

Prof. Dr. Torsten Linß

Modul 61512

Numerische Mathematik II

LESEPROBE

Fakultät für
**Mathematik und
Informatik**

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Kapitel 14

Anfangswertaufgaben für gewöhnliche Differentialgleichungen

Sehr viele Probleme aus den Naturwissenschaften und der Technik führen auf Differentialgleichungen. Daher ist die Entwicklung effizienter Methoden zur Lösung solcher Gleichungen eine Hauptaufgabe der Numerischen Mathematik. Zu diesem Problemkreis gibt es eine Vielzahl von Publikationen, die sowohl sehr allgemeine und für große Klassen von Differentialgleichungen anwendbare Verfahren als auch typabhängige Verfahren, die auf spezielle Differentialgleichungen zugeschnitten sind, behandeln.

14.1 Einleitung, Existenzaussagen

In diesem Kapitel wollen wir uns mit **Anfangswertaufgaben für gewöhnliche Differentialgleichungen erster Ordnung** beschäftigen. Ein einfaches Beispiel ist die Differentialgleichung

$$y'(t) = \lambda y(t), \quad t > t_0, \quad \lambda \in \mathbb{R},$$

die – je nach Vorzeichen der Konstante λ – Wachstums- und Zerfallsprozesse beschreibt, beispielsweise das Wachstum einer Populationen von Lebewesen oder den Zerfall chemischer Substanzen. Durch Vorgabe einer Anfangsbedingung (Größe der Population bzw. Stoffmenge der Substanz) zur Zeit t_0 ist die Lösung der Differentialgleichung eindeutig bestimmt.

Aufgabenstellung (AWA). Gegeben sind eine Funktion $f : D \subset \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ sowie ein Punkt $(t_0, y_0) \in D$. Gesucht ist eine stetig differenzierbare Funktion y , für die

$$y'(t) = f(t, y(t)), \quad t \geq t_0, \tag{14.1a}$$

und

$$y(t_0) = y_0. \tag{14.1b}$$

gilt.

♣

Bemerkung 14.1.1. Die Gleichung (14.1a) wird als **Differentialgleichung** bezeichnet, da sie Ableitungen der zu bestimmenden Funktion y enthält. Es handelt sich dabei um eine **gewöhnliche** Differentialgleichung **erster Ordnung**, weil y eine Funktion einer Veränderlichen ist und die höchste auftretende Ableitung eine Ableitung erster Ordnung ist.

Die Differentialgleichung (14.1a) legt fest, wie sich ein System, das zum Zeitpunkt t durch den Zustand $y(t)$ beschreiben wird, mit der Zeit $t > t_0$ entwickelt, wenn es sich zur Zeit $t = t_0$ im Zustand y_0 befindet. Umgekehrt kann man aber auch nach der Vergangenheit des Systems fragen, also den Zuständen $y(t)$ für $t < t_0$. Diese beiden Aufgaben sind äquivalent, wie man sich durch die Substitution $t \rightarrow \tau = t_0 - (t - t_0)$ überlegt, da $t < t_0$ genau dann, wenn $\tau > t_0$.

Der folgende Satz von PICARD¹ und LINDELÖF² gibt Auskunft über die eindeutige Lösbarkeit der Anfangswertaufgabe (AWA).

Satz 14.1.2. Die Funktion $f : D := (\alpha, \beta) \times (a, b) \rightarrow \mathbb{R}$ sei stetig und genüge bzgl. des zweiten Argumentes einer LIPSCHITZ-Bedingung, d.h., es gebe eine Konstante $L \in \mathbb{R}$ mit

$$|f(x, y) - f(x, z)| \leq L|y - z|, \quad \forall x \in (\alpha, \beta), \quad y, z \in (a, b).$$

Dann existiert ein Intervall $I_\delta := (t_0 - \delta, t_0 + \delta)$, $\delta > 0$, derart, dass (AWA) genau eine Lösung $y \in C^1(I_\delta)$ besitzt, für die

$$y'(t) = f(t, y(t)), \quad t \in I_\delta, \quad \text{und} \quad y(t_0) = y_0$$

gilt.

Bemerkung 14.1.3. Für die Analysis numerischer Verfahren für (AWA) ist es i.Allg. erforderlich, dass die Lösung y höhere Regularität besitzt. Sind die Voraussetzungen des Satzes 14.1.2 erfüllt und ist $f \in C^k(D)$, $k \in \mathbb{N}_0$, dann gilt $y \in C^{k+1}(I_\delta)$.

14.2 Approximationsverfahren für AWA

Für die Lösung y von (AWA) möchten wir auf dem Intervall $[t_0, T]$ Näherungswerte berechnen. Dabei ist die Anfangszeit t_0 durch die Aufgabenstellung gegeben, während $T > t_0$ einen Zeitpunkt darstellt, bis zu dem uns die Lösung von (AWA) interessiert.

Über das Intervall $[t_0, T]$ legen wir ein Gitter mit der **Gitterweite** (oder auch **Schrittweite**) $\tau := (T - t_0)/n$, $n \in \mathbb{N}$:

$$t_0 < t_1 < t_2 < \dots < t_n = T, \quad t_\ell := t_0 + \ell\tau.$$

Auf diesem Gitter möchten wir Werte u_ℓ als Näherungen für die $y(t_\ell)$ berechnen:

$$y(t_\ell) \approx u_\ell, \quad \ell = 0, 1, \dots, n.$$

¹Charles Émile Picard, *24.7.1856 (Paris, Frankreich), †11.12.1941 (Paris, Frankreich)

²Ernst Leonard Lindelöf, *7.3.1870 (Helsingfors, Russland), †4.6.1946 (Helsinki, Finnland)

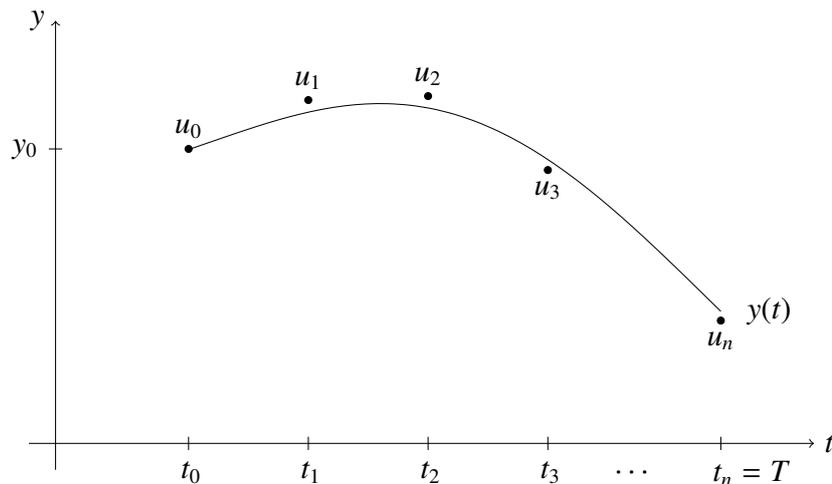


Abbildung 14.1: Punktweise Approximation der Lösung einer Anfangswertaufgabe

Abb. 14.1 illustriert dieses Vorgehen. Dabei setzen wir sinnvollerweise

$$u_0 := y(t_0) = y_0$$

und berechnen sukzessive u_ℓ aus

- der Näherung $u_{\ell-1}$ zur vorherigen Zeit $t_{\ell-1}$ oder
- aus den m , $m > 1$, vorangehenden Näherungen $u_{\ell-1}, u_{\ell-2}, \dots, u_{\ell-m}$.

Im ersten Fall spricht man von **Einschritt-** und im zweiten von **Mehrschrittverfahren**, bzw. konkreter von **m -Schrittverfahren**. Wir werden uns hier auf Einschrittverfahren (ESV) beschränken.

14.3 Einschrittverfahren für AWA

14.3.1 Konstruktion und Struktur von Einschrittverfahren

Es gibt viele Möglichkeiten, ESV zu konstruieren. Eine mit Blick auf Abschnitt 1.3.1 naheliegende Idee ist, die Ableitung in (14.1a) zunächst durch einen Differenzenquotienten zu approximieren:

$$\frac{y(t_\ell) - y(t_{\ell-1})}{\tau} \approx y'(t_{\ell-1}) = f(t_{\ell-1}, y(t_{\ell-1})).$$

Anschließend ersetzen wir $y(t_\ell)$ und $y(t_{\ell-1})$ durch ihre Näherungen u_ℓ bzw. $u_{\ell-1}$ und erhalten so eine Rekursion zur Berechnung von u_ℓ aus $u_{\ell-1}$:

$$u_\ell = u_{\ell-1} + \tau f(t_{\ell-1}, u_{\ell-1}), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0. \quad (14.2)$$

Diese Berechnungsvorschrift wird als **explizites EULER-Verfahren** bezeichnet.

Obiger Differenzenquotient kann auch als Approximation für $y'(t_\ell)$ aufgefasst werden:

$$\frac{y(t_\ell) - y(t_{\ell-1})}{\tau} \approx y'(t_\ell) = f(t_\ell, y(t_\ell)),$$

was uns auf die Rekursion

$$u_\ell = u_{\ell-1} + \tau f(t_\ell, u_\ell), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0, \quad (14.3)$$

führt. Mit (14.3) ist keine explizite Berechnungsvorschrift für u_ℓ gegeben, vielmehr muss eine i.Allg. nichtlineare Gleichung gelöst werden. Daher wird (14.3) auch als **implizites EULER-Verfahren** bezeichnet. Auf den ersten Blick scheint dieses Verfahren daher ungünstiger zu sein als (14.2), es gibt aber wichtige Anwendungen, bei denen das implizite EULER-Verfahren seinem expliziten Pendant deutlich überlegen ist, da auftretende Rundungsfehler nicht verstärkt werden.

Bemerkung 14.3.1. Zur numerischen Lösung von (14.3) kann u_ℓ als Fixpunkt der Funktion φ_ℓ ,

$$\varphi_\ell(x) := u_{\ell-1} + \tau f(t_\ell, x),$$

aufgefasst werden. Genügt f einer LIPSCHITZ-Bedingung bzgl. des zweiten Argumentes mit LIPSCHITZ-Konstante L , dann gilt

$$|\varphi_\ell(x) - \varphi_\ell(z)| = \tau |f(t_\ell, x) - f(t_\ell, z)| \leq \tau L |x - z|.$$

Für $\tau L < 1$ ist φ_ℓ also kontrahierend. Ferner ist φ_ℓ in einer Umgebung von $y(t_\ell)$ eine Selbstabbildung, d.h. es existiert eine Umgebung $U_\varepsilon(y(t_\ell))$ mit $\varphi_\ell(z) \in U_\varepsilon(y(t_\ell))$ für alle $z \in U_\varepsilon(y(t_\ell))$ (ohne Beweis). Somit existiert ein eindeutig bestimmter Fixpunkt von φ_ℓ . Dieser ist u_ℓ . Ist $\tau \ll 1/L$ dann konvergiert die Fixpunktiteration sehr schnell.

Im Abschnitt 1.3.1 – insbesondere in der Aufgabe 1.3.2 – hatten wir gesehen, dass ein symmetrischer Differenzenquotient bessere Approximationen liefert als ein einseitiger:

$$\frac{y(t_\ell) - y(t_{\ell-1})}{\tau} \approx y'(t_{\ell-1/2}) = f(t_{\ell-1/2}, y(t_{\ell-1/2})), \quad (14.4)$$

wobei

$$t_{\ell-1/2} := \frac{t_\ell + t_{\ell-1}}{2} = t_{\ell-1} + \frac{\tau}{2}$$

der Mittelpunkt des Intervalls $[t_{\ell-1}, t_\ell]$ ist. Nun stellt sich allerdings die Frage, wie wir $y(t_{\ell-1/2})$ bzw. $y'(t_{\ell-1/2})$ ersetzen. (Mindestens) drei Möglichkeiten bieten sich an.

(i) Wir berechnen eine Näherung für $y(t_{\ell-1/2})$ mit einem Schritt des expliziten EULER-Verfahren und halbiertes Schrittweite:

$$y(t_{\ell-1/2}) \approx u_{\ell-1} + \frac{\tau}{2} f(t_{\ell-1}, u_{\ell-1})$$

und ersetzen $y(t_{\ell-1/2})$ entsprechend in (14.4):

$$\frac{y(t_\ell) - y(t_{\ell-1})}{\tau} \approx f\left(t_{\ell-1/2}, u_{\ell-1} + \frac{\tau}{2}f(t_{\ell-1}, u_{\ell-1})\right).$$

Hieraus leiten wir die Rekursion

$$u_\ell = u_{\ell-1} + \tau f\left(t_{\ell-1/2}, u_{\ell-1} + \frac{\tau}{2}f(t_{\ell-1}, u_{\ell-1})\right), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0, \quad (14.5)$$

ab, die als **verbessertes EULER-Verfahren** bezeichnet wird.

(ii) Mittelung für y' :

$$y'(t_{\ell-1/2}) \approx \frac{y'(t_{\ell-1}) + y'(t_\ell)}{2} = \frac{f(t_{\ell-1}, y(t_{\ell-1})) + f(t_\ell, y(t_\ell))}{2} \approx \frac{f(t_{\ell-1}, u_{\ell-1}) + f(t_\ell, u_\ell)}{2}.$$

Wir erhalten die sog. **implizite Trapezregel** für Anfangswertaufgaben

$$u_\ell = u_{\ell-1} + \frac{\tau}{2}f(t_{\ell-1}, u_{\ell-1}) + \frac{\tau}{2}f(t_\ell, u_\ell), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0. \quad (14.6)$$

(iii) Mittelung für y :

$$y'(t_{\ell-1/2}) = f(t_{\ell-1/2}, y(t_{\ell-1/2})) \approx f\left(t_{\ell-1/2}, \frac{y(t_{\ell-1}) + y(t_\ell)}{2}\right) \approx f\left(t_{\ell-1/2}, \frac{u_{\ell-1} + u_\ell}{2}\right).$$

Diesmal erhalten wir die Berechnungsvorschrift

$$u_\ell = u_{\ell-1} + \tau f\left(t_{\ell-1/2}, \frac{u_{\ell-1} + u_\ell}{2}\right), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0, \quad (14.7)$$

die sog. **Mittelpunktregel** für Anfangswertaufgaben.

Das verbesserte EULER-Verfahren (14.5) stellt eine explizite Berechnungsvorschrift für die u_ℓ dar, während sowohl bei der Trapez- als auch bei der Mittelpunktregel eine i.Allg. nichtlineare Gleichung zu lösen ist. Diese beiden Verfahren werden daher als **implizite** Verfahren klassifiziert.

Wir möchten es an dieser Stelle bei diesen fünf Beispielen für Einschrittverfahren bewenden lassen. Mit den RUNGE-KUTTA-Verfahren werden wir eine weitere wichtige Klasse im Abschnitt 14.3.3 kennen lernen.

Allgemein besitzen Einschrittverfahren (ESV) die folgende Struktur:

$$u_\ell = u_{\ell-1} + \tau \Phi(t_{\ell-1}, u_{\ell-1}, \tau), \quad \ell = 1, 2, \dots, n, \quad u_0 = y_0. \quad (14.8)$$

Die Funktion Φ wird als **Verfahrensfunktion** oder **Zuwachsfunktion** bezeichnet. Sie ist von f abhängig. Auf eine Verankerung dieser Abhängigkeit in der Notation wollen wir der Einfachheit halber jedoch verzichten.

Aufgabe 14.3.2. Geben Sie die Verfahrensfunktionen für das explizite, das verbesserte und das implizite EULER-Verfahren an! ◇

14.3.2 Konvergenz von Einschrittverfahren

Wir betrachten die Anfangswertaufgabe (AWA) auf einem Intervall $[t_0, T]$, $T > t_0$. Die durch die Verfahrensvorschrift (14.8) erzeugte Folge $(u_\ell)_{\ell=0}^n$ hängt von der Anzahl der Teilintervalle n bzw. der gewählten Gitterweite τ ab. In diesem Abschnitt verdeutlichen wir diese Abhängigkeit durch ein hochgestelltes τ : $(u_\ell^\tau)_{\ell=0}^n$, wobei n und τ gemäß $\tau = (T - t_0)/n$ aneinander gekoppelt sind.

Der Aufwand zur Berechnung einer Approximation zu (AWA) mit einem Einschrittverfahren ist proportional zu n bzw. $1/\tau$. Eine sinnvolle Forderung für das Verfahren ist daher, dass der Fehler der Approximation reduziert wird, wenn man die Schrittweite verkleinert. Uns interessiert deshalb in diesem Abschnitt die Abhängigkeit des Fehlers von der Schrittweite τ , wenn $\tau \rightarrow 0$. Dabei betrachten wir den **globalen Verfahrensfehler** in der Maximumnorm:

$$E(y, u^\tau) := \max_{\ell=0, \dots, n} |y(t_\ell) - u_\ell^\tau|.$$

Die wichtigste Kenngröße für Effizienz eines Einschrittverfahrens ist seine Konvergenzordnung.

Definition 14.3.3. *Es existiere eine Konstante $C \in \mathbb{R}^+$, eine Schrittweite $\tau_0 > 0$ und ein $p > 0$ derart, dass für den **globalen Verfahrensfehler** des ESV gilt*

$$E(y, u^\tau) \leq C\tau^p \quad \text{für alle } \tau \in (0, \tau_0].$$

Dann nennen wir das ESV (14.8) konvergent mit der **Konvergenzordnung** p .

Wichtigstes Hilfsmittel bei der Konvergenzuntersuchung ist der **lokale Verfahrensfehler** η . Er gibt an, wie groß der Fehler nach einem Schritt des Verfahrens ist, wenn mit der exakten Lösung gestartet wurde. Zur Zeit t besitzt y den Wert $y(t)$. Das Paar $(t, y(t))$ setzen wir als Startwerte in die Verfahrensvorschrift (14.8) ein und erhalten den Näherungswert $y(t) + \tau \Phi(t, y(t), \tau)$ für $y(t + \tau)$. Der lokale Fehler ist also

$$\eta(t, \tau) := y(t + \tau) - y(t) - \tau \Phi(t, y(t), \tau).$$

Mittels des lokalen Verfahrensfehlers wird die Konsistenzordnung eines ESV definiert.

Definition 14.3.4. *Es existiere eine Konstante $\Gamma \in \mathbb{R}^+$, eine Schrittweite $\tau_0 > 0$ und ein $p > 0$ derart, dass für den **lokalen Verfahrensfehler** des ESV gilt*

$$\eta(t, \tau) \leq \Gamma\tau^{p+1} \quad \text{für alle } t \in [t_0, T - \tau] \text{ und } \tau \in (0, \tau_0].$$

Dann sagen wir, das ESV (14.8) besitzt die **Konsistenzordnung** p .

Der lokale Verfahrensfehler – und damit die Konsistenzordnung – wird i.Allg. mittels TAYLOR-Entwicklungen untersucht.

Beispiel 14.3.5. Wir wollen den lokalen Fehler des expliziten EULER-Verfahrens untersuchen. Für dieses ist $\Phi(t, u, \tau) = f(t, u)$; vgl. Auf. 14.3.2. Wir nehmen an, dass $y \in C^2[t_0, T]$ sei.

Wir entwickeln y in eine TAYLOR-Reihe:

$$y(t + \tau) = y(t) + \tau y'(t) + \frac{\tau^2}{2} y''(t + \xi), \quad \text{mit einem } \xi \in (0, \tau).$$

Unter Beachtung von $y'(t) = f(t, y(t)) = \Phi(t, y(t), \tau)$ folgt

$$\eta(t, \tau) = y(t + \tau) - y(t) - \tau \Phi(t, y(t), \tau) = \frac{\tau^2}{2} y''(t + \xi).$$

Für den lokalen Fehler erhalten wir die Abschätzung

$$|\eta(t, \tau)| \leq \frac{\tau^2}{2} \|y''\|_{[t_0, T]} \quad \text{für } t_0 \leq t < t + \tau \leq T.$$

Das Verfahren besitzt also die Konsistenzordnung 1. ♠

Aufgabe 14.3.6. Zeigen Sie, dass das verbesserte Eulerverfahren die Konsistenzordnung 2 besitzt, falls f zweimal und y dreimal stetig differenzierbar sind!

Hinweis: Benutzen Sie die TAYLOR-Reihe und entwickeln Sie $y(t + \tau)$ sowie $\Phi_v(t, y(t), \tau)$ an der Stelle $\tau = 0$ bis zum kubischen bzw. quadratischen Glied nach Potenzen von τ ! ◇

Wir wollen nun untersuchen, wie aus dem lokalen auf den globalen Verfahrensfehler geschlossen werden kann. Dazu lösen wir zunächst die folgende Übungsaufgabe, deren Ergebnis wir nutzen werden, um die Akkumulation der lokalen Fehler abzuschätzen.

Aufgabe 14.3.7. Beweisen Sie: Für Zahlen $M > 0$, $a_\ell \geq 0$, $h > 0$ und $b \geq 0$ sei

$$a_\ell \leq (1 + hM) a_{\ell-1} + hb, \quad \ell = 1, 2, \dots, n,$$

erfüllt.

Dann gelten die Abschätzungen

$$a_\ell \leq \frac{e^{M\ell h} - 1}{M} b + e^{M\ell h} a_0, \quad \ell = 0, 1, \dots, n.$$

Hinweis: Benutzen Sie Induktion! ◇

Satz 14.3.8. Das ESV (14.8) besitze Konsistenzordnung $p > 0$. Für den lokalen Verfahrensfehler des ESV (14.8) gelte

$$\eta(t, \tau) \leq \Gamma \tau^{p+1} \quad \text{für alle } t \in [t_0, T - \tau] \text{ und } \tau \in (0, \tau_0]. \quad (14.9)$$

mit einem $\tau_0 > 0$ und einer Konstante Γ . Die Verfahrensfunktion Φ sei LIPSCHITZ-stetig bzgl. ihres zweiten Argumentes mit LIPSCHITZ-Konstante $\Lambda > 0$.

Dann ist das Verfahren konvergent mit der Ordnung p . Für den globalen Verfahrensfehler gilt

$$\max_{\ell=0, \dots, n} |y(t_\ell) - u_\ell^\tau| \leq K \Gamma \tau^p \quad \text{für } \tau \in (0, \tau_0] \quad \text{mit} \quad K := \frac{e^{\Lambda(T-t_0)} - 1}{\Lambda}.$$

Beweis. Wir setzen

$$\begin{aligned} y_\ell &:= y(t_\ell), \quad \ell = 0, \dots, n, && \text{für die Lösung der AWA an der Stelle } t_\ell, \\ e_\ell &:= u_\ell^\tau - y_\ell, \quad \ell = 0, \dots, n, && \text{für den Fehler des Verfahrens an der Stelle } t_\ell \text{ und} \\ \eta_\ell &:= \eta(t_\ell, \tau), \quad \ell = 0, \dots, n-1 && \text{für den lokalen Fehler des Verfahrens.} \end{aligned}$$

Wegen der Verfahrensvorschrift (14.8) gilt

$$u_\ell^\tau = u_{\ell-1}^\tau + \tau \Phi(t_{\ell-1}, u_{\ell-1}^\tau, \tau), \quad \ell = 1, 2, \dots, n,$$

und wegen der Definition des lokalen Fehlers

$$y_\ell = y_{\ell-1} + \tau \Phi(t_{\ell-1}, y_{\ell-1}, \tau) + \eta_\ell, \quad \ell = 1, 2, \dots, n.$$

Somit gilt für den Fehler des Verfahrens die Rekursion

$$e_\ell = e_{\ell-1} + \tau \left(\Phi(t_{\ell-1}, u_{\ell-1}^\tau, \tau) - \Phi(t_{\ell-1}, y_{\ell-1}, \tau) \right) - \eta_\ell, \quad \ell = 1, 2, \dots, n, \quad e_0 = 0.$$

Aus der LIPSCHITZ-Stetigkeit von Φ sowie aus (14.9) folgt

$$|e_\ell| \leq |e_{\ell-1}| + \tau \underbrace{\left| \Phi(t_{\ell-1}, u_{\ell-1}^\tau, \tau) - \Phi(t_{\ell-1}, y_{\ell-1}, \tau) \right|}_{\leq \Lambda |u_{\ell-1}^\tau - y_{\ell-1}|} + |\eta_\ell|$$

Somit gilt

$$|e_\ell| \leq (1 + \tau\Lambda) |e_{\ell-1}| + \Gamma \tau^{p+1}, \quad \ell = 1, 2, \dots, n, \quad e_0 = 0.$$

Wir benutzen Aufg. 14.3.7 mit $h = \tau$, $M = \Lambda$, $b = \Gamma \tau^p$ und $a_0 = 0$, um den Beweis zu kompletieren. \square

Aus Satz 14.3.8 können wir die Faustregel

Konsistenz der Ordnung p impliziert Konvergenz der Ordnung p

für Einschrittverfahren ableiten, die natürlich nur unter gewissen im Satz formulierten Bedingungen gilt.

Beispiel 14.3.9. Es ist Zeit für ein Beispiel. Wir betrachten die AWA

$$y'(t) = 3y(t)^{2/3}, \quad t > 0, \quad y(0) = 1.$$

Die Lösung der AWA ist

$$y(t) = (t + 1)^3, \quad t \geq 0.$$

Wir wollen sie mit dem expliziten und dem verbesserten EULER-Verfahren auf dem Intervall $[0, 1]$ numerisch approximieren.

Die folgende Tabelle enthält die globalen Verfahrensfehler der beiden Verfahren für verschiedene Schrittweiten $\tau = 1/n$.

n	EULER-Verfahren		n	EULER-Verfahren	
	explizit	verbessert		explizit	verbessert
2^5	2.523e-01	3.316e-03	2^{13}	1.005e-03	1.431e-05
2^6	1.280e-01	8.411e-04	2^{14}	4.992e-04	7.629e-06
2^7	6.449e-02	2.108e-04	2^{15}	2.594e-04	4.292e-06
2^8	3.237e-02	4.435e-05	2^{16}	1.464e-04	2.766e-05
2^9	1.621e-02	1.431e-05	2^{17}	1.354e-04	7.248e-05
2^{10}	8.122e-03	3.815e-06	2^{18}	1.249e-04	1.240e-04
2^{11}	4.063e-03	6.199e-06	2^{19}	1.547e-03	1.532e-03
2^{12}	2.048e-03	1.287e-05	2^{20}	6.089e-03	6.081e-03

Für das einfache explizite EULER-Verfahren sichern unsere theoretischen Untersuchungen Konvergenz der Ordnung 1, für das verbesserte EULER-Verfahren hingegen Ordnung 2. Wir erwarten also, dass der globale Fehler proportional zu $1/\tau$ bzw. $1/\tau^2$ ist.

Dieses erwartete Verhalten beobachten wir auch in den Tabellen, solange n einen kritischen Wert, ca. 2^{17} bzw. 2^{10} nicht überschreitet.

Wie bei der numerischen Differentiation (§1.3.1) können Rundungsfehler, die sich aus der Verwendung von Gleitkommazahlen ergeben, die Resultate stark verfälschen. ♣

Zum Abschluss dieses Abschnitts wollen wir uns noch mit der Akkumulation von Rundungsfehlern bei Einschrittverfahren beschäftigen.

Bei der Umsetzung eines ESV in einem Computerprogramm kann die Verfahrensvorschrift (14.8) nicht exakt implementiert werden. Um die Fortpflanzung der Rundungsfehler zu analysieren, nehmen wir an, dass die Rechnung im ℓ -ten Schritt mit einem Fehler ϱ_ℓ behaftet ist:

$$\tilde{u}_\ell = \tilde{u}_{\ell-1} + \tau \Phi(t_{\ell-1}, \tilde{u}_{\ell-1}, \tau) + \varrho_\ell, \quad \ell = 1, 2, \dots, n, \quad \tilde{u}_0 = y_0 + \varrho_0. \quad (14.10)$$

Unter den Voraussetzungen und mit der Notation von Satz 14.3.8 gilt für die Fehler der Näherungen \tilde{u}_ℓ^T :

$$\max_{\ell=0, \dots, n} |y(t_\ell) - \tilde{u}_\ell^T| \leq K \left(\Gamma \tau^p + \tau^{-1} \max_{\ell=1, \dots, n} |\varrho_\ell| \right) + K |\varrho_0|.$$

Bemerkung 14.3.10. Auftretende Fehler bei der Berechnung der u_ℓ werden akkumuliert und mit dem Inversen der Schrittweite τ multipliziert. Hieraus ist abzuleiten, dass die Schritte nicht zu klein gewählt werden dürfen. Außerdem ist die Rechenzeit proportional zur Anzahl der Schritte. Für die Genauigkeit und Effizienz der Verfahren ist daher eine möglichst hohe Konsistenz- und Konvergenzordnung essentiell.

14.3.3 RUNGE-KUTTA-Verfahren

Eine wichtige Klasse von Einschrittverfahren für Anfangswertaufgaben sind die RUNGE-KUTTA³-Verfahren (RKV). Ein s -stufiges RKV, $s \in \mathbb{N}$, hat die Gestalt

$$\begin{aligned} k_1 &= f(t_{\ell-1} + c_1\tau, u_{\ell-1} + \tau(a_{11}k_1 + a_{12}k_2 + \cdots + a_{1s}k_s)) \\ k_2 &= f(t_{\ell-1} + c_2\tau, u_{\ell-1} + \tau(a_{21}k_1 + a_{22}k_2 + \cdots + a_{2s}k_s)) \\ &\vdots \\ k_s &= f(t_{\ell-1} + c_s\tau, u_{\ell-1} + \tau(a_{s1}k_1 + a_{s2}k_2 + \cdots + a_{ss}k_s)) \\ u_\ell &= u_{\ell-1} + \tau(b_1k_1 + b_2k_2 + \cdots + b_s k_s) \end{aligned}$$

Es wird durch den **Knotenvektor** $\mathbf{c} = (c_1, c_2, \dots, c_s)^T$, die **Verfahrensmatrix** $\mathbf{A} = (a_{ij})_{i,j=1}^s$ sowie den Vektor der **Gewichte** $\mathbf{b} = (b_1, b_2, \dots, b_s)^T$ definiert. Sie können in sehr übersichtlicher Form im sog. BUTCHER⁴-Tableau (oder BUTCHER-Schema) angegeben werden:

$$\begin{array}{c|c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b}^T \end{array} \iff \begin{array}{c|ccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array}$$

Einigen Beispielen von RUNGE-KUTTA-Verfahren sind wir schon in Abschnitt 14.3.1 begegnet, so lassen sich sowohl das **explizite** als auch das **implizite EULER-Verfahren** wie folgt als einstufige RKV auffassen:

$$\begin{array}{l} k_1 = f(t_{\ell-1}, u_{\ell-1}) \\ u_\ell = u_{\ell-1} + \tau k_1 \end{array} \quad \text{mit BUTCHER-Tableau} \quad \begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

bzw.

$$\begin{array}{l} k_1 = f(t_{\ell-1} + \tau, u_{\ell-1} + \tau k_1) \\ u_\ell = u_{\ell-1} + \tau k_1 \end{array} \quad \text{mit BUTCHER-Tableau} \quad \begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

während das **verbesserte EULER-Verfahren** ein zweistufiges RKV ist

$$\begin{array}{l} k_1 = f(t_{\ell-1}, u_{\ell-1}) \\ k_2 = f\left(t_{\ell-1} + \frac{\tau}{2}, u_{\ell-1} + \frac{\tau}{2}k_1\right) \\ u_\ell = u_{\ell-1} + \tau k_2 \end{array} \quad \text{mit BUTCHER-Tableau} \quad \begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline 0 & 1 & \end{array}$$

³Martin Wilhelm Kutta, *3.11.1867 (Pitschen, Schlesien) †25.12.1944 (Fürstenfeldbruck, Deutschland)

⁴John Charles Butcher, *31.3.1933 (Auckland, Neuseeland)

Aufgabe 14.3.11. Schreiben Sie die Trapez- und Mittelpunkregel als RUNGE-KUTTA-Verfahren und geben Sie die zugehörigen BUTCHER-Tableaus an! \diamond

Ist die Matrix A eine strikte untere Dreiecksmatrix (oder kann sie durch Zeilen- und Spaltenvertauschungen in eine solche überführt werden), dann handelt es sich um ein explizites Verfahren, andernfalls um ein implizites.

Bemerkung 14.3.12. Bei der programmtechnischen Umsetzung expliziter RKV ist bei jedem Schritt zur Berechnung der k_i , $i = 1, \dots, s$, jeweils genau ein Funktionswert von f zu berechnen. Bei der Umsetzung eines impliziten RKV ist hingegen ein System von s i.Allg. nichtlinearen Gleichungen zu lösen. Ähnlich zum impliziten EULER-Verfahren (Bem. 14.3.1) kann gezeigt werden, dass bei LIPSCHITZ-stetigem f und hinreichend kleiner Schrittweite τ dieses Gleichungssystem genau eine Lösung besitzt und die Fixpunktiteration zu ihrer Berechnung geeignet ist.

Weitere explizite Verfahren sind

- das dreistufige Verfahren von HEUN⁵

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

- das (klassische) vierstufige Verfahren von RUNGE und KUTTA mit dem BUTCHER-Tableau

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

Bemerkung 14.3.13. Bei expliziten RKV werden, wie hier, die Nullen auf und oberhalb der Diagonalen der Verfahrensmatrix üblicherweise nicht angegeben.

Ein s -stufiges RKV wird durch $s(s+2)$ Parameter beschrieben, nämlich den Knotenvektor \mathbf{c} , den Gewichtsvektor \mathbf{b} und die Verfahrensmatrix A . Wichtig ist nun die Frage, wie man diese Parameter bestmöglich wählt, um ein Verfahren möglichst hoher Ordnung zu erhalten. Dazu untersucht man den lokalen Verfahrensfehler mit Hilfe von TAYLOR-Entwicklungen. Man erhält so Bedingungen an die c_i , b_j und a_{ij} , die erfüllt sein müssen damit das Verfahren eine gewisse Ordnung p besitzt. Auf Details wollen wir hier nicht eingehen, sondern verweisen den interessierten Leser auf die Literatur, z.B. [6], bzw. den Kurs 01374.

Für die bisher vorgestellten expliziten RKV kann so gezeigt werden, dass die Anzahl der Stufen identisch mit der Konsistenzordnung ist, vgl. Tab. 14.1.

⁵Karl Heun, *3.4.1859 (Wiesbaden), 10.1.1929 (Karlsruhe)

Verfahren	p	s
explizites EULER-Verfahren	1	1
verbessertes EULER-Verfahren	2	2
Verfahren von HEUN	3	3
klassisches RUNGE-KUTTA-Verfahren	4	4

Tabelle 14.1: Stufenzahl s und Konsistenzordnung p ausgewählter expliziter RKV

Bereits RUNGE hat sich daher mit der Frage beschäftigt, ob es explizite fünfstufige RKV der Ordnung 5 gibt. Diese Frage wurde von BUTCHER negativ beantwortet. Ein explizites 5-stufiges RKV besitzt bestenfalls die Ordnung 4. Anders ist die Situation bei impliziten RKV. Für beliebige vorgegebene Stufenzahl $s \in \mathbb{N}$ können implizite RKV der Ordnung $p = 2s$ konstruiert werden.

14.4 Lösungen der Aufgaben

Lösung 14.3.2. Explizites EULER-Verfahren:

$$\Phi_e(t, u, \tau) = f(t, u).$$

Verbessertes EULER-Verfahren:

$$\Phi_v(t, u, \tau) = f\left(t + \frac{\tau}{2}, u + \frac{\tau}{2}f(t, u)\right).$$

Implizites EULER-Verfahren:

$$\Phi_i : (t, u, \tau) \mapsto z, \text{ und } z \text{ ist die Lösung der Gleichung } z = f(t + \tau, u + \tau z).$$

♣

Lösung 14.3.6. Wir entwickeln zunächst $y(t + \tau)$ nach TAYLOR:

$$y(t + \tau) = y(t) + \tau y'(t) + \frac{\tau^2}{2} y''(t) + \frac{\tau^3}{6} y'''(t + \xi), \quad \xi = \xi(\tau) \in (0, \tau).$$

Zur Untersuchung von $y(t) + \tau \Phi_v(t, y(t), \tau)$ setzen wir zur Abkürzung $z := y(t)$ und definieren die Hilfsfunktion

$$\psi : [0, \tau/2] : \zeta \mapsto \psi(\zeta) := f(t + \zeta, z + \zeta f(t, z)).$$

Offensichtlich ist $\Phi_v(t, z, \tau) = \psi(\tau/2)$.

Wir entwickeln ψ für $\zeta > 0$ nach TAYLOR:

$$\psi(\zeta) = \psi(0) + \zeta \psi'(0) + \frac{\zeta^2}{2} \psi''(\zeta^*), \quad \zeta^* \in (0, \zeta).$$

Für die Ableitungen von ψ erhalten wir zunächst mit der Kettenregel

$$\begin{aligned} \psi(\zeta) &= f(t + \zeta, z + \zeta f(t, z)), \\ \psi'(\zeta) &= f_t(t + \zeta, z + \zeta f(t, z)) + f_y(t + \zeta, z + \zeta f(t, z)) f(t, z), \\ \psi''(\zeta) &= f_{tt}(t + \zeta, z + \zeta f(t, z)) + 2f_{ty}(t + \zeta, z + \zeta f(t, z)) f(t, z) + f_{yy}(t + \zeta, z + \zeta f(t, z)) [f(t, z)]^2 \end{aligned}$$

und hieraus

$$\begin{aligned} \psi(0) &= f(t, z), \\ \psi'(0) &= f_t(t, z) + f_y(t, z) f(t, z), \\ |\psi''(\zeta)| &\leq C_{\psi''} := \|f_{tt}\|_D + 2\|f_{ty}\|_D \|f\|_D + \|f_{yy}\|_D \|f\|_D^2, \end{aligned}$$

wobei D der Definitionsbereich von f und $\|\cdot\|_D$ die Maximumnorm auf D seien.

Beachte, dass aus der Differentialgleichung $y'(t) = f(t, y(t))$ folgt

$$y''(t) = f_t(t, y(t)) + f_y(t, y(t)) f(t, y(t)).$$

Für den lokalen Verfahrensfehler erhalten wir mit obigen TAYLOR-Entwicklungen also

$$\eta(t, \tau) = y(t + \tau) - y(t) - \tau \Phi_v(t, \Phi(t), \tau) = \frac{\tau^3}{6} y'''(t + \xi) - \frac{\tau^3}{8} \psi'''(\xi^*),$$

und hieraus

$$|\eta(t, \tau)| \leq \tau^3 \left(\frac{\|y'''\|_{[t_0, T]}}{6} + \frac{C_{\psi''}}{8} \right).$$

Das verbesserte EULER-Verfahren besitzt somit Konsistenzordnung 2. ♣

Lösung 14.3.7. Voraussetzung: Für Zahlen $M > 0$, $a_\ell \geq 0$, $h > 0$ und $b \geq 0$ gilt

$$a_\ell \leq (1 + hM) a_{\ell-1} + hb, \quad \ell = 1, 2, \dots, n. \quad (14.11)$$

Zu zeigen: Es folgt

$$a_\ell \leq \frac{e^{M\ell h} - 1}{M} b + e^{M\ell h} a_0, \quad \ell = 0, 1, \dots, n. \quad (14.12)$$

Induktionsanfang: Offensichtlich gilt (14.12) für $\ell = 0$, da $e^{M\ell h} = 1$ für $\ell = 0$.

Induktionsschritt ($\ell \rightarrow \ell + 1$): Die Ungleichung (14.12) gelte für ℓ , dann folgt

$$\begin{aligned} a_{\ell+1} &\leq (1 + hM) a_\ell + hb, && \text{wegen (14.11),} \\ &\leq (1 + hM) \left(\frac{e^{M\ell h} - 1}{M} b + e^{M\ell h} a_0 \right) + hb, && \text{wegen (14.12) für } \ell, \\ &= \frac{(1 + hM) e^{M\ell h} - 1}{M} b + (1 + hM) e^{M\ell h} a_0 \\ &\leq \frac{e^{M(\ell+1)h} - 1}{M} b + e^{M(\ell+1)h} a_0, && \text{da } 1 + x < e^x \text{ für } x > 0. \end{aligned}$$

Somit gilt (14.12) auch für $\ell + 1$. ♣

Lösung 14.3.11. implizite Trapezregel: (2-stufiges RKV)

$$\begin{aligned} k_1 &= f(t_{\ell-1}, u_{\ell-1}) \\ k_2 &= f\left(t_{\ell-1} + \tau, u_{\ell-1} + \frac{\tau}{2}(k_1 + k_2)\right) \\ u_\ell &= u_{\ell-1} + \frac{\tau}{2}(k_1 + k_2) \end{aligned} \quad \text{mit BUTCHER-Tableau} \quad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Mittelpunktregel: (1-stufiges RKV)

$$\begin{aligned} k_1 &= f\left(t_{\ell-1} + \frac{\tau}{2}, u_{\ell-1} + \frac{\tau}{2} k_1\right) \\ u_\ell &= u_{\ell-1} + \tau k_1 \end{aligned} \quad \text{mit BUTCHER-Tableau} \quad \begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$$

♣

Inhaltsverzeichnis

10 Eigenwerte und Polynomnullstellen	1
10.1 Eigenwert- und Nullstelleneinschließungen	2
10.1.1 Einschließung aller Nullstellen	2
10.1.2 Einschließung einzelner Nullstellen	4
10.1.3 Einschließungssätze für Eigenwerte	5
10.2 Die Kondition des Eigenwertproblems	7
10.3 STURMSche Ketten	11
10.3.1 Matrixeigenwerte und STURMSche Ketten	13
10.3.2 Der EUKLIDISCHE Divisionsalgorithmus und STURMSche Ketten	14
10.3.3 Wechselzahl und Anzahl der Polynomnullstellen	15
10.4 Lösungen der Aufgaben	19
11 Direkte Verfahren für Eigenwertprobleme	21
11.1 HERMITESCHE Tridiagonalmatrizen	22
11.2 HESSENBERG-Matrizen	27
11.3 Reduktionsverfahren	30
11.3.1 Das Reduktionsverfahren nach HOUSEHOLDER	31
11.3.2 Das Reduktionsverfahren nach GIVENS	34
11.4 Das JACOBI-Verfahren	38
11.5 Lösungen der Aufgaben	45
12 Iterative Verfahren für Eigenwertprobleme	49
12.1 Die Vektoriteration	49
12.1.1 Die Grundform des Verfahrens	50
12.1.2 Deflationsmethoden	55
12.1.3 Die inverse Vektoriteration	58
12.2 Die Teilraumiteration	61
12.3 Das QR-Verfahren für normale Matrizen	63
12.4 Singulärwertzerlegung und Pseudoinverse	65
12.4.1 Die Singulärwertzerlegung	65
12.4.2 Numerischer Rang einer Matrix	68
12.4.3 Die Pseudoinverse	68
12.4.4 Berechnung der Singulärwertzerlegung	70

12.5	Lösungen der Aufgaben	75
13	Zweipunkttrandwertprobleme	81
13.1	Ein einleitendes Beispiel	81
13.1.1	Aussagen zur Analysis	82
13.1.2	Kollokationsverfahren	90
13.2	GALERKIN-Verfahren	107
13.2.1	Quasioptimalität der GALERKIN-Verfahren	109
13.2.2	Polynomiale Ansatzfunktionen	111
13.2.3	Splines als Ansatzfunktionen	114
13.2.4	Beweis von Satz 13.2.9	123
13.3	Differenzenverfahren	128
13.3.1	Das klassische Differenzenverfahren für $-u'' = f$	129
13.3.2	Differenzenverfahren höherer Ordnung	132
13.3.3	Nichtgleichabständige Gitter	135
13.3.4	Allgemeine Differentialgleichungen zweiter Ordnung	144
13.4	Lösungen der Aufgaben	153
14	Anfangswertaufgaben	163
14.1	Einleitung, Existenzaussagen	163
14.2	Approximationsverfahren für AWA	164
14.3	Einschrittverfahren für AWA	165
14.3.1	Konstruktion und Struktur von Einschrittverfahren	165
14.3.2	Konvergenz von Einschrittverfahren	168
14.3.3	RUNGE-KUTTA-Verfahren	172
14.4	Lösungen der Aufgaben	175

