

### 3 Verfahren für steife Differentialgleichungen

#### 3.1 Implizite Runge-Kutta-Verfahren

Im Tableau von expliziten Runge-Kutta-Verfahren aus Definition 2.13 sind nur die Einträge “unter der Diagonalen” von Null verschieden. Füllt man das Tableau voll auf, so erhält man implizite Runge-Kutta-Verfahren:

**Definition 3.1 (Implizite Runge-Kutta-Verfahren)** Eine Inkrementfunktion  $\Phi = \Phi(t, y, h)$  gehört zu einem  $s$ -stufiges implizites Runge-Kutta-Verfahren, falls sie für Zahlen  $a_{ij}, b_i \in \mathbb{R}, c_i \in [0, 1]$  von der folgenden Form ist:

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

wobei die Stufen  $k_i, i = 1, \dots, s$ , das folgende System von nichtlinearen Gleichungen lösen:

$$k_i = f\left(t + c_i h, y + h \sum_{j=1}^s a_{ij} k_j\right), \quad i = 1, \dots, s. \quad (3.1)$$

Zusätzlich müssen die Koeffizienten  $b_i, i = 1, \dots, s$  die Bedingung

$$\sum_{i=1}^s b_i = 1 \quad (3.2)$$

erfüllen.

Explizite RK-Verfahren sind somit Spezialfälle von impliziten RK-Verfahren, bei denen  $a_{ij} = 0$  für  $j \geq i$ . Implizite RK-Verfahren werden genau wie explizite RK-Verfahren als Tableaus notiert: das allgemeine  $s$ -stufige implizite RK-Verfahren ist

$$\begin{array}{c|cccccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s-1} & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s-1} & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss-1} & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array} \quad \text{oder kompakter: } \begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b} \end{array}$$

**Bemerkung 3.2** Die Inkrementfunktion  $\Phi$  von impliziten RK-Verfahren ist

$$\Phi(t, h, y) = \sum_{i=1}^s b_i k_i.$$

Für  $h = 0$  sieht man leicht ein, daß dann  $k_i = f(t, y)$  die (einzige) Lösung für die Stufen ist. Es folgt deshalb  $\Phi(t, 0, y) = f(t, y) \sum_{i=1}^s b_i$ . In Hinblick auf Lemma 2.7 sind implizite RK-Verfahren somit nur konsistent, wenn die Koeffizienten  $b_i$  die Bedingung (3.2) erfüllen. Aus diesem Grund wurde diese Bedingung in die Definition von impliziten RK-Verfahren in Definition 3.1 aufgenommen. ■

Der Grundtyp der impliziten RK-Verfahren ist das sog.  $\theta$ -Schema:

**Beispiel 3.3 ( $\theta$ -Schema)** Für  $\theta \in [0, 1]$  wird das  $\theta$ -Schema durch das Tableau

$$\begin{array}{c|c} \theta & \theta \\ \hline & 1 \end{array}$$

beschrieben, d.h. das Verfahren ist

$$y_{i+1} = y_i + h_i f(t_i + \theta h_i, y_i + \theta(y_{i+1} - y_i))$$

(Übung: man überzeuge sich davon!) Für  $\theta = 0$  erhält man das explizite Eulerverfahren, für  $\theta = 1$  das implizite Eulerverfahren und für  $\theta = 1/2$  die beliebte *implizite Mittelpunktsregel*. Ihre Beliebtheit rührt daher, daß es gute Stabilitätseigenschaften mit der Tatsache kombiniert, daß es ein Verfahren 2. Ordnung ist (vgl. Satz 3.7). ■

Bei impliziten RK-Verfahren können die Stufen  $k_i$  nicht einfach sukzessiv bestimmt werden, sondern sind als Lösung eines (i.a. nichtlinearen) Gleichungssystems definiert. Für hinreichend kleine Schrittweiten  $h$  kann man zeigen, daß die Stufen  $k_i$  eindeutig bestimmt werden können, wenn  $f$  lokal lipschitzstetig im zweiten Argument ist. Der Einfachheit halber zeigen wir folgende vereinfachte Fassung dieser Aussage:

**Satz 3.4 (Lösbarkeit bei impliziten RK-Verfahren)** Sei  $J$  ein Intervall,  $f \in C(J \times \mathbb{R})$  und

$$|f(t, y) - f(t, \hat{y})| \leq L|y - \hat{y}| \quad \forall (t, y), (t, \hat{y}) \in J \times \mathbb{R}$$

für ein  $L \geq 0$ . Für  $s$ -stufige implizite RK-Verfahren von der obigen Form ist dann das Gleichungssystem, welches die Stufen  $k_i$ ,  $i = 1, \dots, s$ , definiert, für jedes  $t \in J$ ,  $y \in \mathbb{R}$  lösbar, falls die Schrittweite  $h > 0$  die Bedingungen

$$\begin{aligned} t + h &\in J, \\ hL \max_{i=1, \dots, s} \sum_{j=1}^s |a_{ij}| &< 1 \end{aligned} \tag{3.3}$$

erfüllt ist.

**Beweis:** Die Bedingung  $t + h \in J$  wurde nur explizit gefordert, damit die Zeitpunkte  $t + c_i h$  im Definitionsbereich von  $f$  liegen können. Für die Lösbarkeit für die Stufen  $k_i$  beobachtet man, daß das Gleichungssystem (3.1) ein Fixpunktproblem ist. Die Bedingung (3.3) zeigt dann, daß die zugehörige Abbildung eine Kontraktion auf  $\mathbb{R}^s$  ist. Die eindeutige Lösbarkeit folgt dann aus dem Banachschen<sup>1</sup> Fixpunktsatz, wenn man den  $\mathbb{R}^s$  mit der Maximumsnorm versieht. □

**Bemerkung 3.5** Satz 3.4 beruht auf dem Banachschen Fixpunktsatz. Dies legt nahe, die Stufen  $k_i$  mit Hilfe der Fixpunktiteration aus dem Beweis des Banachschen Fixpunktsatzes zu bestimmen. In der Praxis wird jedoch das Newtonverfahren<sup>2</sup> eingesetzt. Der Grund ist, daß implizite Verfahren in erster Linie für steifen Differentialgleichungen eingesetzt, bei denen die

---

<sup>1</sup>Banach, Stefan 1892–1945

<sup>2</sup>Newton, Sir Isaac 1643–1727

Lipschitzkonstante  $L$  groß ist. Damit dann die Bedingung (3.3) erfüllt ist, müßten dann die Schrittweiten  $h$  sehr klein sein; somit würde ein Vorteil von impliziten Verfahren (nämlich im Vergleich zu expliziten Verfahren große Schrittweiten) zunichte gemacht werden (Übung: Betrachten Sie das Anfangswertproblem  $y' = \lambda y$ ,  $y_0 = 1$  mit  $\lambda = 10$ . Es soll ein Schritt des impliziten Eulerverfahrens gemacht werden. Überlegen Sie sich, daß das Newtonverfahren die benötigte Stufe  $k_1$  in einem Schritt liefert. Überlegen Sie sich, daß die Fixpunktiteration des Banachschen Fixpunktsatzes für  $k_1$  nur konvergiert, wenn  $h\lambda < 1$ , wenn als Startwert die sinnvolle Wahl  $\lambda y_0$  getroffen wird.) Ein weiterer Vorteil des Newtonverfahrens ist natürlich die quadratische Konvergenz im Unterschied zur linearen Konvergenz der Fixpunktiteration. Ferner sei darauf hingewiesen, daß die Bedingung (3.3) nur eine hinreichende Bedingung für Lösbarkeit des Systems (3.1) ist; in der Praxis hofft man, daß selbst bei großem  $L$  Lösbarkeit für vergleichsweise große  $h$  gegeben ist (für das Beispiel  $f = \lambda y$  ist für jedes  $h$  mit  $\lambda h \neq 1$  Lösbarkeit gegeben!). ■

**Übung 3.6 (Wohldefiniertheit von IRK-Verfahren)** Zeigen Sie folgende Verallgemeinerung von Satz 3.4. Sei  $G \subset \mathbb{R}^2$  offen,  $f \in C^1(G)$ . Sei  $K \subset G$  eine kompakte Teilmenge. Sei weiter ein implizites RK-Verfahren mit  $s$  Stufen gegeben. Dann gibt es ein  $h_0 > 0$ , so daß für alle  $h \in (0, h_0]$  und jedes  $(t, y) \in K$  die Stufen  $k_i$  des impliziten RK-Verfahren eindeutig bestimmt werden können. Zudem existiert ein  $C > 0$ , welches nur von  $K$ ,  $f$  und  $h_0$  abhängt, so daß für alle Stufen  $k_i$ ,  $i = 1, \dots, s$  gilt:  $\text{dist}(k_i, f(K)) \leq Ch$ .

Hinweis: Wählen Sie eine kompakte Teilmenge  $K'$  von  $G$  mit  $K \subset K' \subset G$  und  $\text{dist}(\partial K, \partial K') > 0$ . Um den Banachschen Fixpunktsatz anwenden zu können, muß noch die Menge  $\tilde{K} \subset \mathbb{R}^s$  definiert werden, in der die Stufen  $k_i$ ,  $i = 1, \dots, s$  gesucht werden. Eine sinnvolle Wahl ist die kompakte Menge

$$\tilde{K} := \underbrace{f(K') \times \dots \times f(K')}_{s \text{ mal}}.$$

Überlegen Sie sich nun, daß für  $h$  hinreichend klein der Banachsche Fixpunktsatz zur Bestimmung der Stufen  $k_i$  für jedes  $(t, y) \in K$  anwendbar ist. ■

Der nächste Satz illustriert das prinzipielle Vorgehen beim Nachprüfen der Konsistenzordnung bei impliziten Verfahren am Beispiel des  $\theta$ -Schemas. Formal kann man mit Hilfe des Satzes über implizite Funktionen auch implizite Einschrittverfahren als explizite Verfahren auffassen. Diese Sichtweise wird am Beispiel des  $\theta$ -Schemas im folgenden Satz vorgeführt.

**Satz 3.7 (Konsistenzfehler des  $\theta$ -Schemas)** Sei  $G \subset \mathbb{R}^2$  ein Gebiet,  $f \in C^2(G)$  und  $K \subset G$  eine kompakte Teilmenge. Dann gibt es ein  $h_0 > 0$  und ein  $C > 0$ , so daß für jedes  $h \in (0, h_0]$  gilt:

(i) (Wohldefiniertheit des  $\theta$ -Schemas) Für jedes  $(t_0, y_0) \in K$  und  $h \in (0, h_0]$  hat die Gleichung

$$\hat{y} = y_0 + hf(t_0 + \theta h, y_0 + \theta(\hat{y} - y_0)) \quad (3.4)$$

eine eindeutige Lösung  $\hat{y}$ .

(ii) (Konsistenz) Es gilt für den Konsistenzfehler für jedes  $(t_0, y_0) \in K$

$$|y_{t_0, y_0}(t_0 + h) - \hat{y}| \leq C \begin{cases} h^2 & \text{für } \theta \neq 1/2 \\ h^3 & \text{für } \theta = 1/2. \end{cases}$$

**Beweis:** Die Wohldefiniertheit folgt aus Übung 3.6. Wir wenden uns deshalb dem Konsistenzfehler zu. Dabei gehen wir in 2 Schritten vor.

1. *Schritt:* In diesem Schritt nehmen wir an, daß  $f \in C^2(\mathbb{R}^2)$  ist. Im folgenden (rein technischen) 2. Schritt werden wir uns dann dem Fall  $f \in C^2(G)$  zuwenden.

Sei  $f \in C^2(\mathbb{R}^2)$ . Zudem nehmen wir an, daß alle Ableitungen von  $f$  bis zur 2. Ordnung beschränkt sind:

$$\sup_{(t,y) \in \mathbb{R}^2} |f(t,y)| + |f_t(t,y)| + |f_{tt}(t,y)| + |f_{ty}(t,y)| + |f_{yy}(t,y)| =: M < \infty.$$

Dies impliziert insbesondere, daß  $f$  Lipschitzstetig (mit Lipschitzkonstante  $L \leq M$ ) im 2. Argument ist:

$$|f(t,y) - f(t,z)| \leq L|y - z| \quad \forall (t,y), (t,z) \in \mathbb{R}^2.$$

Sei nun  $(t_0, y_0) \in \mathbb{R}^2$  beliebig. Dann existiert ein  $h_0 > 0$  (welches nur von  $M$  und dem IRK abhängt), so daß für alle  $0 < h \leq h_0$  nach Satz 3.4 die Lösung  $\hat{y}$  des diskreten Problems existiert; weiterhin ist nach Satz 1.3  $y_{t_0, y_0} \in C^1(\mathbb{R})$  und, weil  $y_{t_0, y_0}$  die Differentialgleichung  $y' = f(t, y)$  mit  $f \in C^2$  erfüllt, erhalten wir  $y_{t_0, y_0} \in C^3(\mathbb{R})$  und

$$\|y_{t_0, y_0}^{(3)}\|_{C([t_0 - 2h_0, t_0 + 2h_0])} \leq M' := 2M^3 + 3M^2 + M.$$

Unser Ziel ist, eine Gleichung für den Konsistenzfehler für  $h \in (0, h_0)$  herzuleiten. Hierzu führen wir den Abschneidefehler  $r(h)$  durch die Gleichung

$$y_{t_0, y_0}(t_0 + h) = y_{t_0, y_0}(t_0) + hf(t_0 + \theta h, y_{t_0, y_0}(t_0) + \theta[y_{t_0, y_0}(t_0 + h) - y_{t_0, y_0}(t_0)]) + r(h)$$

ein. Da  $y_{t_0, y_0}(t_0) = y_0$  gilt, haben wir folgende zwei Gleichungen zur Verfügung:

$$\begin{aligned} y_{t_0, y_0}(t_0 + h) &= y_0 + hf(t_0 + \theta h, y_0 + \theta[y_{t_0, y_0}(t_0 + h) - y_0]) + r(h) \\ \hat{y} &= y_0 + hf(t_0 + \theta h, y_0 + \theta[\hat{y} - y_0]) \end{aligned}$$

Subtraktion beider Gleichungen ergibt eine Gleichung für den Konsistenzfehler  $y_{t_0, y_0}(t_0 + h) - \hat{y}$ :

$$\begin{aligned} |y_{t_0, y_0}(t_0 + h) - \hat{y}| &\leq h \left| f(t_0 + \theta h, y_0 + \theta[y_{t_0, y_0}(t_0 + h) - y_0]) - f(t_0 + \theta h, y_0 + \theta[\hat{y} - y_0]) \right| + |r(h)| \\ &\leq hL\theta|y_{t_0, y_0}(t_0 + h) - \hat{y}| + |r(h)|. \end{aligned} \quad (3.5)$$

Wählen wir nun  $h_0$  zusätzlich so klein, daß  $hL \leq h_0L \leq 1/2$ , dann erhalten wir für den Konsistenzfehler  $\tau(t_0, y_0, h)$ :

$$|y_{t_0, y_0}(t_0 + h) - \hat{y}| \leq \frac{1}{1 - h\theta L} |r(h)| \leq 2|r(h)|. \quad (3.6)$$

Es bleibt, die Größe  $r(h)$  abzuschätzen. Dies erfolgt nun durch Taylorentwicklung. Wir führen das Vorgehen für den Fall  $\theta = 1/2$  vor; der Fall  $\theta \in [0, 1] \setminus \{1/2\}$  wird durch ähnliche Überlegungen behandelt (Übung).

Wir entwickeln die Funktion  $y_{t_0, y_0}$  nach Taylor um den Mittelpunkt  $\hat{t} = t_0 + h/2$  des Intervalls  $[t_0, t_0 + h]$ . Es gilt

$$\begin{aligned} y_{t_0, y_0}(t_0) &= y_{t_0, y_0}(\hat{t}) - \frac{h}{2} y'_{t_0, y_0}(\hat{t}) + \frac{1}{2} \left(\frac{h}{2}\right)^2 y''_{t_0, y_0}(\hat{t}) + O(h^3) \\ y_{t_0, y_0}(t_0 + h) &= y_{t_0, y_0}(\hat{t}) + \frac{h}{2} y'_{t_0, y_0}(\hat{t}) + \frac{1}{2} \left(\frac{h}{2}\right)^2 y''_{t_0, y_0}(\hat{t}) + O(h^3). \end{aligned}$$

wobei die Fehlerterme, die sich hinter  $O(h^3)$  verbergen, abgeschätzt werden können durch

$$\frac{1}{6} \|y_{t_0, y_0}'''\|_{C([t_0-h_0, t_0+h_0])} h^3 \leq Ch^3;$$

die Konstante  $C = M'/6 > 0$  hängt über  $M'$  nur von  $M$  ab. Damit gilt

$$y_{t_0, y_0}(t_0 + h) + y_{t_0, y_0}(t_0) = 2y_{t_0, y_0}(\hat{t}) + \left(\frac{h}{2}\right)^2 y_{t_0, y_0}''(\hat{t}) + O(h^3), \quad (3.7)$$

$$y_{t_0, y_0}(t_0 + h) - y_{t_0, y_0}(t_0) = hy_{t_0, y_0}'(\hat{t}) + O(h^3). \quad (3.8)$$

Somit gilt für  $r$

$$\begin{aligned} r(h) &= y_{t_0, y_0}(t_0 + h) - \{y_{t_0, y_0}(t_0) + hf(t_0 + h/2, [y_{t_0, y_0}(t_0) + y_{t_0, y_0}(t_0 + h)]/2)\} \\ &= \{y_{t_0, y_0}(t_0 + h) - y_{t_0, y_0}(t_0) - hf(\hat{t}, y_{t_0, y_0}(\hat{t}))\} \\ &\quad - h \{f(t_0 + h/2, [y_{t_0, y_0}(t_0) + y_{t_0, y_0}(t_0 + h)]/2) - f(\hat{t}, y_{t_0, y_0}(\hat{t}))\} \end{aligned} \quad (3.9)$$

Wegen der Differentialgleichung  $y_{t_0, y_0}'(\hat{t}) = f(\hat{t}, y_{t_0, y_0}(\hat{t}))$  und (3.8) folgt damit, daß die erste Klammer in (3.9) die Abschätzung

$$|y_{t_0, y_0}(t_0 + h) - y_{t_0, y_0}(t_0) - hf(\hat{t}, y_{t_0, y_0}(\hat{t}))| \leq Ch^3,$$

erfüllt, wobei die Konstante  $C > 0$  unabhängig von  $h$  ist (wie oben gesehen, hängt sie nur von  $M$  ab). Für die zweite Klammer in (3.9) nutzen wir die Lipschitzstetigkeit von  $f$  im zweiten Argument:

$$\begin{aligned} &|hf(t_0 + h/2, [y_{t_0, y_0}(t_0) + y_{t_0, y_0}(t_0 + h)]/2) - f(\hat{t}, y_{t_0, y_0}(\hat{t}))| \\ &\leq hL |[y_{t_0, y_0}(t_0) + y_{t_0, y_0}(t_0 + h)]/2 - y_{t_0, y_0}(\hat{t})| \leq Ch^3, \end{aligned}$$

wobei die Konstante  $C$  aus (3.7) bestimmt werden kann. Damit haben wir gezeigt, daß  $|r(h)| = O(h^3)$ . Insgesamt haben wir erzielt:

$$|y_{t_0, y_0}(t_0 + h) - \hat{y}| \leq Ch^{2+1}, \quad (3.10)$$

d.h. für  $\theta = 1/2$  hat das  $\theta$ -Schema die Konsistenzordnung 2. Die Konstante  $C > 0$  in (3.10) hängt nur von  $M$  ab.

**2. Schritt:** Für einen vollständigen Beweis, daß das  $\theta$ -Schema die im Satz angegebene Ordnung hat, müssen wir die Voraussetzung  $f \in C^2(\mathbb{R}^2)$  zu  $f \in C^2(G)$  abschwächen und zeigen, daß für jedes Kompaktum  $K \subset G$  sich ein  $C' > 0$  und ein  $h_0 > 0$  finden läßt, so daß für alle  $(t_0, y_0) \in K$  und  $h \in (0, h_0)$  die Abschätzung (3.10) (im Fall  $\theta = 1/2$ ) mit dieser Konstanten  $C' > 0$  gilt. Wir versuchen hierzu, das Ergebnis des 1. Schrittes zu verwenden.

Sei  $K \subset G$  kompakt. Wähle  $K'$  kompakt mit  $K \subset K' \subset G$  und  $\text{dist}(K, \partial K') > 0$ . Sei<sup>3</sup>  $\chi \in C_0^\infty(\mathbb{R}^2)$  mit  $\text{supp} \chi \subset G$  und  $\chi \equiv 1$  auf  $K'$ . Definiere<sup>4</sup>  $\tilde{f} := \chi \cdot f$ . Bezeichne mit  $\tilde{y}_{t_0, y_0}$  die Lösung des AWP  $\tilde{y}_{t_0, y_0}' = \tilde{f}$  und  $\tilde{y}_{t_0, y_0}(t_0) = y_0$ . Die Funktion  $\tilde{f}$  erfüllt  $f \in C^2(\mathbb{R}^2)$  und wir können  $M$  finden, so daß

$$\sup_{(t, y) \in \mathbb{R}^2} |\tilde{f}(t, y)| + |\tilde{f}_t(t, y)| + |\tilde{f}_{tt}(t, y)| + |\tilde{f}_y(t, y)| + |\tilde{f}_{ty}(t, y)| + |\tilde{f}_{yy}(t, y)| \leq M < \infty.$$

<sup>3</sup>es reicht  $\chi \in C^2(\mathbb{R}^2)$  zu fordern—für eine explizite Konstruktion, siehe den 3. Schritt unten

<sup>4</sup>genauer:  $\tilde{f}(t, y) := \chi(t, y)f(t, y)$  für  $(t, y) \in G$  und  $\tilde{f}(t, y) = 0$  sonst

Dann schließen wir aus dem ersten Schritt die Existenz von  $h_0 > 0$  (welches nur von  $M$  abhängt) und eine Konstante  $C' > 0$  (welche ebenfalls nur von  $M$  abhängt), so daß für alle  $(t_0, y_0) \in K$  der Konsistenzfehler des  $\theta$ -Schemas für die Funktion  $\tilde{f}$  die Abschätzung

$$\left| \tilde{y}_{t_0, y_0}(t_0 + h) - \hat{\tilde{y}} \right| \leq C' h^3$$

erfüllt (hier ist  $\hat{\tilde{y}}$  die numerische Approximation, die zum Verfahren mit  $\tilde{f}$  gehört).

Weiters folgt aus dem Satz von Picard-Lindelöf (Satz 1.3) und der Kompaktheit von  $K'$  und  $\text{dist}(K, \partial K')$ , daß es ein  $h_0 > 0$  gibt, welches nur von  $K$  und  $f$  abhängt, so daß für jedes  $(t_0, y_0) \in K$  die Funktion  $y_{t_0, y_0} \in C^1(J)$  mit  $J = (t_0 - 2h_0, t_0 + 2h_0)$ . Zudem können wir  $h_0$  so klein wählen<sup>5</sup>, daß  $\{(t, y_{t_0, y_0}(t)) \mid t \in J\} \subset K'$ . Aus  $|\tilde{f}(t, y)| \leq M$  folgt, daß es ein  $h_0 > 0$  gibt, so daß für jedes  $(t_0, y_0) \in K$  gilt:  $\{(t, \tilde{y}_{t_0, y_0}(t)) \mid t \in [t_0, t_0 + h_0]\} \subset K'$ . Aufgrund der Eindeutigkeitsaussage von Picard-Lindelöf und weil  $f|_{K'} = \tilde{f}|_{K'}$  folgt dann, daß  $y_{t_0, y_0}|_J = \tilde{y}_{t_0, y_0}|_J$ . Weiterhin folgt aus der Tatsache, daß  $f$  und  $\tilde{f}$  auf  $K'$  übereinstimmen, daß für  $(t_0, y_0) \in K$  und  $h_0$  hinreichend klein (abhängig nur  $\text{dist}(K, \partial K')$  und  $M$ ), daß auch die Stufe  $k_1$ , die im  $\theta$ -Schema bestimmt wird, für die Funktionen  $f$  und  $\tilde{f}$  übereinstimmt. Also gilt auch  $\hat{y} = \hat{\tilde{y}}$ . Damit ergibt sich also aus dem ersten Schritt, daß das  $\theta$ -Schema die angegebene Konsistenzordnung hat.

**3. Schritt (Anhang):** Die Konstruktion der Funktion  $\chi$  im 2. Schritt kann auf verschiedene Arten realisiert werden—wir verweisen z.B. auf die Vorlesungen Analysis<sup>6</sup> oder Partielle Differentialgleichungen. Eine konkrete "elementare" Realisierung für den Fall  $\chi \in C^2(\mathbb{R}^2)$  ist wie folgt.

1. sei  $\psi \in C^2(\mathbb{R})$  mit  $\text{supp} \psi \subset [-1, 1]$  und  $\psi \geq 0$  auf  $\mathbb{R}$ . Z.B. kann konkret  $\psi(r) = \cos^4(r\pi/2)$  für  $|r| \leq 1$  und  $\psi(r) = 0$  für  $|r| > 1$  gewählt werden.
2. Definiere für  $\varepsilon > 0$  die Funktion  $\rho_\varepsilon \in C^2(\mathbb{R}^2)$  durch  $\rho_\varepsilon(x) := \lambda \varepsilon^{-2} \rho(\|x\|_2^2/\varepsilon)$ , wobei  $\lambda > 0$  durch die Bedingung  $\int_{\mathbb{R}^2} \rho_1(x) dx = 1$  festgelegt ist. D.g.:  $\int_{\mathbb{R}^2} \rho_\varepsilon(x) dx = 1$  für alle  $\varepsilon > 0$  und  $\text{supp} \rho_\varepsilon \subset B_\varepsilon(0)$ .
3. Wähle nun eine weitere kompakte Menge  $K''$  mit  $K' \subset K'' \subset G$  und  $\text{dist}(K', \partial K'') > 0$ . Wähle weiters  $\varepsilon > 0$ , so daß  $\varepsilon < \text{dist}(K'', \partial G)$  und  $\varepsilon < \text{dist}(K', \partial K'')$ . Definiere für die charakteristische Funktion  $\chi_{K''}$  des Kompaktums  $K''$ :

$$\chi(x) := (\rho_\varepsilon * \chi_{K''})(x) := \int_{\mathbb{R}^2} \rho_\varepsilon(x - y) \chi_{K''}(y) dy = \int_{K''} \rho_\varepsilon(x - y) dy.$$

Die Vorlesung Analysis zeigt nun, daß  $\chi \in C^2(\mathbb{R}^2)$ . Weiters folgt aus  $\text{supp} \rho_\varepsilon \subset B_\varepsilon(0)$  und  $\text{dist}(K'', \partial G) > \varepsilon$ , daß  $\text{supp} \chi \subset G$ . Ferner folgt für  $x \in K'$  und  $\text{supp} \rho_\varepsilon \subset B_\varepsilon(0)$  und  $\text{dist}(K', \partial K'') > \varepsilon$ , daß  $B_\varepsilon(x) \subset K''$ . Aus  $\int_{\mathbb{R}^2} \rho_\varepsilon(y) dy = 1$  schließen wir damit, daß für  $x \in K'$  gilt:

$$\chi(x) = \int_{K''} \rho_\varepsilon(x - y) dy = \int_{B_\varepsilon(x)} \rho_\varepsilon(y) dy = \int_{\mathbb{R}^2} \rho_\varepsilon(y) dy = 1.$$

□

**Bemerkung 3.8** Der Beweis von Satz 3.7 zeigt, daß der Konsistenzfehler des  $\theta$ -Schemas im wesentlichen durch die Größe

$$r(h) = y_{t_0, y_0}(t + h) - \{y_{t_0, y_0}(t) - hf(t + \theta h, y_{t_0, y_0}(t) + \theta[y_{t_0, y_0}(t + h) - y_{t_0, y_0}(t)])\} \quad (3.11)$$

<sup>5</sup>auch dies ist Teil der Aussage von Picard-Lindelöf

<sup>6</sup>z.B. im Wesentlichen Lemma 13.4.1 des Ana III Skripts von Kaltenbäck

bestimmt wird. Allgemeiner schätzt man in der Numerik oft den Konsistenzfehler für ein implizites Verfahren der Form  $y_{i+1} = y_i + h_i \Phi(t_i, h_i, y_i, y_{i+1})$  ab, indem man für eine hinreichend glatte Lösung  $y$  der Differentialgleichung die Größe

$$r(h) = y(t+h) - \{y(t) + h\Phi(t, h, y(t), y(t+h))\}$$

durch Taylorentwicklung abschätzt. Im Beweis von Satz 3.7 (der Schritt von (3.5) auf (3.6)) wird am Beispiel des  $\theta$ -Schemas vorgeführt, warum  $r(h)$  tatsächlich im wesentlichen der Konsistenzfehler im Sinne von Definition 2.6 ist. ■

### 3.1.1 Bemerkungen zu impliziten RK-Verfahren

#### Gaußverfahren

Bei expliziten RK-Verfahren haben wir in Satz 2.17 gesehen, daß ein  $s$ -stufiges Verfahren maximal die Konvergenzordnung  $p = s$  erreichen kann (für  $s \geq 5$  waren die Butcherschranken sogar noch restriktiver). Für impliziten RK-Verfahren kann man zeigen, daß die maximal erreichbar Ordnung  $p = 2s$  ist (siehe, z.B. [2, Lemma 6.34]). Diese maximale erreichbare Ordnung wird auch tatsächlich von den sog. Gaußverfahren erreicht.

Wichtige Vertreter impliziter RK-Verfahren sind die Gaußregeln<sup>7</sup>, und die verwandten Radau- und Lobattoregeln<sup>8</sup>. Die Namen dieser impliziten RK-Verfahren rühren daher, daß für rechte Seiten  $f$ , welche nicht von  $y$  abhängen, die bekannten Gauß-, Radau- und Lobattoquadraturregeln entstehen.

1. *Gaußregeln*: Die Stützstellen  $c_i$  und die Koeffizienten  $b_i$  entsprechen den Stützstellen und Gewichten für die klassische Gaußquadratur (natürlich bezogen auf das Intervall  $[0, 1]$ ). Insbesondere reduziert sich für rechte Seiten  $f$ , welche nur von  $t$  abhängen, eine Gaußregel auf die klassische Gaußquadratur. Eine Gaußregel mit  $s$  Stufen hat Konsistenzordnung  $p = 2s$ .

Beispiel für  $s = 2$  und  $p = 4$ :

$\frac{3-\sqrt{3}}{6}$	$\frac{1}{4}$	$\frac{3+2\sqrt{3}}{12}$
$\frac{3+\sqrt{3}}{6}$	$\frac{3+2\sqrt{3}}{12}$	$\frac{1}{4}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

2. *Radauregeln*: Bei Radauregeln ist entweder der linke Endpunkt  $t$  oder der rechte Endpunkt  $t+h$  ein Quadraturpunkt, d.h. entweder ist  $c_1 = 0$  oder  $c_s = 1$ . Radauregeln mit  $s$  Stufen haben Konsistenzordnung  $p = 2s - 1$ . Die einfachsten Beispiele sind das explizite und das implizite Eulerverfahren ( $s = 1$  und  $p = 1$ ). Weitere Beispiele sind z.B.

Beispiele für  $s = 2$  und  $p = 3$ :

$0$	$0$	$0$
$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\frac{1}{4}$	$\frac{3}{4}$	$\frac{3}{4}$

$\frac{1}{3}$	$\frac{1}{3}$	$0$
$1$	$1$	$0$
$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

3. *Lobattoregeln*: Bei Lobattoregeln ist sowohl der linke Endpunkt  $t$  als auch der rechte Endpunkt  $t+h$  ein Quadraturpunkt, d.h.  $c_1 = 1$  und  $c_s = 0$ . Lobattoregeln mit  $s$  Stufen haben Konsistenzordnung  $p = 2s - 2$ . Ein Vorteil ist, daß sie so konstruiert werden

<sup>7</sup>Gauß, Carl Friedrich, 1777–1855

<sup>8</sup>Rodolphe Radau (1835–1911), Reuel Lobatto (1797–1866)

können, daß die erste Zeile und die letzte Spalte des Butcherschemas nur Nullen enthält, d.h. die Stufen  $k_1$  und  $k_s$  können explizit bestimmt werden; damit reduziert sich die Anzahl impliziter Gleichungen für die Stufen.

Beispiele für  $s = 2$  und  $p = 2$ , bzw.  $s = 3$ ,  $p = 4$ :

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 1 & 0 & 1 & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

### 3.1.2 Diagonal implizite und linear implizite Verfahren

Bei impliziten RK-Verfahren muß ein (nichtlineares) Gleichungssystem für die Stufen  $k_i$  gelöst werden. Da dies kompliziert bzw. aufwendig sein kann (Newtonverfahren!), hat man vereinfachte Verfahren entwickelt. Eine Vereinfachungsmöglichkeit sind die *diagonal impliziten Verfahren*, bei denen die Koeffizienten  $a_{ij}$  aus dem Tableau die Bedingung  $a_{ij} = 0$  für  $j > i$  erfüllen (es sollte natürlich mindestens ein  $a_{ii} \neq 0$ , da sonst ein explizites Verfahren vorliegt). Diese sind numerisch etwas einfacher zu behandeln, da die Stufen  $k_i$  sukzessive bestimmt werden können. Um das Lösen einer nichtlinearen Gleichung (oder eines nichtlinearen Gleichungssystems im Fall von Systemen von Differentialgleichungen) vermeiden, wurden *linear-implizite RK-Verfahren* (auch nach Rosenbrock *Rosenbrock-Verfahren* genannt) entwickelt. Wir führen die Grundidee von linear-impliziten RK-Verfahren für autonome Differentialgleichungen der Form

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

vor. Die Differentialgleichung wird in der Form

$$y'(t) = J y(t) + \left[ f(y(t)) - J y(t) \right]$$

geschrieben, wobei  $J = \partial_y f(y_0)$  gesetzt wird. Für geeignete Parameter  $b_i$ ,  $i = 1, \dots, s$ , und  $\alpha_{ij}$ ,  $\beta_{ij}$ ,  $i, j = 1, \dots, s$ , werden dann die Stufen  $k_i$  definiert als Lösungen von

$$k_i = J \left( y_0 + h \sum_{j=1}^i \beta_{ij} k_j \right) + \left( f \left( y_0 + h \sum_{j=1}^{i-1} \alpha_{ij} k_j \right) - J \left( y_0 + h \sum_{j=1}^{i-1} \alpha_{ij} k_j \right) \right), \quad i = 1, \dots, s$$

und anschließend  $\hat{y} = y_0 + h \sum_{i=1}^s b_i k_i$  gesetzt. Man beachte, daß die Summen bis  $i$  bzw.  $i - 1$  laufen. Daraus ergibt sich, daß die Stufen  $k_i$  einfach sukzessive als Lösungen von linearen Gleichungen bestimmt werden:

(i)  $J := \partial_y f(y_0)$

(ii)  $(\text{Id} - h\beta_{ii}J)k_i = h \sum_{j=1}^{i-1} (\beta_{ij} - \alpha_{ij})Jk_j + f(y_0 + h \sum_{j=1}^{i-1} \alpha_{ij} k_j), \quad i = 1, \dots, s$

(iii)  $\hat{y} := y_0 + h \sum_{i=1}^s b_i k_i.$

(Im zweiten Schritt haben wir  $\text{Id}$  geschrieben, um anzudeuten, wie im Falle von vektorwertigen Funktion  $f$  vorzugehen ist; in diesem Fall ist  $J$  eine Matrix und die Stufen  $k_i$  sind Vektoren.)



## 3.2 Steife Differentialgleichungen

### 3.2.1 Einführung in die Problematik

#### Einige Beispiele

Gewöhnliche Differentialgleichungen, die z.B. in der chemischen Reaktionskinetik auftreten, können typischerweise nicht mit vertretbarem Aufwand mit expliziten Verfahren behandelt werden. Um die typische Schwierigkeit zu demonstrieren, betrachten wir die folgenden zwei Beispiele:

**Beispiel 3.9** Wir betrachten für  $\lambda < 0$  das Anfangswertproblem

$$y'(t) = \lambda y(t), \quad \text{auf } [0, T], \quad y(0) = y_0.$$

Wendet man das explizite Eulerverfahren mit konstanter Schrittweite  $h > 0$  an, so ergibt sich als Iterationsvorschrift

$$y_{i+1} = y_i + \lambda h y_i, \quad i = 0, 1, \dots \quad (3.12)$$

und damit

$$y_i = y_0(1 + \lambda h)^i = y_0(1 + \lambda h)^{t_i/h}, \quad t_i = ih.$$

Für  $|\lambda h| \geq 2$  ist  $|1 + \lambda h| > 1$ , so daß  $|y_i| \rightarrow \infty$  für  $i \rightarrow \infty$ . Da für die gesuchte Lösung  $y$  gilt:  $\lim_{t \rightarrow \infty} y(t) = 0$ , reflektiert das numerische Verfahren nicht einmal das qualitative Verhalten der Lösung wie auch in Fig. 3.1 illustriert. Das Konvergenzresultat Satz 2.10 garantiert natürlich Konvergenz für  $h \rightarrow 0$ ; die numerische Approximation beginnt aber erst das qualitative Verhalten der Lösung widerzuspiegeln, wenn  $|\lambda h| < 2$ . Ist  $|\lambda|$  groß, erzwingt die Bedingung  $|\lambda h| < 2$  sehr kleine Schrittweiten  $h$ , d.h. man muß einen hohen Aufwand treiben.

Das implizite Eulerverfahren verhält sich hier wesentlich günstiger: Die Iterationsvorschrift  $y_{i+1} = y_i + \lambda h y_{i+1}$  kann zu

$$y_{i+1} = \frac{1}{1 - \lambda h} y_i, \quad i = 0, 1, \dots \quad (3.13)$$

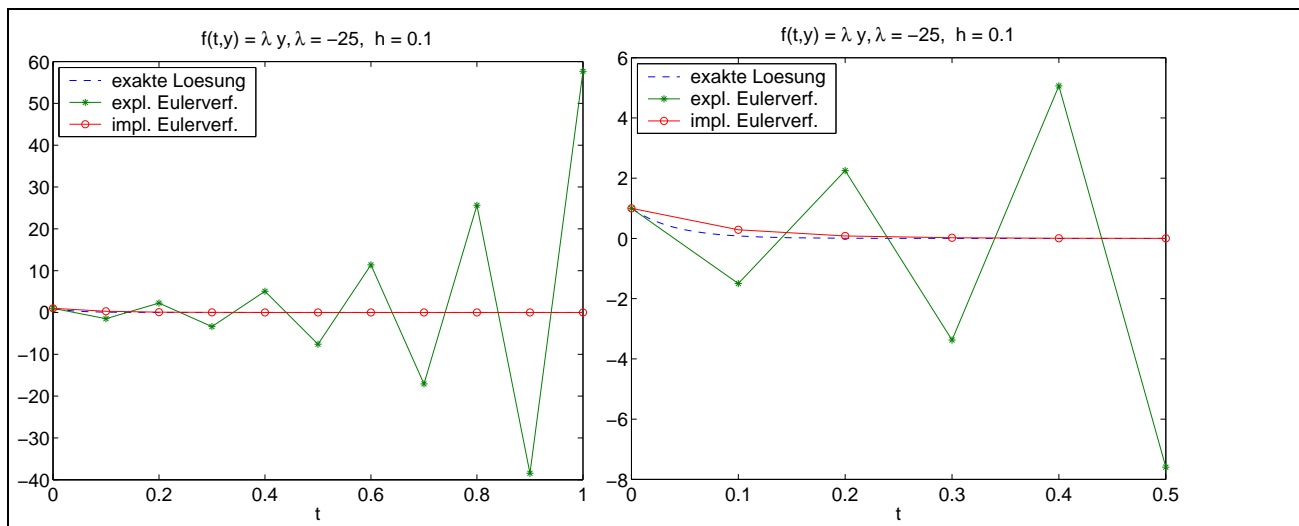
umgeformt werden und liefert damit in geschlossener Form

$$y_i = y_0 \frac{1}{(1 - \lambda h)^i} = y_0 (1 - \lambda h)^{-t_i/h}.$$

Man erkennt  $y \rightarrow 0$  für  $i \rightarrow \infty$ , welches das qualitative Verhalten der Lösung ist (vgl. Fig. 3.1). Auch das implizite Eulerverfahren ist nach Satz 3.7 ein Verfahren erster Ordnung und konvergiert für festes  $t_i$  gegen die korrekte Lösung. ■

**Beispiel 3.10** Wir betrachten das Lösen des Anfangswertproblems

$$y' = Ay, \quad y(0) = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad A = \begin{pmatrix} -21 & 19 & -20 \\ 19 & -21 & 20 \\ 40 & -40 & -40 \end{pmatrix}.$$



**Figur 3.1:** (vgl. Beispiel 3.9) Funktionsplot bei Verwenden des expliziten und des impliziten Eulerverfahrens für  $y'(t) = \lambda y(t)$ ,  $y(0) = 1$ . Links: auf  $[0, 1]$ . Rechts: Ausschnitt auf  $[0, 0.5]$ .

Die Eigenwerte von  $A$  sind gegeben durch  $\lambda_1 = -2$ ,  $\lambda_2 = -40(1 + \mathbf{i})$ ,  $\lambda_3 = -40(1 - \mathbf{i})$ . Die gesuchte Lösung ist

$$\begin{aligned} y_1(t) &= \frac{1}{2}e^{-2t} + \frac{1}{2}e^{-40t}(\cos 40t + \sin 40t), \\ y_2(t) &= \frac{1}{2}e^{-2t} - \frac{1}{2}e^{-40t}(\cos 40t + \sin 40t), \\ y_3(t) &= -e^{-40t}(\cos 40t - \sin 40t). \end{aligned}$$

Im Bereich  $0 \leq t \leq 0.1$  variieren alle drei Lösungskomponenten schnell, so daß die Notwendigkeit einer kleinen Schrittweite  $h \ll 1$  plausibel erscheint. Für  $t > 0.1$  sind die Komponenten  $y_1$  und  $y_2$  nahezu identisch und variieren sehr langsam (der hochoszillatorische Anteil der Lösung ist wegen des Faktors  $e^{-40t}$  weggedämpft worden) und  $y_3$  ist fast Null. Es bietet sich deshalb an, größere Schritte zu machen. Allerdings zeigt es sich in Fig. 3.2, daß z.B. für  $h = 0.05$  das explizite Eulerverfahren völlig inakzeptable Resultate liefert. In der Tat werden wir später sehen, daß das explizite Eulerverfahren die *Stabilitätsbedingung*

$$|1 + hz| \leq 1 \quad z \in \{\lambda_1, \lambda_2, \lambda_3\}$$

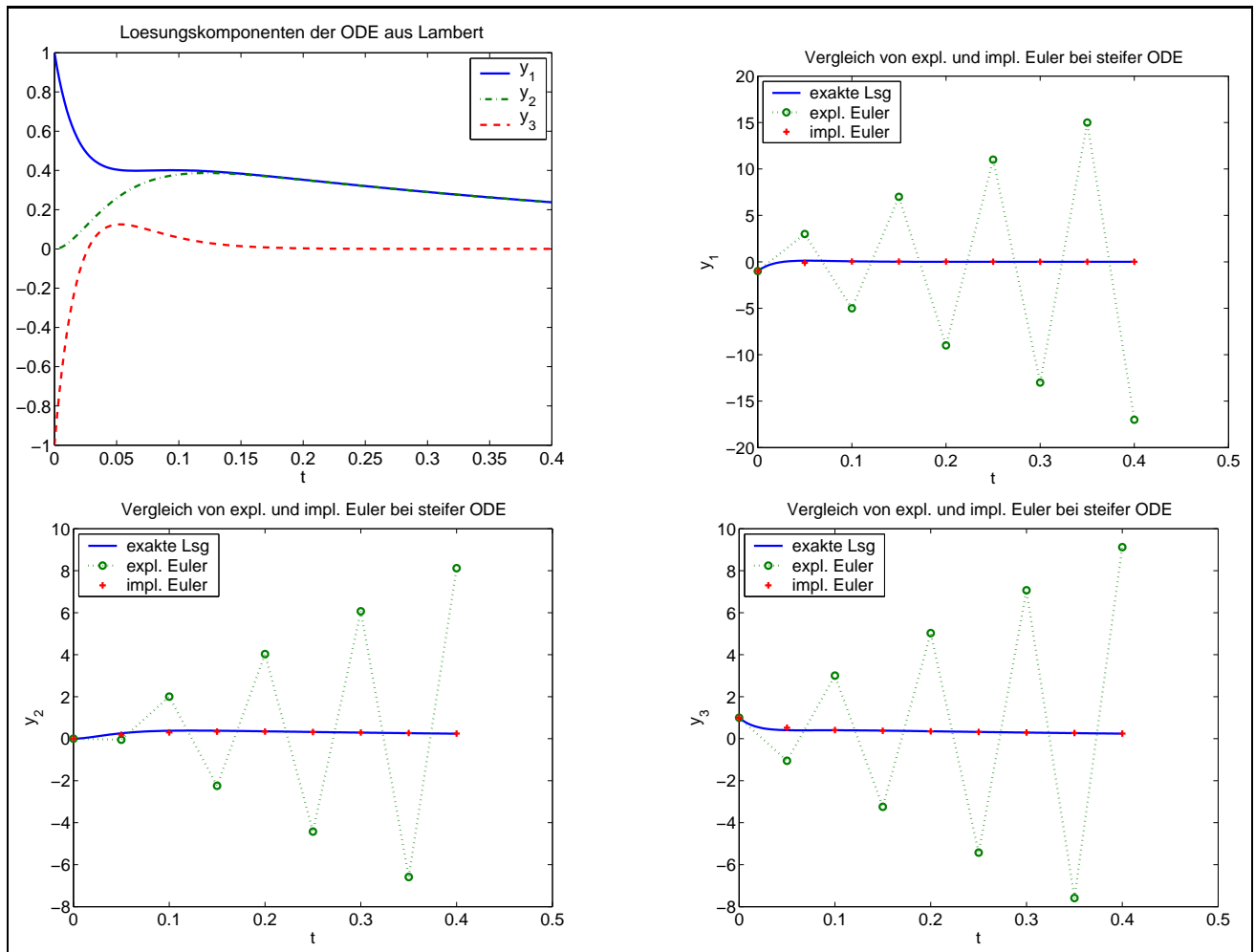
d.h.,  $h \leq \frac{1}{40} = 0.025$  erfüllen muß, damit eine sinnvolle Approximation erwartet werden kann. Das implizite Eulerverfahren, welches wir ebenfalls in Fig. 3.2 sehen, verhält sich hier wesentlich besser. ■

**Beispiel 3.11** Für  $\lambda_1, \lambda_2 < 0$  wird die Lösung des folgenden Systems von Differentialgleichungen gesucht:

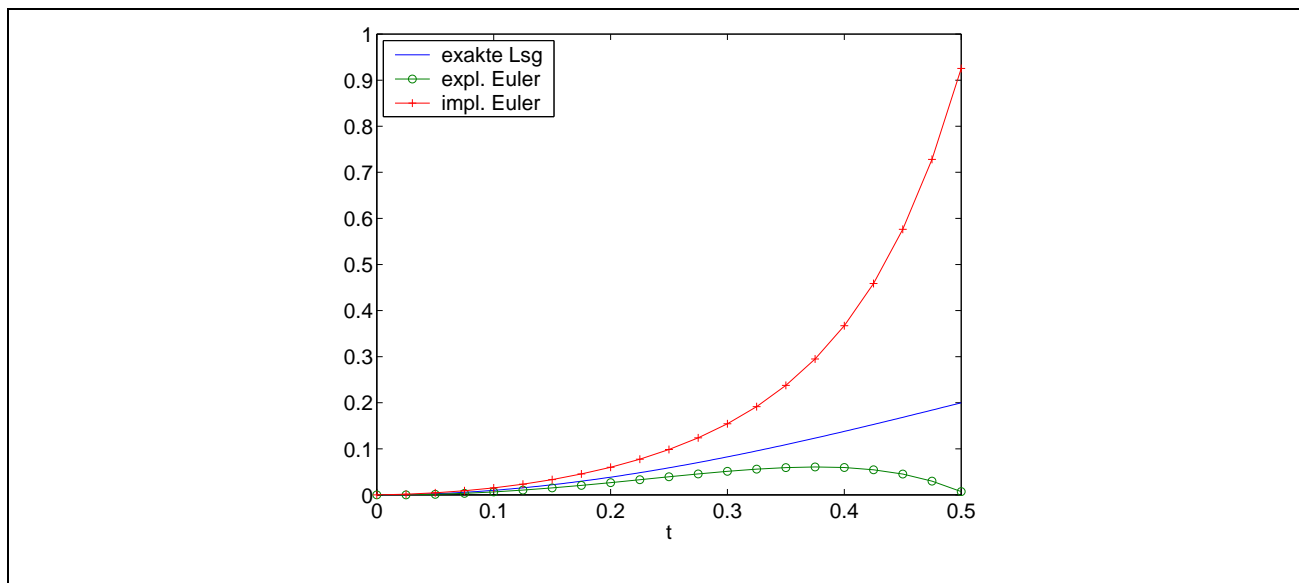
$$\underline{y}'(t) = A\underline{y}(t), \quad A = \begin{pmatrix} \frac{\lambda_1 + \lambda_2}{2} & \frac{\lambda_1 - \lambda_2}{2} \\ \frac{\lambda_1 - \lambda_2}{2} & \frac{\lambda_1 + \lambda_2}{2} \end{pmatrix}.$$

Die allgemeine Lösung  $\underline{y}(t) = (y_1(t), y_2(t))^T$  ist

$$\begin{aligned} y_1(t) &= C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}, \\ y_2(t) &= C_1 e^{\lambda_1 t} - C_2 e^{\lambda_2 t}. \end{aligned}$$



**Figur 3.2:** Vergleich von explizitem und implizitem Eulerverfahren für das steife Problem aus Beispiel 3.10: exakte Lösung(links oben) und numerische Approximationen.



**Figur 3.3:** Vergleich von explizitem und implizitem Eulerverfahren für Beispiel 3.12.

Wendet man das explizite Eulerverfahren mit konstanter Schrittweite  $h$  auf dieses System an, so ergibt sich numerische Approximation

$$\begin{aligned} y_{1,i} &= C_1(1 + \lambda_1 h)^i + C_2(1 + \lambda_2 h)^i = C_1(1 + \lambda_1 h)^{t_i/h} + C_2(1 + \lambda_2 h)^{t_i/h}, \\ y_{2,i} &= C_1(1 + \lambda_1 h)^i - C_2(1 + \lambda_2 h)^i = C_1(1 + \lambda_1 h)^{t_i/h} - C_2(1 + \lambda_2 h)^{t_i/h}. \end{aligned}$$

Wir beobachten die gleiche Schwierigkeit wie in Beispiel 3.9: Im Falle  $C_1, C_2 \neq 0$  muß  $|\lambda_1|h < 2$  und  $|\lambda_2|h < 2$  sein, damit die numerische Lösung wenigstens qualitativ das Verhalten der exakten Lösung widerspiegelt. Ist  $|\lambda_1| \gg |\lambda_2|$ , so ist das explizite Eulerverfahren extrem ineffizient: Obwohl der Lösungsanteil  $C_1 e^{\lambda_1 t}$  zur Gesamtlösung für festes  $t > 0$  fast nicht beiträgt, verlangt das explizite Eulerverfahren, daß die Schrittweite  $h$  die Bedingung  $|\lambda_1|h < 2$  erfüllt. Verhältnisse  $|\lambda_1|/|\lambda_2|$  von der Größenordnung  $10^3$ – $10^6$  treten in der Praxis durchaus auf. Wir beobachten also das “widersinnige” Verhalten, daß die Lösungskomponente  $e^{\lambda_1 t}$ , die eigentlich im Vergleich zur Komponenten  $e^{\lambda_2 t}$  vernachlässigbar ist, vorschreibt, welchen Aufwand man mit dem expliziten Verfahren treiben muß, um eine vernünftige Approximation zu erhalten.

Genau wie in Beispiel 3.9 kann man sehen, daß der Einsatz des impliziten Eulerverfahrens ein Verfahren liefert, daß auch für (relativ) große Schrittweiten  $h$  das qualitativ richtige Verhalten der Lösung widerspiegelt. (Übung: Geben Sie die Approximationen des impliziten Eulerverfahrens in geschlossener Form an. Hinweis: Hierzu müssen die Matrizen  $(\text{Id}_2 - \lambda h A)^{-i}$ ,  $i \in \mathbb{N}$ , mithilfe von Diagonalisierung geschlossen dargestellt werden). ■

**Beispiel 3.12** Wir betrachten noch einmal das Problem aus Beispiel 1.6 mit  $y_0 = 0$ . Das Problem ist von der Bauart  $y' = \lambda y + f(t)$ , wobei  $\lambda > 0$  (vergleichsweise) groß ist. Wir hatten in Beispiel 1.6 gesehen, daß das Problem sehr sensitiv bzgl. Störungen der Anfangsdaten ist. Solche Probleme sind naturgemäß für jedes numerische Verfahren hart, und es ist nicht klar, daß ein implizites Verfahren besser ist. In Fig. 3.3 sehen wir in der Tat qualitativ, daß das implizite Eulerverfahren nicht besser als das explizite Eulerverfahren ist; beide Verfahren liefern für  $h = 0.025$  ungenügende Approximationen. ■

## Diskussion der Beispiele

Das unterschiedliche Verhalten des expliziten und impliziten Eulerverfahrens wird mit dem Begriff der Stabilität näher gefaßt. Die Idee der A-Stabilität (Definition 3.17) von numerischen

Verfahren ist, daß einige *qualitative Eigenschaften* der exakten Lösung auch vom numerischen Verfahren erhalten werden. Hierzu betrachten wir das folgende System

$$y' = Ay, \quad y(0) = y_0, \quad (3.14)$$

wobei  $A \in \mathbb{R}^{n \times n}$ . Wir nehmen nun an, daß  $A$  diagonalisierbar ist. Dann hat die Lösung die Form

$$y(t) = \sum_{i=1}^n c_i e^{\lambda_i t} v^{(i)},$$

wobei die  $\lambda_i \in \mathbb{C}$  die Eigenwerte von  $A$  sind und die  $v^{(i)}$  die zugehörigen Eigenvektoren; die Koeffizienten  $c_i$  sind geeignete reelle Zahlen. Die Lösungskomponenten  $c_i e^{\lambda_i t} v^{(i)}$  haben folgende Verhalten abhängig davon, ob sich  $\lambda_i$  in der linken oder rechten komplexen Halbebene befindet:

- Die Lösungskomponenten deren zugehörige Eigenwerte  $\operatorname{Re} \lambda_i > 0$  sind, wachsend für  $t \rightarrow \infty$ . Diese Lösungskomponenten von  $y$  dominieren die Lösung  $y$  (falls die entsprechenden Koeffizienten  $c_i \neq 0$  sind) und müssen von numerischen Verfahren erfaßt werden. Um diese adäquat numerisch zu approximieren, muß ggf. die Schrittweite klein gewählt werden, d.h.  $|\lambda_i| h$  sollte hinreichend klein sein. Diese Schrittweitenbeschränkung ist unvermeidbar, wenn man die Lösung  $y$  von (3.14) numerisch sinnvoll approximieren will.
- Die Lösungskomponenten, deren zugehörige Eigenwerte  $\operatorname{Re} \lambda_i < 0$  sind, gehen gegen 0 für  $t \rightarrow \infty$ . Ist  $\operatorname{Re} \lambda_i$  sehr stark negativ, dann werden diese Lösungskomponenten sehr schnell weggedämpft. Diese Lösungskomponenten tragen kaum zur Gesamtlösung (zumindest für  $t \rightarrow \infty$ ) bei und brauchen in vielen Anwendungen nicht genau bestimmt werden. Sind sie vom Approximationsstandpunkt vernachlässigbar, so will man keine Schrittweitenbeschränkung hinnehmen, die daher kommt, daß diese Lösungskomponenten aufgelöst werden müssen, obwohl sie fast nichts zur Gesamtlösung beitragen. Beispiele 3.9, 3.10, 3.11 zeigen, daß man bei expliziten Verfahren wie dem Eulerverfahren eine Schrittweitenbeschränkung hinnehmen muß, um vernünftige Lösungen zu erhalten.

Will man die vernachlässigbaren Lösungskomponenten nicht auflösen, so muß man Verfahren verwenden, die doch wenigstens einige qualitative Eigenschaften der Lösungskomponenten widerspiegeln. Eine Minimalforderung für Probleme der Form (3.14) wird sein, daß die numerischen Approximationen zu Lösungskomponenten zu Eigenwerten  $\lambda_i$  mit  $\operatorname{Re} \lambda_i \leq 0$  beschränkt bleiben. Diese Minimalforderung führt dann auf den Begriff der A-Stabilität in Definition 3.17, der sie wenigstens für diagonalisierbare Matrizen  $A$  sicherstellt. Um eine einfachere Theorie zu entwickeln, vereinfachen wir diese Forderung mit Hilfe folgender Übung:

**Übung 3.13** Sei  $A$  eine diagonalisierbare Matrix. Dann ist die Anwendung eines RK-Verfahrens auf (3.14) äquivalent zur Anwendung desselben RK-Verfahrens auf die diagonalisierte Form; in der diagonalisierten Form zerfällt dann das Verfahren auf die *komponentenweise* Anwendung des RK-Verfahrens, d.h. in RK-Verfahren für Probleme der Form

$$\tilde{y}'_i = \lambda_i \tilde{y}_i, \quad i = 1, \dots, n, \quad (3.15)$$

wobei die  $\lambda_i \in \mathbb{C}$  die Eigenwerte von  $A$  sind. ■