgende Blockstruktur:

$$F'(z) \quad = \quad \begin{pmatrix} Y_1 & -I_n & & & 0 \\ & Y_2 & -I_n & & \\ & & \ddots & \ddots & \\ 0 & & & Y_{m-2} & -I_n \\ B_a & 0 & \cdots & 0 & B_b Y_{m-1} \end{pmatrix} \tag{8.70}$$

wobei

$$\begin{aligned} Y_j &:= \frac{\partial}{\partial\, z_j}\, y(t_{j+1}; t_j, z_j), \qquad j = 1, \ldots, m-1 \\ B_a &:= \frac{\partial}{\partial\, z_1}\, r(z_1, y(t_m; t_{m-1}, z_{m-1})) \\ B_b &:= \frac{\partial}{\partial\, z_m}\, r(z_1, y(t_m; t_{m-1}, z_{m-1})) \quad . \end{aligned} \tag{8.71}$$

Die Berechnung der Variations- oder auch Propagationsmatrizen Y_j, $j = 1, \ldots, m-1$, kann wie beim einfachen Schießverfahren mittels numerischer Differentiation erfolgen. Der folgende Satz belegt die Regularität dieser Matrizen.

Satz (8.72)

Ist f eine C^1–Funktion und sind die $z_j \in \mathbb{R}^n$ so gewählt, dass die AWA (8.67) eine Lösung $y(t; t_j, z_j)$ besitzt, die im gesamten Intervall $t_j \le t \le t_{j+1}$ existiert, so ist $Y_j \in \mathbb{R}^{(n,n)}$ regulär.

Beweis:

$y(\cdot; t_j, z)$ ist zugleich Lösung der AWA $y' = f(t,y)$, $y(t_{j+1}) = y(t_{j+1}; t_j, z)$, also: $z = y(t_j; t_{j+1}, y(t_{j+1}; t_j, z))$. Dies gilt für alle z in einer Umgebung von z_j.

Die Differentiation dieser Identität nach z und Einsetzen von $z = z_j$ liefert:

$$I_n \quad = \quad \left[\frac{\partial}{\partial z_{j+1}} \; y\,(t_j; t_{j+1}, z_{j+1}) \right] \cdot Y_j \; .$$

Hieraus folgt die Regularität von Y_j. □

Zur Anwendung des Newton–Verfahrens wird nun aber die Regularität der Jacobi–Matrix benötigt. Hierzu lässt sich der folgende Zusammenhang aufzeigen, der im Wesentlichen besagt, dass aus der Regularität der Jacobi-Matrix für das einfache Schießverfahren auf die Regularität der Mehrzielmatrix geschlossen werden kann.

Satz (8.73)

Die RWA (8.5) habe eine Lösung y^*. Ferner sei $z_j^* := y^*(t_j)$ und $z^* = (z_1^*, \ldots, z_{m-1}^*)^{\mathrm{T}}$.

Die Abbildung $\Psi(z_1) := r(z_1, y(t_m; t_1, z_1))$ ist dann in einer Umgebung von z_1^* wohldefiniert. Die Abbildung sei regulär, d.h., sie besitze eine reguläre Jacobi–Matrix:

$$E^* \quad := \quad \Psi'(z_1^*) \;\in\; \mathbb{R}^{(n,n)} \; .$$

Dann ist auch die Mehrzielmatrix $F'(z^*)$ regulär.

Beweis

Wir betrachten das lineare Gleichungssystem (Newton-Gleichung) $F'(z)\,\Delta z = -F(z)$ und schreiben die Variablen in der Reihenfolge:

$$(\Delta z_2,\ \Delta z_3,\ \ldots,\ \Delta z_{m-1},\ \Delta z_1).$$

Man erhält dann die folgende Blockstruktur des Gleichungssytems

$$\begin{pmatrix} -I & & & 0 & Y_1 \\ Y_2 & -I & & & \\ & Y_3 & -I & & \\ & & \ddots & \ddots & \\ & & Y_{m-2} & -I & \mathbf{0} \\ \mathbf{0} & & & (B_b\, Y_{m-1}) & B_a \end{pmatrix} \begin{pmatrix} \Delta z_2 \\ \Delta z_3 \\ \vdots \\ \Delta z_{m-1} \\ \Delta z_1 \end{pmatrix} \;=\; - \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ \vdots \\ F_{m-1} \end{pmatrix} .$$

Hierauf wenden wir nun Block–Gauß–Elimination an.

Die Pivotelemente sind dabei gerade die $(-I)$–Matrizen. Es folgt:

$$\begin{pmatrix} (-I) & & \mathbf{0} & & Y_1 \\ & (-I) & & & [Y_2\, Y_1] \\ & & \ddots & & \vdots \\ & & & (-I) & [Y_{m-2}\, \ldots\, Y_1] \\ \mathbf{0} & & & & \{B_a + B_b\, Y_{m-1}\, \ldots\, Y_1\} \end{pmatrix} \begin{pmatrix} \Delta\, z_2 \\ \Delta\, z_3 \\ \vdots \\ \Delta\, z_{m-1} \\ \Delta\, z_1 \end{pmatrix} = - \begin{pmatrix} \widetilde{F_1} \\ \widetilde{F_2} \\ \vdots \\ \vdots \\ \widetilde{F}_{m-1} \end{pmatrix}$$

mit

$$\begin{aligned} \widetilde{F}_1 &:= F_1 \\ \widetilde{F}_2 &:= F_2 + Y_2\, F_1 \\ \widetilde{F}_3 &:= F_3 + Y_3\, F_2 + Y_3\, Y_2\, F_1 \\ &\vdots \\ \widetilde{F}_{m-1} &:= F_{m-1} + Y_{m-1}\, F_{m-2} + \ldots + Y_{m-1}\, \ldots\, Y_2\, F_1\,. \end{aligned}$$

Da die Gauß–Elimination den (absoluten) Wert der Determinante erhält, folgt hiermit:

$$\det\,(JF(z)) \quad = \quad \pm\, \det\,(B_a + B_b\, Y_{m-1} \cdot Y_{m-2} \cdot \ldots \cdot Y_1) \;=\; \pm\, \det\; E.$$

Setzt man hierin $z_j = z_j^*$ ein, so folgt schließlich:

$$\begin{aligned} E &= \frac{\partial\, r}{\partial\, z_1} + \frac{\partial\, r}{\partial\, z_m} \cdot \left(\frac{\partial\, y^*(t_m)}{\partial\, z_{m-1}}\right) \cdot \left(\frac{\partial\, y^*(t_{m-1})}{\partial\, z_{m-2}}\right) \cdot \ldots \cdot \left(\frac{\partial\, y^*(t_2)}{\partial\, z_1}\right) \\ &= \frac{\partial}{\partial\, z_1}\, [r(z_1, y(t_m; t_1, z_1))]\Big|_{z_1 = z_1^*} \\ &= E^* \quad \text{regulär!} \end{aligned}$$

□

Der Beweis zu Satz (8.73) liefert zugleich eine schnelle und sparsame Methode zur Berechnung der Newton–Korrekturen $\Delta\, z_j$:

$$\begin{aligned} &\text{a)} \quad \text{Löse} \qquad E \cdot \Delta\, z_1 \;=\; \widetilde{F}_{m-1} \\ &\text{b)} \quad \text{Berechne rekursiv:} \\ &\qquad\qquad \Delta\, z_{j+1} \;=\; Y_j\, \Delta\, z_j + F_j \quad (j = 1, \ldots, m-2)\,. \end{aligned} \tag{8.74}$$

Der obige Algorithmus (manchmal auch „condensing algorithm“ genannt) ist jedoch nur dann numerisch brauchbar, falls das resultierende kleine lineare Gleichungssystem (8.72a) nicht zu schlecht konditioniert ist. Andernfalls ist der Algorithmus instabil, und es empfiehlt sich, statt dessen die direkte Auflösung des „großen“ linearen Gleichungssystems $F'(z) = -F(z)$ – etwa mittels QR–Zerlegung – vorzunehmen.

Aus dem Beweis zu (8.73) lässt sich auch umgekehrt folgern: Ist die Mehrzielmatrix $F'(z)$ in Lösungspunkt z^* regulär, so ist auch die Matrix $E^* = \Psi'\,(z_1^*)$ regulär.

Nach dem Umkehrsatz existieren daher Umgebungen

$$\begin{aligned} U_y &\subset \mathbb{R}^n \quad \text{von} \quad z_1^* \\ U_0 &\subset \mathbb{R}^n \quad \text{von} \quad 0\,, \end{aligned}$$

so dass $\Psi|_{U_y} : U_y \to U_0$ sogar bijektiv ist. Dies bedeutet, dass die RWA eine „isolierte“ Lösung besitzt und sogar bei Störung der Randdaten immer noch (lokal) eindeutig lösbar ist.

Dies ist die wesentliche Aussage des folgenden Einbettungssatzes, der somit unmittelbar aus dem Beweis des Satzes (8.73) folgt.

Satz (8.75) (Einbettungssatz)

Die RWA besitze eine Lösung y^*. Zu einer vorgegebenen Intervallunterteilung (8.66) und $z_j^* := y^*(t_j),\ j = 1, \ldots, m-1$ sei die Mehrzielmatrix $F'(z^*)$ regulär.

Dann existieren Umgebungen $U_y \subset \mathbb{R}^n$ von $\mathbf{y}^*(a)$ und U_0 von $0 \in \mathbb{R}^n$, so dass für jeden Vektor $\varepsilon \in U_0$ die „gestörte“ RWA

$$y' \;=\; f(t,y); \qquad r(y(a), y(b)) \;=\; \varepsilon$$

genau eine Lösung $y^*(\cdot;\varepsilon)$ mit $y^*(a,\varepsilon) \in U_y$ besitzt.

Insbesondere ist die Lösung der ursprünglichen RWA lokal eindeutig.

Beispiel (8.76)

Wir betrachten nochmals die lineare RWA aus Beispiel (8.58)

$$\begin{aligned} y'' &= 12y + y' \\ y(0) &= y(10) = 1\,. \end{aligned}$$

Akzeptiert man eine lokale Fehlerverstärkung von $\mathrm{e}^{L(t_{j+1}-t_j)} \approx 10^5$, so genügt es offenbar, die Intervallänge auf $|t_{j+1} - t_j| = 1$ zu reduzieren.

In der Tat liefert die Mehrzielmethode für dieses Beispiel mit $m = 11$ äquidistanten Mehrzielknoten, die in der folgenden Tabelle angegebenen Approximationen für $y(t_j)$ und $y'(t_j),\ j = 1, \ldots, 11,$ die auf etwa fünf Dezimalstellen mit der exakten Lösung übereinstimmen.

T(J)	Y(1,J)	Y(2,J)
0	0.1000 0000E+01	-0.3000 0000E+01
1	0.4978 7068E-01	-0.1493 6121E+00
2	0.2478 7522E-02	-0.7436 2565E-02
3	0.1234 0980E-03	-0.3702 2941E-03
4	0.6144 2501E-05	-0.1843 2486E-04

```
 5       0.3079 6345E-06     -0.9094 6245E-06
 6       0.1277 6450E-06      0.4044 4808E-06
 7       0.6144 9072E-05      0.2457 4329E-04
 8       0.3354 6162E-03      0.1341 8460E-02
 9       0.1831 5503E-01      0.7326 2024E-01
10       0.9999 9267E+00      0.3999 9707E+01
```

Es sei abschließend bemerkt, dass für gewisse Probleme der optimalen Steuerung RWAen auftreten, die nur eine stückweise stetige rechte Seite besitzen und bei denen gewisse Randbedingungen gerade diese (a priori unbekannten) Unstetigkeitsstellen (*Schaltpunkte*) festlegen.

Auch solche *Randwertaufgaben mit Schaltbedingungen* lassen sich mit einer Variante der Mehrzielmethode (Routine BNDSCO) für Mehrpunkt–Randwertaufgaben effizient numerisch lösen.