

Satz 4.20 (Stabilitätssatz für Differenzgleichungen) Sei $[t_0, T] \subset J \subset \mathbb{R}$ und $f \in C(J, \mathbb{R})$ lipschitzstetig im zweiten Argument mit Lipschitzkonstante $L > 0$. Sei ein nullstabiles LMM gegeben. Seien drei Folgen $(y_i)_{i=0}^N, (\tilde{y}_i)_{i=0}^N, (\varepsilon_i)_{i=0}^N$ gegeben mit

$$\begin{aligned} \tilde{y}_i &= y_i + \varepsilon_i, & i = 0, \dots, k-1, \\ \sum_{j=0}^k \alpha_{k-j} \tilde{y}_{i-j} &= h \sum_{j=0}^k \beta_{k-j} f(t_{i-j}, \tilde{y}_{i-j}) + \varepsilon_i, & i = k, \dots, N, \\ \sum_{j=0}^k \alpha_{k-j} y_{i-j} &= h \sum_{j=0}^k \beta_{k-j} f(t_{i-j}, y_{i-j}), & i = k, \dots, N. \end{aligned}$$

Dann gilt unter der Voraussetzung $h < 1/(L|\beta_k|)$ (falls $\beta_k \neq 0$), daß

$$|\tilde{y}_i - y_i| \leq C e^{\gamma(t_i - t_0)} \left\{ \max_{0 \leq j \leq k-1} |\varepsilon_j| + \sum_{j=k}^i |\varepsilon_j| \right\}, \quad (4.23)$$

wobei die Konstanten $C, \gamma > 0$ nur von der Lipschitzkonstanten L und den Koeffizienten α_j, β_j des LMM abhängen.

Beweis: Die wesentliche Beweisidee ist, das Vorgehen beim Einschrittverfahren zu wiederholen, indem man eine Rekurrenzrelation für die Differenz

$$e_i := \tilde{y}_i - y_i$$

hergeitet und dann mit dem diskreten Gronwall-Lemma (Lemma 2.11) Abschätzungen für die $|e_i|$ erhalten. Technisch besteht der ‘‘Trick’’ darin, das LMM in ein Vektor-Einschrittverfahren (vgl. (4.25)) umzuformulieren. ⁷

Offensichtlich ist $e_i = \varepsilon_i$ für $i = 0, \dots, k-1$. Für $i \geq k$ schreiben wir wegen $\alpha_k = 1$

$$e_i = - \sum_{j=1}^k \alpha_{k-j} e_{i-j} + h \beta_k [f(t_i, \tilde{y}_i) - f(t_i, y_i)] + h \underbrace{\sum_{j=1}^k \beta_{k-j} [f(t_{i-j}, \tilde{y}_{i-j}) - f(t_{i-j}, y_{i-j})]}_{=: b_i} + \varepsilon_i.$$

Weiter definieren wir eine Hilfsgröße σ_i durch

$$\sigma_i := \begin{cases} \frac{f(t_i, \tilde{y}_i) - f(t_i, y_i)}{e_i} & \text{falls } e_i \neq 0, \\ 0 & \text{falls } e_i = 0. \end{cases}$$

Damit gilt dann

$$|\sigma_i| \leq L \quad (4.24)$$

⁷ Es ist sinnvoll, sich den Beweis beim Einschrittverfahren (Satz 2.10) zu vergegenwärtigen: Dort war der entscheidende Schritt, daß für die Fehler e_i eine Rekursion der Form $|e_{i+1}| \leq (1 + O(h))|e_i| + \text{‘‘klein’’}$ erzeugt werden konnte, um dann das Gronwall-Lemma anwenden zu können. Der Beweis beim Mehrschrittverfahren geht genauso vor: Die Nullstabilität des Verfahrens liefert, daß man eine Norm $\|\cdot\|$ finden kann, so daß $\|E_{i+1}\| \leq (1 + O(h))\|E_i\| + \text{‘‘klein’’}$ —vgl. (4.27)

sowie

$$(1 - h\beta_k\sigma_i)e_i = -\sum_{j=1}^k \alpha_{k-j}e_{i-j} + b_i.$$

Diese Beziehung zusammen mit den trivialen Beziehungen

$$(1 - h\beta_k\sigma_{i-j})e_{i-j} = (1 - h\beta_k\sigma_{i-j})e_{i-j}, \quad j = 1, \dots, k-1,$$

liefert dann die Rekurrenzrelation

$$D_i E_i = C_i A E_{i-1} + B_i, \quad i \geq k,$$

wobei wir

$$E_i = \begin{pmatrix} e_{i-k+1} \\ e_{i-k+2} \\ \vdots \\ e_i \end{pmatrix} \in \mathbb{R}^k, \quad B_i = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ b_i \end{pmatrix} \in \mathbb{R}^k$$

sowie

$$D_i = \begin{pmatrix} 1 - h\beta_k\sigma_{i-k+1} & & 0 \\ & \ddots & \\ 0 & & 1 - h\beta_k\sigma_i \end{pmatrix} \in \mathbb{R}^{k \times k}, \quad A = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & & \\ & & 0 & 1 \\ -\alpha_0 & -\alpha_1 & \cdots & -\alpha_{k-1} \end{pmatrix} \in \mathbb{R}^{k \times k},$$

$$C_i = \begin{pmatrix} 1 - h\beta_k\sigma_{i-k+1} & & 0 \\ & \ddots & \\ 0 & & 1 - h\beta_k\sigma_{i-1} & 0 \\ & & 0 & 1 \end{pmatrix} \in \mathbb{R}^{k \times k}$$

geschrieben haben. Aufgrund unserer Voraussetzung $h < 1/(L|\beta_k|)$ sind die Diagonalmatrizen D_i invertierbar, und wir erhalten die gewünschte Rekursion

$$E_i = D_i^{-1} (C_i A E_{i-1} + B_i), \quad i \geq k. \quad (4.25)$$

Wir wollen nun diese Rekursion mit Hilfe des diskreten Gronwall-Lemmas abschätzen. Dazu müssen wir geeignete Normen einführen. Durch Entwickeln nach der letzten Zeile sehen wir, daß die Matrix A das charakteristische Polynom

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

hat (vgl. z.B. [9, Abschnitt 6.3] zum Thema ‘‘Frobeniussche Normalform’’ einer Matrix). Aus der geforderten Nullstabilität des Mehrschrittverfahrens und Satz 4.19 folgt damit die Existenz eine Norm $\|\cdot\|$ auf dem \mathbb{R}^k , derart, daß die zugehörige Matrixnorm $\|\cdot\|$ die Bedingung

$$\|A\| = \rho(A) \leq 1$$

erfüllt. Damit können wir aus (4.25) abschätzen:

$$\|E_i\| \leq \|D_i^{-1}\| (\|C_i\| \|E_{i-1}\| + \|B_i\|). \quad (4.26)$$

Unser Ziel ist nun, die Abschätzung auf die Form

$$\|E_i\| \leq (1 + \delta)\|E_{i-1}\| + \eta_i \quad (4.27)$$

für ein $\delta = O(h)$ zu bringen, um das Gronwall-Lemma anwenden zu können. Hierzu schätzen wir nun weiter die Terme $\|D_i^{-1}\|$, $\|C_i\|$ und $\|B_{i-1}\|$ ab. Dazu beobachten wir zuerst, daß es aufgrund der Äquivalenz von Normen auf dem \mathbb{R}^k eine Konstante $\kappa > 0$ gibt, so daß

$$\kappa^{-1}\|z\| \leq \|z\|_1 := \sum_{j=1}^k |z_j| \leq \kappa\|z\| \quad \forall z \in \mathbb{R}^k.$$

Damit können wir den Term $\|B_i\|$ wie folgt behandeln: Wir setzen $\beta := \max_{j=1, \dots, k} |\beta_{k-j}|$ und erhalten

$$|b_i| \leq h\beta L \sum_{j=1}^k |e_{i-j}| + |\varepsilon_i| \leq h\beta L \kappa \|E_{i-1}\| + |\varepsilon_i|.$$

Folglich gilt

$$\|B_i\| \leq \kappa |b_i| \leq h\beta L \kappa^2 \|E_{i-1}\| + \kappa |\varepsilon_i|. \quad (4.28)$$

Wir betrachten nun $\|D_i^{-1}\|$. Hierzu beobachten wir

$$\frac{1}{1-a} = 1 + \frac{a}{1-a} \quad \text{für } |a| < 1.$$

Somit erhalten wir die Darstellung

$$D_i^{-1} = \text{Id} + \tilde{D}_i, \quad \tilde{D}_i = \text{diag}_{j=k-1, \dots, 0} \left(\frac{h\beta_k \sigma_{i-j}}{1 - h\beta_k \sigma_{i-j}} \right)$$

Mithin können wir abschätzen

$$\begin{aligned} \|D_i^{-1}\| &\leq \|\text{Id}\| + \|\tilde{D}_i\| \leq 1 + \kappa^2 \|\tilde{D}_i\|_1 \\ &\leq 1 + \kappa^2 \max_{j=0, \dots, k-1} \left| \frac{h\beta_k \sigma_{i-j}}{1 - h\beta_k \sigma_{i-j}} \right| \leq 1 + h \frac{\kappa^2 |\beta| L}{1 - h|\beta| L}, \end{aligned} \quad (4.29)$$

wobei wir (4.24) und die Beobachtung

$$\|\tilde{D}_i\| = \sup_{0 \neq x \in \mathbb{R}^k} \frac{\|\tilde{D}_i x\|}{\|x\|} \leq \sup_{0 \neq x \in \mathbb{R}^k} \frac{\kappa \|\tilde{D}_i x\|_1}{\kappa^{-1} \|x\|_1} = \kappa^2 \|\tilde{D}_i\|_1$$

verwendet haben. Weiter erhalten wir mit den gleichen Techniken die Abschätzung

$$\|C_i\| \leq 1 + h|\beta_k| L \kappa^2. \quad (4.30)$$

Setzen wir die Abschätzungen (4.28), (4.29), (4.30) in (4.26) ein, so erhalten wir

$$\begin{aligned} \|E_i\| &\leq \left(1 + h \frac{\kappa^2 |\beta| L}{1 - h|\beta| L} \right) \{ (1 + h|\beta| L \kappa^2) \|E_{i-1}\| + h\beta L \kappa^2 \|E_{i-1}\| + \kappa |\varepsilon_i| \} \\ &\leq (1 + \gamma h) \|E_{i-1}\| + K |\varepsilon_i|, \end{aligned}$$

wobei die Konstanten γ , K durch

$$\begin{aligned}\gamma &= \frac{\kappa^2 |\beta_k| L}{1 - h\beta_k L} (1 + h\kappa^2 |\beta| L) + \kappa^2 L(\beta + \beta), \\ K &= \kappa \left(1 + h \frac{\kappa^2 \beta L}{1 - h\beta L} \right)\end{aligned}$$

gegeben sind. Das diskrete Gronwall-Lemma (Lemma 2.11) liefert dann

$$\|E_i\| \leq e^{\gamma(t_i - t_0)} \left\{ \|E_{k-1}\| + K \sum_{j=1}^i |\varepsilon_j| \right\},$$

was schließlich die gewünschte Aussage ergibt. \square

Satz 4.21 (Konvergenz von Mehrschrittverfahren) *Ein konsistentes und stabiles LMM ist konvergent.*

Genauer: Sei $G \subset \mathbb{R}^2$ offen, $f \in C^p(G)$ und $y \in C^{p+1}(J)$ mit $[t_0, T] \subset J$ Lösung des Anfangswertproblems $y' = f(t, y)$, $y(t_0) = y_0$. Sei ein nullstabiles LMM von der Form (4.6) mit Konsistenzordnung p gegeben. Dann existieren $\underline{h} > 0$, $\bar{\varepsilon} > 0$ und $C > 0$, so daß für $0 < h \leq \underline{h}$ und Anfangsfehler

$$\max_{i=0, \dots, k-1} |y_i - y(t_i)| =: \varepsilon \leq \bar{\varepsilon}$$

gilt:

$$\max_{i=0, \dots, N} |y_i - y(t_i)| \leq C(h^p + \varepsilon). \quad (4.31)$$

Beweis: Wir betrachten den vereinfachten Fall, daß $f \in C^p(J, \mathbb{R})$ und eine Lipschitzbedingung im zweiten Argument mit Lipschitzkonstante $L > 0$ erfüllt. Der Fall $f \in C^p(G)$ erfolgt mit den Techniken, die wir bei Einschrittverfahren im Beweis von Satz 2.10 kennengelernt haben.

Die wesentliche Idee ist, das Stabilitätsresultat Satz 4.20 anzuwenden (dessen wesentliche Idee wiederum ist, das Mehrschrittverfahren als Einschrittverfahren für ein geeignetes System aufzufassen). Schreibt man kurz

$$\tilde{y}_i = y(t_i),$$

so erfüllen die exakten Werte \tilde{y}_i die ‘‘gestörte’’ Differenzgleichung

$$\begin{aligned}\tilde{y}_i &= y_i + (\tilde{y}_i - y_i), \quad i = 0, \dots, k-1, \\ \sum_{j=0}^k \alpha_{k-j} \tilde{y}_{i-j} &= h \sum_{j=0}^k \beta_{k-j} \tilde{y}_{i-j} + R(t_i, y, h), \quad i = k, \dots, N,\end{aligned}$$

wobei der Abschneidefehler $R(t_i, y, h)$ in (4.9) definiert ist. Satz 4.20 liefert dann

$$|y(t_i) - y_i| = |\tilde{y}_i - y_i| \leq C e^{\gamma(t_i - t_0)} \left\{ \max_{j=0, \dots, k-1} |y_i - \tilde{y}_i| + \sum_{j=k}^N |R(t_j, y, h)| \right\}.$$

Nutzt man nun $y \in C^{p+1}(J)$ aus, so können wir den Abschneidefehler aufgrund der Annahme, daß ein LMM der Ordnung p vorliegt, gleichmäßig in j kontrollieren durch $|R(t_j, y, h)| \leq Ch^{p+1}$

mit einer Konstanten $C > 0$, welche nicht von j und $h \leq \underline{h}$ abhängt. Damit ergibt sich schließlich die gewünschte Behauptung. \square

Satz 4.21 zeigt, daß Konsistenz und Stabilität die Konvergenz eines LMM garantieren. Diese beiden Bedingungen sind auch notwendig, wenn man eine Klasse von Anfangswertaufgaben betrachtet, die die “trivialen” Differentialgleichungen $y' = 0$ und $y' = 1$ enthält. Es gilt nun:

Satz 4.22 *Sei ein Mehrschrittverfahren der Form (4.6) gegeben. Dann gilt:*

- (i) *damit das Mehrschrittverfahren konvergent (im Sinn von Definition 4.18) für das Anfangswertproblem $y' = 0$, $y(t_0) = 0$ ist, muß es nullstabil sein;*
- (ii) *damit das Mehrschrittverfahren konvergent (im Sinn von Definition 4.18) für die Anfangswertprobleme $y' = 0$ mit $y(t_0) = 0$ und $y' = 1$ mit $y(t_0) = 0$ ist, muß es konsistent sein.*

Beweis: Wir können o.B.d.A. annehmen, daß $t_0 = 0$ ist. *Beweis von (i):* Annahme: es gibt eine Nullstelle λ von ρ , so daß $|\lambda| > 1$ oder $|\lambda| = 1$ und λ ist mehrfache Nullstelle. Im ersten Fall betrachten wir die Folge $(y_i^{(1)})_{i=0}^N$, im zweiten Fall die Folge $(y_i^{(2)})_{i=0}^N$ gegeben durch

$$y_i^{(1)} = \sqrt{h}\lambda^i, \quad y_i^{(2)} = (ih)\lambda^i/\sqrt{h}.$$

Dann gilt: Die Folgen $(y_i^{(1)})_{i=0}^N$ bzw. $(y_i^{(2)})_{i=0}^N$ sind Lösungen des Differenzenverfahrens, es gilt

$$\begin{aligned} \max_{i=0,\dots,k-1} |y_i^{(1)} - y(t_i)| &= \max_{i=0,\dots,k-1} |y_i^{(1)}| \rightarrow 0 && \text{für } h \rightarrow 0, \\ \max_{i=0,\dots,k-1} |y_i^{(2)} - y(t_i)| &= \max_{i=0,\dots,k-1} |y_i^{(2)}| \rightarrow 0 && \text{für } h \rightarrow 0, \end{aligned}$$

aber an der Stelle T gilt

$$|y_N^{(1)}| \rightarrow \infty, \quad |y_N^{(2)}| \rightarrow \infty, \quad \text{für } h \rightarrow 0,$$

d.h. es liegt keine Konvergenz vor.

Beweis von (ii): Das Mehrschrittverfahren konvergiert gegen die Lösung $y(t) = t$. Da die Folge $(y_i)_{i=0}^N$ die Differenzengleichung erfüllt, muß wegen $f(t, y) = 1$ für $i \in \{k, \dots, N\}$ gelten

$$\sum_{j=0}^k \alpha_{k-j} y_{i-j} = h \sum_{j=0}^k \beta_{k-j} = h\sigma(1). \quad (4.32)$$

Läßt man nun $h \rightarrow 0$, so folgt aus der Konvergenz von y_{N-j} für $j = 0, \dots, k$, daß $\rho(1) = 0$ sein muß. Aus dem Teil (i) wissen wir zudem, daß diese Nullstelle $\lambda = 1$ eine einfache Nullstelle ist. Wir definieren deshalb $K := \sigma(1)/\rho'(1)$. Mit Hilfe von $\rho(1) = 0$ überzeugt man sich leicht davon, daß die Folge $(y_i)_{i=0}^N$ mit $y_i = Kih$ eine Lösung der Differenzengleichung (4.32) ist. Weiterhin gilt für die Anfangswerte y_j , $j = 0, \dots, k-1$, daß $\lim_{h \rightarrow 0} |y_j - jh| = 0$. Aus der vorausgesetzten Konvergenz des Verfahrens folgt damit an der Stelle $t_N = 1$, daß $1 = \lim_{h \rightarrow 0} y_N = \lim_{h \rightarrow 0} KNh = K$. Wir erhalten also $1 = K = \sigma(1)/\rho'(1)$, welches $\sigma(1) = \rho'(1)$. Nach Satz 4.8 folgt also die Konsistenz. \square

In Satz 4.21 hatten wir bereits gesehen, daß Konsistenz und Nullstabilität die Konvergenz eines LMM nach sich ziehen. Satz 4.22 zeigt nun, daß Konsistenz und Nullstabilität auch notwendig sind, um für interessante Klassen von Problemen Konvergenz von LMM zu erhalten. Diese Beobachtung wird in der Literatur oft als “Gleichung”

$$\text{Konvergenz} = \text{Konsistenz} + \text{Stabilität}$$

geschrieben und ist als *Äquivalenzprinzip von Peter Lax* bekannt.

finis 12.D

4.4 Bemerkungen zu Mehrschrittverfahren

Für die praktische Umsetzung von Mehrschrittverfahren müssen eine Reihe von Aspekten beachtet werden. Wir führen hier einige auf:

1. *Anlaufrechnung*: Um ein k -Schnittverfahren einzusetzen, müssen die Startwerte y_0, \dots, y_{k-1} bekannt sein. Da exakte Werte nicht bekannt sind, werden hinreichend genaue Approximationen mit Hilfe einer Anlaufrechnung bestimmt. Dies kann z.B. mit einem Einschrittverfahren geschehen. Dabei muß die Genauigkeit des Einschrittverfahrens an die Ordnung und Schrittweite des Mehrschrittverfahrens angepaßt sein. Satz 4.21 zeigt, daß bei Mehrschrittverfahren der Ordnung p und Schrittweite h ein Fehler $\varepsilon \approx h^p$ für die Approximationen y_i , $i = 0, \dots, k - 1$ akzeptabel ist. Dies kann z.B. dadurch erreicht werden, daß ein Einschrittverfahren der Konsistenzordnung $p - 1$, dessen *lokaler* Fehler ja $O(h^p)$ ist (und damit ist der Fehler für die ersten $k - 1$, Werte ebenfalls gleich dem lokalen Fehler $O(h^p)$), eingesetzt wird. Für ein Mehrschrittverfahren der Ordnung 2 würde z.B. das explizite Eulerverfahren hinreichend genaue Approximationen an die Startwerte liefern. Verwendet man Einschrittverfahren niedriger Ordnung, muß man ggf. die Schrittweite des Einschrittverfahrens hinreichend klein wählen. Angesichts der Tatsache, daß die Anlaufrechnung oft nur einen kleinen Anteil der Gesamtkosten ausmacht, wird man typischerweise versuchen, die Startwerte y_i , $i = 0, \dots, k - 1$, möglichst genau zu bestimmen.

Eine Alternative ergibt sich im Kontext von Mehrschrittverfahren, die mit Schritt- und Ordnungssteuerung arbeiten. Hier könnte man das Mehrschrittverfahren sich selbst starten lassen, indem man mit einem Einschrittverfahren und hinreichend kleiner Schrittweite startet (Schrittweite so klein, daß die Toleranzbedingungen erfüllt sind) und dann, sobald es möglich ist, langsam die Ordnung des Verfahrens erhöht.

2. *Schrittweitensteuerung*: Schrittweitensteuerung ist bei LMM nicht sehr einfach. Wir verweisen auf [6, Chap. III.5], wo dies ausführlich beschrieben ist. Eine Möglichkeit, Verfahren variabler Schrittweite zu erzeugen, ist, die Herleitung in Abschnitt 4.1 zu analysieren und dann Koeffizienten α_j , β_j zu bestimmen, die von verwendeten Punkten t_{i+1-j} , $j = 0, \dots, k$ explizit abhängen. Eine andere Möglichkeit besteht darin, aus den bekannten Werten y_{i-j} , $j = 0, \dots, k - 1$, die zu dem Gitter mit Schrittweite h gehören, durch Interpolation neue Werte $y_{i-j,neu}$, $j = 0, \dots, k - 1$, zu berechnen, die zu dem neuen Gitter mit Schrittweite h_{neu} gehören.

In der Praxis stellt sich heraus, daß man die Schrittweite aus Stabilitätsgründen nur vergleichsweise langsam verändern kann. Hier liegt somit ein Unterschied zu den Ein-

schrittverfahren vor, die ggf. auf “Unvorhergesehenes” mit einer drastischen Schrittweitenreduktion reagieren können. Bei zu starker Schrittweitenänderung müßte ein Mehrschrittverfahren mit einer erneuten Anlaufrechnung neu gestartet werden.

3. *Fehlerschätzer*: Für Schrittweitensteuerung benötigt man eine Abschätzung für den Fehler. Da man bei Mehrschrittverfahren eine ganze Familie von Verfahren verschiedener Ordnung zu Verfügung hat, ist das Schätzen des Fehlers relativ einfach und analog zu dem Vorgehen bei eingebetteten RK-Verfahren.

4.4.1 Prädiktor-Korrektor Methoden

Verwendet man implizite Mehrschrittverfahren, so muß ein nichtlineares Gleichungssystem gelöst werden. Bei nicht-steifen (oder nur schwach steifen) Problemen kann man das Gleichungssystem mit der Fixpunktiteration des Banachschen Fixpunktsatzes lösen (für steife würde man das Newtonverfahren einsetzen!). Einen Startwert für die Fixpunktiteration liefert dann ein explizites Mehrschrittverfahren, welches der *Prädiktor* genannt wird. Das implizite Mehrschrittverfahren wird dann der *Korrektor* genannt. Wir führen das Vorgehen am folgenden Beispiel vor.

Beispiel 4.23 Als *Prädiktor* wird das Adams-Bashforth-Verfahren der Ordnung 4 eingesetzt:

$$y_i = y_{i-1} + \frac{h}{24} [55f_{i-1} - 59f_{i-2} + 37f_{i-3} - 9f_{i-4}]. \quad (4.33)$$

Der *Korrektor* ist das Adams-Moulton-Verfahren der Ordnung 4 eingesetzt:

$$y_i = y_{i-1} + \frac{h}{24} [9f_i + 19f_{i-1} - 5f_{i-2} + f_{i-3}]. \quad (4.34)$$

Löst man (4.34) mit einer Fixpunktiteration, so ergibt sich die Iteration

$$y_i^{(n+1)} = y_{i-1} + \frac{h}{24} [9f_i^{(n)} + 19f_{i-1} - 5f_{i-2} + f_{i-3}], \quad (4.35)$$

wobei $f_i^{(n)} = f(t_i, y_i^{(n)})$ gesetzt wird. Diese Fixpunktiteration führt man nun wie folgt aus:

- (P) bestimme $y_i^{(0)}$ mit der Formel (4.33), d.h. der Startwert wird durch das explizite Verfahren erhalten (P steht für *predict*);
- (E) setze $f_i^{(n)} := f(t_i, y_i^{(n)})$ (E steht für *evaluate*);
- (C) bestimme $y_i^{(n+1)}$ mit (4.35) (C steht für *correct*).

Nach dem Startschritt (P) wiederholt man die Schritte (E) und (C) $m \in \mathbb{N}$ mal (oder, wenn man es wirklich als Fixpunktiteration auffassen will, bis eine Abbruchbedingung erfüllt ist). In der Praxis gibt es zwei Varianten der Prädiktor-Korrektor-Verfahren, die kompakt als

$$P(EC)^m \quad \text{und} \quad P(EC)^m E$$

notiert werden. Bei der ersten Variante $P(EC)^m$ wird die letzte Approximation $f_i^{(m)}$ als Wert für f_i im nächsten Schritt des LMM eingesetzt; bei der zweiten Variante wird zum Abschluß noch ein *Evaluate*-Schritt durchgeführt und somit $f_i^{(m+1)}$ als Wert für f_i im nächsten Schritt des LMM verwendet. ■

In der Praxis sind $m = 1$ oder $m = 2$ üblich, d.h. man iteriert nicht “bis zur Konvergenz”. Die so erhaltenen Verfahren $P(EC)^m$ und $P(EC)^m E$ sind nicht mehr LMM von der Form (4.6), denn durch die *Evaluate*-Schritte entstehen geschachtelte f -Auswertungen. Sie ähneln so den RK-Verfahren. Man kann folgendes Resultat zeigen:

Satz 4.24 (Ordnung des Prädiktor-Korrektor-Verfahrens) Sei $m^{(P)}$ die Ordnung des verwendeten Prädiktorverfahrens und $m^{(C)}$ die Ordnung des verwendeten Korrektorverfahrens. Dann gilt für die Ordnung p_m der Verfahren $P(EC)^m$ und $P(EC)^m E$

$$p_m = \min\{m^{(C)}, m^{(P)} + m\}.$$

Beweis: Übung □

4.4.2 A-Stabilität von linearen Mehrschrittverfahren

Für die Behandlung von steife Differentialgleichungen hatten wir bereits in Abschnitt 3.2 den Begriff der A-Stabilität herausgearbeitet. Dort hatten wir ein numerisches Verfahren als A-stabil bezeichnet, wenn es, angewandt auf die Modellgleichung

$$y' = \lambda y, \tag{4.36}$$

beschränkte Lösungen für alle $\lambda \in \mathbb{C}^- = \{z \in \mathbb{C} \mid \operatorname{Re} z \leq 0\}$ liefert. Wir übernehmen nun diese Sichtweise für unsere Definition von A-Stabilität von LMM. Um die Beschränktheit von Lösungen von LMM zu überprüfen, verfahren wir wie in Abschnitt 4.3.1: Das LMM angewandt auf die Modellgleichung (4.36) führt auf die Differenzgleichung

$$\sum_{j=0}^k \alpha_{k-j} y_{i+1-j} = z \sum_{j=0}^k \beta_{k-j} y_{i+1-j}, \quad z = \lambda h. \tag{4.37}$$

Die allgemeine Lösung dieser Differenzgleichung bestimmen wir wieder als Linearkombination von Folgen der Form $(\xi^i)_{i \in \mathbb{N}_0}$. Setzt man diesen Ansatz in (4.37) ein und kürzt mit ξ^{i-k} , so ergibt sich als Bestimmungsgleichung für ξ :

$$\sum_{j=0}^k \alpha_{k-j} \xi^{k-j} = z \sum_{j=0}^k \beta_{k-j} \xi^{k-j}, \quad z = \lambda h, \tag{4.38}$$

d.h. die Werte ξ sind die Nullstellen von

$$\rho(\xi) = z\sigma(\xi). \tag{4.39}$$

Genau wie beim Begriff der Nullstabilität muß für die Beschränktheit der Lösungen der Differenzgleichung (4.37) gelten, daß *alle* Nullstellen ξ die folgenden Bedingungen erfüllen: $|\xi| \leq 1$ und, falls ξ eine mehrfache Nullstellen ist, so muß $|\xi| < 1$ gelten. Offensichtlich hängen die Nullstellen von $z \in \mathbb{C}$ ab. Das Stabilitätsgebiet ist nun definiert als die Werte von $z \in \mathbb{C}$, für die die obige Stabilitätsbedingung erfüllt ist:

Definition 4.25 (Stabilitätsgebiete von LMM) Sei ein LMM von der Form (4.6) gegeben. Dann ist das Stabilitätsgebiet $S \subset \mathbb{C}$ definiert als

$$S := \{z \in \mathbb{C} \mid \text{alle Nullstellen } \xi \text{ von } \rho(\xi) - z\sigma(\xi) = 0 \text{ erfüllen } |\xi| \leq 1 \text{ und} \quad (4.40) \\ \text{mehrfache Nullstellen } \xi \text{ erfüllen } |\xi| < 1\}.$$

LMM, welche $\mathbb{C}^- \subset S$ erfüllen heißen A-stabil. LMM, deren Stabilitätsgebiet die Menge $C(\alpha) = \{re^{i\varphi} \in \mathbb{C} \mid r > 0, \varphi \in (\pi - \alpha, \pi + \alpha)\}$ enthält, heißen $A(\alpha)$ stabil.

Bemerkung 4.26 Nullstabilität (siehe Def. 4.13) ist der Spezialfall $z = 0$, d.h. für nullstabile Verfahren muß $0 \in S$ sein. ■

Mit den Gauß- und Radau-Verfahren sind A-stabile Einschrittverfahren beliebig hoher Ordnung verfügbar. Eine naheliegende Frage ist, ob es A-stabile LMM höherer Ordnung gibt. Die zweite Dahlquistschranke zeigt, daß bei LMM sich die Forderung nach A-Stabilität nicht mit der Forderung nach hoher Ordnung verträgt:

Satz 4.27 (zweite Dahlquistschranke) Es gibt keine A-stabilen LMM der Ordnung $p \geq 3$. A-stabile Verfahren der Ordnung zwei sind z.B. die Trapezregel (Adams-Moulton mit $k = 1$, vgl. Beispiel 4.1) und das BDF-2 Verfahren (vgl. Beispiel 4.2 mit $k = 2$).

Beweis: Siehe z.B. [7, Sec. V.1]. □

Will man Verfahren der Ordnung $p \geq 3$ einsetzen, so muß man auf A-Stabilität verzichten. Die BDF-Verfahren (bis Ordnung 6) sind null-stabil und sogar $A(\alpha)$ -stabil, wobei der Parameter $\alpha \in (0, \pi/2)$ von der Ordnung k abhängt. Für viele praktische Zwecke reicht das Stabilitätsgebiet der BDF-Verfahren aus.

5 Randwertprobleme

Bei den bisher betrachteten Problemen handelte es sich um Anfangswertprobleme. In der Praxis treten, insbesondere bei Differentialgleichungen höherer Ordnung, auch *Randwertprobleme* auf. Bei solchen Problemen ist eine Funktion $t \mapsto y$ gesucht, die zum einen eine Differentialgleichung erfüllt, zum anderen Bedingungen für Funktionswerte (oder Ableitungen) zu zwei verschiedenen Zeitpunkten t_0, T .

Wir führen die Problematik an den folgenden, einfachen Fällen vor:

Beispiel 5.1 Finde eine Funktion $y \in C^1(J, \mathbb{R}^n)$ (hier ist $n \geq 2$), so daß

$$y' = f(t, y), \quad t \in J, \quad (5.1a)$$

$$\mathbf{A}y(t_0) + \mathbf{B}y(T) = \mathbf{c}, \quad (5.1b)$$

wobei die Matrizen $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ und der Vektor $\mathbf{c} \in \mathbb{R}^n$ gegeben sind. Z.B. können bei verschiedenen Komponenten von y die Funktionswerte zum Zeitpunkt t_0 und T gegeben sein. ■

Eine wichtige Klasse von Randwertproblemen entsteht bei der Behandlung von Differentialgleichungen höherer Ordnung:

Beispiel 5.2 Finde eine Funktion $t \mapsto y(t)$, so daß

$$y'' = f(t, y, y') \quad \text{für } t \in J, \quad (5.2a)$$

$$y(t_0) = y_0, \quad y(T) = y_T \quad (5.2b)$$

zu gegebener Funktion f und Werten y_0, y_T . ■

Bemerkung 5.3 Die Randbedingung (5.2b) ist nur eine mögliche Randbedingung. Man spricht im vorliegenden Fall von *separierten* Randbedingungen, weil die Werte von y (oder von y') an den Stellen t_0 und T nicht gekoppelt sind. Allgemeinere Randbedingungen wären analog zu (5.1b)

$$\mathbf{A} \begin{pmatrix} y(t_0) \\ y'(t_0) \end{pmatrix} + \mathbf{B} \begin{pmatrix} y(T) \\ y'(T) \end{pmatrix} = \mathbf{c}$$

für Matrizen $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{2 \times 2}$ und $\mathbf{c} \in \mathbb{R}^2$. Auch nichtlineare Kopplungen treten in der Praxis auf. ■

Während wir mit dem Satz von Peano (Satz 1.1) unter schwachen Annahmen an die rechte Seite f Existenz von Lösungen und mit dem Satz von Picard-Lindelöf (Satz 1.3) auch (lokale) Eindeutigkeit erhalten haben, ist die Situation bei Randwertproblemen schwieriger, wie die folgenden Beispiele zeigen:

Beispiel 5.4 1. Das Randwertproblem

$$y'' + y = 0 \quad \text{auf } [0, \pi/2], \quad y(0) = 0, \quad y(\pi/2) = 1,$$

hat die eindeutige Lösung $y(t) = \sin t$.

2. Das Randwertproblem

$$y'' + y = 0 \quad \text{auf } [0, \pi], \quad y(0) = 0, \quad y(\pi) = 0,$$

wird von jeder Funktion der Form $y(t) = c \sin t$, $c \in \mathbb{R}$ gelöst.

3. Das Randwertproblem

$$y'' + y = 0 \quad \text{auf } [0, \pi], \quad y(0) = 0, \quad y(\pi) = 1,$$

besitzt keine Lösung.

■

Wir werden im folgenden nicht die Frage nach Existenz und Eindeutigkeit vertiefen. Für die vorzustellenden Algorithmen nehmen wir Existenz und Eindeutigkeit als gegeben an.

Randwertprobleme können mit zwei verschiedenen Typen von Techniken gelöst werden:

1. *Schießverfahren* (engl.: *shooting methods*) bauen auf dem Lösen von Anfangswertproblemen auf und führen das Lösen von Randwertproblemen auf Nullstellensuche von (nichtlinearen) Funktionen zurück.
2. Bei *Differenzenverfahren/Projektionsverfahren* werden die Unbekannten Funktionswerte y_i , $i = 0, \dots, N$, an den Stellen t_i direkt als Unbekannte angesetzt. Es entsteht dann ein großes (i.a. nichtlineares) Gleichungssystem, welches gelöst werden muß.

Wir werden beide Verfahren kurz ansprechen.

5.1 Schießverfahren

Wir führen die Ideen des Schießverfahrens für das Randwertproblem (5.2) vor. Wir formulieren wir die Differentialgleichung als ein System von ODEs erster Ordnung: Mit der Definition

$$\mathbf{y} := \begin{pmatrix} y(t) \\ y'(t) \end{pmatrix} \quad (5.3)$$

erhalten wir die Aufgabe

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad \mathbf{f}(t, \mathbf{y}) = \begin{pmatrix} \mathbf{y}_2 \\ f(t, \mathbf{y}_1, \mathbf{y}_2) \end{pmatrix} \quad (5.4)$$

Weiter schreiben wir

$$\mathbf{y}_0 := \begin{pmatrix} y_0 \\ s_0 \end{pmatrix}. \quad (5.5)$$

Ist $s_0 = y'(t_0)$ bekannt, dann kann die Lösung von (5.2) einfach als Lösung $\mathbf{y}_{t_0, \mathbf{y}_0}$ des Anfangswertproblems

$$\mathbf{y}' = \mathbf{f}(t, \mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad (5.6)$$

bestimmt werden. Zur Vereinfachung der Notation schreiben wir die Komponenten des Lösungsvektors $\mathbf{y}_{t_0, \mathbf{y}_0}$ als

$$\mathbf{y}_{t_0, \mathbf{y}_0}(t) = \begin{pmatrix} y(t, t_0, y_0, s_0) \\ y'(t, t_0, y_0, s_0) \end{pmatrix}.$$

Die unbekannte Größe s_0 , die die zweite Komponente von \mathbf{y}_0 darstellt, ergibt sich aus der zweiten Randbedingung (5.2b), d.h. aus der Bedingung

$$y(T, t_0, y_0, s_0) \stackrel{!}{=} y_T.$$

Somit erhalten wir:

$$\text{Finde } s_0 \in \mathbb{R}, \text{ so da\ss } y(T, t_0, y_0, s_0) - y_T = 0. \quad (5.7)$$

Dies ist nun eine (i.a. nichtlineare) Gleichung, die mit bekannten Techniken gelöst werden kann. Wir betrachten hier die Möglichkeit, die Gleichung (5.7) mit Hilfe des Newtonverfahrens zu lösen. Hierzu benötigen wir die Ableitung $\partial_{s_0} y(T, t_0, y_0, s_0)$. Wir zeigen nun, da\ss $\partial_{s_0} y(T, t_0, y_0, s_0)$ als Lösung eines Anfangswertproblems bestimmt werden kann (vgl. auch Übung 1.7):

Lemma 5.5 *Sei $y(t, t_0, y_0, s_0)$ die Lösung des Anfangswertproblems*

$$y''(t) = f(t, y(t), y'(t)), \quad y(t_0) = y_0, \quad y'(t_0) = s_0.$$

Dann ist die Funktion $v : t \mapsto \partial_{s_0} y(t, t_0, y_0, s_0)$ Lösung des Anfangswertproblems

$$\begin{aligned} v''(t) &= f_y(t, y(t, t_0, y_0, s_0), y'(t, t_0, y_0, s_0))v(t) + f_{y'}(t, y(t, t_0, y_0, s_0), y'(t, t_0, y_0, s_0))v'(t) \\ v(t_0) &= 0, \quad v'(t_0) = 1. \end{aligned} \quad (5.8a) \quad (5.8b)$$

Beweis: Die Lösung $t \mapsto y(t, t_0, y_0, s_0)$ erfüllt

$$y''(t, t_0, y_0, s_0) = f(t, y(t, t_0, y_0, s_0), y'(t, t_0, y_0, s_0)), \quad y(t_0, t_0, y_0, s_0) = y_0, \quad y'(t_0, t_0, y_0, s_0) = s_0.$$

Differenziert man diese Gleichung nach s_0 und wendet die Kettenregel an, so ergibt sich, da\ss die Funktion $v(t) := \partial_{s_0} y(t, t_0, y_0, s_0)$ die Gleichung

$$v''(t) = f_y(t, y(t, t_0, y_0, s_0), y'(t, t_0, y_0, s_0))v(t) + f_{y'}(t, y(t, t_0, y_0, s_0), y'(t, t_0, y_0, s_0))v'(t)$$

erfüllt. Als Anfangsbedingungen haben wir für v wegen

$$\begin{aligned} v(t_0) &= \partial_{s_0} y(t_0, t_0, y_0, s_0) = \partial_{s_0} y_0 = 0, \\ v'(t_0) &= \partial_{s_0} y'(t_0, t_0, y_0, s_0) = \partial_{s_0} s_0 = 1. \end{aligned}$$

□

Lemma 5.5 zeigt, da\ss die gesuchte Funktion $s \mapsto y(T, t_0, y_0, s)$ sogar als Lösung *linearer* Anfangswertproblems bestimmt werden kann. Insgesamt ergibt sich folgender Algorithmus für Randwertprobleme der Form (5.2):

Algorithmus 5.6 (einfaches Schießverfahren) % input: Startwerte $y_0, s_0^{(0)}$ $n := 0$

1. bestimme (numerisch) $y(t, t_0, y_0, s_0^{(n)})$ durch Lösen des Anfangswertproblems (5.6)
2. bestimme (numerisch) $\partial_{s_0} y(T, t_0, y_0, s_0^{(n)})$ durch Lösen des Anfangswertproblems (5.8)
3. Führe einen Newtonschritt zur Lösung von $y(T, t_0, y_0, s_0^{(n)}) - y_T = 0$ durch:

$$(a) \quad s_0^{(n+1)} := s_0^{(n)} - \left(\partial_{s_0} y(T, t_0, y_0, s_0^{(n)}) \right)^{-1} (y(T, t_0, y_0, s_0^{(n)}) - y_T)$$

$$(b) \quad n := n + 1$$

4. Prüfe Abbruchbedingungen des Newtonverfahrens (im allereinfachsten Fall, ob $|y(T, t_0, y_0, s_0^{(n)}) - y_T|$ hinreichend klein ist). Falls nicht, gehe zu 1.

Beispiel 5.7 Wir betrachten das Randwertproblem

$$y'' = \lambda y + y', \quad (5.9a)$$

$$y(0) = 1, \quad y(T) = 1, \quad (5.9b)$$

wobei $\lambda > 0$ ein Parameter ist. Für $\lambda = 10$ und $T = 1$ bzw. $T = 10$ ist die gesuchte Lösung in Tabelle 5.1 gezeichnet. Wir verwenden nun Algorithmus 5.6, um das Randwertproblem zu lösen. Die erste Tabelle in Tabelle 5.1 zeigt das Verhalten von Algorithmus 5.6, wenn als numerisches Verfahren zum Lösen der Anfangswertproblem das RK4-Verfahren mit $N = 10$ Schritten verwendet wird. Der Algorithmus liefert nach einem Newtonschritt die gewünschte Approximation (dies sollte auch so sein, da im vorliegenden Fall die Funktion $s \mapsto y(T, t_0, y_0, s)$ ein Polynom ersten Grades ist!).

Wir wenden uns nun dem Fall $T = 10$ zu. Die zweite Tabelle in Tabelle 5.1 zeigt das Verhalten von Algorithmus 5.6 (wieder mit dem RK4-Verfahren und $N = 10$ Schritten). Hier beobachten wir, daß der Fehler $y(T) - y_T$ nicht unter $0.5 \cdot 10^{-2}$ gedrückt werden kann. Auch eine Erhöhung der Rechengenauigkeit ($N = 1000$) in der letzten Tabelle erbringt nicht das gewünschte Ergebnis. Das Verhalten kann man mit einer Sensitivitätsanalyse erklären. Aus Satz 1.5 erhalten wir die Abschätzung

$$|y(T, t_0, y_0, s) - y(T, t_0, y_0, s + \varepsilon)| \leq C e^{LT} \varepsilon, \quad (5.10)$$

wobei L die Lipschitzkonstante ist. Im vorliegenden Fall ist $L = \lambda = 10$ und $T = 10$, so daß der Verstärkungsfaktor $e^{LT} = e^{100} = 2.7 \cdot 10^{43}$. Zwar ist dies nur eine Abschätzung, aber im vorliegenden Fall können wir uns davon überzeugen, daß die Aussage qualitativ richtig ist (siehe unten). Wir erhalten damit, daß das Schießverfahren im vorliegenden Fall extrem sensitiv auf Störungen der Anfangssteigung s reagiert. Diese Sensitivität ist zu groß für eine Rechengenauigkeit von ungefähr 10^{-16} .

Die allgemeine Lösung der Differentialgleichung (5.9a) ist von der Form

$$y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t},$$

wobei die Parameter λ_1, λ_2 die Nullstellen

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 + 4\lambda}}{2}$$

der quadratischen Gleichung $x^2 - x - \lambda = 0$ sind. Die Lösung des Randwertproblems läßt sich über die Lösungsformel einfach bestimmen. Für den Fall $T = 10$ und $\lambda = 10$ erhalten wir z.B.

$$\mathbf{y}(t) = \frac{e^{110} - 1}{e^{110} - e^{-110}} e^{-10t} \begin{pmatrix} 1 \\ -10 \end{pmatrix} + \frac{1 - e^{-110}}{e^{110} - e^{-110}} e^{11t} \begin{pmatrix} 1 \\ -10 + \frac{21(1 - e^{-100})}{e^{110} - e^{-100}} \end{pmatrix}. \quad (5.11)$$

Weiter können wir die Lösung $y(T, t_0, y_0, s)$ für $T = 10, y_0 = 1$ und $s \in \mathbb{R}$ zu

$$y(t, t_0, y_0, s) = \frac{11y_0 - s}{21} e^{-10t} + \frac{10y_0 + s}{21} e^{11t} \quad (5.12)$$

Der gesuchte Wert s_{exakt} , der zu $y(T, t_0, y_0, s) = y_T$ führt, ist damit

$$s_{\text{exakt}} = -10 + \frac{21(1 - e^{-100})}{e^{110} - e^{-110}}. \quad (5.13)$$

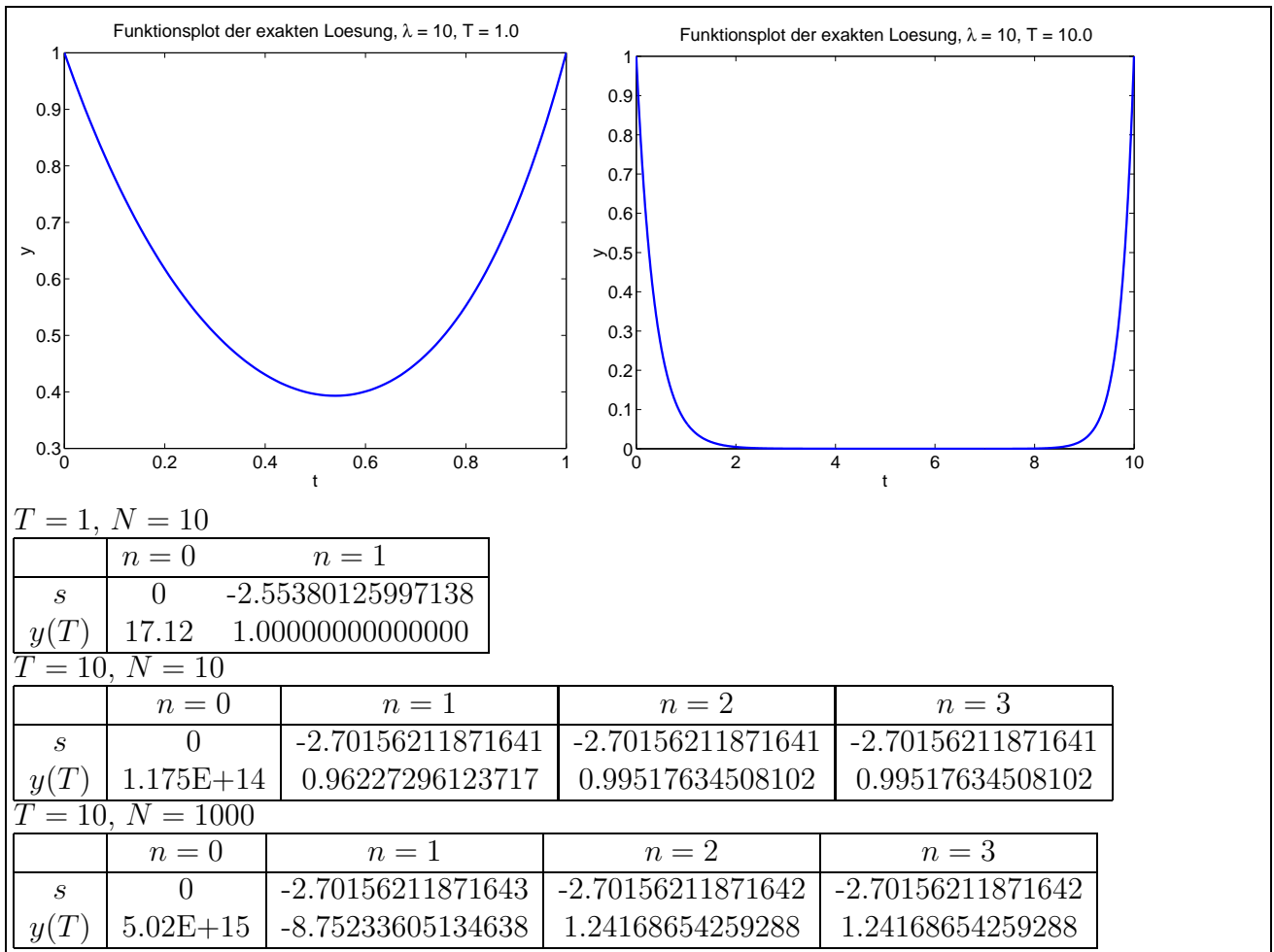


Tabelle 5.1: Beispiel für das Schießverfahren

Hieraus folgt für $y_0 = 1$, $t = T = 10$ und $s = s_{\text{exakt}} + 10^{-16}$

$$y(T, t_0, y_0, s) \approx y_T + 10^{-16} e^{11T} \approx y_T + 6 \cdot 10^{31}$$

■

Bemerkung 5.8 Die in Beispiel 5.7 beobachtete Sensitivität zeigt, daß auch beim iterativen Lösen Vorsicht geboten ist:

- Wegen der nur lokalen Konvergenz des Newtonverfahrens wird man in der Praxis das gedämpfte Newtonverfahren einsetzen, weil der korrekte Wert der Steigung s nicht bekannt ist.
- Falls die Ableitungen von f nicht zur Verfügung stehen (hier könnten auch Techniken des automatischen Differenzieren, [4] eingesetzt werden), muß numerisch differenziert werden. Bei hohen Genauigkeitsanforderungen (z.B. wenn $L|T - t_0|$ groß ist) kann dies hohe Rechenanforderungen an die zu lösenden Anfangswertprobleme stellen.

■

Beispiel 5.7 zeigt eine wesentliche Schwäche des einfachen Schießverfahrens: Der Verstärkungsfaktor $e^{L|T-t_0|}$, der sich aus Satz 1.5 ergibt, kann dazu führen, daß das Schießverfahren aufgrund zu großer Sensitivität der Funktion $s \mapsto y(T, t_0, y_0, s)$ versagt. Um diese Sensitivität zu verringern muß der Faktor $L|T-t_0|$ verkleinert werden. Da L vorgegeben ist, muß also das Intervall verkleinert werden. Das führt auf das *Mehrfachschießverfahren* (engl.: *multiple shooting method*, auch Mehrzielverfahren): Das Intervall $[t_0, T]$ wird in $M \in \mathbb{N}$ Teilintervalle zerlegt mit $t_0 < t_1 < \dots < t_M = T$. Die Teilintervalle werden so gewählt, daß die zugehörigen Verstärkungsfaktoren $e^{L|t_{i+1}-t_i|}$ moderat sind. Beim Mehrfachschießverfahren werden

$$\text{die Werte } y_i = y(t_i) \quad \text{und die Steigungen } s_i = y'(t_i), \quad i = 0, \dots, M,$$

als gesuchte Parameter angesetzt. Die Idee ist, auf jedem Intervall (t_i, t_{i+1}) die Lösung $y(t, t_i, y_i, s_i)$ zu berechnen und dann die zu bestimmenden Werte y_i, s_i durch die Bedingungen

$$y(t_{i+1}, t_i, y_i, s_i) \stackrel{!}{=} y_{i+1}, \quad y'(t_{i+1}, t_i, y_i, s_i) \stackrel{!}{=} s_{i+1}, \quad i = 0, \dots, M-2$$

festzulegen; dies ist gerade die stetige Differenzierbarkeit der Lösung $t \mapsto y(t)$ an den "inneren" Stützstellen $t_i, i = 1, \dots, M-1$. Die Randbedingungen $y(t_0) = y_0$ und $y(T) = y_T =: y_M$ kommen als zwei weitere Bedingungen hinzu. Somit erhalten wir als das Mehrfachschießverfahren:

gegeben y_0 und $y_M := y_T$,

finde $y_i, i = 1, \dots, M-1$ und $s_i, i = 0, \dots, M-1$, so daß

$$y(t_{i+1}, t_i, y_i, s_i) = y_{i+1}, \quad i = 0, \dots, M-1, \quad (5.14a)$$

$$y'(t_{i+1}, t_i, y_i, s_i) = s_{i+1}, \quad i = 0, \dots, M-2. \quad (5.14b)$$

Die Bedingungen (5.14) stellen ein (i.a. nichtlineares) Gleichungssystem dar. Löst man dieses mit dem Newtonverfahren, so müssen die Ableitungen

$$\partial_{s_i} y(t, t_i, y_i, s_i), \quad \partial_{y_i} y(t, t_i, y_i, s_i), \quad \partial_{y_i} y'(t, t_i, y_i, s_i) \quad \partial_{s_i} y(t, t_i, y_i, s_i)$$

bestimmt werden. Diese können wie in Lemma 5.5 als Lösungen von geeigneten Anfangswertproblemen identifiziert werden (Übung: geben Sie die Anfangswertprobleme an). Auch hier gelten die Kommentare aus Bemerkung 5.8 sinngemäß.

5.2 Differenzenverfahren

Ein ganz anderer Zugang zum Lösen von Randwertproblemen ist mit Techniken der *finiten Differenzenverfahren* bzw. *finiten Elementeverfahren* gegeben. Wir wollen hier die Grundzüge des finiten Differenzenverfahrens für das Randwertproblem (5.2) vorstellen.

Sei der Einfachheit halber ein uniformes Gitter

$$t_0 < t_1 < \dots < t_N = T \quad \text{mit } t_i = t_0 + ih$$

gegeben. Die Lösung $t \mapsto y(t)$ approximieren wir in den Gitterpunkten t_i durch zu bestimmende Werte y_i . Ferner approximieren wir erste und zweite Ableitungen durch Differenzenquotienten:

$$y'(t_i) \approx D_i^{sym} y := \frac{y(t_{i+1}) - y(t_{i-1}))}{t_{i+1} - t_{i-1}} \approx \frac{y_{i+1} - y_{i-1}}{2h}, \quad (5.15)$$

$$y''(t_i) \approx \frac{y(t_{i+1}) - 2y(t_i) + y(t_{i-1}))}{(t_{i+1} - t_i)(t_i - t_{i-1})} \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}. \quad (5.16)$$

Fordert man, daß die Differentialgleichung in den *inneren Punkten* t_i , $i = 1, \dots, N - 1$, (approximativ) gelten soll, so erhalten wir als Gleichungssystem:

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = f\left(t_i, y_i, \frac{y_{i+1} - y_{i-1}}{2h}\right), \quad i = 1, \dots, N - 1. \quad (5.17)$$

Weiter haben wir die Werte y_0 und $y_N := y_T$ gegeben. Damit stellt (5.17) ein großes Gleichungssystem dar, welches gelöst werden muß. Unter geeigneten Voraussetzungen an die Funktion f kann dann gezeigt werden, daß die Lösung des Gleichungssystems (5.17) gegen die exakte Lösung konvergiert:

$$|y(t_i) - y_i| \rightarrow 0 \quad \text{für } h \rightarrow 0.$$

Man kann z.B. folgendes zeigen:

Satz 5.9 Sei $y \in C^4([0, 1])$ die Lösung von

$$-y'' + \alpha y = f(t) \quad \text{auf } [0, 1], \quad y(0) = y(1) = 0,$$

wobei α eine Funktion mit $\alpha(x) \geq 0$ ist. Sei $(y_i)_{i=0}^N$ die Lösung, die sich aus dem Differenzenverfahren ergibt. Dann gilt:

$$\max_{i=0, \dots, N} |y(t_i) - y_i| \leq \frac{1}{24} h^2 \|y^{(4)}\|_{C([0,1])},$$

wobei die Konstante $C > 0$ nur von α abhängt.

Bemerkung 5.10 Wir haben oben die Ableitung $y'(t_i)$ durch die *symmetrische* Differenz

$$\frac{y_{i+1} - y_{i-1}}{2h}$$

approximiert. Man kann natürlich auch *einseitige* Differenzenquotienten

$$\frac{y_i - y_{i-1}}{h} \quad \text{oder} \quad \frac{y_{i+1} - y_i}{h}$$

verwenden. Der Grund für die Verwendung des symmetrischen ist, daß bei hinreichend glatter Lösung die symmetrische Differenz eine bessere Approximation an die Ableitung liefert als die einseitige (Übung: überlegen Sie sich mit Hilfe einer Taylorentwicklung, was die Fehler bei den Verfahren sind). ■

A Appendices

A.1 dies und das

A.1.1 Verfahren der Ordnung 4

Das klassische Runge-Kutta-Verfahren vierter Ordnung ist nicht das einzige 4-stufige Verfahren, das vierte Ordnung erreicht, wie das folgende Beispiel belegt:

Beispiel A.1 (3/8-Regel und Formel von Kuntzmann) Zwei weitere 4-stufige explizite Runge-Kutta-Verfahren, die vierter Ordnung sind, sind das sog. 3/8-Verfahren und das Verfahren von Kuntzmann, welche gegeben sind durch folgende Tableaus:

$$\begin{array}{c|ccc}
 0 & & & \\
 \frac{1}{3} & \frac{1}{3} & & \\
 \frac{2}{3} & -\frac{1}{3} & 1 & \\
 1 & 1 & -1 & 1 \\
 \hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
 \end{array}
 \qquad
 \begin{array}{c|ccc}
 0 & & & \\
 \frac{2}{5} & \frac{2}{5} & & \\
 \frac{3}{5} & -\frac{3}{20} & \frac{3}{4} & \\
 1 & \frac{19}{44} & -\frac{15}{44} & \frac{40}{44} \\
 \hline
 & \frac{55}{360} & \frac{125}{360} & \frac{125}{360} & \frac{55}{360}
 \end{array}$$

Für Funktionen f , die nur von t abhängen, ist das 3/8-Verfahren gerade die 3/8-Regel. ■

A.1.2 zum Schätzen des Konsistenzfehlers

Die “Herleitung” von (2.19) basiert auf der Annahme, daß $y_{t_0, y_0}(t_0 + H) - \hat{y}$ wesentlich kleiner ist als $y_{t_0, y_0}(t_0 + H) - \tilde{y}$. Es ist möglich, die Differenz $y_{t_0, y_0}(t_0 + H) - \hat{y}$ etwas genauer zu quantifizieren, wie wir nun vorführen. Daraus ergibt sich dann die Formel (2.21). Wir schreiben

$$\begin{aligned}
 y_{t_0, y_0}(t_0 + H) - \hat{y} &= y_{t_0, y_0}(t_0 + H) - y_{t_0 + H/2, y_{1/2}}(t_0 + H) \\
 &\quad + y_{t_0 + H/2, y_{1/2}}(t_0 + H) - \left[y_{1/2} + \frac{H}{2} \Phi\left(t_0 + \frac{H}{2}, y_{1/2}, \frac{H}{2}\right) \right] \\
 &= y_{t_0, y_0}(t_0 + H) - y_{t_0 + H/2, y_{1/2}}(t_0 + H) + \tau(t_0 + H/2, y_{1/2}, H/2) \\
 &= \left[y_{t_0 + H/2, y_{t_0, y_0}(t_0 + H/2)}(t_0 + H) - y_{t_0 + H/2, y_{1/2}}(t_0 + H) \right] + \tau(t_0 + H/2, y_{1/2}, H/2).
 \end{aligned}$$

Aus Notationsgründen ist es nun einfacher, die Funktion $(t, t_0, y_0) \mapsto y_{t_0, y_0}(t)$ als $y(t_0, y_0, t)$ zu schreiben. Weiterhin führen wir die Abkürzungen $t_{1/2} := t_0 + H/2$ und $\bar{y}_{1/2} := y_{t_0, y_0}(t_0 + H/2)$ ein. Wir bemerken: $\bar{y}_{1/2} - y_{1/2} = \tau(t_0, y_0, H/2)$. Dann ist:

$$y_{t_0, y_0}(t_0 + H) - \hat{y} = \underbrace{\left[y(t_{1/2}, \bar{y}_{1/2}, t_0 + H) - y(t_{1/2}, y_{1/2}, t_0 + H) \right]}_{=: F} + \tau(t_0 + H/2, y_{1/2}, H/2). \quad (\text{A.1})$$

Der erste Term ist Differenz von exakten Lösungen zu verschiedenen Anfangswerten (sie repräsentiert die “Fortpflanzung” des Fehlers aus dem ersten Teilschritt im zweiten Teilschritt); der zweite Term ist der Konsistenzfehler im zweiten Teilschritt. Aus den Übungen wissen wir, daß Anfangswertproblem stetig differenzierbar von den Anfangsdaten abhängen. Genauer: Die Funktion $t \mapsto R_{t_{1/2}, y_{1/2}}(t) := \partial_z y_{t_{1/2}, z}(t)|_{z=y_{1/2}}$ erfüllt

$$R_{t_{1/2}, y_{1/2}}(t_{1/2}) = 1, \quad R'_{t_{1/2}, y_{1/2}}(t) = f_y(t, y_{t_{1/2}, y_{1/2}}(t)) R_{t_{1/2}, y_{1/2}}(t),$$

Damit können wir den ersten Term in (A.1) mittels des Taylorschen Satzes so entwickeln:

$$\begin{aligned}
F &= \partial_z y(t_{1/2}, z, t_0 + H)|_{z=y_{1/2}} (\bar{y}_{1/2} - y_{1/2}) + O(|\bar{y}_{1/2} - y_{1/2}|^2) \\
&= R_{t_{1/2}, y_{1/2}}(t_0 + H) (\bar{y}_{1/2} - y_{1/2}) + O(|\bar{y}_{1/2} - y_{1/2}|^2) \\
&= \left[\underbrace{R_{t_{1/2}, y_{1/2}}(t_{1/2})}_{=1} + O(H) \right] (\bar{y}_{1/2} - y_{1/2}) + O(|\bar{y}_{1/2} - y_{1/2}|^2).
\end{aligned}$$

Wir erhalten also

$$F = (1 + O(H))\tau(t_0, y_0, H/2) + O(|\tau(t_0, y_0, H/2)|^2).$$

Zusammenfassen haben wir erhalten (wir machen nun die Annahme, daß H klein und $\tau(t_0, y_0, H/2)$ klein sind):

$$y_{t_0, y_0}(t_0 + H) - \hat{y} \approx \tau(t_0, y_0, H/2) + \tau(t_{1/2}, y_{1/2}, H/2). \quad (\text{A.2})$$

Nun ist τ stetig¹. Damit kann man annehmen, daß $\tau(t_{1/2}, y_{1/2}, H/2) \approx \tau(t_0, y_0, H/2)$, so daß sich ergibt:

$$y_{t_0, y_0}(t_0 + H) - \hat{y} \approx 2\tau(t_0, y_0, H/2). \quad (\text{A.3})$$

Differenzbildung von (2.18) und (A.3) und ersetzen von \approx durch $=$ liefert damit

$$\hat{y} - \tilde{y} = \tau(t_0, y_0, H) - 2\tau(t_0, y_0, H/2) = \gamma(t_0, y_0)H^{p+1} (1 - 2 \cdot 2^{-(p+1)}) = \gamma(t_0, y_0)H^{p+1}(1 - 2^{-p}).$$

Wegen des Ansatzes $\tau(t_0, y_0, h) = \gamma(t_0, y_0)h^{p+1}$ ergibt sich damit gerade (2.21).

Wir bemerken, daß (A.2) zeigt, daß bei wenigen (hier: 2) Schritten, der dominante Anteil des Gesamtfehlers nur die Summe der Konsistenzfehler ist.

¹dies folgt im Wesentlichen aus der Darstellung $\tau(t_0, y_0, h) = y_{t_0, y_0}(t_0 + h) - (y_0 + h\Phi(t_0, y_0, h))$, aus der Stetigkeit von Φ und der stetigen Abhängigkeit der Lösung y_{t_0, y_0} vom Anfangsdatum y_0

Bibliographie

- [1] P. Deuffhard. Order and stepsize control in extrapolation methods. *Numer. Math.*, 41:399–422, 1983.
- [2] P. Deuffhard and A. Bornemann. *Numerische Mathematik II*. deGruyter, 1994.
- [3] J.R. Dormand and P.J. Prince. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.*, 6:19–26, 1980.
- [4] A. Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, 2000.
- [5] A. Griewank and G.F. Corliss, editors. *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*. SIAM, 1991.
- [6] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations, Part 1*. Springer-Verlag, 2 edition, 1993.
- [7] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations, Part 2*. Springer-Verlag, 2 edition, 1996.
- [8] H. Heuser. *Lehrbuch der Analysis*. Teubner, 1986.
- [9] J. Stoer and R. Bulirsch. *Einführung in die Numerische Mathematik II*. Springer-Verlag, 1973.
- [10] W. Walter. *Gewöhnliche Differentialgleichungen*. Springer-Verlag, 1972.