3. Eigenschaften von Ein- und Mehrschrittverfahren

3.1 Konsistenz und Konvergenz von Einschrittverfahren

Wir betrachten $y' = f(x, y)$, $y(t_0) = y_0$, im Fall $n = 1$.
Die Funktion f sei hinreichend oft stetig differenzierbar.
Man fragt sich, ob und wie schnell die errechneten Näherungen y_k die Werte $y(t_k)$ der exakten Lösung $y(t)$ approximieren.

3.1.1 Definition: globaler Diskretisierungsfehler

Zur AWA $y' = f(t, y)$, $y(t_0) = y_0$, bezeichnet

$$E_h := \max_{k=0,\ldots,N} \|y(t_k) - y_k\|$$

den globalen Diskretisierungsfehler des ESVs mit der Zuwachsfunktion Φ.

3.1.2 Definition: Konvergenzordnung

Ein Einschrittverfahren mit der Zuwachsfunktion Φ heißt konvergent von der Ordnung p, falls gilt

$$E_h = \max_{k=0,\ldots,N} |y(t_k) - y_k| = O(h^p), \quad h \to 0.$$

Der globale Fehler E_h entsteht durch Akkumulation von lokalen Fehlern an den Stellen $t_0, \ldots, t_{k-1}$. Um das präziser zu beschreiben, braucht man den lokalen Abbruchfehler.

3.1.3 Definition: lokaler Abbruchfehler

Sei $z(t)$ die Lösung der AWA

$$z'(t) = f(t, z), \quad z(t_k) = y_k.$$

Dann heißt

$$\tau_h(t_k) := z(t_k + h) - (z(t_k) + h\Phi(f, t_k, z(t_k), h))$$

der lokale Abbruchfehler im Intervall $[t_k, t_k + h]$.

Ein wesentliches Qualifikationskriterium eines Verfahrens stellt der folgende Begriff der Konsistenzordnung dar, der als Maß für die Größe des lokalen Abbruchfehlers dient.

3.1.4 Definition: Konsistenzordnung

Ein ESV heißt mit der AWA $y' = f(t, y)$, $y(t_0) = y_0$, konsistent von der Ordnung p, falls

$$\frac{\tau_h(s)}{h} = \mathcal{O}(h^p), \quad h \to 0,$$

für alle $p-$mal stetig differenzierbaren Funktionen $f : I \times G \to \mathbb{R}$ und alle $(s, y) \in I \times G$ gilt.

3.1.5 Bemerkung:

(a) Die Konstante in dem $\mathcal{O}-$Term ist unabhängig von s und $y(s)$.

(b) Bei konsistenten Verfahren (d.h. $p = 1$) soll gelten

$$\lim_{h \to 0} \Phi(f, t, y, h) = f(t, y), \quad t \geq t_0.$$

3.1.6 Beispiel: Konsistenzordnung expliziter Verfahren Das Euler-Verfahren hat mindestens die Konsistenzordnung $p = 1$.

Nachweis: Die Differentialgleichung $y' = f(t, y)$ mit $f \in C^1(I \times G)$ sei gegeben. Zu festem $(s, y) \in I \times G$ sei z die Lösung der AWA $z' = f(t, z)$, $z(s) = y(s)$.

Taylorentwicklung von z mit Entwicklungspunkt s ergibt für $h \to 0$

$$\begin{aligned} \frac{z(s+h) - z(s)}{h} &= z'(s) + \mathcal{O}(h) \\ &= f(s, y) + \mathcal{O}(h). \end{aligned}$$

Daraus folgt

$$\frac{\tau_h(s)}{h} = f(s, y) + \mathcal{O}(h) - f(s, y) = \mathcal{O}(h)$$

als $h \to 0$.

Eigentlich ist man an der Abschätzung des globalen Diskretisierungsfehlers interesiert.

3.1.7 Satz:

Sei f wie in 3.1.5. Die Zuwachsfunktion $\Phi(f, t, y, h)$ sei Lipschitz-beschränkt bzgl. y mit Lipschitz-Konstante $L > 0$. Dann gilt

$$\text{Konsistenz der Ordnung } p \quad \Rightarrow \quad \text{Konvergenz der Ordnung } p.$$

Weiterhin gilt

$$E_h \leq e^{L(t_N - t_0)} \left(\|y(t_0) - y_0\| + \sum_{k=0}^{N-1} \|\tau_h(t_k)\| \right).$$

Beweis: J. Stoer, R. Bulrisch, *Einführung in die Numerische Mathematik*, 1983

3.1.8 Bemerkung:

(a) Sicherung einer gewünschten Genauigkeit der numerischen Lösung läßt sich im gesamten Integrationsintervall auf einfache Konsistenzbetrachtung reduzieren.

(b) Bei ESV werden die lokalen Abbruchfehler $\|\tau_h(t_k)\|$ sich schlimmstenfalls aufsummieren. Diese kontrollierte Fehlerfortpflanzung beim ESV gilt nicht nur für $\|\tau_h(t_k)\|$ und für Datenfehler $\|y(t_0) - y_0\|$, sondern auch für die Fehler $\|R_k\|$,

die bei der Auswertung

$$\widetilde{y}_{k+1} = \widetilde{y}_k + h \cdot \Phi(f, t_k, \widetilde{y}_k, h) + R_k$$

von Φ auftretten können. Es gilt

$$\max_{k=0,\ldots,N} \|y(t_k) - \widetilde{y}_k\|_\infty \leq e^{L(t_N - t_0)} \left(\|y(t_0) - y_0\| + \sum_{k=0}^{N-1} (\|\tau_h(t_k)\| + \|R_k\|) \right).$$

ESV sind in diesem Sinne numerisch stabil.

(c) **Satz:**

Die Zuwachsfunktion $\Phi(f, t, y, h)$, $h > 0$, und $f : I \times G \to \mathbb{R}$ sei ($p-$mal) stetig differenzierbar. Dann ist ein ESV mit der Zuwachsfunktion Φ in $I \times G$ konsistent von der Ordnung p genau dann, wenn es in $I \times G$ konvergent von der Ordnung p ist.

Beweis: Vorlesung *Numerik II*

(d)

Verfahren	Konsistenzordnung	Konvergenzordnung
Euler-verfahren	1	1
verb. Polygonzugverf.	2	2
Verf. von Heun 2. Ordnung	2	2
klassisches RK-Verf.	4	4
implizites Euler-Verf.	1	1
Trapezregel	2	2
implizite Mittelpunktsregel	2	2

3.1.9 Bemerkung: Praktische Bedeutung der Konvergenzordnung

Falsch: Man kann eine gewünschte Zielgenauigkeit einer Näherungslösung dadurch realisieren, dass man die Schrittweite h hinreichend klein macht.

Korrekt: Verfahren niedriger Ordnung p sind für Anwendungen nicht geeignet, die bei längeren Zeitintervallen hohe Genauigkeit verlangen. Hier kommen durchaus Verfahren der Ordnung 4 bis 8 zum Einsatz, denn:

Die Abschätzung aus Satz 1.4.8 mit $t_N - t_0 = 1$ und unter der Annahme der $m-$stelligen Genauigkeit besagt, dass für den Fehler E nach $N = h^{-1}$ Schritten

$$E \sim h^{-1}(h^{p+1} + 10^{-m}) = h^p + h^{-1} \cdot 10^{-m}$$

gilt. Die rechte Seite wird minimal für $h = \left(\frac{1}{p}10^{-m}\right)^{\frac{1}{p+1}}$. Setzt man diesen Wert ein, ergibt sich

$$E \sim 10^{-\frac{mp}{p+1}} \qquad (*).$$

Bei $m = 8$ und $p = 1$, besagt $(*)$, dass man bestenfalls eine Genauigkeit von der Ordnung 10^{-4} erreichen kann, da bei kleinerer Schrittweite als 10^{-4} die Rundungsfehler dominieren:

$$E \sim 10^{-4} + \underbrace{10^4 \cdot 10^{-8}}_{\textit{Rundungsfehler}}$$

3.1.10 Beispiel: Die AWA

$$y' = -200ty^2, \quad y(0) = 1,$$

hat die exakte Lösung

$$y(t) = \frac{1}{1 + 100t^2}.$$

Um Näherungen von $y(2) = 1/401$ zu erhalten, verwenden wir das klassische Runge-Kutta-Verfahren der Ordnung 4 und das Verfahren von Heun (Ordnung 2) zu verschiedenen Schrittweiten und berechnen den Fehler an der Stelle $t = 2$.

	Runge-Kutta 4. Ordnung		Heun 2. Ordnung	
h	Auswertungen von f	Fehler	Auswertungen von f	Fehler
10^{-1}	80	$0.21 \cdot 10^{-5}$	40	$0.25 \cdot 10^{-2}$
10^{-2}	800	$-0.20 \cdot 10^{-9}$	400	$-0.65 \cdot 10^{-6}$
10^{-3}	8000	$-0.19 \cdot 10^{-13}$	4000	$-0.62 \cdot 10^{-8}$
10^{-4}	80000	$-0.12 \cdot 10^{-16}$	40000	$-0.62 \cdot 10^{-10}$
10^{-5}	800000	$-0.12 \cdot 10^{-16}$	400000	$-0.62 \cdot 10^{-12}$

Man erkennt die Konvergenzordnungen anhand der globalen Fehler, da die Verkleinerung der Schrittweite um den Faktor $1/10$ jeweils eine Verkleinerung des Fehlers um 10^{-4} (Runge-Kutta) bzw. 10^{-2} (Heun) bewirkt. Typisch tritt dieser Effekt nur bei "mittleren" Schrittweiten auf und wird bei zu großem oder zu kleinem h von anderen Fehlern überdeckt (z.B. Rundungsfehler, Terme höherer Ordnung in h).

3.1.11 Bemerkung: Die Parameter des R-stufigen Runge-Kutta Verfahrens

$$\begin{array}{c|ccccc} 0 & & & & & \\ a_2 & b_{21} & & & & \\ a_3 & b_{31} & b_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ a_R & b_{R1} & b_{R2} & \cdots & b_{R,R-1} & \\ \hline 1 & c_1 & c_2 & \cdots & c_{R-1} & c_R \end{array}$$

sind so zu wählen, dass das Verfahren möglichst hohe Konsistenzordnung hat und Φ Lipschitz-beschränkt bzgl. y ist. Dabei stehen in der ersten Spalte aufgrund der inneren

$$\sum_{j=1}^{k-1} b_{kj} = a_k, \qquad 2 \leq k \leq R,$$

und äußeren Konsistenzbedingung

$$\sum_{k=1}^{R} c_k = 1$$

jeweils die Zeilensummen.

3.1.12 Lemma

Das R-stufige Runge-Kutta Verfahren ist konsistent und es gilt

$$\lim_{h \to 0} K_k(f, t, y, h) = f(t, y), \qquad 1 \leq k \leq R.$$

3.1.13 Bemerkung:

(a) Für alle $R \geq 10$ ist als obere Schranke für die maximal erreichbare Konsistenzordnung $p_R \leq R - 2$ nachgewiesen. Folgende Tabelle gibt die maximal erzielbare Konsistenzordnung (=Konvergenzordnung) p_R für kleine R an:

R	1	2	3	4	5	6	7	8	9
p_R	1	2	3	4	4	5	6	6	7

(b) P. Albrecht (*The Runge-Kutta theory in a nutshell*, s. Bibl. b352/Albr) hat eine Rekursion zur Berechnung der Parameter des R-stufigen Runge-Kutta Verfahrens zur Erzielung der optimalen Konvergenzordnung angegeben.

(c) Die $R(R+1)$ Parameter eines impliziten RK-Verfahrens lassen sich so festlegen, dass die Konsistenzordnung und Konvergenzordnung $2R$ erzielt wird (Butcher, 1963).

3.2 Konsistenz und Konvergenz von Mehrschrittverfahren

Eine zentrale Rolle nimmt wieder der Begriff des *lokalen Abbruchsfehlers* ein.

3.2.1 Definition: Konsistenz, Konsistenzordnung

Das k-Schrittverfahren besitzt *Konsistenzordnung* $p \geq 1$, wenn

$$\left\| \frac{\tau_h(s)}{h} \right\| = \mathcal{O}(h^p) \qquad \text{für} \quad h \to 0.$$

für alle $f \in C^{p+1}(I \times G)$ und alle $(s, y) \in I \times G$ gilt.

Bemerkung: Die Konsistenzordnung von k-Schrittverfahren wird wie in 3.1 durch Taylor-Entwicklung von $z(s + k \cdot h)$ um s bestimmt. Es gilt
$\tau_h(s) = z(s + k \cdot h) - z_h(s + k \cdot h)$, wobei $z_h(s + k \cdot h)$ der Wert $y_{\ell+k}$, den das $k-$Schrittverfahren bei Verwendung der exakten Werte $y_{\ell+k-j} = z(t_{\ell+k-j})$, $j = 1, \ldots, k$, liefert.

3.2.2 Satz: Konsistenzordnung linearer $k-$Schrittverfahren

Die folgenden Aussagen sind für ein k-Schrittverfahren

$$\sum_{j=0}^{k} a_j y_{\ell+j} = h \sum_{j=0}^{k} b_j f_{\ell+j}, \quad \ell = 0, 1, \ldots, \quad a_k = 1,$$

äquivalent:

(i) Das Verfahren besitzt die Konsistenzordnung $p \in \mathbb{N}$.

(ii) Die Zahlen $d_0, \ldots, d_p$

$$d_0 := \sum_{j=0}^{k} a_j \quad \text{und} \quad d_\beta := \sum_{j=0}^{k} \left(j^\beta a_j - \beta j^{\beta-1} b_j \right), \quad \beta = 1, \ldots, p,$$

sind alle gleich Null.

3.2.3 Folgerung:

Ein $k-$Schrittverfahren besitzt Konsistenzordnung 1 genau dann, wenn für

das *erste charakteristische Polynom* $\rho(z) := \sum_{j=0}^{k} a_j z^j$ und

das *zweite charakteristische Polynom* $\sigma(z) := \sum_{j=0}^{k} b_j z^j$.

$\rho(1) = 0$ und $\rho'(1) = \sigma(1)$ gilt.

3.2.4 Korollar:

- Die expliziten k-Schrittverfahren von Adams-Bashforth ($k \geq 1$) und Nyström ($k \geq 2$) haben die Konsistenzordnung k.
- Die impliziten k-Schritt Verfahren von Adams-Moulton ($k \geq 1$) und Milne-Simpson ($k \geq 2$) haben die Konsistenzordnung $k + 1$.
- Die impliziten k-Schritt BDF-Verfahren haben die Konsistenzordnung k.

Die angegebene Ordnung ist jeweils exakt, mit Ausnahme des Milne-Simpson Verfahrens mit $k = 2$, das die Konsistenzordnung 4 besitzt.

3.2.5 Beispiel: Die allgemeine (implizite) lineare 2-Schrittmethode hat 6 Koeffizienten a_j, b_j, $0 \le j \le 2$, mit $a_2 = 1$. Um die Konsistenzordnung 3 zu erzielen, muss $d_0 = d_1 = d_2 = d_3 = 0$ gelten. Dies wird mit $a := a_0$ erfüllt, falls wir

$$a_1 = -1 - a,\ b_0 = -\frac{1}{12}(1+5a),\ b_1 = \frac{2}{3}(1-a),\ b_2 = \frac{1}{12}(5+a)$$

setzen. Das allgemeine implizite 2-Schrittverfahren der Ordnung 3 hat also die Form

$$y_{\ell+2} = (1+a)y_{\ell+1} - ay_\ell + \frac{h}{12}\Big((5+a)f_{\ell+2} + 8(1-a)f_{\ell+1} - (1+5a)f_\ell\Big).$$

Es folgt:

(i) Für $a = -1$ ergibt sich das Simpson-Milne Verfahren

$$y_{\ell+2} = y_\ell \frac{h}{3}(f_{\ell+2} + 4f_{\ell+1} + f_\ell)$$

der Konsistenzordnung 4. Ihre Fehlerkonstante ist $d_5 = -1/90$.

(ii) Für $a \neq -1$ ist die exakte Konsistenzordnung 3. Z.B. ergibt $a = 0$ das implizite 2-Schrittverfahren von Adams-Moulton und $a = -5$ ergibt das explizite Verfahren

$$y_{\ell+2} = 4y_{\ell+1} + 5y_\ell + h(4f_{\ell+1} + 2f_\ell).$$

Dieses Verfahren hat eine attraktive Form, da pro Schritt nur eine Auswertung von f erfolgt. Es ist jedoch praktisch unbrauchbar, da eine wichtige Eigenschaft der "Null-Stabilität" fehlt, die wir im Folgenden analysieren wollen. Kleine Rundungsfehler der Anfangswerte y_0, y_1 werden in jedem Schritt so verstärkt, dass nach wenigen Schritten vollkommen unbrauchbare Werte y_ℓ vorliegen (siehe Beispiel 3.2.6).

Das numerische Verhalten von $k-$Schrittverfahren unter dem Einfluss von Rundungsfehlern wird durch die Analyse der linearen Differenzengleichungen

$$\sum_{j=0}^{k} a_j y_{\ell+j} = 0$$

vollständig erfasst.

3.2.6 Beispiel: Die (homogene) lineare Differenzengleichung

$$y_{\ell+2} + 4y_{\ell+1} - 5y_\ell = 0$$

mit den Anfangswerten $y_0 = 0$ und $y_1 = 0$ hat die triviale Lösung $y_\ell = 0$ für alle $\ell \in \mathbb{N}_0$. Eine kleine Störung der Anfangswerte $y_0 = \delta$, $y_1 = \epsilon$ liefert hingegen

$$y_2 = 5\delta - 4\epsilon, \quad y_3 = -20\delta + 21\epsilon, \quad y_4 = 105\delta - 104\epsilon, \quad \ldots$$

und allgemein

$$y_\ell = A + B(-5)^\ell \quad \text{mit} \quad B = \frac{1}{6}(\delta - \epsilon), \quad A = \frac{1}{6}(5\delta + \epsilon).$$

Diese Darstellung erhält man durch Berechnen der Nullstellen -5 und 1 des *charakteristischen Polynoms* $\rho(z) = z^2 + 4z - 5 = (z+5)(z-1)$ der Differenzengleichung. Sie zeigt, dass das Differenzenverfahren *numerisch instabil* ist, da Rundungsfehler sich sehr schnell verstärken.

Im Beispiel 3.2.6 sieht man, dass kleine Störungen der Anfangswerte "explodieren", wenn die Differenzengleichung unbeschränkte Fundamentallösungen besitzt.

3.2.7 Satz: Fundamentallösungen homogener linearen Differenzengleichung

Gegeben sei die *homogene lineare Differenzengleichung*

$$\sum_{j=0}^{k} a_j y_{\ell+j} = 0, \quad \ell = 0, 1, \ldots \qquad (a_k = 1). \tag{$*$}$$

Die *linear unabhängigen Fundamentallösungen* $\left(y_\ell^\mu\right)_{\ell\geq 0}$, $1 \leq \mu \leq k$, sind:

- Ist $\lambda \in \mathbb{C}$ eine s-fache Nullstelle von dem ersten charakteristischen Polynom ρ, so setze

$$\begin{aligned}
(y_\ell^{(\mu_1)})_{\ell\geq 0} &= (\lambda^k)_{\ell\geq 0}, \\
(y_\ell^{(\mu_2)})_{\ell\geq 0} &= (\ell\lambda^{\ell-1})_{\ell\geq 0}, \\
&\vdots \\
(y_\ell^{(\mu_s)})_{\ell\geq 0} &= \left(\frac{\ell!}{(\ell-s+1)!}\lambda^{\ell-s+1}\right)_{\ell\geq 0}, \quad s \in \mathbb{N}.
\end{aligned}$$

3.2.8 Folgerung: Die allgemeine Lösung von $(*)$ ist gegeben durch

$$(y_\ell)_{\ell\geq 0} = \sum_{\mu=1}^{m} C_\mu \cdot \left(y_\ell^\mu\right)_{\ell\geq 0}, \quad C_\mu \in \mathbb{C}.$$

Die Koeffizienten C_μ sind durch die Angabe von $y_0, \ldots, y_{k-1}$ eindeutig bestimmt.

Um die Konvergenzordnung von einem $k-$Schrittverfahren zu definiern, brauchen wir einen entsprechenden Stabilitätsbegriff.

3.2.9 Definition: Null-Stabilität

Das lineare $k-$Schrittverfahren heißt *Null-stabil*, wenn alle Fundamentallösungen der homogenen Differenzengleichung $\sum_{j=0}^{k} a_j y_{\ell+j} = 0$, $\ell \geq 0$, beschränkte Folgen sind.

Es ergibt sich die folgende Charakterisierung der Null-Stabilität.

3.2.10 Satz: Wurzelbedingung

Das $k-$Schrittverfahren ist genau dann Null-stabil, wenn keine Nullstelle des ersten charakteristischen Polynoms ρ einen Betrag größer als 1 hat und wenn alle Nullstellen vom Betrag 1 einfach sind.

Für die bereits vorgestellten Verfahren erhalten wir:

3.2.11 Korollar:

Die Adams-Bashforth, Adams-Moulton, Nyström und Milne-Simpson Verfahren sind Null-stabil. Die BDF-Verfahren sind Null-stabil nur für $1 \leq k \leq 6$.

Der Konvergenzbegriff wird wie bei den Einschrittverfahren in 3.1 definiert.

3.2.12 Konvergenzsatz von Dahlquist

Ein lineares k-Schritt Verfahren ist genau dann konvergent (hat Konvergenzordnung 1), wenn es konsistent (hat Konsistenzordnung 1) und Null-stabil ist.

3.2.13 Korollar

- Die expliziten k-Schritt Adams-Bashforth ($k \geq 1$) und Nyström-Verfahren ($k \geq 2$) haben die Konvergenzordnung k.
- Die impliziten Adams-Moulton ($k \geq 1$) und Milne-Simpson-Verfahren ($k \geq 3$) haben die Konvergenzordnung $k+1$, das Milne-Simpson-Verfahren mit $k = 2$ sogar die Konvergenzordnung 4.
- Die impliziten BDF-Verfahren mit $k \leq 6$ haben die Konvergenzordnung k.

3.2.14 Bemerkung: Dahlquist (Convergene and stability in the numerical integration of ordinary differential equations, Math. Scand. 4 (1956) 33-53) hat gezeigt, dass die maximal erreichbare Konsistenzordnung von Null-stabilen m-Schritt-Verfahren gegeben ist durch $k+1$, falls k ungerade, und $k+2$, falls m gerade ist. Demnach sind die impliziten Verfahren von Adams-Moulton und Milne-Simpson mit ungeradem k optimal. Weiterhin ist die Simpson Regel ($k = 2$) optimal.

Durch geeignete Wahl zweier Verfahren (P) und (C) passender Konvergenzordnung kann die Anzahl N der Iterationsschritte (und der Auswertungen von f) gering gehalten werden. Die Begründung im letzten Abschnitt ergibt folgendes Resultat.

Satz: Konvergenzordnung des Prädiktor-Korrektor-Verfahrens

Mit $p^{(P)}$ und $p^{(C)}$ bezeichnen wir die Konvergenzordnungen des Prädiktors bzw. Korrektors. Dann ist die Konvergenzordnung p des Prädiktor-Korrektor-Verfahrens in der $P(EC)^M$ bzw. $P(EC)^M E$-Form mindestens

$$p = \min\{p^{(C)}, p^{(P)} + M\}.$$

In der Praxis verwendet man häufig die Adams-Bashforth (P) und Adams-Moulton Verfahren (C) der gleichen Konvergenzordnung und führt nur $M = 1$ Iterationsschritt durch.

Bemerkung:

Scheinbar sind die Prädiktor-Korrektor-Verfahren (mit $M = 1$) den expliziten Einschrittverfahren überlegen, da pro Schritt zur Schrittweite h mit nur 2 Auswertungen von f eine "beliebig" hohe Konvergenzordnung erzielt werden kann. Die genaue Fehleranalyse zeigt jedoch, dass zur Erzielung gleicher Genauigkeit die Mehrschrittverfahren mit deutlich kleinerer Schrittweite operieren müssen als z.B. das klassische Runge-Kutta-Verfahren der Ordnung 4.

3.3 Automatische Schrittweitenkontrolle

Große Schrittweiten h haben den Vorteil, dass man mit wenigen Schritten Näherungen der Lösung $y(t)$ an Stellen bekommt, die weit vom Anfangspunkt t_0 entfernt sind. Andererseits führt zu großes h auf zu große (lokale und globale) Diskretisierungsfehler.

3.3.1 Beispiel: Die exakte Lösung der AWA $y' = -10y$, $y(0) = 1$, ist $y(t) = e^{-10t}$.

(a) Das explizite Euler-Verfahren zur Schrittweite h liefert

$$y_{k+1} = y_k + h \cdot f(t_k, y_k) = (1 - 10h)y_k \quad \Rightarrow y_k = (1 - 10h)^k y_0.$$

Für $h > 1/5$ ist y_k unbeschränkt, im Gegensatz zur exakten Lösung y.

(b) Das implizite Euler-Verfahren zur Schrittweite h liefert

$$y_{k+1} = y_k + h \cdot f(t_{k+1}, y_{k+1}) = y_k - 10hy_{k+1}.$$

Wir lösen diese lineare Gleichung nach y_{k+1} auf und erhalten

$$y_{k+1} = \frac{1}{1 + 10h}\, y_k. \qquad (*)$$

Also ist

$$y_k = (1 + 10h)^{-k} y_0.$$

Für beliebige $h > 0$ (insbesondere für große Schrittweiten h) zeigt y_k dasselbe asymptotische Verhalten $\lim_{k\to\infty} y_k = 0$ wie die exakte Lösung y.

Explizites Euler-Verfahren liefert für $h > 1/5$ unbrauchbare Werte y_k; implizites Euler-Verfahren liefert für beliebige $h > 0$ ein *stabiles Verfahren* $(*)$.

In der Praxis wird man durch Anpassen der Schrittweite h in jedem Iterationsschritt versuchen, die Anzahl der Schritte zu kontrollieren. Dann hat das Einschrittverfahren zur Zuwachsfunktion Φ die Form

> Für $k = 0, 1, \ldots$
> bestimme die Schrittweite h_k und setze
> $t_{k+1} := t_k + h_k, \quad y_{k+1} := y_k + h_k \Phi(f, t_k, y_k, h_k).$

Wahl von h_k: Um für ein gegebenes Integrationsintervall $[t_0, T]$ die Zielgenaugkeit $TOL_E := (T - t_0) \cdot \epsilon$ (oft mit $\epsilon = 10^{-6}$) für den Fehler $\|y(T) - y_h(T)\|$ mit wenigen Schritten zu erreichen, d.h. soll gelten

$$\|y(T) - y_h(T)\| \leq \sum_{k=0}^{N-1} \|\tau_h(t_k)\| \leq TOL_E.$$

Die obige Ungleichung ist erfüllt, falls $\|\tau_h(t_k)\| \leq (t_{k+1} - t_k) \cdot \epsilon$ gilt, da

$$\sum_{k=0}^{N-1} \|\tau_h(t_k)\| \leq \sum_{k=0}^{N-1} (t_{k+1} - t_k) \cdot \epsilon = (T - t_0) \cdot \epsilon.$$

3.3.2 Kriterium zur Schrittweitensteuerung: Ist $h = t_{k+1} - t_k$, so darf der lokale Fehler $\|\tau_h(t_k)\|$ im Schritt $k \to k + 1$ höchstens $h \cdot \epsilon$ sein. Anderseits sollte diese Spannweite möglichst ausgeschöpft sein, d.h. wähle $h_k > h$.

Zur Schrittweitensteuerung kann man entweder die Methode der Schrittweitenhalbierung oder die Methode der eingebetteten Einschrittverfahren verwenden.

3.3.3 Methode der eingebetteten Einschrittverfahren: Wir verwenden zwei Einschrittverfahren mit Zuwachsfunktionen Φ_A und Φ_B und Konsistenzordnungen $1 \leq p_A < p_B$. Die Verfahren werden so ausgewählt, dass die Berechnungen von Φ_B nur noch wenige zusätzliche Auswertungen von f erfordern.
Bezeichne H die vorgesehene Schätzschrittweite (etwa $H = h_{k-1}$). Dann wird der Fehler $r_{A,H}(t_k)$ geschätzt durch

$$r_{A,H}(t_k) \approx \rho := \Phi_B(f, t_k, y_k, H) - \Phi_A(f, t_k, y_k, H).$$

Unter Beachtung der Konsistenzordnung p_A des eingebetteten Verfahrens Φ_A verfährt man wie folgt:

1. Falls $\rho > \epsilon$ gilt, halbiere H und berechne damit den neuen Schätzwert für den Diskretisierungsfehler. (Beende die Rechnung, falls $H < h_{\min}$!. In diesem Fall vermutet man, dass die Lösung eine Singularität aufweist.)
2. Falls in ν (etwa $\nu = 3$) aufeinanderfolgenden Schritten des Einschrittverfahrens $\rho < TOL/2^{p_A}$ gilt, verdopple H und berechne damit den neuen Schätzwert für den Diskretisierungsfehler.
3. Im Bereich $TOL/2^{p_A} \leq \rho \leq TOL$ wird $t_{k+1} := t_k + H$, $y_{k+1} := y_k + H\Phi_A(f, t_k, y_k, H)$ akzeptiert.

3.3.4. Bemerkung: Beispiele von passenden Verfahren sind RK-Verfahren 2. und 3. Ordnung; das Verfahren höherer Ordnung benötigt nur eine weitere Auswertung von f.

a)

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

(verb. Euler)

0			
$\frac{1}{2}$	$\frac{1}{2}$		
1	-1	2	
1	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

(Kutta 3. Ordnung)

b)

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

(Heun 2. Ordnung)

0			
1	1		
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{3}$

(weiteres Kutta 3. Ordnung)

c) Weitere eingebettete Verfahren 4. Ordnung in Verfahren 5. Ordnung wurden von England (1969) und Fehlberg (1964–1970) vorgeschlagen. Siehe H. R. Schwarz, *Numerische Mathematik* (Bibl. Sign. b350/Schw).

3.3.5 Algorithmus: Verbessertes Euler-Verfahren Φ_A der Ordnung 2, Einbettung in Runge-Kutta-Verfahren 3. Ordnung Φ_B:

```
function  [x,y]=schrittweite_euler(x0,y0,xend,h0,hmin,TOL,f)
% Loesung der AWA  y'=f(x,y),  y(x0)=y0  im Intervall x0<=x<=xend
% mit automatischer Schrittweitenkontrolle
xk=x0;   yk=y0;    h=h0;    x=[x0];    y=[y0];
nu=0;  % Positiv-Zaehler zur Verkleinerung der Schrittweite
while(xk<xend)
    h=min(h,xend-xk);
    k1=feval(f,xk);                    % verb. Euler und
    k2=feval(f,xk+h/2,yk+h/2*k1);      % Kutta-Verf. 3. Ordnung
    k3=feval(f,xk+h,yk+h*(-k1+2*k2));
    phiA=k2;
    phiB=(k1+4*k2+k3)/6;
    rho=phiB-phiA;
    if (abs(rho)>TOL)                  % Schrittweite zu gross
        h=h/2;   nu=0;
        if (h<hmin)
            error('Schrittweite wird zu klein')
        end
    elseif (nu<3)|(abs(rho)>(TOL/4)) % verwende Schrittweite h
        xk=xk+h;   yk=yk+h*phiA;    x=[x,xk];   y=[y,yk];
        if abs(rho)>(TOL/4)  % Normalfall, gute Schrittweite
            nu=0;
        else
            nu=nu+1;  % evtl. Schrittweite demnaechst verdoppeln
        end
    else               % rho<TOL/4 ist zum 3. Mal erfuellt
        h=2*h;         % verdopple Schrittweite
        nu=0;
    end
end
```

3.3.6 Beispiel: AWA: $y' = (3-t)y$, $y(0) = 1$, mit $y(t) = e^{3t-t^2/2}$.
Was ergibt die Schrittweitenkontrolle?

- Zum Wert $\mathrm{TOL} = 10^{-4}$ benötigt das Verfahren auf $[0, 100]$ genau 4899 Schritte. Zu Beginn stellen sich Schrittweiten $h = 2^{-8} \approx 0.004$ und $2^{-9} \approx 0.002$ ein, um den "Buckel" der Lösung mit der gewünschten Genauigkeit zu berechnen.
- Der globale Fehler wächst auf ca. 0.0019 bei $x = 3$ und nimmt dann wieder ab.
- Ab ca. $x = 6.4$ wird h mehrmals verdoppelt bis zur Maximalschrittweite $h = 0.5$ bei $x = 9.5$. Diese große Schrittweite tritt nur sporadisch auf und wird umgehend wieder verkleinert: bei $x = 100$ ist man wieder bei $h = 1/64$.

Ersetze nun Φ_A und Φ_B durch Verfahren aus 1.3.15 (a) bzw. aus 1.3.15 (b).

- Zum Wert $\mathrm{TOL} = 10^{-4}$ benötigt das Verfahren auf dem Intervall $[0, 100]$ genau 2039 Schritte. Zu Beginn stellen sich auch Schrittweiten $h = 2^{-8} \approx 0.004$ und $2^{-9} \approx 0.002$ ein, um den "Buckel" der Lösung mit der gewünschten Genauigkeit zu berechnen. Dies liegt daran, dass hier wie oben $p_A = 2$ gilt.
- Der globale Fehler wächst auf ca. 0.0018 bei $x = 3$ und nimmt dann wieder ab.
- Ab ca. $x = 6.2$ wird h mehrmals verdoppelt. Die Schrittweite $h = 0.5$ wird erstmals bei $x = 9.3$ angenommen.
- Die Schätzungen des lokalen Diskretisierungsfehlers bleiben nun deutlich unter der Toleranz und könnten in jedem Schritt weiter verdoppelt werden. Nur der Zähler ν verhindert dies. Es kommt zur Verdopplung von h in jedem 3. Schritt bis $h = 16$, der globale Fehler fällt unter 10^{-9}.
- Rechnet man bis $x = 1000$ weiter, braucht man nur 10 Schritte: es entstehen Schrittweiten von $h = 256$ bei weiterhin hoher Genauigkeit.

3.3.7 Beispiel (MATLAB-Funktion *ode*45): Betrachte die Dgl. zweiter Ordnung

$$\begin{aligned} \phi''(t) &= 8(1-\phi^2(t))\phi'(t) - \phi(t), \quad t \in [0,30], \\ \phi(0) &= 2, \quad \phi'(0) = 0. \end{aligned}$$

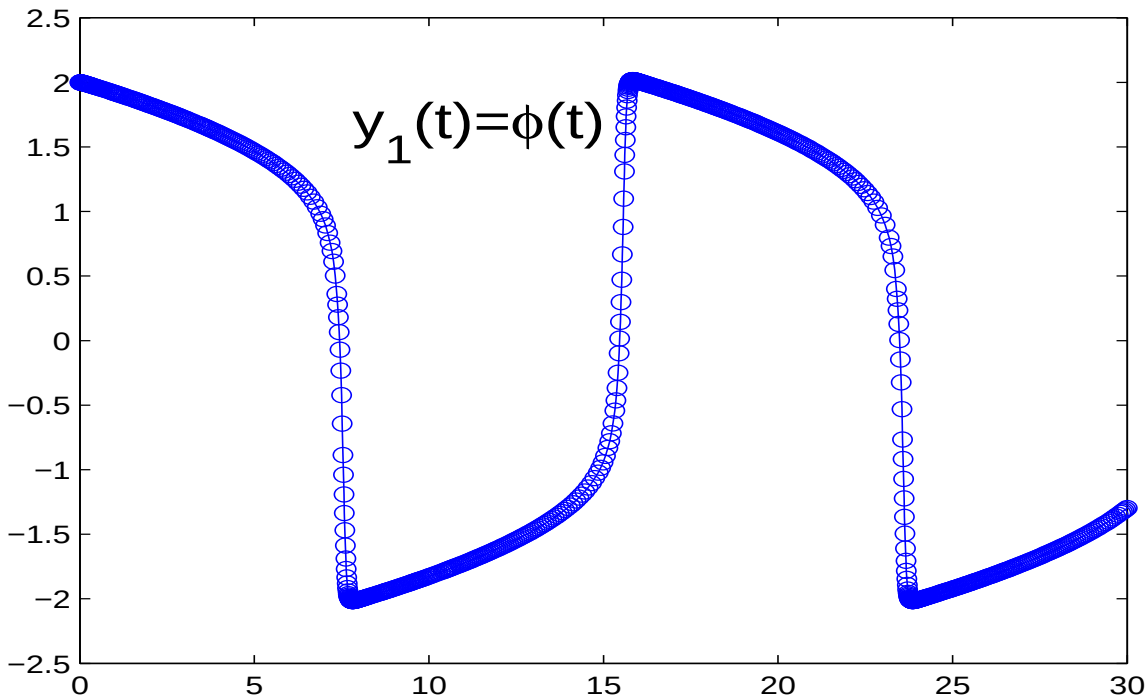
Diese Dgl. läßt sich schreiben als System

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \end{bmatrix} = \begin{bmatrix} y_2(t) \\ 8(1-y_1^2(t))y_2(t) - y_1(t) \end{bmatrix}$$

erster Ordnung mit dem Anfansgwert

$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Verwendet man die eingebettete Verfahren aus 1.6.5 (c), bekommt man die Näherungen y_k:



3.3.8 Beispiel (MATLAB-Funktion *ode15s*): Betrachte die Dgl. zweiter Ordnung

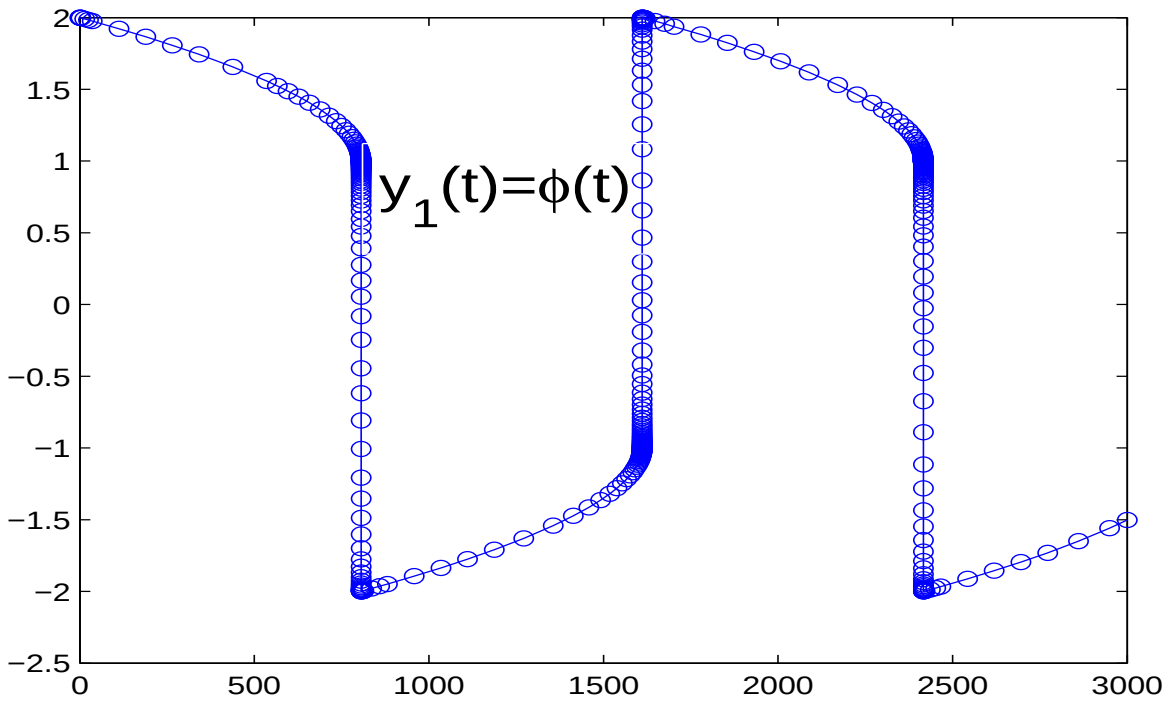
$$\begin{aligned} \phi''(t) &= 1000(1-\phi^2(t))\phi'(t) - \phi(t), \quad t \in [0, 3000], \\ \phi(0) &= 2, \quad \phi'(0) = 0. \end{aligned}$$

Diese Dgl. läßt sich schreiben als System

$$\begin{bmatrix} y_1'(t) \\ y_2'(t) \end{bmatrix} = \begin{bmatrix} y_2(t) \\ 1000(1-y_1^2(t))y_2(t) - y_1(t) \end{bmatrix}$$

erster Ordnung mit dem Anfansgwert

$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$



3.4 Steife Systeme

3.4.1 Definition: steife AWA

Eine AWA

$$y'(t) = \begin{pmatrix} y_1'(t) \\ \vdots \\ y_n'(t) \end{pmatrix} = f(t,y) = \begin{pmatrix} f_1(t, y_1, \ldots, y_n) \\ \vdots \\ f_n(t, y_1, \ldots, y_n) \end{pmatrix}, \; y(t_0) = y_0 \in \mathbb{R}^n,$$

heißt steif, falls alle Komponeneten $y_1, \ldots, y_n$ der Lösung $y(t)$ für wachsendes t abklingen, dies jedoch mit sehr unterschiedlicher Geschwindigkeit.

3.4.2 Definition: Jacobi-Matrix einer AWA

Die $n \times n$ Funktionalmatrix

$$\frac{\partial f}{\partial y}(t,y) := \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & & \\ \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{pmatrix}, \quad t \geq t_0,$$

heißt Jacobi–Matrix der entsprechenden AWA.

3.4.3 Definition: Steifigkeitsrate

Der Quotient

$$\kappa(t) := \frac{\max\{|\text{Re } \lambda(t)| \ : \ \lambda(t) \text{ Eigenwert von } \frac{\partial f}{\partial y} \text{ mit Re } \lambda(t) < 0\}}{\min\{|\text{Re } \lambda(t)| \ : \ \lambda(t) \text{ Eigenwert von } \frac{\partial f}{\partial y} \text{ mit Re } \lambda(t) < 0\}}$$

heißt die Steifigkeitrate der AWA an der Stelle $t \geq t_0$.

3.4.4 Bemerkung: Für steife AWA gilt $\kappa \gg 1$.

3.4.5 Beispiele:

(a) Die Linien Methode für das Wärmeleitungsproblem in 1.1.1 ergibt das System der Dgl. ersster Ordnung

$$y'(t) = A \cdot y(t) \quad \text{mit} \quad A = \frac{\partial f}{\partial y} = \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}$$

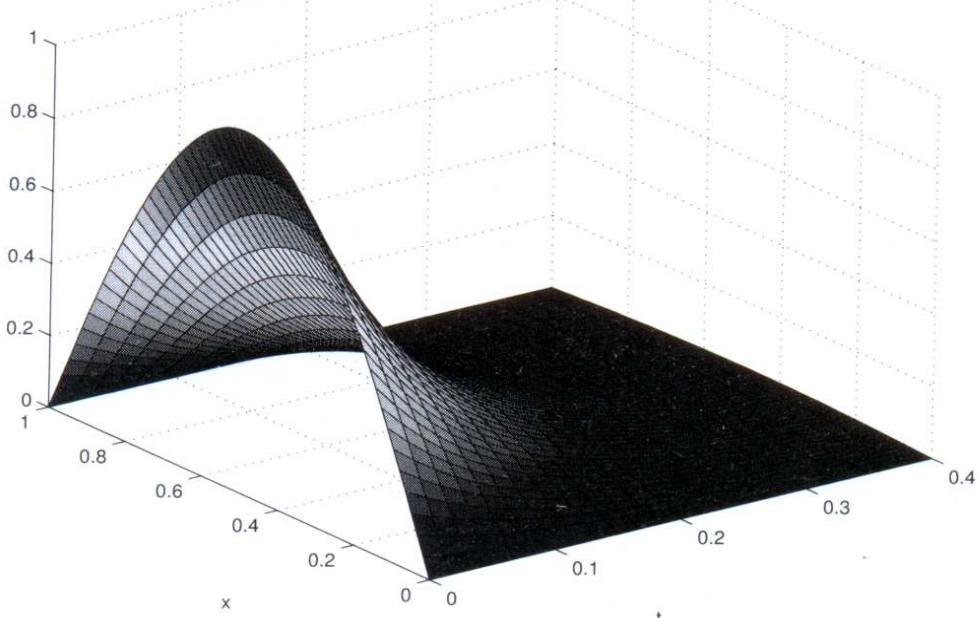
Die Eigenwerte λ_j von A sind

$$\lambda_j = -\frac{4k}{h_x^2} \sin^2\left(\frac{j\pi}{2n_x}\right), \quad j = 1, \ldots, n_x - 1.$$

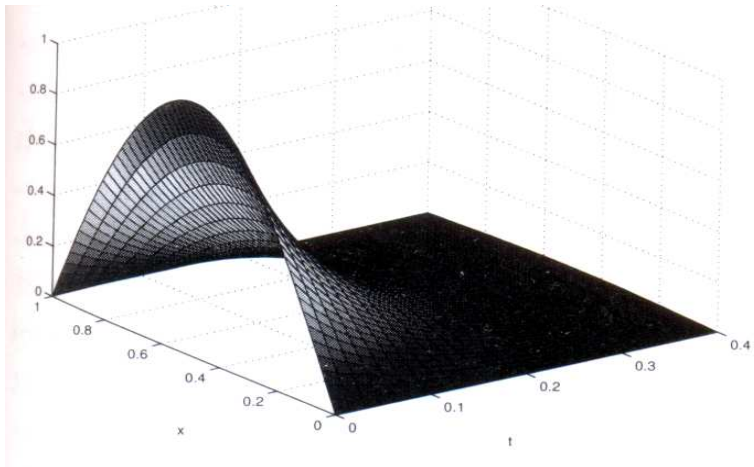
Daraus folgt

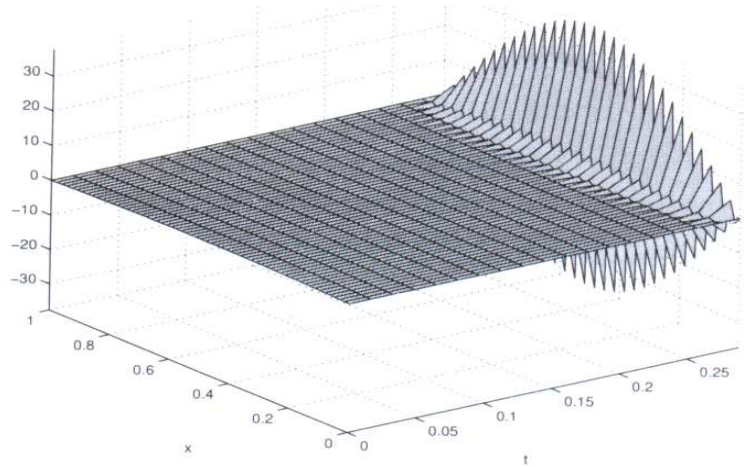
$$\kappa(t) = \frac{|\lambda_{n_x-1}|}{|\lambda_1|} = \frac{\sin^2\left(\frac{\pi}{2} - \frac{\pi}{2n_x}\right)}{\sin^2\left(\frac{\pi}{2n_x}\right)} \leq \frac{4}{\pi^2} n_x^2 \quad \text{und} \quad \kappa \gg 1 \quad \text{für} \quad n_x \gg 1.$$

Betrachte das Wärmeleitungsproblem mit $k = 1$, $\ell = 1$, $f(x) = \sin(\pi x)$ und $h_x = 60^{-1}$. Für die Eigenwerte von A gilt $\lambda_1 = -9.87$ und $\lambda_{n_x-1} = -14390$.



Numerische Lösung $T(j/60, t)$, $j = 1, \ldots, 24$, mit dem impliziten A-stabilien 2-BDF-Verfahren, $h = h_x = 1/60$.





Numerische Lösung mit dem verbesserten Euler-Verfahren, $h = 1.38 \cdot 10^{-4}$ (links) und $h = 1.4 \cdot 10^{-4}$ (rechts).

(b) Die chemische Reaktionsprozesse, bei denen die Reaktionsgeschwindigkeitskonstanten k_j stark unterschiedliche Größenordnungen haben führen auf ein steifes System von Dgl. In Beispiel 1.1.1 mit

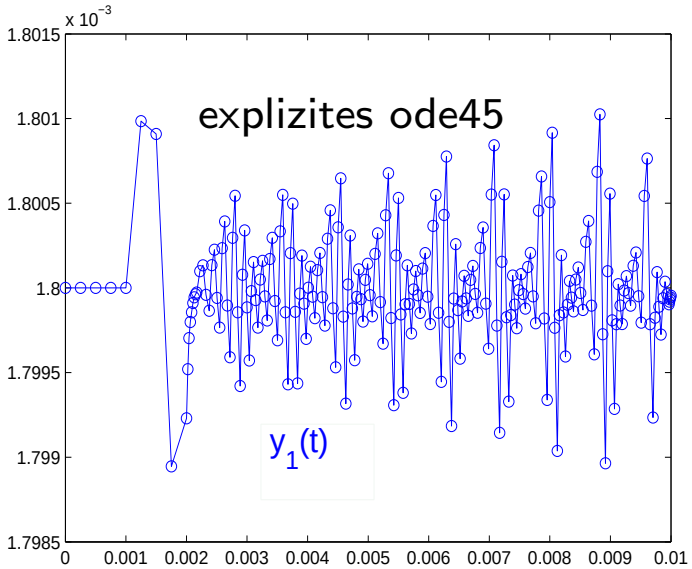
$$k_1 = 7.9 \cdot 10^{-10},\ k_2 = 1.1 \cdot 10^{9},\ k_3 = 1.1 \cdot 10^{7} \text{ und } k_4 = 1.1 \cdot 10^{3}$$

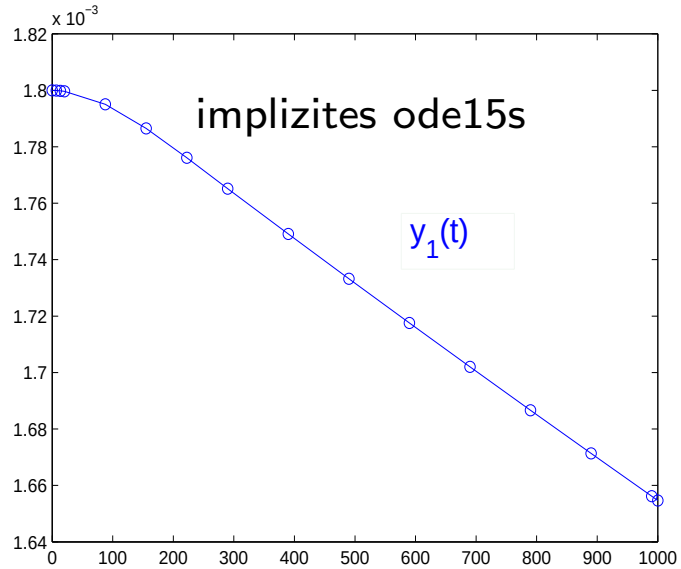
hat die Jacobi-Matrix

$$\frac{\partial f}{\partial y}(0, y) = \begin{pmatrix} -k_1 & 0 & -k_3 y_1(0) & 0 & 0 & 0 \\ k_1 & 0 & 0 & 0 & 0 & 0 \\ k_1 & 0 & -k_3 y_1(0) & k_4 & 0 & 0 \\ 0 & 0 & k_3 y_1(0) & -k_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_4 & 0 & 0 \end{pmatrix}, \quad y(0) = \begin{pmatrix} 1.8 \cdot 10^{3} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

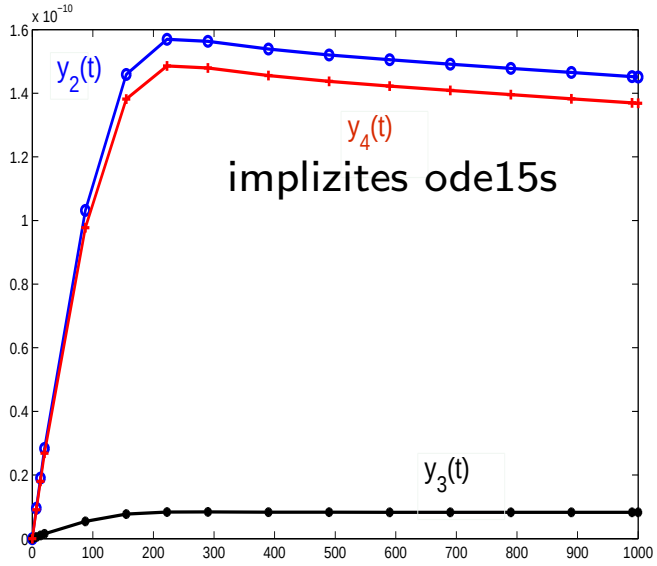
die Eigenwerte

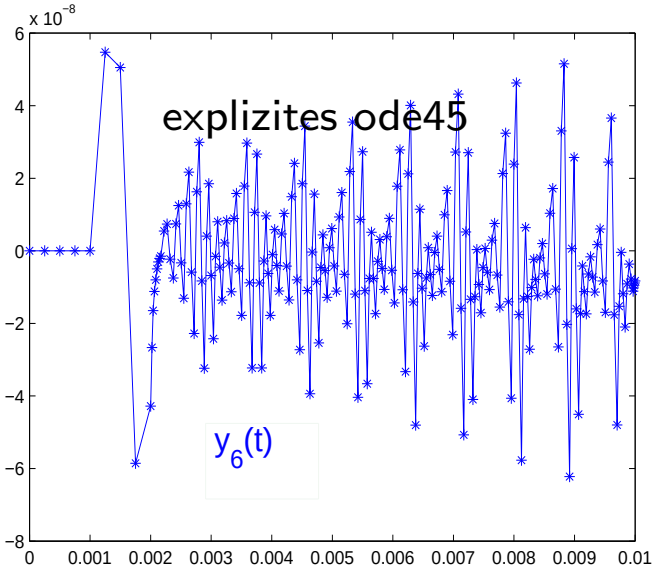
$$\lambda_1(0) = \lambda_2(0) = \lambda_3(0) = 0, \quad \lambda_4(0) = -2.1 \cdot 10^{4}, \quad \lambda_{5,6} = -7.5 \cdot 10^{-10} \pm 9.1 \cdot 10^{-4} i$$

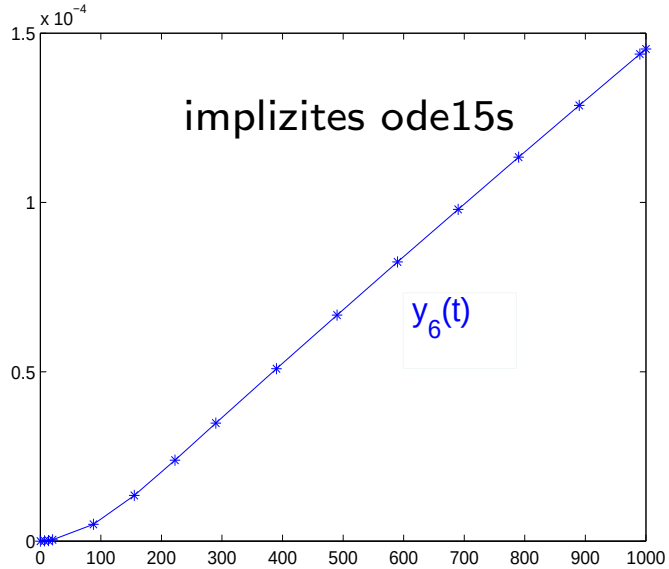




. . .


x 10^-10
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
y2(t)
y4(t)
implizites ode15s
y3(t)
0
100
200
300
400
500
600
700
800
900
1000

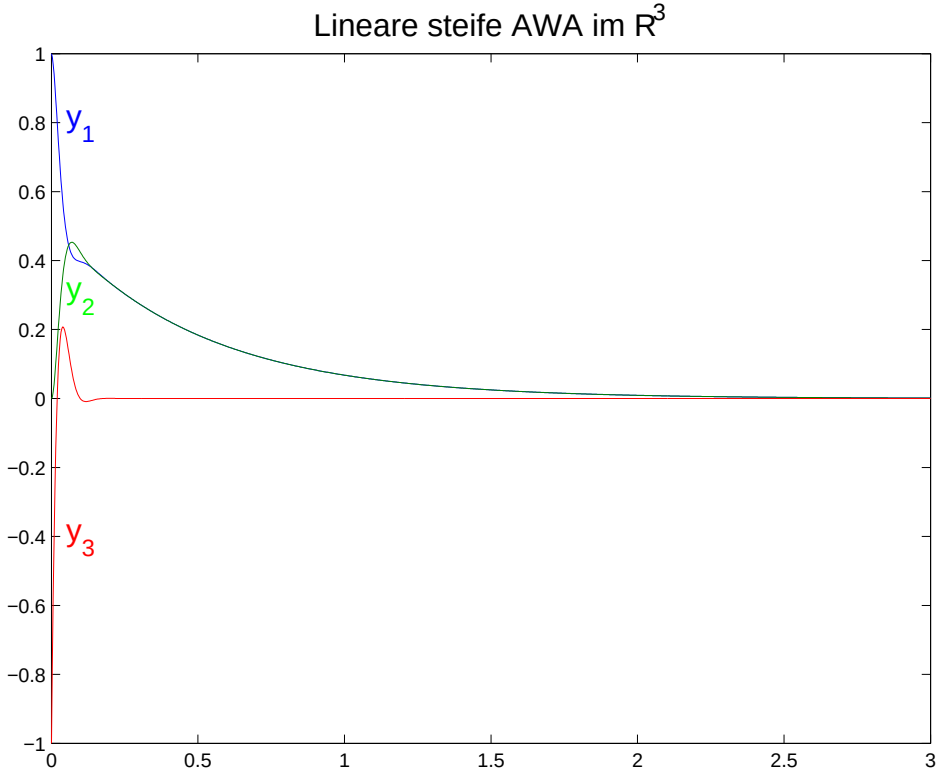

x 10^-8
6
4
2
0
−2
−4
−6
−8
explizites ode45
y6(t)
0
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.01


x 10^-4
1.5
1
0.5
0
implizites ode15s
y6(t)
0
100
200
300
400
500
600
700
800
900
1000

(d) Betrachte die AWA

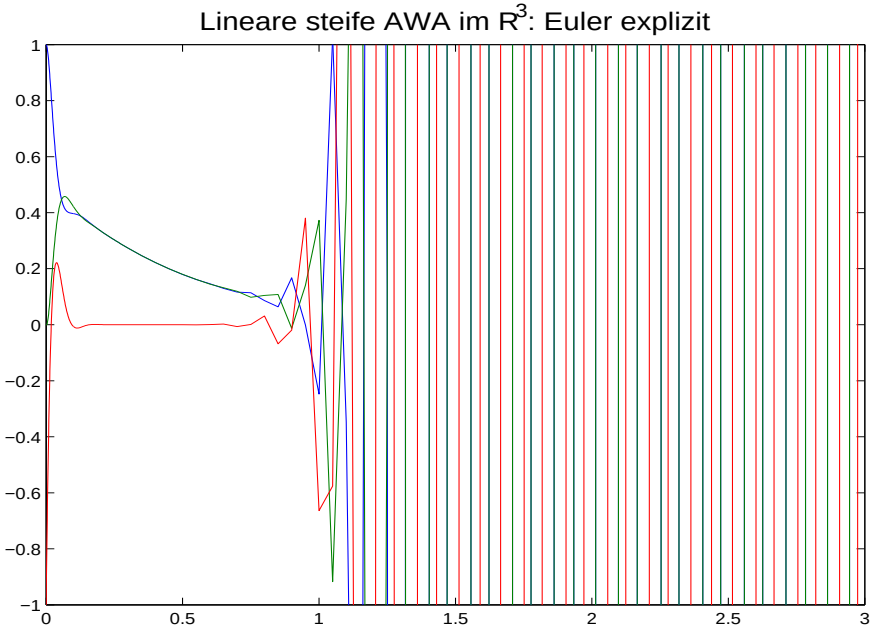
$$y' = Ay, \quad y(0) = [1, 0, -1]^T \qquad \text{mit} \quad A = \begin{bmatrix} -21 & 19 & -20 \\ 19 & -21 & 20 \\ 40 & -40 & -40 \end{bmatrix}.$$

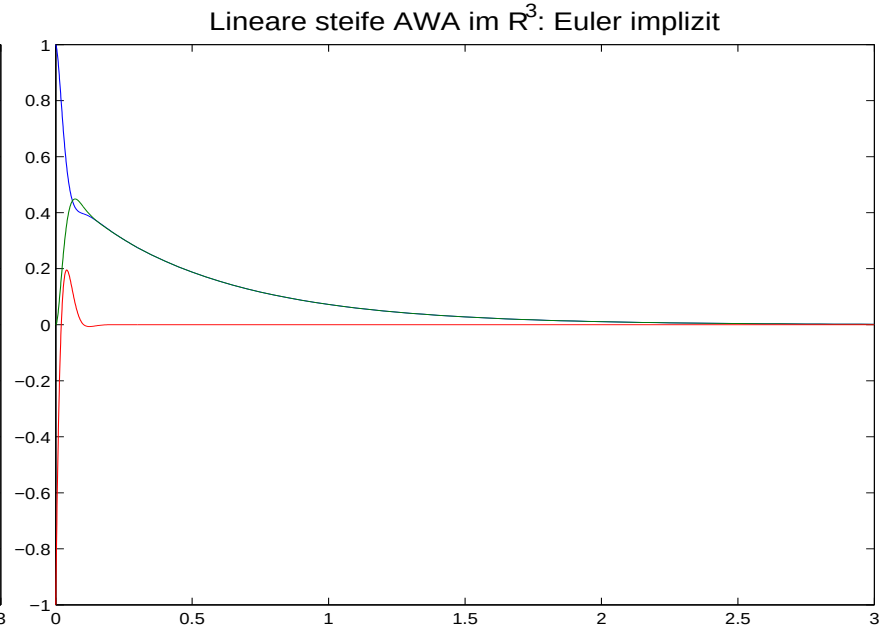
Die Eigenwerte von A sind $\lambda_1 = -2$, $\lambda_{2,3} = -40 \pm 40i$ und die exakte Lösung ist



$$y(t) = \frac{1}{2}\begin{bmatrix} e^{-2t} + e^{-40t}(\cos 40t + \sin 40t) \\ e^{-2t} - e^{-40t}(\cos 40t + \sin 40t) \\ -2e^{-40t}(\cos 40t - \sin 40t) \end{bmatrix}$$

Im Bereich $0 \leq t \leq 0.1$ haben die Lösungskomponenten y_k, $k = 1, 2, 3$, noch große Steigungen, also wählen wir eine kleine Schrittweite $h = 0.001$ für explizites und implizites Euler-Verfahren. Für $0.1 \leq t \leq 0.3$ liegt weniger Variation der y_k vor, wir arbeiten mit $h = 0.005$, und für $t \geq 0.3$ sogar mit $h = 0.05$. Das implizite Euler-Verfahren liefert eine gute Näherung, während das explizite Verfahren bereits ab $t = 0.6$ zu unbrauchbaren Schwingungen der Lösung führt

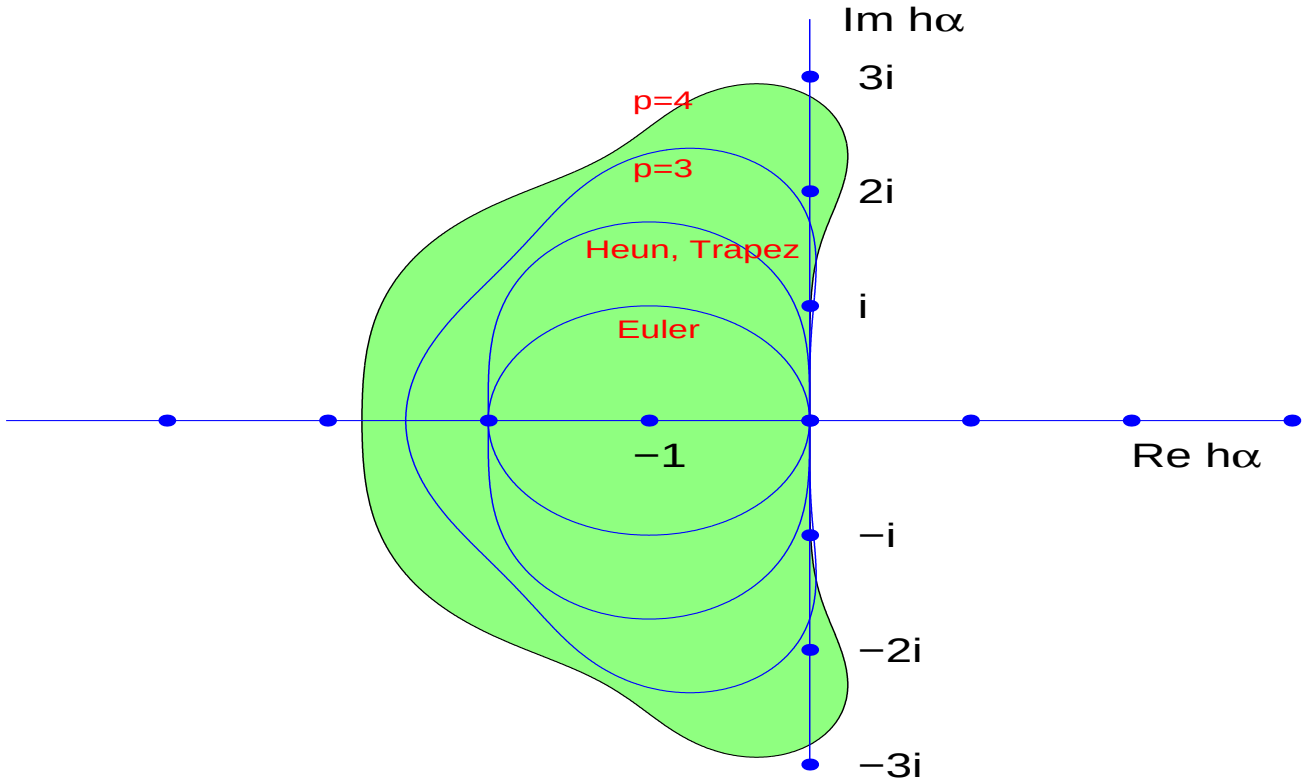




3.4.6 Stabilitätsgebiete skalare $k-$Schrittverfahren: Alle betrachteten expliziten und impliziten ESVerfahren liefern für die Modellgleichung $y' = \alpha y$, $\alpha \in \mathbb{C}$, $y(t_0) = y_0$, zur Schrittweite $h \neq 0$ die Werte

$$y_\ell = [g(h\alpha)]^\ell y_0, \qquad \ell \in \mathbb{N}.$$

Man bezeichnet $g(h\alpha)$ als den "Dämpungs- bzw. Verstärkungsfaktor", und definiert den Stabilitätsgebiet für die ESVen als $\mathrm{SG} := \{z \in \mathbb{C} \setminus \{0\};\ |g(z)| \leq 1\}$. Für die $k-$SVen mit $k > 1$ gilt $\mathrm{SG} := \{z \in \mathbb{C} \setminus \{0\};\ k-\text{Schrittverfahren ist stabil für } z\}$.



3.4.7 Definition: A-Stabilität

Ein $m-$Schrittverfahren heißt *A-stabil*, wenn gilt $\{z \in \mathbb{C} \setminus \{0\};\ \mathrm{Re}\, z \leq 0\} \subset \mathrm{SG}$.

3.4.8 Beispiele (A-Sabilität von ESV):

(i) Implizites Euler-Verfahren: $g(z) = \frac{1}{1-z}$ liefert sofort

$$\mathrm{SG} = \{z \in \mathbb{C} \setminus \{0\};\ |z-1| \geq 1\}.$$

Das Verfahren ist A-stabil. VORSICHT: Exakte Lösungen für $\mathrm{Re}\,\alpha > 0$ wachsen exponentiell, während $|y_\ell|$ für Schrittweiten $h > 0$ mit $h\alpha \in \mathrm{SG}$ beschränkt ist oder sogar exponentiell gegen Null konvergiert. Das Verfahren "täuscht" für $\mathrm{Re}\,\alpha > 0$ ein falsches asymptotisches Verhalten vor.

(ii) Implizite Trapezmethode: $g(z) = \frac{2+z}{2-z}$ liefert

$$\mathrm{SG} = \{z \in \mathbb{C} \setminus \{0\};\ \mathrm{Re}\, z \leq 0\}.$$

Das Verfahren ist ebenfalls A-stabil.

(iii) Implizite Mittelpunktsregel: $g(z) = -\frac{2+z}{2-z}$ liefert das gleiche Ergebnis wie in (ii).

(iv) Implizite 2-stufige RK-Methode:

$$g(z) = \frac{12 + 6z + z^2}{12 - 6z + z^2} = \frac{(z+\eta)(z+\overline{\eta})}{(z-\eta)(z-\overline{\eta})}$$

mit $\eta = 3 + i\sqrt{3}$ hat das gleiche Stabilitätsgebiet wie die Mittelpunktsregel, also ist das Verfahren ebenfalls A-stabil.

3.4.9: Für reelle α ergeben sich die folgenden *Stabilitätsintervalle* $\mathrm{SG} \cap \mathbb{R}$:

Verfahren	Stabilitätsintervall	$A-$ *stabil*
Euler $-$ *Verfahren*	$(-2, 0)$	*nein*
Verb.Euler $-$ *Verfahren*	$(-2, 0)$	*nein*
klassischesRK $-$ *Ver.*	$(-2.78, 0)$	*nein*
$2-$ *Schritt* $-$ *Adams* $-$ *Bashforth*	$(-1, 0)$	*nein*
$4-$ *Schritt* $-$ *Adams* $-$ *Bashforth*	$(-0.3, 0)$	*nein*
$3-$ *Schritt* $-$ *Adams* $-$ *Moulton*	$(-3, 0)$	*nein*
$BDF-2$	$(-\infty, 0)$	*ja*
$BDF-3,4,5,6$	$(-\infty, 0)$	*nein*
implizites Euler	$(-\infty, 0)$	*ja*
implizite Trapez $-$ *Methode*	$(-\infty, 0)$	*ja*
implizites RK $-$ *Gauß* $-$ *Verf.*	$(-\infty, 0)$	*ja*

Bemerkung:

(a) Der Einsatz expliziter Verfahren bei steifen Systemen ist nicht sinnvoll. Normalerweise verwendet man die BDF-k-Verfahren bis $k = 6$ oder A-stabile implizite RK-Gauß- Verfahren höhere Ordnung.

(b) Falls die Eigenwerte von $\frac{\partial f}{\partial y}$ alle reel sind, oder weit von der imaginären Achse entfernt sind, sind auch A(θ)-stabile Verfahren verwendbar.

3.4.10 Definition: A(θ)-Stabilität

Ein $k-$Schrittverfahren heißt *A(θ)-stabil*, wenn gilt

$$\{z \in \mathbb{C} \setminus \{0\} \mid |\arg(z) - \pi| < \theta\} \subset \mathrm{SG}.$$