Kapitel 6

Numerische Behandlung von Anfangswertproblemen gewöhnlicher Differentialgleichungen

6.1 Einführung

Viele Anwendungen aus Naturwissenschaft, Technik und Wirtschaft führen auf Anfangswertprobleme für gewöhnliche Differentialgleichungen.

Anfangswertproblem: Gegeben sei eine Funktion $f : [a,b] \times \mathbb{R}^n \to \mathbb{R}^n$ und ein Anfangswert $y_0 \in \mathbb{R}^n$. Gesucht ist eine Funktion $y : [a,b] \to \mathbb{R}^n$, deren Ableitung y' eine *gewöhnlichen Differentialgleichung* der Form

$$y'(t) = f(t, y(t)), \quad t \in [a,b]$$

erfüllt und die zudem der *Anfangsbedingung*

$$y(a) = y_0$$

genügt. Also kurz

$$\text{(AWP)} \qquad \begin{aligned} &y'(t) = f(t, y(t)), \quad t \in [a,b] && (6.1)\\ &y(a) = y_0 && (6.2) \end{aligned}$$

In vielen Fällen bezeichnet t die Zeit, was die Bezeichnung Anfangswertproblem rechtfertigt.

Anwendungen: Bewegungsgleichungen (z.B. Fahrdynamik, Planetenbewegung), Reaktionskinetik, Schaltkreissimulation, etc.

Grundlegend für die Existenz und Eindeutigkeit einer Lösung von (AWP) ist der folgende

Satz 6.1.1 (Existenz- und Eindeutigkeitssatz)
$f : [a,b] \times \mathbb{R}^n \to \mathbb{R}^n$ sei stetig. Ferner gebe es eine feste Zahl $L > 0$ mit

$$\|f(t,y) - f(t,z)\| \le L\|y - z\| \quad \textit{für alle } t \in [a,b] \textit{ und } y, z \in \mathbb{R}^n \quad \textit{(Lipschitz-Bedingung).}$$

Dann gilt:

a) *(Picard/Lindelöf) Zu jedem $y_0 \in \mathbb{R}^n$ besitzt (AWP) genau eine Lösung $y \in C^1([a,b];\mathbb{R}^n)$.*

b) *Sind y, z Lösungen zu den Anfangswerten $y(a) = y_0$ bzw. $z(a) = z_0$, dann gilt*

$$\|y(t) - z(t)\| \le e^{L(t-a)}\|y_0 - z_0\| \quad \forall\, t \in [a,b]. \tag{6.3}$$

Für einen Beweis siehe z.B. Heuser [Heu89] oder Walter [Wa86]. Teil b) ist eine Folge des Lemmas von Gronwall.

Bemerkung: Teil b) besagt, dass die Lösung *stetig* vom Anfangswert y_0 abhängt.

6.1.1 Grundkonzept numerischer Verfahren

Zur numerischen Lösung von (AWP) zerlegen wir das Intervall $[a,b]$ in Teilintervalle:

$$t_j = a + jh, \quad j = 0, 1, \ldots, N, \quad h = \frac{b-a}{N}.$$

Durch Integration von (AWP) erhält man mit der Abkürzung $y_j = y(t_j)$

$$y_{j+1} = y_j + \int_{t_j}^{t_{j+1}} y'(t)\,dt = y_j + \int_{t_j}^{t_{j+1}} f(t, y(t))\,dt. \tag{6.4}$$

Das Integral rechts kann nicht exakt berechnet werden, da $y(t)$ unbekannt ist. Wir approximieren daher das Integral durch interpolatorische Quadratur und erhalten hieraus einen numerischen Algorithmus zur Berechnungen von Näherungen

$$u_j \approx y(t_j), \quad j = 1, \ldots, N, \quad u_0 = y_0.$$

Den Fehler

$$e_j = y(t_j) - u_j$$

bezeichnet man als *Diskretisierungsfehler*.

6.1.2 Einige wichtige Verfahren

Approximiert man das Integral in (6.4) durch die Rechtecksregel, wobei wir das linke Intervallende als Stützpunkt verwenden, also

$$\int_{t_j}^{t_{j+1}} f(t, y(t))\, dt \approx h f(t_j, y_j),$$

dann erhalten wir das

Explizite Euler-Verfahren:

$$\begin{aligned} u_0 &:= y_0 \\ u_{j+1} &:= u_j + h f(t_j, u_j), \quad j = 0, \ldots, N-1. \end{aligned} \tag{6.5}$$

Verwenden wir zur Approximation des Integrals die Rechtecksregel mit dem rechten Randpunkt t_{j+1} als Stützstelle, dann erhalten wir das

Implizite Euler-Verfahren:

$$\begin{aligned} u_0 &:= y_0 \\ u_{j+1} &:= u_j + h f(t_{j+1}, u_{j+1}), \quad j = 0, \ldots, N-1. \end{aligned} \tag{6.6}$$

Hierbei ist zu beachten, dass für jedes j die Gleichung nach u_{j+1} aufgelöst werden muss.

Approximiert man das Integral in (6.4) durch die Trapezregel, dann erhält man

$$u_{j+1} = u_j + \frac{h}{2}\left(f(t_j, u_j) + f(t_{j+1}, u_{j+1})\right).$$

Die rechte Seite hängt von u_{j+1} ab, das Verfahren ist also implizit. Ersetzt man rechts u_{j+1} durch den expliziten Euler-Schritt $u_{j+1} = u_j + h f(t_j, u_j)$, dann ergibt sich das

Verfahren von Heun, erstes Runge-Kutta-Verfahren 2. Ordnung: (Heun, 1900)

$$u_0 = y_0, \quad u_{j+1} = u_j + \frac{h}{2}\left(f(t_j, u_j) + f(t_{j+1}, u_j + h f(t_j, u_j))\right), j = 0, \ldots, N-1.$$

Das Verfahren kann auch geschrieben werden als

$$u_{j+1} = u_j + \frac{h}{2}(k_1 + k_2)$$

mit $k_1 = f(t_j, u_j)$, $k_2 = f(t_{j+1}, u_j + h k_1)$.

Approximieren wir das Integral durch die Mittelpunktsregel und $u_{j+1/2}$ durch den Euler-Schritt $u_j + h/2 f(t_j, u_j)$, dann ergibt sich das

Modifizierte Euler-Verfahren, zweites Runge-Kutta-Verfahren 2. Ordnung: (Runge, 1895)

$$u_0 = y_0, \quad u_{j+1} = u_j + hf(t_j + h/2, u_j + (h/2)\, f(t_j, u_j)), \quad j = 0, \ldots, N-1.$$

Das Verfahren kann auch geschrieben werden als

$$u_{j+1} = u_j + hk_2$$

mit $k_1 = f(t_j, u_j)$, $k_2 = f(t_j + h/2, u_j + (h/2)\, k_1)$.

Wenden wir schließlich die Simpson-Regel an und ersetzen $u_{j+1/2}$, u_{j+1} geeignet durch Taylorentwicklungen, dann ergibt sich das sehr genaue und beliebte

Klassische Runge-Kutta-Verfahren 4. Ordnung (RK4)

$$\begin{aligned}
u_0 &= y_0 \\
u_{j+1} &= u_j + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4), \quad j = 0, \ldots, N-1 \\
\text{mit } \; k_1 &= f(t_j, u_j) \\
k_2 &= f(t_j + h/2, u_j + (h/2)\, k_1) \\
k_3 &= f(t_j + h/2, u_j + (h/2)\, k_2) \\
k_4 &= f(t_{j+1}, u_j + hk_3)
\end{aligned}$$

6.1.3 Konvergenz und Konsistenz

Wir wollen nun die vorgestellten Verfahren auf ihre praktische Brauchbarkeit und Genauigkeit hin untersuchen. Die Verfahren lassen sich in der allgemeinen Form

$$\begin{aligned}
u_0 &= y_0 \\
u_{j+1} &= u_j + h\, \phi(t_j, h; u_j, u_{j+1}), \quad j = 0, \ldots, N-1,
\end{aligned} \tag{6.7}$$

schreiben,

Definition 6.1.2 *Die Funktion $\phi(t, h; u, v)$ in* (6.7) *heißt* Verfahrensfunktion. *Hängt ϕ nicht von v ab, dann heißt das Verfahren* explizit, *sonst* implizit.

Die Größe

$$\begin{aligned}
\tau(t, h) &= \frac{1}{h}\left(y(t+h) - y(t) - h\phi(t, h; y(t), y(t+h))\right), \quad h > 0, \;\; t \in [a, b-h], \\
&= 1/h \times \textit{ Defekt bei Einsetzen der Lösung in das Verfahren}
\end{aligned}$$

heißt der lokale Abbruchfehler *oder* Konsistenzfehler *des Verfahrens* (6.7) *für (AWP) an der Stelle t.*

Definition 6.1.3 *Das Verfahren* (6.7) *heißt zu (AWP) konsistent von der Ordnung p, falls es Konstanten $C > 0$ und $\bar{h} > 0$ gibt mit*

$$\|\tau(t,h)\| \le Ch^p \quad \textit{für alle } 0 < h \le \bar{h} \textit{ und alle } t \in [a, b-h].$$

Das Verfahren (6.7) *heißt* stabil*, falls eine Konstante $K > 0$ existiert mit*

$$\|\phi(t,h;u,v) - \phi(t,h;\tilde{u},\tilde{v})\| \le K\left(\|u-\tilde{u}\| + \|v-\tilde{v}\|\right) \quad \textit{für alle } t \in [a,b]\ u,v,\tilde{u},\tilde{v} \in \mathbb{R}^n.$$

Das Verfahren (6.7) *heißt* konvergent von der Ordnung p*, falls mit Konstanten $M > 0, H > 0$ gilt*

$$\|e_j\| = \|y(t_j) - u_j\| \le Mh^p, \quad \textit{für } j = 0,\dots,N \textit{ und alle } h = \tfrac{b-a}{N} \le H.$$

Besipiel: Explizites Euler-Verfahren Das Euler-Verfahren hat Konsistenzordnung 1.

Nachweis: Sei $f \in C^1([a,b] \times \mathbb{R}^n; \mathbb{R}^n)$ und y Lösung von $y' = f(t,y)$. Dann ist $y' \in C^1([a,b];\mathbb{R}^n)$, also $y \in C^2([a,b];\mathbb{R}^n)$ und Taylorentwicklung liefert komponentenweise mit einem $\xi_i \in [0,1]$

$$y(t+h) = y(t) + y'(t)h + \frac{1}{2}(y_i''(t+\xi_i h))_{1\le i\le n}h^2 = y(t) + f(t,y(t))h + \frac{1}{2}(y_i''(t+\xi_i h))_{1\le i\le n}h^2.$$

Also ergibt sich

$$\begin{aligned}\|\tau(t,h)\| &= \left\|\frac{1}{h}(y(t+h) - y(t) - hf(t,y(t)))\right\| = \frac{1}{2}\|(y_i''(t+\xi_i h))_{1\le i\le n}\|h \\ &\le \frac{1}{2}\|(\sup_{s\in[a,b]} |y_i''(s)|)_{1\le i\le n}\|h.\end{aligned}$$

Damit hat das Euler-Verfahren Konsistenzordnung 1. □

Verfahren	Konsistenzordnung
Expl. Euler	1
Impl. Euler	1
Heun	2
Mod. Euler	2
RK4	4

6.1.4 Ein Konvergenzsatz

Wir beweisen nun einen grundlegenden Konvergenzsatz für explizite Einschrittverfahren.

Satz 6.1.4 *Sei $y \in C^1([a,b];\mathbb{R}^n)$ Lösung von (AWP). Das Verfahren* (6.7) *sei konsistent von der Ordnung p und stabil. Dann ist das Verfahren konvergent von der Ordnung p. Genauer gibt es $H > 0$, so dass für den globalen Diskretisierungsfehler gilt*

$$\|e_j\| = \|y(t_j) - u_j\| \le \frac{e^{4K|t_j-a|}-1}{4K} 2Ch^p \quad \text{für } j = 0,\dots,N \text{ und alle } h = \tfrac{b-a}{N} \le H.$$

Beweis: (für Interessierte) Setze

$$y_j = y(t_j), \quad e_j = y_j - u_j, \quad j = 0,\dots,N.$$

Dann gilt für $j = 0,\dots,N-1$ nach Definition des Verfahrens (6.7) und des lokalen Diskretisierungsfehlers

$$\begin{aligned} u_{j+1} &= u_j + h\phi(t_j,h;u_j,u_{j+1}),\\ y_{j+1} &= y_j + h\phi(t_j,h;y_j,y_{j+1}) + h\tau(t_j,h). \end{aligned}$$

Subtraktion der ersten von der zweiten Gleichung ergibt

$$e_{j+1} = e_j + h(\phi(t_j,h;y_j,y_{j+1}) - \phi(t_j,h;u_j,u_{j+1})) + h\tau(t_j,h).$$

Sei nun $0 < h = (b-a)/N \le \bar{h}$. Wegen $t_j \in [a, b-h]$ liefert die Konsistenzbedingung $\|\tau(t_j,h)\| \le Ch^p$. Zusammen mit der Stabilität des Verfahrens erhalten wir daher mit der Dreiecksungleichung

$$\|e_{j+1}\| \le (1+hK)\|e_j\| + hK\|e_{j+1}\| + hCh^p$$

Wähle nun $0 < H \le \bar{h}$ so klein, dass gilt $HK \le 1/2$. Dann ergibt sich für alle $0 < h = (b-a)/N \le H$

$$\|e_{j+1}\| \le \frac{1+hK}{1-hK}\|e_j\| + h2Ch^p \le (1+h4K)\|e_j\| + h2Ch^p$$

Das nachfolgende Lemma liefert nun wegen $e_0 = 0$

$$\|e_{j+1}\| \le \frac{e^{4K|t_{j+1}-a|}-1}{4K} 2Ch^p.$$

Damit ist der Satz bewiesen. □

Wir benötigen zur Vervollständigung des Beweises noch das folgende *diskrete Gronwall-Lemma* zur Abschätzung der Fehlerakkumulation.

Lemma 6.1.5 *Für Zahlen $L > 0$, $a_j \ge 0$, $h_j > 0$ und $b \ge 0$ sei*

$$a_{j+1} \le (1 + h_j L)a_j + h_j b, \quad j = 0,1,\dots,n-1.$$

Dann gilt

$$a_j \le \frac{e^{Lt_j}-1}{L} b + e^{Lt_j} a_0 \quad \text{mit } t_j := \sum_{i=0}^{j-1} h_i.$$

Beweis: (für Interessierte) Für $j = 0$ ist die Behauptung klar. Der Induktionsschritt $j \to j+1$ ergibt sich aus

$$
\begin{aligned}
a_{j+1} &\le \underbrace{(1 + h_j L)}_{\le e^{h_j L}} \left(\frac{e^{Lt_j} - 1}{L} b + e^{Lt_j} a_0 \right) + h_j b \\
&\le \left(\frac{e^{L(t_j+h_j)} - 1 - h_j L}{L} + h_j \right) b + e^{L(t_j+h_j)} a_0 \\
&= \frac{e^{Lt_{j+1}} - 1}{L} b + e^{Lt_{j+1}} a_0
\end{aligned}
$$

□

6.1.5 Explizite Runge-Kutta-Verfahren

Verfahren hoher Konsistenzordnung kann man durch eine Verallgemeinerung des Ansatzes beim RK4-Verfahren gewinnen:

r-stufiges explizite Runge-Kutta-Verfahren:

Hier wählt man die Verfahrensfunktion

$$
\begin{aligned}
k_i &= f\left(t + \gamma_i h, u + h \sum_{j=1}^{i-1} \alpha_{ij} k_j \right), \quad i = 1, \dots, r, \\
\phi(t, h; u) &= \sum_{i=1}^{r} \beta_i k_i.
\end{aligned}
\tag{6.8}
$$

Hierbei heißt $k_i = k_i(t, u, h)$ die *i-te Stufe*. Zur kompakten Beschreibung von expliziten Runge-Kutta-Verfahren notiert man die Koeffizienten in einem Tableau, dem sogenannten

Butcher-Schema:

$$
\begin{array}{c|ccccc}
\gamma_1 & 0 & & & & \\
\gamma_2 & \alpha_{21} & 0 & & & \\
\gamma_3 & \alpha_{31} & \alpha_{32} & 0 & & \\
\vdots & \vdots & \vdots & \ddots & \ddots & \\
\gamma_r & \alpha_{r1} & \cdots & \cdots & \alpha_{r,r-1} & 0 \\
\hline
 & \beta_1 & \beta_2 & \cdots & \beta_{r-1} & \beta_r
\end{array}
$$

Beispiele für Butcher-Schemata:

Explizites Euler-Verfahren:

$$
\begin{array}{c|c}
0 & 0 \\
\hline
 & 1
\end{array}
$$

Modifiziertes Euler-Verfahren:

$$
\begin{array}{c|cc}
0 & 0 & \\
1/2 & 1/2 & 0 \\
\hline
 & 0 & 1
\end{array}
$$

Verfahren von Heun:

$$
\begin{array}{c|cc}
0 & 0 & \\
1 & 1 & 0 \\
\hline
 & 1/2 & 1/2
\end{array}
$$

Mit diesem Ansatz kann man Verfahren beliebiger Konsistenzordnung p erzeugen. Man muss hierzu die Stufenzahl r groß genug wählen. Taylorentwicklung des lokalen Abbruchfehlers liefert dann Gleichungen für die Koeffizienten.

Durch Taylorentwicklung läßt sich der folgende Satz beweisen.

Satz 6.1.6 *Betrachte ein Runge-Kutta Verfahren* (6.7) *mit Verfahrensfunktion* (6.8) *mit*

$$\gamma_i = \sum_{j=1}^{r} \alpha_{ij} \quad i = 1, \ldots, r.$$

Es besitzt genau dann für jede rechte Seite $f \in C^p([a,b] \times \mathbb{R})$ *die Konsistenzordnung* $p = 1$, *falls die Koeffizienten der Gleichung*

$$\sum_{i=1}^{r} \beta_i = 1$$

genügen; genau dann die Konsistenzordnung $p = 2$, *falls die Koeffizienten zusätzlich der Gleichung*

$$\sum_{i=1}^{r} \beta_i \gamma_i = 1/2$$

genügen; genau dann die Konsistenzordnung $p = 3$, *falls die Koeffizienten zusätzlich den Gleichungen*

$$\sum_{i=1}^{r} \beta_i \gamma_i^2 = 1/3$$
$$\sum_{i,j=1}^{r} \beta_i \alpha_{ij} \gamma_j = 1/6$$

genügen; genau dann die Konsistenzordnung $p = 4$, *falls die Koeffizienten zusätzlich den Gleichungen*

$$\sum_{i=1}^{r} \beta_i \gamma_i^3 = 1/4$$
$$\sum_{i,j=1}^{r} \beta_i \gamma_i \alpha_{ij} \gamma_j = 1/8$$
$$\sum_{i,j=1}^{r} \beta_i \alpha_{ij} \gamma_j^2 = 1/12$$
$$\sum_{i,j,k=1}^{r} \beta_i \alpha_{ij} \alpha_{jk} \gamma_k = 1/24$$

genügen.

Beweis: Siehe zum Beispiel Deuflhard und Bornemann [DB02]. □

6.2 Steife Differentialgleichungen

In zahlreichen Anwendungen (z.B. beim Ablauf chemischer Rekationen), aber auch bei Semidiskretisierung partieller Differentialgleichungen, treten *steife Systeme* auf. Obwohl es sich auch um Anfangswertprobleme handelt, erzwingen sie bei vielen – aber nicht bei allen – Verfahren inakzeptabel kleine Schrittweiten h, um eine genaue Lösung zu erhalten.

Ausgangspunkt ist ein Anfangswertproblem für ein System n gewöhnlicher Differentialgleichungen:

$$\text{(AWPn)} \qquad \begin{array}{l} y'(t) = f(t, y(t)), \quad t \in [a, b] \\ y(a) = y_0 \end{array}$$

mit $f : [a, b] \times \mathbb{R}^n \to \mathbb{R}^n$, $y_0 \in \mathbb{R}^n$.

Der Begriff "steifes System" ist in der Literatur nicht ganz einheitlich definiert. Der wesentlich Punkt ist, dass die Lösung zusammengesetzt ist aus einem langsam veränderlichen Teil (der meist abklingt) und einem Anteil, die im Allgemeinen sehr schnell gedämpft wird.

Wir betrachten den Spezialfall, dass (AWPn) linear ist:

$$\text{(LAWPn)} \qquad \begin{array}{l} y'(t) = Ay(t) + b, \quad t \in [a, b] \\ y(a) = y_0 \end{array}$$

mit einer Matrix $A \in \mathbb{R}^{n,n}$ und einem Vektor $b \in \mathbb{R}^n$.

Sei zudem $A \in \mathbb{R}^{n,n}$ diagonalisierbar mit zugehörigen Eigenwerten λ_i sowie Eigenvektoren v_i. Mit einer partikulären Lösung y_P ist dann die allgemeine Lösung von der Form

$$y(t) = y_H(t) + y_P(t), \quad y_H(t) = \sum_{i=1}^{n} C_i e^{\lambda_i t} v_i.$$

Ist nun $\mathrm{Re}(\lambda_i) < 0$ für $i = 1, \ldots, n$, so gilt

$$\lim_{t \to \infty} y_H(t) \to 0,$$

alle Lösungen nähern sich also y_P an. Hierbei klingen die Summanden in y_H mit $\mathrm{Re}(\lambda_i) \ll -1$ sehr schnell und Summanden mit $\mathrm{Re}(\lambda_i) \not\ll -1$ deutlich langsamer ab. Gibt es Eigenwerte mit $\mathrm{Re}(\lambda_i) \ll -1$ und Eigenwerte mit schwach negativem Realteil, so nennt man das System *steif*.

Beispiel: Betrachte zum Beispiel das Problem

$$y' = Ay, \quad y(0) = y_0 := \begin{pmatrix} C_1 + C_2 \\ C_1 - C_2 \end{pmatrix}$$

mit $C_1, C_2 \in \mathbb{R}$ und

$$A = \begin{pmatrix} \frac{\lambda_1+\lambda_2}{2} & \frac{\lambda_1-\lambda_2}{2} \\ \frac{\lambda_1-\lambda_2}{2} & \frac{\lambda_1+\lambda_2}{2} \end{pmatrix}.$$

A hat die Eigenwerte λ_1, λ_2 mit zugehörigen Eigenvektoren $\binom{1}{1}$ bzw. $\binom{1}{-1}$. Ist zum Beispiel $\lambda_1 = -1$ und $\lambda_2 = -1000$, dann lautet die Lösung

$$y(t) = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-t} + C_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-1000t}.$$

Der zweite Term spielt nach kürzester Zeit so gut wie keine Rolle mehr. Der erste Term ist bestimmend und konvergiert für $t \to \infty$ ebenfalls gegen 0. Von einem geeigneten Integrationsverfahren wird man erwarten, dass es ohne große Einschränkungen an die Schrittweite Näherungen u_j liefert mit

$$\lim_{j \to \infty} u_j = 0.$$

Betrachten wir jedoch zum Beispiel die Anwendung des expliziten Euler-Verfahrens, so ergibt sich mit $u_0 = y_0 = C_1 \binom{1}{1} + C_2 \binom{1}{-1}$

$$u_1 = (I + hA)u_0 = C_1(1 + h\lambda_1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2(1 + h\lambda_2) \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

und nun induktiv

$$u_j = C_1(1 + h\lambda_1)^j \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2(1 + h\lambda_2)^j \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Ist $C_2 \neq 0$, so müssen wir $|1 + h\lambda_2| < 1$, also $-h\lambda_2 = 1000h < 2$ wählen, damit gilt $\lim_{j \to \infty} u_j = 0$. Ein geeignetes Verfahren sollte dies möglichst für alle $h > 0$ sicherstellen. $\square$

Das Euler-Verfahren benötigt also sehr kleine Schrittweiten, obwohl sich die Lösung kaum ändert. Man nennt die Differentialgleichung dann *steif*. Die formale Definition ist uneinheitlich. Folgende Definition ist am weitesten verbreitet.

Definition 6.2.1 *Ein Anfangswertproblem (LAWPn) heißt* steif, *wenn A Eigenwerte mit* $\mathrm{Re}(\lambda_i) \ll -1$ *und Eigenwerte λ_i mit schwach negativem Realteil besitzt.*

Wir kommen nun zur numerischen Behandlung steifer Differentialgleichungen.

Um eine einfache Modellgleichung für steife Differentialgleichungen herzuleiten, betrachten wir zunächst (LAWPn) mit $b = 0$, also

$$y' = Ay, y(0) = y_0. \tag{6.9}$$

A sei diagonalisierbar. Dann existiert $M \in \mathbb{R}^{n,n}$ mit

$$MAM^{-1} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$$

mit den Eigenwerten $\lambda_1, \ldots, \lambda_n$ von A. Setzen wir $z = My$, dann gilt also

$$z' = MAy = MAM^{-1}z = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)z, \quad z(0) = My_0.$$

Die Komponenten z_i von $z = My$ erfüllen also

$$z_i' = \lambda_i z_i, \quad z_i(0) = (My_0)_i. \tag{6.10}$$

Für eine steife Differentialgleichung gilt zudem $\mathrm{Re}(\lambda_i) \le 0$, wobei einige $\mathrm{Re}(\lambda_i)$ stark, andere schwach negativ sind.

Beobachtung: Verhält sich ein numerisches Verfahren für alle Differentialgleichungen (6.10) gutartig, dann liefert es auch für das steife System (6.9) gute Ergebnisse.

Um Verfahren für steife Differentialgleichungen zu bewerten und zu analysieren, betrachtet man daher nach Dahlquist (1963) die

Modellgleichung

$$y' = \lambda y, \quad y(0) = 1, \quad \text{mit } \lambda \in \mathbb{C}, \ \mathrm{Re}(\lambda) < 0. \tag{6.11}$$

Die Lösung ist

$$y(t) = e^{\lambda t}$$

und wegen $\mathrm{Re}(\lambda) < 0$ gilt

$$\lim_{t\to\infty} y(t) = 0. \tag{6.12}$$

Die Lösung fällt also je nach größe von $|\mathrm{Re}(\lambda)|$ sehr unterschiedlich stark ab. Damit ein Verfahren gut für steife Differentialgleichungen geeignet ist, hat sich folgende Anforderung bewährt:

Forderung: Die numerische gewonnene Näherungslösung von (6.11) soll die Eigenschaften von der Lösung $y(t) = e^{\lambda t}$, also insbesondere (6.12), möglichst gut widerspiegeln.

Dies motiviert folgende

Definition 6.2.2 (A-stabil (absolut stabil), L-stabil)
Ein Verfahren heißt

a) absolut stabil (A-stabil)*, wenn seine Anwendung auf das Modellproblem* (6.11) *für jede Schrittweite $h > 0$ eine Folge $\{u_j\}_{j\in\mathbb{N}_0}$ produziert mit*

$$|u_{j+1}| \le |u_j| \quad \forall\, j \ge 0.$$

b) L-stabil*, wenn es A-stabil ist und zudem gilt*

$$\lim_{j\to\infty} u_j = 0.$$

Bei vielen Einschrittverfahren gilt bei Anwendung auf das Modellproblem (6.11) die Beziehung

$$u_{j+1} = R(q)u_j \quad \text{mit } q = \lambda h$$

und einer Funktion $R : D \to \mathbb{C}$, $0 \in D \subset \mathbb{C}$.

Beispiel: Wendet man das explizite Euler-Verfahren auf das Modellproblem (6.11) an, dann erhält man

$$u_{j+1} = u_j + h\lambda u_j = (1 + \lambda h)u_j = (1 + q)u_j,$$

also ist für das explizite Euler-Verfahren $R(q) = 1 + q$. □

Definition 6.2.3 *Man nennt R die* Stabilitätsfunktion *des Einschrittverfahrens. Die Menge*

$$S = \{q \in \mathbb{C} : |R(q)| < 1\}.$$

heißt Stabilitätsgebiet *des Einschrittverfahrens.*

Offensichtlich gilt

$$\text{A-stabil} \iff |R(q)| \le 1 \quad \forall\, q \in \mathbb{C},\ \text{Re}(q) < 0.$$

$$\text{L-stabil} \iff |R(q)| < 1 \quad \forall\, q \in \mathbb{C},\ \text{Re}(q) < 0 \iff S \supset \{q \in \mathbb{C} : \text{Re}(q) < 0\}.$$

6.2.1 Stabilitätsgebiete einiger Verfahren

Explizites Euler-Verfahren

Anwendung des expliziten Euler-Verfahrens auf das Modellproblem (6.11) ergibt

$$u_{j+1} = u_j + h\lambda u_j = (1 + \lambda h)u_j,$$

die Stabilitätsfunktion ist daher $R(q) = 1 + q$. Das Stabilitätsgebiet ist also

$$S = \{q \in \mathbb{C} : |1 + q| < 1\}.$$

Bemerkung: Man kann leicht zeigen, dass alle expliziten Runge-Kutta-Verfahren nicht A-stabil sind!

Implizites Euler-Verfahren

Das implizite Euler-Verfahren liefert für das Modellproblem (6.11)

$$u_{j+1} = u_j + h\lambda u_{j+1}$$

und somit

$$u_{j+1} = \frac{1}{1-\lambda h} u_j.$$

Dies ergibt die Stabilitätsfunktion $R(q) = \frac{1}{1-q}$, $q \neq 1$, und das Stabilitätsgebiet

$$S = \{q \in \mathbb{C} \,:\, |1-q| > 1\} \supset \{q \in \mathbb{C} \,:\, \mathrm{Re}(q) < 0\}\,.$$

Das implizite Euler-Verfahren ist also A-stabil, sogar L-stabil!

Implizite Trapezregel

Die Verfahrensgleichung lautet

$$u_{j+1} = u_j + \frac{h}{2}(f(u_j) + f(u_{j+1})).$$

Wir erhalten für das Modellproblem (6.11)

$$u_{j+1} = u_j + \frac{h}{2}\lambda(u_j + u_{j+1})$$

und somit

$$u_{j+1} = \frac{1+\lambda h/2}{1-\lambda h/2} u_j.$$

Daher gilt $R(q) = \frac{1+q/2}{1-q/2}$, $q \neq 2$, und das Stabilitätsgebiet ist

$$S = \{q \in \mathbb{C} \,:\, |1+q/2| < |1-q/2|\} = \{q \in \mathbb{C} \,:\, \mathrm{Re}(q) < 0\}\,.$$

Implizite Runge-Kutta-Verfahren

Besonders gut geeignet für steife Differentialgleichungen sind implizite Runge-Kutta-Verfahren.

Implizite Runge-Kutta-Verfahren erhält man durch Butcher-Schemata, bei denen die Koeffizienten α_{ij} *keine* strikte untere Dreiecksmatrix bilden. Die Verfahrensgleichung ist gegeben durch

$$\begin{aligned} k_i &= f\left(t + \gamma_i h, u + h\sum_{l=1}^{r} \alpha_{il} k_l\right), \quad i = 1, \ldots, r, \\ \phi(t, h; u) &= \sum_{i=1}^{r} \beta_i k_i. \end{aligned} \tag{6.13}$$

(beachte die Summation bis r anstelle $i-1$). Ein implizites Runge-Kutta-Verfahren ist ein explizites Einschrittverfahren, lediglich die Stufen k_i sind als Lösung eines nichtlinearen Gleichungssystems gegeben. Man kann nun die Koeffizienten α_{ij}, β_i, γ_i tatsächlich so wählen, dass ein L-stabiles Verfahren der Ordnung $p = 2r$ ensteht.