Kapitel 4

Berechnung von Lösungen

4.1 Die Euler'sche Polygonzugmethode

Die Grundlage dieser Methode ist die einfache Beobachtung, dass $f(u, t)$ in der Nähe eines Punktes als nahezu konstant angesehen werden kann. Man wählt daher eine (kleine) Schrittweite h und ersetzt $u(t)$ durch die Approximation

$$u_h(t_n) = u_h(t_{n-1}) + hf(u_h(t_{n-1}), t_{n-1})$$

Diese Methode ist einfach durchzuführen. Natürlich kann man keine übermäßig gute Approximation erwarten. Jedoch hat man den folgenden Satz, der zumindest noch theoretische Konvergenz vorhersagt. Man beachte, dass die Näherungslösung nicht für alle t im Existenzintervall definiert ist. Deshalb muß man nach Wahl einer Zahl t im Existenzintervall der Lösung sich auf Schrittweiten

$$h_n = (t - t_0)/n$$

beschränken.

Satz 4.1.1 *Sei $u(t, t_0, u_0)$ Lösung eines Anfangswertproblems der Form (2.1.2), (2.1.3) mit einer Lipschitz stetigen Funktion $f : W \times \mathbb{R} \to \mathbb{R}^n$. Es sei $u_{h_n}(t_k, t_0, u_0)$ der Wert, den das Euler'sche Verfahren nach k Schritten liefert. Dann konvergiert $u_{h_n}(t_n, t_0, u_0)$ für $n \to \infty$, d.h.*

$$u_{h_n}(t_n, t_0, u_0) \to_{n\to\infty} u(t, t_0, u_0).$$

Bevor wir diesen Satz beweisen, geben wir noch einen Hilfssatz an, der eine wichtige Abschätzung liefert.

Hilfssatz 4.1.2 *Seien $\{\zeta_n\}_{n\in\mathbb{N}_0}$ eine Folge reeller Zahlen, $\alpha \geq 0$ und $\beta > 0$ mit*

$$|\zeta_{n+1}| \leq \alpha + (1 + \beta)|\zeta_n| \; n \in \mathbb{N}_0. \tag{4.1.3}$$

Dann genügt die Folge $\{|\zeta_n|\}_{n\in\mathbb{N}}$ der Abschätzung

$$|\zeta_n| \leq e^{n\beta}|\zeta_0| + \alpha\frac{e^{n\beta} - 1}{\beta}. \tag{4.1.4}$$

Beweis des Hilfssatzes 4.1.2. Durch eine einfache Induktion erhält man die Aussage

$$|\zeta_n| \leq (1+\beta)^n |\zeta_0| + \alpha \frac{(1+\beta)^n - 1}{\beta}. \tag{4.1.5}$$

Für $n = 1$ ist dies genau die Voraussetzung (4.1.3). Angenommen die Gleichung (4.1.5) ist für n gültig. Dann folgt daraus, dass

$$|\zeta_{n+1}| \leq \alpha + (1+\beta)^{n+1}|\zeta_0| + \alpha(1+\beta)\frac{(1+\beta)^n - 1}{\beta}$$

und daher folgt

$$|\zeta_{n+1}| \leq (1+\beta)^{n+1}|\zeta_0| + \alpha \frac{(1+\beta)^{n+1} - (1+\beta) + \beta}{\beta}.$$

Dies ist (4.1.5) für $n+1$. Weil $(1+\beta)^n \leq e^{n\beta}$ ist (sogar für $\beta > -1$) folgt aus (4.1.5) die Aussage des Hilfssatzes. □

Beweis des Satzes 4.1.1. Sei $G = \{(u(s), s) \mid s \in [t_0, t] \text{ (oder } [t, t_0])\}$. O.B.d.A. dürfen wir annehmen, dass $t > t_0$ ist. Die Modifikationen für den anderen Fall sind nahezu offensichtlich. Sei W eine Umgebung von G mit kompaktem Abschluß $\overline{W}$, so dass für f bezüglich der ersten Variablen auf W eine Lipschitz Konstante M_W existiert, d.h. eine Konstante mit der die Ungleichung (2.1.6) erfüllt ist. Wie schon zuvor setzen wir wie schon zuvor $M = \max\{1, M_W\}$, $N = \sup\{|f(u,s)| \,|\, (u,s) \in W\} + 1$ und $\Delta = \operatorname{dist}(G, \partial U)$. Für $\varepsilon < \frac{\Delta}{2M}$ sei

$$G_\varepsilon = \{(u,t) \in W \mid \operatorname{dist}((u,t), G) < \varepsilon\}.$$

Für $n \in \mathbb{N}$, $t_k = t_0 + kh_n$ und $(u_1, t_k) \in G_\varepsilon$ mit $|u_1 - u(t_k, t_0, u_0)| < \varepsilon$ berechnen wir die Distanz von $(u_1, t_k) + h_n(f(u_1, t_k), 1)$ zu $u(t_{k+1}, t_0, u_0)$. Sei $u^k = u(t_k, t_0, u_0)$. Damit ist $u^{k+1} = u(t_{k+1}, t_0, u_0) = u(t_k + h_n, t_k, u^k)$. Es ergibt sich die Formel

$$u^{k+1} = u^k + \int_{t_k}^{t_k+h_n} f(u(s, t_k, u^k), s)ds = \int_0^{h_n} f(u(t_k + s, t_k, u^k), t_k + s)ds.$$

Für die Differenz erhält man daher

$$\begin{aligned}
|(u_1, t_k) + h_n(f(u_1, t_k), 1) - (u^{k+1}, t_{k+1})| &\leq |u_1 - u^k| \\
&\quad + \int_0^{h_n} |f(u_1, t_k) - f(u(t_k + s, t_k, u^k), t_k + s)|ds \\
&\leq |u_1 - u^k| + h_n|f(u_1, t_k) - f(u^k, t_k)| + h_n s(h_n) \\
&\leq (1 + Mh_n)|u_1 - u^k| + h_n s(h_n),
\end{aligned}$$

wobei $s(.)$ eine Funktion ist, mit der Eigenschaft $s(h) \to 0$ für $h \to 0$. Die Existenz einer solchen Funktion erhält man einfach:

$$\begin{aligned}
\int_0^{h_n} |f(u_1, t_k) - f(u(t_k + s, t_k, u^k), t_k + s)|ds &\leq h_n|f(u_1, t_k) - f(u^k, t_k)| \\
+ \int_0^{h_n} |f(u^k, t_k) - f(u(t_k + s, t_k, u^k), t_k + s)|ds.&
\end{aligned}$$

Aus der Differenzierbarkeit von f folgt die Zwischenbehauptung. Ist nun u_1 der Punkt der durch k Schritte des Euler Verfahrens definiert wurde, so gilt für den Abstand $\zeta_k = |u_{h_n}(t_k, t_0, u_0) - u^k|$ eine Abschätzung der Form

$$\zeta_{k+1} \leq (1 + Mh_n)|\zeta_k| + s(h_n)h_n.$$

Mit $\alpha = s(h_n)h_n$ und $\beta = Mh_n$ und dem Hilfssatz 4.1.2 schließt man, dass der Fehler nach k Schritten abzuschätzen ist durch

$$\zeta_k \leq e^{k\beta}\zeta_0 + \alpha\frac{e^{k\beta} - 1}{\beta}. \tag{4.1.6}$$

Da $\zeta_0 = 0$ ist erhält man die Abschätzung

$$\zeta_k \leq \alpha\frac{e^{k\beta} - 1}{\beta}.$$

Wir müssen noch zeigen, dass der Ausdruck rechts durch die Wahl von h_n für alle $k \leq n$ kleiner als ε wird. Dann ist der Satz bewiesen. Eine einfache Rechnung zeigt aber, dass der Ausdruck rechts mit h_n gegen Null konvergiert:

$$\begin{aligned}\zeta_k &\leq h_n s(h_n)\frac{e^{kMh_n} - 1}{Mh_n} \\ &\leq s(h_n)\frac{e^{nMh_n} - 1}{M} \\ &\leq s(h_n)\frac{e^{M(t-t_0)}}{M} \rightarrow_{h_n \to 0} 0.\end{aligned}$$

□

In Worten besagt dieser Satz, dass man auf endlichen Zeitintervallen, durch Wahl einer genügend kleinen Schrittweite die Lösung mit dem Euler Verfahren approximieren kann. Leider kann man genötigt werden, die Schrittweite sehr klein zu wählen, wobei man in der Praxis natürlich schnell an die Grenzen stößt. Wir werden dies in der Aufgabe 4.4.3 einsehen. Ein zweiter gewichtiger Nachteil ist, dass gewisse qualitative Phänomene nicht erkannt werden können. Wir werden dies an einem Beispiel sehen, vgl. Aufgabe 4.4.2. Man beachte, dass beim numerischen Rechnen noch zusätzlich Rundungsfehler auftreten.

Trotz dieser wichtigen Aussage, haben wir am Beispiel des harmonischen Oszillators gesehen, dass das Euler'sche Polygonzugverfahen ein falsches qualitatives Verhalten vorhersagt, insbesondere finden wir keine periodischen Lösungen. Wir benötigen daher andere Verfahren.

4.2 Das implizite Euler-Verfahren

Hier wird zumindest eines der beiden genannten Probleme vermieden. Natürlich kann man nicht erwarten, dass alle Mängel bei der Beschreibung gewisser qualitativer Phänomene behoben werden. Hier bestimmt man die Näherungslösung nicht dadurch, dass man tangential um die Schrittweite h weitergeht, sondern man bestimmt die $(k+1)$-Näherung aus der k-ten durch die *implizite Bedingung*

$$y_{k+1} = y_k + hf(y_{k+1}, x_{k+1}).$$

Dies bedeutet, dass man in jedem Schritt ein Gleichungssystem lösen muß. Deswegen ist die praktische Durchführung mit einem wesentlich größeren Rechenaufwand verbunden. Hat man eine lineare Abbildung $A : \mathbb{R}^n \to \mathbb{R}^n$, so lautet die Bedingung

$$y_{k+1} = y_k + hAy_{k+1}$$

und daraus erhält man

$$(\mathbb{1} - hA)y_{k+1} = y_k.$$

Ist die Schrittweite h genügend klein, so läßt sich die Matrix $\mathbb{1} - hA$ invertieren und die Iteration hat die Darstellung

$$y_{k+1} = (\mathbb{1} - hA)^{-1}y_k.$$

(Für allgemeines f (stetig, bzgl. der ersten Komponente Lipschitz stetig, folgt die eindeutige Lösbarkeit des Gleichungssystems

$$u_{k+1} - hf(u_{k+1}, t_{k+1}) = u_k$$

aus dem Satz über implizite Funktionen.) An dieser Darstellung erkennt man, dass dieses Verfahren, zumindest wenn alle Eigenwerte von A negativen Realteil haben, das Verhalten der Lösungen zufriedenstellend wiedergibt. Ist nämlich y_0 Eigenvektor von A zum Eigenwert λ mit $\mathrm{Re}\lambda < 0$ so lautet die Folge der y_k

$$y_k = \left(\frac{1}{1-h\lambda}\right)^k y_0.$$

Da der auftretende Bruch betragsmäßig kleiner als 1 ist, konvergiert die Folge der y_k gegen Null, wie es auch die wahre Lösung tut.

Definition 4.2.1 *Ein Verfahren (zur numerischen Lösung von Anfangswertproblemen) heißt* absolut stabil, *wenn für lineare Systeme, deren Eigenwerte alle negativen Realteil besitzen, jede Näherungslösung für $t \to \infty$ gegen Null konvergiert.*

Lemma 4.2.2 *Das implizite Euler-Verfahren ist absolut stabil*

Beweis. Folgt sofort aus dem zuvor gesagten. □

Tatsächlich hat man Konvergenz gegen Null für manche A mit Eigenwerten mit positiven Realteilen. Konvergenz gegen Null tritt genau dann auf, wenn $\sigma(\mathbb{1}-hA)$ außerhalb des Einheitskreises liegt. Dies ist genau dann der Fall, wenn die Eigenwerte von A außerhalb eines Kreises vom Radius $1/h$ um den Punkt $(1/h, 0)$ liegen.

4.3 Das Runge-Kutta Verfahren

Das Runge-Kutta Verfahren ist wie das Euler'sche Polygonzugverfahren ein Einschrittverfahren, d.h. von der Form

$$y_{n+1} = y_n + h\,\Phi(y_n, t, h),$$

nur, dass in diesem Fall der Schritt auf relativ komplizierte Weise berechnet wird. Wir wollen zunächst das Verfahren beschreiben und danach theoretische Eigenschaften herleiten. Soll deutlich werden, dass Φ auch von f abhängt, so schreiben wir $\Phi(t, y, h, f)$. Wir setzen

$$\Phi(t, y, h) = \frac{1}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right), \tag{4.3.1}$$

wobei die k_i durch

$$\left\{ \begin{array}{rcl} k_1 &=& f(t, y) \\ k_2 &=& f\left(t + \frac{h}{2}, y + \frac{1}{2}hk_1\right) \\ k_3 &=& f\left(t + \frac{h}{2}, y + \frac{1}{2}hk_2\right) \\ k_4 &=& f\left(t + h, y + hk_3\right) \end{array} \right. \tag{4.3.2}$$

definiert werden. Erste Tests zeigen uns, dass man mit Hilfe des Runge-Kutta-Verfahrens qualitative Eigenschaften, wie z.B. periodische Lösungen beim mathematischen Pendel, gut wiederfindet. Es stellt sich die Frage nach einer theoretischen Begründung dieser angenehmen Eigenschaften.

Dazu brauchen wir ein paar Begriffe und Definitionen. Eine wichtige Rolle spielt die Approximation des Differenzenquotienten der wahren Lösung an einer Stelle durch Φ. Wir setzen

$$\Delta(t, y, h, f) = \left\{ \begin{array}{cc} \frac{y(t+h)-y(t)}{h} & h \neq 0 \\ f(t, y) & h = 0 \end{array} \right. , \tag{4.3.3}$$

wobei $y(.)$ eine Lösung mit $y(t) = y$ ist.

Definition 4.3.4 *Der* lokale Diskretisierungsfehler *ist die Differenz*

$$\tau(t, y, h) = \Delta(t, y, h, f) - \Phi(t, y, h, f).$$

Offenkundig ist $\lim_{h\to 0} \Delta(t, y, h, f) = f(t, y)$. Also konvergiert $\Phi(t, y, h, f)$ genau dann gegen $f(t, y)$, wenn $\lim_{h\to 0} \tau(t, y, h) = 0$.

Damit sehen wir, dass der lokale Diskretisierungsfehler für die Konvergenz des Verfahrens eine große Rolle spielen wird.

Definition 4.3.5 *Ein Einschrittverfahren heißt* konsistent, *falls* $\lim_{h\to 0}\tau(t,y,h) = 0$ *erfüllt ist.*

Beispiel 4.3.6 *Das Euler'sche Polygonzugverfahren ist konsistent. Dort ist* $\Phi(t,y,h,f) = f(t,y)$*, und wegen der Konvergenz von* Δ *gegen* f *konvergiert* $\tau \to 0$.

Natürlich interessiert nicht nur, ob die Größe τ konvergiert, sondern auch die Konvergenzrate. Um diese auszurechnen, benötigen wir höhere Differenzierbarkeit von f.

Definition 4.3.7 *Wir nennen ein Einschrittverfahren von der* Ordnung $p \in \mathbb{N}$, *wenn für jedes* $f \in C^{p-1}(I \times W, \mathbb{R}^n)$ *gilt*

$$\tau(t,y,h,f) = O(|h|^p).$$

Beispiel 4.3.8 *Wir kehren zum Euler'schen Polygonzugverfahren zurück und berechnen die Ordnung des Verfahrens.*

Dazu bemerken wir, dass für eine Lösung (mit $y(t_0) = y_0$*) gilt*

$$y(t_0+h) = y_0 + h\dot{y}(t_0) + \frac{h^2}{2}\ddot{y}(t_0) + \dots. \tag{4.3.9}$$

Wegen $\dot{y}(t_0) = f(t_0,y_0)$ *und* $\dot{y} = f(t,y(t))$ *errechnet sich* $\ddot{y}(t_0)$ *zu*

$$\ddot{y}(t_0) = \frac{d}{dt}f(t,y(t))_{|t=t_0} = f_t(t_0,y_0) + f_y(t_0)\dot{y}(t_0) = f_t(t_0,y_0) + f_y(t_0,y_0)f(t_0,y_0).$$

Den lokalen Diskretisierungsfehler beim Euler'schen Polygonzugverfahren für $h \neq 0$ *erhält man nun durch Entwicklungen von* Δ *und* Φ *nach Potenzen von* h*. Zunächst ergibt sich für* $h \neq 0$

$$\begin{aligned}
\Delta(t_0,y_0,h,f) &= \frac{1}{h}\left(y(t_0+h) - y(t_0)\right) \\
&= \frac{1}{h}\left(y_0 + hf(t_0,y_0) + \frac{h^2}{2}(f_t(t_0,y_0) + f_y(t_0,y_0)f(t_0,y_0)) + \dots - y_0\right) \\
&= f(t_0,y_0) + O(|h|).
\end{aligned}$$

Wir notieren Δ *bis zur ersten Ordnung in* h *als*

$$\Delta(t_0,y_0,h,f) = f(t_0,h_0) + \frac{h}{2}(f_t(t_0,y_0) + f_y(t_0,y_0)f(t_0,y_0)). \tag{4.3.10}$$

Wegen $\Phi(t_0,y_0,h,f) = f(t_0,y_0)$ *ergibt sich für den lokalen Diskretisierungsfehler* $O(|h|)$:

$$\begin{aligned}
\tau(t_0,y_0,h,f) &= \Delta(t_0,y_0,h,f) - \Phi(t_0,y_0,h,f) \\
&= f(t_0,y_0) + \frac{h}{2}(f_t(t_0,y_0) + f_y(t_0,y_0)f(t_0,y_0)) - f(t_0,y_0) \\
&= \frac{h}{2}(f_t(t_0,y_0) + f_y(t_0,y_0)f(t_0,y_0)).
\end{aligned}$$

Also ist das Euler Verfahren ein Verfahren erster Ordnung.

Satz 4.3.11 *Das Verfahren von Runge-Kutta ist konsistent und von vierter Ordnung.*

Beweis. Aus der Definition von Φ mittels der k_j erhalten wir mit $k_0 = 0 = h_0$

$$\begin{aligned} k_j &= f(t+h_j, y+h_j k_{j-1}) \\ &= f(t,y) + f_t(t,y)h_j + f_y(t,y)h_j k_{j-1} + f_{tt}(t,y)\frac{h_j^2}{2} + \\ &\qquad + f_{ty}(t,y)h_j^2 k_{j-1} + \frac{1}{2}h_j^2 f_{yy}(t,y)[k_{j-1}, k_{j-1}] + O(|h|^3). \end{aligned}$$

Damit ergibt sich unter Ausnutzung, dass jedes der k_j in nullter Ordnung gleich $f(t,y)$ ist

$$\begin{aligned} \Phi(t,y,h,f) &= \frac{1}{6}\Big(f(t,y) + 2f(t,y) + 2f(t,y) + f(t,y) + \frac{h}{2}(2f_t(t,y) + \\ &\qquad +2f_y(t,y)k_1 + 2f_t(t,y) + 2f_y(t,y)k_2 + f_t(t,y) + f_y(t,y)k_3) + O(|h|^2)\Big) \\ &= \frac{h}{6}(3f_t(t,y) + 3f_y(t,y)f(t,y) + O(|h|^2) \\ &= \frac{1}{2}h(f_t(t,y) + f_y(t,y)f(t,y)) + O(|h|^2). \end{aligned}$$

Aufgrund der obigen Formel (4.3.10) zur Berechnung des (in h) linearen Terms des Differenzenquotienten erhalten wir

$$\tau(t,y,h,f) = O(|h|^2).$$

Eine etwas umfangreichere Rechnung ergibt auch noch die Übereinstimmung der kubischen Terme von Δ und Φ, und damit ist das Runge-Kutta-Verfahren konsistent und von vierter Ordnung. □

Bemerkung 4.3.12 Verfahren, die in der Praxis Verwendung finden, benutzen eine Schrittweitensteuerung (Verfahren von Runge-Kutta-Fehlberg). Damit läßt sich der Aufwand reduzieren. Der im Matlab-Paket integrierte Löser für gewöhnliche Differentialgleichungen beruht auf dem Runge-Kutta-Fehlberg-Verfahren.

4.4 Aufgaben

Aufgabe 4.4.1 Man schreibe die Differentialgleichung

$$\begin{aligned} \dot{u}_1 &= -u_2 \\ \dot{u}_2 &= u_1 \end{aligned}$$

in Polarkoordinaten und löse die Gleichung.

Aufgabe 4.4.2 Man schreibe Matlab-Programme die die Gleichung aus Aufgabe 4.4.1 für einen oder mehrere Anfangswerte $u_0 \neq 0$ nach dem Euler'schen Polygonzugverfahren und dem Verfahren von Runge-Kutta lösen. Man wähle verschiedene Schrittweiten und interpretiere das Ergebnis.

Aufgabe 4.4.3 Man schreibe ein Matlab-Programm das die Gleichung

$$\begin{aligned} \dot{u}_1 &= -u_1 \\ \dot{u}_2 &= -au_2 \end{aligned}$$

für die Werte $a = 1, 10, 100, 1000$ und $h = 10^{-2}$ für Anfangswerte auf der Winkelhalbierenden nach dem Euler'schen Polygonzugverfahren löst. Wie sehen die exakten Lösungen aus? wie sieht dies beim impliziten Euler Verfahren aus?

Aufgabe 4.4.4 Man untersuche die logistische Gleichung mit dem Euler'schen Polygonzugverfahren.

Aufgabe 4.4.5 (a) Eine Teilmenge $G \subset \mathbb{R}$ heißt *abgeschlossene Untergruppe* von $\mathbb{R}$, wenn G eine abgeschlossene Teilmenge und gleichzeitig, bezüglich der Addition, eine Untergruppe ist. Man beschreibe alle abgeschlossenen Untergruppen von $\mathbb{R}$.

(b) Gegeben sei eine stetige Funktion $x : \mathbb{R} \to \mathbb{R}^n$. Eine Zahl $T \in \mathbb{R}$ heißt Periode von x, wenn für alle $t \in \mathbb{R}$ gilt $x(t+T) = x(t)$. Man zeige, die Menge

$$P_x = \{T \in \mathbb{R} \mid T \text{ ist Periode von } x\}$$

ist eine abgeschlossene Untergruppe von $\mathbb{R}$.

(c) Man zeige, gibt es für die Lösung $u(t, t_0, u_0)$ des autonomen Anfangswertproblems

$$\dot{u} = f(u),\ u(t_0) = u_0$$

zwei verschiedene Zeitpunkte t_1, t_2 mit $u(t_1, t_0, u_0) = u(t_2, t_0, u_0)$, so existiert die Lösung für alle $t \in \mathbb{R}$ und ist periodisch, d.h. es existiert eine Zahl $T > 0$ mit $u(t+T, t_0, u_0) = u(t, t_0, u_0)$.