# Gewöhnliche Differentialgleichungen und ihre numerische Behandlung

Jochen Werner

Wintersemester 2001/2002

# Inhaltsverzeichnis


**1 Einführung, Beispiele, Grundlagen** **1**
1.1 Wachstumsmodelle . . . 3
1.1.1 Populationsmodelle . . . 3
1.1.2 Das Räuber-Beute-Modell . . . 6
1.1.3 Aufgaben . . . 13
1.2 Beispiele aus der Physik . . . 16
1.2.1 Das mathematische Pendel und andere schwingende Systeme . . . 16
1.2.2 Planetenbahnen . . . 22
1.2.3 Aufgaben . . . 31
1.3 Elementar lösbare Differentialgleichungen . . . 34
1.3.1 Differentialgleichung mit getrennten Veränderlichen . . . 36
1.3.2 Lineare Differentialgleichungen erster Ordnung . . . 39
1.3.3 Bernoullische Differentialgleichung . . . 42
1.3.4 Riccatische Differentialgleichung . . . 43
1.3.5 Aufgaben . . . 44
1.4 Funktionalanalytische Grundlagen . . . 46
1.4.1 Der Fixpunktsatz für kontrahierende Abbildungen . . . 47
1.4.2 Der Brouwersche und der Schaudersche Fixpunktsatz . . . 49
1.4.3 Der Satz von Arzela-Ascoli . . . 52
1.4.4 Aufgaben . . . 53

**2 Die Theorie gewöhnlicher Anfangswertaufgaben** **57**
2.1 Existenz- und Eindeutigkeitsaussagen . . . 58
2.1.1 Der Satz von Picard-Lindelöf . . . 58
2.1.2 Der Satz von Peano . . . 64
2.1.3 Das Lemma von Gronwall . . . 67
2.1.4 Aufgaben . . . 69
2.2 Lineare Differentialgleichungssysteme . . . 72
2.2.1 Lineare Systeme mit variablen Koeffizienten . . . 72
2.2.2 Lineare Systeme mit konstanten Koeffizienten . . . 76
2.2.3 Periodische lineare Systeme . . . 90
2.2.4 Zweidimensionale autonome lineare Systeme . . . 92
2.2.5 Aufgaben . . . 97
2.3 Stabilität . . . 100
2.3.1 Definitionen . . . 100

2.3.2 Stabilität bei linearen Systemen mit konstanten Koeffizienten . . 101
2.3.3 Stabilität bei linearen Systemen mit variablen Koeffizienten . . . 103
2.3.4 Periodische lineare Systeme . . . . . 105
2.3.5 Stabilität bei nichtlinearen Systemen . . . . . 106
2.3.6 Aufgaben . . . . . 112

**3 Die numerische Behandlung gewöhnlicher Anfangswertaufgaben 115**
3.1 Einschrittverfahren . . . . . 116
3.1.1 Beispiele von Einschrittverfahren . . . . . 116
3.1.2 Konsistenz von Einschrittverfahren . . . . . 123
3.1.3 Konvergenz von Einschrittverfahren . . . . . 126
3.1.4 Einschrittverfahren und Extrapolation . . . . . 129
3.1.5 Schrittweitensteuerung . . . . . 132
3.1.6 Aufgaben . . . . . 134
3.2 Mehrschrittverfahren . . . . . 135
3.2.1 Beispiele von Mehrschrittverfahren . . . . . 136
3.2.2 Konsistenz, Konvergenz und Stabilität: Definitionen . . . . . 141
3.2.3 Der Äquivalenzsatz . . . . . 147
3.2.4 Lineare Mehrschrittverfahren . . . . . 155
3.2.5 Aufgaben . . . . . 160
3.3 MATLAB-Funktionen für nicht-steife Differentialgleichungen . . . . . 162
3.3.1 ODE45 . . . . . 163
3.3.2 ODE23 . . . . . 166
3.3.3 Aufgaben . . . . . 167
3.4 Steife Differentialgleichungen . . . . . 168
3.4.1 Beispiele, Motivation . . . . . 168
3.4.2 Stabilitätsgebiet expliziter Runge-Kutta-Verfahren . . . . . 172
3.4.3 Stabilitätsgebiet linearer Mehrschrittverfahren . . . . . 175
3.4.4 BDF-Methoden . . . . . 179
3.4.5 Implizite Runge-Kutta-Verfahren . . . . . 181
3.4.6 MATLAB-Funktionen . . . . . 188
3.4.7 Aufgaben . . . . . 195

**4 Die Theorie gewöhnlicher Rand- und Eigenwertaufgaben 199**
4.1 Sturmsche Randwertaufgaben . . . . . 199
4.1.1 Beispiele, Definitionen . . . . . 199
4.1.2 Existenz, Eindeutigkeit . . . . . 202
4.1.3 Die Greensche Funktion . . . . . 205
4.1.4 Aufgaben . . . . . 213
4.2 Das Sturm-Liouvillesche Eigenwertproblem . . . . . 214
4.2.1 Problemstellung, Beispiele . . . . . 214
4.2.2 Der Existenzsatz und der Trennungssatz . . . . . 216
4.2.3 Aufgaben . . . . . 224

**5 Lösungen zu den Aufgaben 225**
5.1 Aufgaben zu Kapitel 1 . . . 225
5.1.1 Aufgaben zu Abschnitt 1.1 . . . 225
5.1.2 Aufgaben zu Abschnitt 1.2 . . . 235
5.1.3 Aufgaben zu Abschnitt 1.3 . . . 243
5.1.4 Aufgaben zu Abschnitt 1.4 . . . 253
5.2 Aufgaben zu Kapitel 2 . . . 264
5.2.1 Aufgaben zu Abschnitt 2.1 . . . 264
5.2.2 Aufgaben zu Abschnitt 2.2 . . . 274
5.2.3 Aufgaben zu Abschnitt 2.3 . . . 283
5.3 Aufgaben zu Kapitel 3 . . . 291
5.3.1 Aufgaben zu Abschnitt 3.1 . . . 291
5.3.2 Aufgaben zu Abschnitt 3.2 . . . 298
5.3.3 Aufgaben zu Abschnitt 3.3 . . . 309
5.3.4 Aufgaben zu Abschnitt 3.4 . . . 312
5.4 Aufgaben zu Kapitel 4 . . . 323
5.4.1 Aufgaben zu Abschnitt 4.1 . . . 323
5.4.2 Aufgaben zu Abschnitt 4.2 . . . 327

# Kapitel 1

# Einführung, Beispiele, Grundlagen

Unter einer (reellen) *gewöhnlichen Differentialgleichung* versteht man eine Gleichung, in der *eine* reelle unabhängige Variable $t$, eine reellwertige Funktion $x(\cdot)$ dieser unabhängigen Variablen und Ableitungen von $x$ vorkommen:

$$(*) \qquad F(t, x, x', \ldots, x^{(m)}) = 0.$$

Wir wählen die Bezeichnung $t$ (und nicht $x$, wie in vielen Lehrbüchern) für die unabhängige Variable, da zumindestens bei *Anfangswertaufgaben*, die uns bei weitem am meisten beschäftigen werden, die unabhängige Variable für die Zeit steht ($t$ wie **t**ime). Differentiation nach $t$ wird dann häufig durch einen Punkt gekennzeichnet. Wir werden daher gleichberechtigt die folgenden Bezeichnungen benutzen:

$$\frac{d}{dt}x(t) = x'(t) = \dot{x}(t).$$

In $(*)$ heißt $m$ die *Ordnung* der Differentialgleichung. Eine Funktion $x \in C^m(I)$, also eine $m$-mal auf dem Intervall $I \subset \mathbb{R}$ differenzierbare Funktion, heißt *Lösung* der Differentialgleichung $(*)$ auf dem Intervall $I$, wenn

$$F(t, x(t), x'(t), \ldots, x^{(m)}(t)) = 0 \quad \text{für alle } t \in I.$$

Eine Differentialgleichung $m$-ter Ordnung heißt *explizit*, wenn man sie nach $x^{(m)}$ auflösen kann, sie also die Form

$$x^{(m)} = f(t, x, \ldots, x^{(m-1)})$$

hat. Neben Differentialgleichungen der Form $(*)$, bei der eine reellwertige Funktion $x(\cdot)$ einer reellen Variablen $t$ gesucht wird, werden wir allgemeiner meistens *gewöhnliche Differentialgleichungssysteme* untersuchen. Ein Differentialgleichungssystem von $k$ Differentialgleichungen $m$-ter Ordnung für $(x_1, \ldots, x_n)$ hat die Form

$$(**) \qquad F_i(t, x_1, \ldots, x_n, x_1', \ldots, x_n', x_1^{(m)}, \ldots, x_n^{(m)}) = 0, \qquad i = 1, \ldots, k.$$

Wir interessieren uns vor allem für explizite Differentialgleichungssysteme von $n$ Differentialgleichungen erster Ordnung. Diese haben die Form

$$\begin{aligned} x_1' &= f_1(t, x_1, \ldots, x_n) \\ &\vdots \qquad\qquad \vdots \\ x_n' &= f_n(t, x_1, \ldots, x_n) \end{aligned}$$

bzw. in kompakterer Vektorschreibweise

$$x' = f(t, x).$$

Eine explizite Differentialgleichung $n$-ter Ordnung

$$x^{(n)} = f(t, x, \ldots, x^{(n-1)})$$

lässt sich als ein System von $n$ Differentialgleichungen $n$-ter Ordnung schreiben:

$$\begin{pmatrix} x_1' \\ \vdots \\ x_{n-1}' \\ x_n' \end{pmatrix} = \begin{pmatrix} x_2 \\ \vdots \\ x_n \\ f(t, x_1, \ldots, x_n) \end{pmatrix}.$$

Was sind die wichtigsten Fragen im Zusammenhang mit gewöhnlichen Differentialgleichungen bzw. Differentialgleichungssystemen? Dies sind vor allem:

- Existiert eine Lösung? Auf welchem Intervall?
- Ist eine Lösung durch geeignete Zusatzbedingungen eindeutig festgelegt?
- Wie wirken sich Änderungen in den Daten der gegebenen Differentialgleichung aus?
- Welche qualitativen Aussagen können über eine Lösung gemacht werden, etwa über ihr asymptotisches Verhalten (Aussagen über die ferne Zukunft), ihre "Stabilität"?
- Wie berechnet man eine Lösung?

Auf alle diese Punkte werden wir ausführlich eingehen. Zunächst aber soll der Begriff einer gewöhnlichen Differentialgleichung bzw. eines gewöhnlichen Differentialgleichungssystem anhand von Beispielen verdeutlicht werden. Anschließend werden wir (verhältnismäßig kurz, da wir es als ziemlich uninteressant ansehen) auf elementar integrierbare Differentialgleichungen erster Ordnung eingehen. Zum Schluss dieses ersten Kapitels werden wir die insbesondere für Existenzfragen so wichtigen Fixpunktsätze behandeln, nämlich den Kontraktionssatz und den Schauderschen Fixpunktsatz.

Im Gegensatz zu früheren Vorlesungen mit einem entsprechenden Titel werden in dieser Vorlesung mathematische Anwendersysteme (Maple und MATLAB) eine wesentliche Rolle spielen. Hierbei wird Maple vor allem für symbolische Berechnungen, MATLAB vor allem für numerische Rechnungen eingesetzt, beide für Visualierungen. Kenntnisse hierzu werden nicht vorausgesetzt. Wichtiger ist es, die nötige Neugier zu entwickeln und an hand der in diesem Skript angegebenen Beispiele selbst Erfahrungen mit den mathematischen Anwendersystemen zu machen.

# 1.1 Wachstumsmodelle

## 1.1.1 Populationsmodelle

In einem Populationsmodell bezeichne $p(t)$ die Population einer gewissen Spezies (etwa der Menschen auf der Erde oder der Lachse in der Weser) zur Zeit $t$. Geht man wie T. R. Malthus (1766-1834) von einer konstanten Geburtenrate $\gamma$ und Sterberate $\delta$ pro Kopf der Bevölkerung und Zeiteinheit aus und nimmt man an, dass sich die Bevölkerungszahl innerhalb eines Zeitraum $\Delta t$ von $t$ bis $t + \Delta t$ gemäß der Formel

$$p(t + \Delta t) = p(t) + \lambda p(t)\Delta t$$

verändert, wobei wir

$$\lambda := \gamma - \delta$$

gesetzt haben, so erhält man aus

$$\frac{p(t + \Delta t) - p(t)}{\Delta t} = \lambda p(t)$$

durch den Grenzübergang $\Delta t \to 0$ die *Differentialgleichung*

$$\frac{dp}{dt}(t) = \lambda p(t)$$

mit der eindeutigen Lösung $p(t) = p_0 e^{\lambda(t-t_0)}$, wenn man noch $p(t_0) = p_0$ vorgibt[1]. Wenn man die Zeit $T$ kennt, in der sich die Bevölkerung verdoppelt, so kann man unter Zugrundelegung des Malthus-Wachstumsmodells auf die Wachstumsrate $\lambda$ schließen. Denn aus

$$p_0 e^{\lambda(t+T-t_0)} = p(t + T) = 2p(t) = 2p_0 e^{\lambda(t-t_0)}$$

erhält man

$$\lambda = \frac{\log 2}{T}.$$

**Beispiel:** Bei M.Braun (1983)[2] wird bemerkt, dass die Weltbevölkerung 1961 mit 3 060 000 000 geschätzt wurde. Geht man von einer Wachstumsrate von 2%/Jahr (bzw.

[1] Dass es sich bei $p(t) = p_0 e^{\lambda(t-t_0)}$ um eine Lösung der Differentialgleichung $p' = \lambda p$ mit der Anfangsbedingung $p(t_0) = p_0$ handelt, ist klar. Weshalb aber ist es die einzige? Sei hierzu $p$ eine Lösung von $p' = \lambda p$. Dann ist

$$p'(t)e^{-\lambda(t-t_0)} = \lambda e^{-\lambda(t-t_0)} p(t)$$

bzw.

$$\frac{d}{dt}[e^{-\lambda(t-t_0)} p(t)] = 0.$$

Folglich ist $e^{-\lambda(t-t_0)} p(t)$ konstant. Für eine Lösung $p$ von $p' = \lambda p$, welche der Anfangsbedingung $p(t_0) = p_0$ genügt, ist diese Konstante gleich $p_0$, so dass also $p(t) = p_0 e^{\lambda(t-t_0)}$ notwendigerweise die einzige Lösung der gestellten Anfangswertaufgabe ist.

[2] M. Braun (1983) "Single species population models". In: *Differential Equation Models* (eds. M. Braun et al.), Springer-Verlag, New York-Heidelberg-Berlin.

einer Verdoppelung der Weltbevölkerung alle 35 Jahre) aus, so erhielte man unter Zugrundelegung des Malthus-Ansatzes

$$p(t) = (3.06)10^9 e^{0.02(t-1961)}.$$

Dies scheint mit den Zahlen der Jahre 1700 bis etwa 1960 gut im Einklang zu sein. □

In der fernen Zukunft ist das Malthus-Modell unrealistisch, da bei großen Populationen die Wachstumsrate von der Populationsgröße abhängen wird. Die nächst einfache Annahme ist, dass die Wachstumsrate linear von der Populationsgröße abhängt und kleiner wird, wenn die Population wächst. Dies führt auf das sogenannte Verhulst-Modell, genannt nach P. F. Verhulst (1804-1849). Hiernach berechnet sich die Population $p$ als Lösung der Anfangswertaufgabe

$$p' = ap - bp^2, \qquad p(t_0) = p_0.$$

Hierbei ist $p_0 > 0$ und $0 < b \ll a$, so dass bei kleinen Populationen $bp^2$ gegenüber $ap$ vernachlässigt werden kann. Diese (nichtlineare) Differentialgleichung (man nennt sie auch Gleichung des beschränkten Wachstums oder logistische Differentialgleichung) mit Anfangsbedingung (man spricht dann, wie wir es schon getan haben, von einer *Anfangswertaufgabe*) kann man geschlossen lösen, und zwar ist durch

$$p(t) = \frac{ap_0}{bp_0 + (a - bp_0)\,\exp[-a(t - t_0)]}$$

die Lösung der gestellten Anfangswertaufgabe gegeben. Wie kommt man auf diese Lösung? Hierzu schreiben wir die gegebene Differentialgleichung in der Form

$$\frac{p'}{(a - bp)p} = 1.$$

Wir definieren

$$F(p) := \int_{p_0}^{p} \frac{1}{(a - br)r}\, dr,$$

so dass $F$ eine Stammfunktion von $1/(a - bp)p$ ist, welche für $p = p_0$ verschwindet. Ist $p$ eine Lösung der gegebenen Anfangswertaufgabe, so ist offenbar

$$\frac{d}{dt} F(p(t)) = 1,$$

daher $F(p(t)) = t + c$ mit einer Konstanten $c$. Setzt man $t = t_0$, so erhält man $c = -t_0$. Also ist $F(p(t)) = t - t_0$. Nun kann man aber $F(\cdot)$ geschlossen ausrechnen. Hierzu mache man die Partialbruchzerlegung

$$\frac{1}{(a - br)r} = \frac{1}{ar} + \frac{b}{a(a - br)}.$$

Wir setzen $p_0 \neq 0$ (andernfalls ist $p(t) \equiv 0$ Lösung der gestellten Anfangswertaufgabe) und $p_0 \neq a/b$ voraus (andernfalls ist $p(t) \equiv a/b$ Lösung). Folglich ist

$$F(p) = \frac{1}{a}\, \ln\Big(\frac{p(a - bp_0)}{p_0(a - bp)}\Big).$$

Wegen $F(p(t)) = 1$ erhält man hieraus

$$a(t - t_0) = \ln\Big(\frac{p(t)\,(a - bp_0)}{p_0\,(a - bp(t))}\Big) \quad \text{bzw.} \quad e^{a(t-t_0)} = \frac{p(t)\,(a - bp_0)}{p_0\,(a - bp(t))}.$$

Auflösen der letzten Gleichung nach $p(t)$ ergibt die behauptete Darstellung der Lösung. Durch Einsetzen kann man leicht nachprüfen, dass dies wirklich eine Lösung ist. Da eine Lösung, wie wir gesehen haben, notwendigerweise die angegebene Form hat, ist sie auch eindeutig.

**Beispiel:** Wir wollen einmal die Fähigkeiten von Maple testen, Integrale zu berechnen und (gewöhnliche) Differentialgleichungen geschlossen zu lösen.

Hierzu aktivieren wir Maple zunächst durch `xmaple&`. Als Resultat von

```
int(1/((a-b*p)*p),p)
```

zur Berechnung des unbestimmten Integrals

$$\int^{p} \frac{1}{(a - bp)p}\, dp$$

erhalten wir

$$-\frac{\ln(-a + bp)}{a} + \frac{\ln(p)}{a},$$

der anschließende Befehl `simplify(%)` führt auf

$$\frac{-\ln(-a + bp) + \ln(p)}{a}.$$

Will man die Anfangswertaufgabe

$$p' = ap - bp^2, \qquad p(t_0) = p_0$$

mit Maple lösen, zunächst für allgemeine Daten $a, b, t_0, p_0$, danach für spezielle Daten, so kann dies mittels

```
> ode:=diff(p(t),t)=a*p(t)-b*p(t)^2;
```

$$ode := \tfrac{\partial}{\partial t}\,\mathrm{p}(t) = a\,\mathrm{p}(t) - b\,\mathrm{p}(t)^2$$

```
> initial:=p(t_0)=p_0;
```

$$initial := \mathrm{p}(t\_0) = p\_0$$

```
> sol:=dsolve({ode,initial},p(t));
```

$$sol := \mathrm{p}(t) = \frac{a}{b + \dfrac{e^{(-a\,t)}\,(a - p\_0\,b)}{p\_0\,e^{(-a\,t\_0)}}}$$

```
> simplify(%);
```

$$\mathrm{p}(t) = \frac{a\,p\_0}{p\_0\,b + e^{(-a\,(t-t\_0))}\,a - e^{(-a\,(t-t\_0))}\,p\_0\,b}$$

```
> par:={a=100,b=0.1,t_0=0,p_0=10};
```

$$par := \{a = 100,\ b = .1,\ t\_0 = 0,\ p\_0 = 10\}$$

```
> solspecial:=dsolve(subs(par,{ode,initial}),p(t));
```

$$solspecial := \mathrm{p}(t) = 1000\,\frac{1}{1 + 99\,e^{(-100\,t)}}$$

geschehen. □

Bemerkenswert ist die Folgerung, die wir aus der angegebenen Lösung der logistischen Differentialgleichung ziehen können. Es ist nämlich $\lim_{t\to\infty} p(t) = a/b$, die Gesamtpopulation strebt also mit wachsender Zeit der Grenzpopulation $\xi := a/b$ zu. Dies geschieht monoton fallend, wenn $p_0 > \xi$ und monoton wachsend, wenn $0 < p_0 < \xi$, wobei der letztere Fall sicher der interessantere ist. In Abbildung 1.1 zeigen wir die typische S-Form der Lösung der logistischen Differentialgleichung. Den Plot haben wir

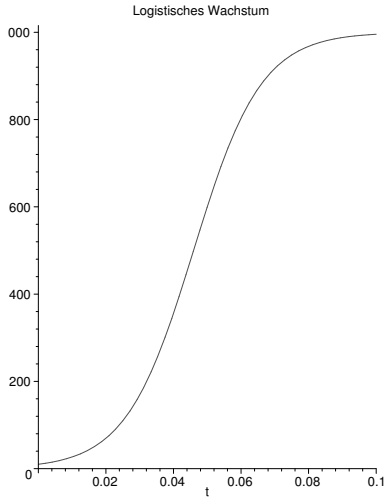


Abbildung 1.1: Eine Lösung der logistischen Differentialgleichung

in Maple durch

```
plot(rhs(solspecial),t=0..0.1,title="Logistisches Wachstum");
```

erzeugt.

### 1.1.2 Das Räuber-Beute-Modell

Im letzten Unterabschnitt wurden mathematische Modelle zum Wachstum einer einzigen Population aufgestellt, Wechselwirkungen mit anderen Species wurden ebenso vernachlässigt wie "Ein- oder Auswanderungen". Nun wollen wir das Wachstum von zwei Arten untersuchen, die sich gegenseitig beeinflussen und die wir Räuber und Beute (engl.: predator, prey) nennen wollen. Man stelle sich etwa Raub- und Beutefische in der Adria vor. Dies war der Ausgangspunkt für die Untersuchungen von V. Volterra (1860-1940) und A. J. Lotka (1880-1949). Sei $x(t)$ die Population der Beute, $y(t)$ die Population der Räuber zur Zeit $t$. Falls genügend Nahrung für die Beute vorhanden ist, so dass sich diese nicht gegenseitig das Futter wegzunehmen brauchen, und keine Räuber vorhanden sind, ist $x' = ax$ mit einer positiven Konstante $a$, wenn vom Malthus-Modell

ausgegangen wird. Andererseits ist die Anzahl der "Kontakte" zwischen Räuber und Beute proportial zu $xy$, so dass $x' = ax - bxy$ mit einer positiven Konstanten $b$. Ist dagegen keine Beute vorhanden, so sterben die Räuber aus: $y' = -cy$. Andererseits ist die Zuwachsrate proportional zu $xy$, so dass die zeitliche Änderung der Räuberpopulation durch $y' = -cy + dxy$ gegeben ist. Insgesamt erhält man ein System von zwei Differentialgleichungen erster Ordnung, das sogenannte Lotka-Volterra-System:

$$\begin{aligned} x' &= ax - bxy, \\ y' &= -cy + dxy, \end{aligned}$$

wobei $a, b, c, d$ positive Konstanten sind. Sind positive Anfangspopulationen $x_0$, $y_0$ zur Zeit $t = 0$ vorgegeben, so können die zukünftigen Räuber- bzw. Beute-Populationen $x(t)$ bzw. $y(t)$ durch (numerisches) Lösen der Anfangswertaufgabe

$$\begin{aligned} x' &= ax - bxy, & x(0) &= x_0, \\ y' &= -cy + dxy, & y(0) &= y_0 \end{aligned}$$

bestimmt werden.

**Beispiel:** Wir wollen die Lösung von

$$(*) \qquad \begin{aligned} x' &= 2x - 0.01xy, & x(0) &= 300, \\ y' &= -y + 0.01xy, & y(0) &= 150 \end{aligned}$$

veranschaulichen. Diese Anfangswertaufgabe für ein Differentialgleichungssystem von zwei Differentialgleichungen erster Ordnung ist nicht geschlossen lösbar, so dass wir auf numerische Methoden angewiesen sind. Wir geben nach dem Maple-Prompt jeweils ein:

```
par:=a=2,b=0.01,c=1,d=0.01,x_0=300,y_0=150;
eqn:=diff(x(t),t)=a*x(t)-b*x(t)*y(t),diff(y(t),t)=-c*y(t)+d*x(t)*y(t);
initial:=x(0)=x_0,y(0)=y_0;
sol:=dsolve(subs(par,{eqn,initial}),{x(t),y(t)},type=numeric);
plots[odeplot](sol,[[t,x(t)],[t,y(t)]],0..10,title="Lotka-Volterra");
```

und erhalten den Plot in Abbildung 1.2. Ganz erstaunlich ist nun, dass eine Konstante $T > 0$ existiert mit

$$(x(t), y(t)) = (x(t + T), y(t + T))$$

für alle $t$. D. h. die Populationen der Räuber und der Beute haben eine Periode $T$, so dass nach der Zeit $T$ der Anfangszustand wieder erreicht wird. Anders gesagt: In der sogenannten Phasenebene, der $(x, y)$-Ebene, beschreibt $(x(\cdot), y(\cdot))$ eine geschlossene Bahn. Durch

```
with(plots);
odeplot(sol,[x(t),y(t)],0..5,title="Geschlossene
Phasenbahn",labels=["Beute","Raeuber"]);
```

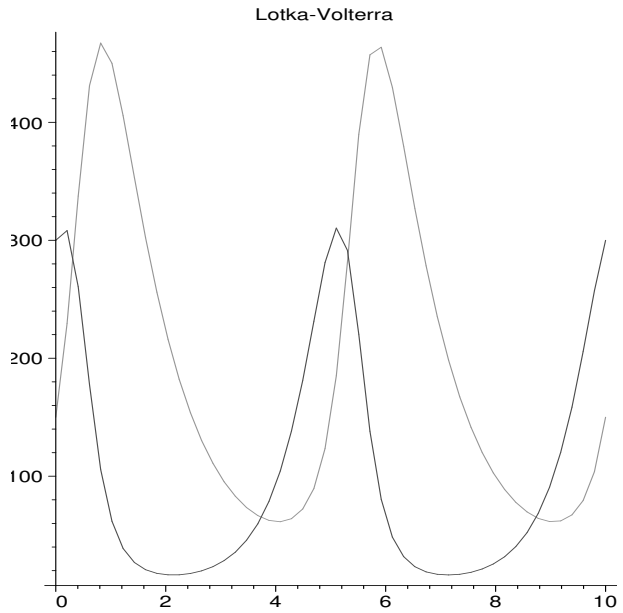


Abbildung 1.2: Die Population der Beute $x$ und der Räuber $y$

wird der entsprechende Plot hergestellt, der in Abbildung 1.3 zu sehen ist. □

In dem Beispiel hatten wir beobachtet, dass sich (zumindestens in einem speziellen Fall) beim Lotka-Volterra-System in der Phasenebene geschlossene Bahnen ergeben. Dies zu beweisen, ist noch außerhalb unserer Reichweite. Unter der Annahme, dass Anfangswertaufgaben für das Lotka-Volterra-System (vorgegebene Anfangszeit und vorgegebener Anfangszustand) eindeutig lösbar sind, ist es klar, dass eine Lösung, die im Inneren des ersten Quadranten startet (d. h. die Anfangspopulationen sind positiv), dort auch bleibt. Im folgenden Lemma wird gezeigt, dass die *Phasenbahn* $\{(x(t), y(t)) : t \geq 0\}$ zum Lotka-Volterra-System auf einer geschlossenen Kurve im Inneren des ersten Quadranten im $\mathbb{R}^2$ liegt. Mit weiteren Hilfsmitteln kann hieraus auf die Periodizität von $(x(\cdot), y(\cdot))$ geschlossen werden.

**Lemma 1.1** *Sei $(x(\cdot), y(\cdot))$ bei vorgegebenen positiven Konstanten $a, b, c, d$ und vorgegebener positiver Anfangspopulation $(x_0, y_0)$ die Lösung des Lotka-Volterra-Systems*

$$\begin{array}{rclcrcl} x' & = & ax - bxy, & & x(0) & = & x_0, \\ y' & = & -cy + dxy, & & y(0) & = & y_0. \end{array}$$

*Dann liegt die Phasenbahn $C := \{(x(t), y(t)) : t \geq 0\}$ auf einer geschlossenen Kurve im Inneren des ersten Quadranten im $\mathbb{R}^2$.*

**Beweis:** Wir definieren die Abbildung $V : \mathbb{R}^2_+ \longrightarrow \mathbb{R}$ durch

$$V(x, y) := c \log x - dx + a \log y - by.$$

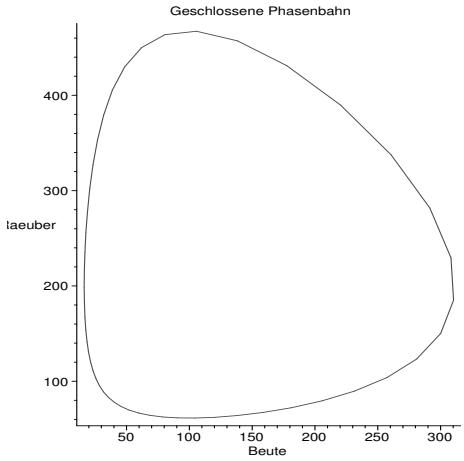


Abbildung 1.3: Beute und Räuber in der Phasenebene

Dann ist

$$\begin{aligned}
\frac{d}{dt}V(x(t),y(t)) &= V_x(x(t),y(t))x'(t) + V_y(x(t),y(t))y'(t) \\
&= \Big(\frac{c}{x(t)} - d\Big)x'(t) + \Big(\frac{a}{y(t)} - b\Big)y'(t) \\
&= \Big(\frac{c}{x(t)} - d\Big)[ax(t) - bx(t)y(t)] + \Big(\frac{a}{y(t)} - b\Big)[-cy(t) + dx(t)y(t)] \\
&= c[a - by(t)] - d[ax(t) - bx(t)y(t)] \\
&\qquad + a[-c + dx(t)] - b[-cy(t) + dx(t)y(t)] \\
&= 0.
\end{aligned}$$

Daher ist $V(x(\cdot), y(\cdot))$ konstant und damit die Phasenbahn $C$ enthalten in $K := \{(x,y) \in \mathbb{R}^2_+ : V(x,y) = V(x_0,y_0)\}$. Wir überlegen uns, dass $K$ eine geschlossene Kurve ist. Offenbar ist (man wende auf beide Seiten der Gleichung $V(x,y) = V(x_0,y_0)$ die Exponentialfunktion an)

$$K = \Big\{(x,y) \in \mathbb{R}^2_+ : \Big(\frac{x^c}{e^{dx}}\Big)\Big(\frac{y^a}{e^{by}}\Big) = \exp(V(x_0,y_0))\Big\}.$$

Zur Abkürzung definieren wir $f(x) := x^c/e^{dx}$ und $g(y) := y^a/e^{by}$, jeweils auf $\mathbb{R}_+$. Die Funktionen $f, g$ sind für $(a,b,c,d) := (2,1,0.01,0.01)$ in Abbildung 1.4 angegeben. Allgemeiner betrachten wir für $\gamma > 0$ die Menge

$$K_\gamma := \{(x,y) \in \mathbb{R}_+ : f(x)g(y) = \gamma\}.$$

In Abbildung 1.5 sind für verschiedene $\gamma$ die entsprechenden Mengen abgebildet. (Diese wurden folgendermaßen hergestellt. Zunächst wurden durch

```
f:=x->x/exp(0.01*x);
g:=y->y^2/exp(0.01*y);
```

die Funktionen $f$ und $g$ definiert. Die Abbildungen wurde durch

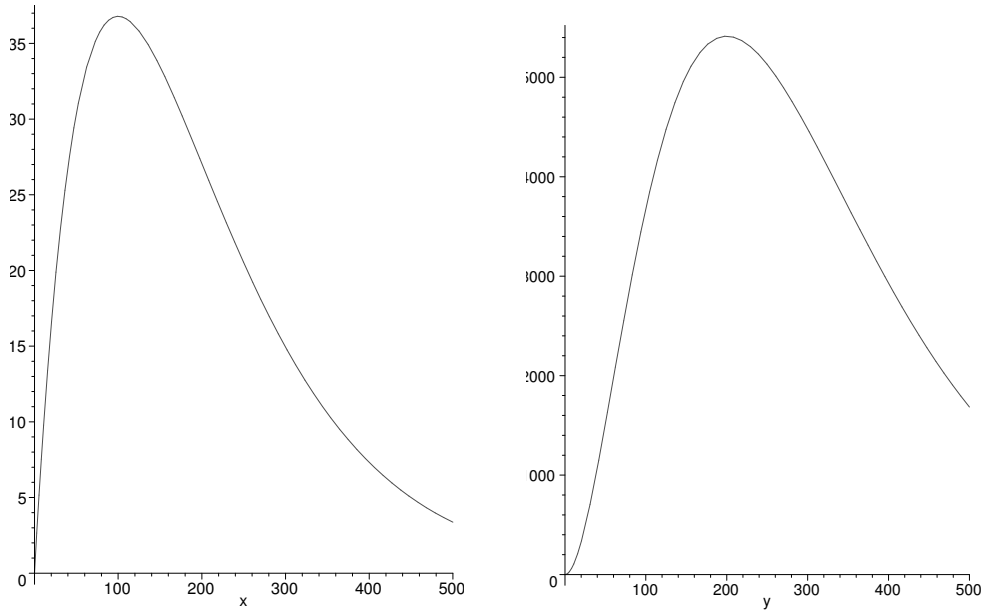


Abbildung 1.4: Die Funktionen $f(x) := x^c/e^{dx}$ und $g(y) := y^a/e^{by}$

```
contourplot(f(x)*g(y),x=10..500,y=10..500,filled=true);
contourplot(f(x)*g(y),x=10..500,y=10..500);
```

erzeugt.) Offensichtlich ist $f(0) = 0$, $\lim_{x\to\infty} f(x) = 0$ und $f(x) > 0$ auf $(0, \infty)$. Wegen

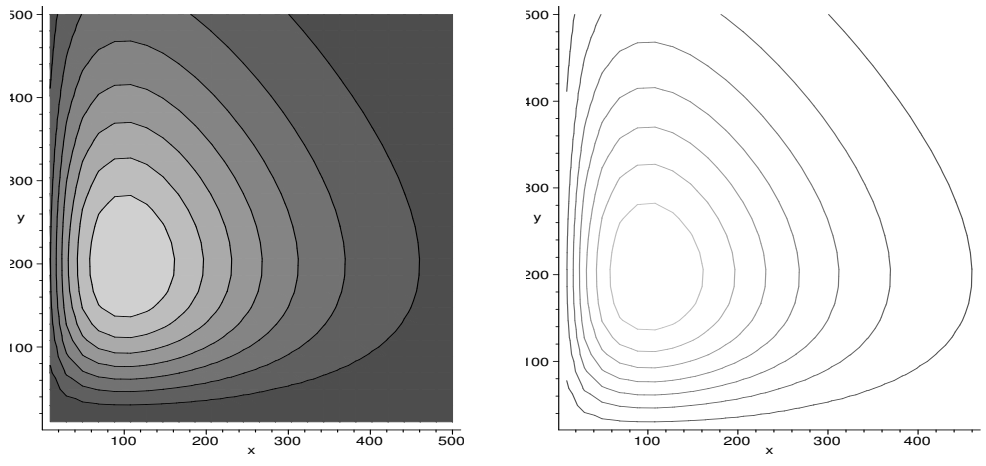


Abbildung 1.5: Die Niveaulinien von $f(x)g(y)$

$$f'(x) = \frac{x^{c-1}(c - dx)}{e^{dx}}$$

besitzt $f$ ein Maximum in $c/d$ und ist auf $(0, c/d)$ monoton wachsend, auf $(c/d, \infty)$ monoton fallend. Entsprechendes gilt für die Funktion $g$, die in $a/b$ ihr Maximum annimmt. Sei $m_x := f(c/d)$, $m_y := g(a/b)$. Dann ist klar, dass $K_\gamma = \emptyset$ für $\gamma > m_x m_y$. Ist $\gamma = m_x m_y$, so ist $K_\gamma = \{(c/d, a/b)\}$. Wir nehmen daher jetzt an, es sei $\gamma = \lambda m_y$

mit $\lambda \in (0, m_x)$. Die Gleichung $f(x) = \lambda$ besitzt zwei Lösungen $x_1 < x_2$, von denen eine kleiner und die andere größer als $c/d$ ist. Bei gegebenem $x \in (0, \infty)$ untersuchen wir nun die Existenz einer Lösung $y$ der Gleichung

$$(*) \qquad f(x)g(y) = \gamma.$$

Für $x \notin [x_1, x_2]$ ist $f(x) < \lambda$, so dass $(*)$ keine Lösung $y$ besitzt. Ist $x = x_1$ oder $x = x_2$, so besitzt $(*)$ die Lösung $y = a/b$. Für $x \in (x_1, x_2)$ hat $(*)$ genau zwei Lösungen $y_1(x) < a/b < y_2(x)$. Da $y_1$ und $y_2$ mit $x \to x_1+$ und $x \to x_2-$ gegen $a/b$ konvergieren, ist durch $K_\gamma$ eine geschlossene Kurve gegeben. □

**Bemerkung:** Sei $(x(t), y(t))$ eine $T$-periodische Lösung des Lotka-Volterra-Systems. Durch

$$\overline{x} := \frac{1}{T}\int_0^T x(t)\,dt, \qquad \overline{y} := \frac{1}{T}\int_0^T y(t)\,dt$$

sind die mittleren Populationen der Beute bzw. der Räuber über das Periodenintervall $[0, T]$ gegeben. Es ist überraschend, dass man diese Mittelwerte berechnen kann, ohne die Populationen $x$ bzw. $y$ zu kennen. Denn wegen

$$\begin{aligned} 0 &= \frac{1}{T}[\log y(t) - \log y(0)] \\ &= \frac{1}{T}\int_0^T \frac{d}{dt}\log y(t)\,dt \\ &= \frac{1}{T}\int_0^T \frac{y'(t)}{y(t)}\,dt \\ &= \frac{1}{T}\int_0^T [-c + dx(t)]\,dt \\ &= -c + d\overline{x}, \end{aligned}$$

so dass $\overline{x} = c/d$. Entsprechend erhält man $\overline{y} = a/b$. Hieraus kann eine interessante Folgerung gezogen werden. Bei den Räubern und der Beute handele es sich jeweils um Fische. Es sollen Aussagen über die Auswirkung des Fischfangs auf die jeweiligen Populationen gemacht werden, wobei davon ausgegangen wird, dass die Fischer beim Fang zwischen Räubern und Beute machen können. Durch die Konstante $\epsilon \geq 0$ werde die "Intensität" des Fischfangs angegeben. Das modifizierte Differentialgleichungssystem lautet dann

$$\begin{aligned} x' &= (a - \epsilon)x - bxy, \\ y' &= -(c + \epsilon)y + dxy. \end{aligned}$$

Ist nun $\epsilon \in (0, a)$, erhält man nach obiger Überlegung als mittlere Populationen

$$\overline{x} = \frac{c + \epsilon}{d}, \qquad \overline{y} = \frac{a - \epsilon}{b}.$$

Ein nicht zu intensiver Fischfang erhöht also die mittlere Population der Beute und erniedrigt die der Räuber. □

**Bemerkung:** Will man bei einem Räuber-Beute-Modell auch noch den Konkurrenzkampf innerhalb der Arten berücksichtigen, also von einem logistischen Wachstum ausgehen, so kommt man zu einem Differentialgleichungssystem

$$\begin{aligned} x' &= ax - bxy - ex^2, \\ y' &= -cy + dxy - fy^2 \end{aligned}$$

mit positiven Konstanten $a, b, c, d, e, f$. Auch hier bleibt eine im Inneren des ersten Quadranten startende Phasenbahn dort, es wird sich i. Allg. aber keine[3] periodische Bahn einstellen. Für

$$(a, b, c, d, e, f) := (2, 0.01, 1, 0.01, 0.001, 0.001)$$

und den Anfangszustand $(x_0, y_0) := (300, 150)$ geben wir in Abbildung 1.6 die zugehörige Phasenbahn an. Man ahnt, dass die Phasenbahn einem Punkt zustrebt, und zwar

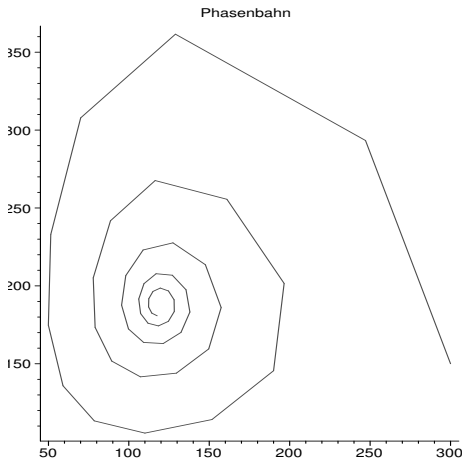


Abbildung 1.6: Eine Phasenbahn zum logistischen Räuber-Beute-Modell

der positiven Lösung des Gleichungssystems

$$\begin{aligned} ax - bxy - ex^2 &= 0, \\ -cy + dxy - fy^2 &= 0, \end{aligned}$$

also

$$(\hat{x}, \hat{y}) = \frac{1}{ef + bd}(af + bc, da - ec) = (118.812, 188.119),$$

einem sogenannten *Gleichgewichtspunkt*. Um dies zu verstehen, benötigt man Kenntnisse über das Stabilitätsverhalten autonomer Systeme von zwei Differentialgleichungen erster Ordnung in der Nähe eines Gleichgewichtspunktes. Im obigen Fall ist der Gleichgewichtspunkt ein *Attraktor*. In Aufgabe 11 in Abschnitt 2.3 kommen wir auf dieses Beispiel zurück. Übrigens kann man die (symbolischen) Lösungen eines Gleichungssystems (wir konzentrieren uns auf obiges Beispiel) mit Maple folgendermaßen erhalten:

[3] Dies kann präzisiert werden. Siehe

C. S. Coleman (1983) "Quadratic population models: almost never any cycles". In: *Differential Equation Models* (eds. M. Braun et al.), Springer-Verlag, New York-Heidelberg-Berlin.

```
eqn:={a*x-b*x*y-e*x^2=0,-c*y+d*x*y-f*y^2=0};
sol:=solve(eqn,{x,y});
```

Ergebnis in diesem Falle ist

$$sol := \{x=0, y=0\}, \{x=0, y=-\frac{c}{f}\}, \{y=0, x=\frac{a}{e}\}, \{x=\frac{fa+bc}{fe+bd}, y=\frac{-ce+da}{fe+bd}\}.$$

Nach

```
par:={a=2,b=0.01,c=1,d=0.01,e=0.001,f=0.001};
numsol:=solve(subs(par,eqn),{x,y});
```

erhält man die entsprechenden numerischen Lösungen ((118.8118812, 188.1188119) ist hierbei die uns unteressierende). □

### 1.1.3 Aufgaben

1. Man betrachte eine große Population von $N$ Individuen. Geburten, "natürliche Tode", Ein- und Auswanderungen mögen vernachlässigt werden. Es grassiere eine Krankheit, die sich durch Kontakt zwischen Individuen ausbreitet. Diese Krankheit sei so beschaffen, dass ein Individuum entweder durch sie stirbt oder nach einer Genesung immun gegen sie wurde. Die Population kann dann in drei Klassen eingeteilt werden.

   - In der Klasse $S$ sind die anfälligen (**s**usceptibles) zusammengefasst, also diejenigen, die die Krankheit noch nicht bekommen haben und nicht gegen sie immun sind. Ihre Zahl zur Zeit $t$ sei $S(t)$.
   - In der Klasse $I$ sind die **i**nfizierten enthalten, also diejenigen, die die Krankheit haben und andere anstecken können. Zur Zeit $t$ sei ihre Zahl $I(t)$.
   - Zur Klasse $R$ gehört der Rest (**r**emoved), genauer also diejenigen, die tot, isoliert oder immun sind. $R(t)$ sei die Anzahl der Individuen der Klasse $R$ zur Zeit $t$.

   Die Krankheit genüge der folgenden Gesetzmäßigkeit.

   (a) Die Änderungsrate der anfälligen Population ist proportional zur Anzahl der Kontakte zwischen anfälliger und infizierter Population. Wir nehmen daher an, es sei

   $$S' = -\beta SI$$

   mit einer Konstanten (der sogenannten Infektionsrate) $\beta > 0$.

   (b) Individuen werden aus der Klasse $I$ der Infizierten mit einer Rate entfernt (sie sterben, werden isoliert oder immun), die proportional zu ihrer Anzahl ist. Daher ist

   $$I' = \beta SI - \gamma I, \qquad R' = \gamma I.$$

   Mit $S_0, I_0$ seien die positiven Populationen der Klassen $S$ und $I$ zur Anfangszeit $t = 0$ bezeichnet. Zu dieser Zeit sei noch niemand an der Krankheit gestorben bzw. ihretwegen isoliert oder immun. Man hat daher die Anfangswertaufgabe

   $$\text{(P)} \qquad \begin{array}{rclrcl} S' &=& -\beta SI, & S(0) &=& S_0, \\ I' &=& \beta SI - \gamma I, & I(0) &=& I_0, \\ R' &=& \gamma I, & R(0) &=& 0. \end{array}$$

Dies ist das sogenannte Kermack-McKendrick-Modell für die Ausbreitung ansteckender Krankheiten. Wir gehen davon aus, dass obige Anfangswertaufgabe eine eindeutige Lösung $(S, I, R)$ auf $[0, \infty)$ besitzt. Man zeige (die ersten beiden Aussagen sind anschaulich völlig trivial, müssen aber trotzdem bewiesen werden):

(a) Es sind $I(\cdot)$ und $S(\cdot)$ auf $[0, \infty)$ positiv.

(b) Es ist $S(\cdot)$ auf $[0, \infty)$ monoton fallend. Daher existiert $S_\infty := \lim_{t\to\infty} S(t)$.

(c) Es ist $S(t) + I(t) - (\gamma/\beta) \ln S(t) = \text{const}$ für alle $t$.

(d) Ist $S_0 > \gamma/\beta$, so kommt es zu einer Epidemie in dem Sinne, dass es ein $t > 0$ mit $I(t) > I_0$ gibt. Weiter gibt es ein $t^* > 0$ derart, dass $I(\cdot)$ auf $[0, t^*]$ monoton wachsend und auf $[t^*, \infty)$ monoton fallend ist. Es ist $\lim_{t\to\infty} I(t) = 0$ und $S_\infty$ ist die eindeutige Lösung der transzendenten Gleichung

$$S_0 \exp\left(-\frac{(N-x)\beta}{\gamma}\right) - x = 0.$$

(e) Ist $S_0 < \gamma/\beta$, so ist $I(\cdot)$ auf $[0, \infty)$ monoton fallend und $\lim_{t\to\infty} I(t) = 0$. Es kommt also zu keiner Epidemie und die Krankheit verschwindet letztendlich.

Hinweis: Es kann zweckmäßig sein, zunächst die folgende Aussage zu beweisen:

- Sei $h\colon [0, \infty) \longrightarrow \mathbb{R}$ stetig. Dann besitzt die Anfangswertaufgabe $x' = h(t)x$, $x(0) = x_0$ die eindeutige Lösung

$$x(t) = x_0 \exp\left(\int_0^t h(\tau)\, d\tau\right).$$

2. Sei[4] $p$ die Lösung der Anfangswertaufgabe für die logistische Differentialgleichung

$$p' = ap - bp^2, \qquad p(t_0) = p_0,$$

wobei $a, b, p_0$ positive Konstanten mit $p_0 < \frac{1}{2}(a/b)$ sind.

(a) Seien $t_1 < t_2$ mit $t_1 > t_0$ und $t_1 - t_0 = t_2 - t_1$ gegeben. Man zeige, dass $a$ und $b$ eindeutig durch $p_0 = p(t_0), p(t_1), p(t_2)$ bestimmt sind. Dies bedeutet: Legt man das logistische Wachstumsmodell zugrunde und sind die Populationen $p_0, p_1, p_2$ zu äquidistanten Zeiten $t_0, t_1, t_2$ bekannt, so sind hierdurch die Parameter $a, b$ im Modell eindeutig festgelegt.

(b) Man zeige, dass genau ein $t^* > t_0$ mit $p(t^*) = \frac{1}{2}(a/b)$ existiert und die Darstellung

$$p(t) = \frac{a/b}{1 + e^{-a(t-t^*)}}$$

gilt.

[4]Diese Aufgabe findet man auf S. 88 von

M. Braun (1983) "Single species population models". In: *Differential Equation Models* (eds. M. Braun et al.), Springer-Verlag, New York-Heidelberg-Berlin.

(c) Aus

| $k$ | $t_k$ | $p(t_k)$ |
|---|---|---|
| 0 | 1790 | 3 929 000 |
| 1 | 1850 | 23 192 000 |
| 2 | 1910 | 91 972 000 |

bestimme man $a$ und $b$. Anschließend berechne man $t^*$ mit $p(t^*) = \frac{1}{2}(a/b)$.

Hinweis: Es darf Maple eingesetzt werden.

3. Gegeben sei die Anfangswertaufgabe

$$\begin{aligned} x' &= 2x - 0.01xy, & x(0) &= 300, \\ y' &= -y + 0.01xy, & y(0) &= 150. \end{aligned}$$

Aus Abbildung 1.2 kann man ablesen, dass die Lösung $(x(\cdot), y(\cdot))$ eine Periode $T \approx 5$ besitzt. Man berechne (wie auch immer) eine verbesserte Näherung.

4. Das Lotka-Volterra-System

$$\begin{aligned} x' &= ax - bxy, \\ y' &= -cy + dxy \end{aligned}$$

mit positiven Konstanten $a, b, c, d$ besitzt den Gleichgewichtspunkt $(c/d, a/b)$. Man mache die Variablentransformation $u = x - c/d$, $v = y - a/b$ und stelle für $u, v$ ein Differentialgleichungssystem auf. Man löse das durch Weglassen der nichtlinearen Terme entstehende System, indem man nachweist, dass $u$ und $v$ der Differentialgleichung zweiter Ordnung $w'' + acw = 0$ genügen. Für die Anfangswertaufgabe

$$\begin{aligned} x' &= ax - bxy, & x(0) &= x_0, \\ y' &= -cy + dxy, & y(0) &= y_0 \end{aligned}$$

mit $x_0 \approx c/d$ und $y_0 \approx a/b$ berechne man hierdurch eine Näherungslösung.

5. Das Wachstumgesetz von B. Gompertz (1779-1865) soll das Wachsen von Tumoren gut beschreiben. Es basiert auf der Anfangswertaufgabe

$$V' = -rV \ln\left(\frac{V}{K}\right), \qquad V(0) = V_0.$$

Hierbei sind $r$ und $K$ gegebene Konstanten, $V(t)$ die Größe des Tumors zur Zeit $t$ und $V_0$ der Anfangszustand. Mit Hilfe von Maple löse man diese Anfangswertaufgabe.

6. Wir[5] machen über die Population $p$ einer Spezies mit $p(0) = p_0$ die folgenden Annahmen:

   - Die Population hängt nur vom Vorhandensein eines Grundstoffs $R$ ab.

[5]Diese und einige weitere Aufgaben haben wir dem Skript "Einführung in die Theorie der Differentialgleichungen" von H. Behncke (Universität Osnabrück) entnommen. Sehr viele Beispiele sind übrigens bei

H. Heuser (1989) *Gewöhnliche Differentialgleichungen*. B. G. Teubner, Stuttgart

enthalten.

- Die Population verbraucht laufend diesen Grundstoff, genauer sei

$$R(t) = R_0 - b\int_0^t p(s)\,ds.$$

- Die Wachstumsrate $p'(t)$ der Population ist proportional zu $p(t)$ und $R(t)$.

Mit positiven Konstanten $R_0, b, c$ ist also eine Lösung $p$ von

$$p'(t) = cp(t)\Big(R_0 - b\int_0^t p(s)\,ds\Big), \qquad p(0) = p_0$$

zu bestimmen. Dies ist keine Differentialgleichung, sondern eine Integro-Differentialgleichung für $p$. Man stelle für $z(t) := \int_0^t p(s)\,ds$ eine Anfangswertaufgabe erster Ordnung auf. Für $p_0 := 1, R_0 := 1, b := 0.002$ und $c := 1$ berechne man (mit Maple) $z$ und $p$, ferner plotte man beide Funktionen über dem Intervall $[0, 10]$.

## 1.2 Beispiele aus der Physik

### 1.2.1 Das mathematische Pendel und andere schwingende Systeme

Ein Massenpunkt[6] $M$ der Masse $m$ sei durch eine masselose Stange der Länge $l$ an einem festen Punkt drehbar aufgehängt. Von der Reibung im Aufhängepunkt und vom Luftwiderstand wird abgesehen. Auf $M$ wirkt als bewegende Kraft also lediglich die Schwerkraft, genauer: ihre tangentiale Komponente $-mg\sin\phi$, wobei $\phi$ der Winkel zwischen der Pendelstange und der Vertikalen, $g$ die Erdbeschleunigung ist. Eine solche Vorrichtung heißt ein *mathematisches Pendel*. In Abbildung 1.7 haben wir versucht, dies zu verdeutlichen[7]. Da die Bogenlänge $s$ vom Ruhe- oder Tiefstpunkt $R$ bis zum Massenpunkt $M$ gemessen gleich $l\phi$ ist, erhält man aus dem Newtonschen Bewegungsgesetz

$$\text{Kraft} = \text{Masse} \times \text{Beschleunigung},$$

dass (zeitliche Ableitungen werden durch einen hochgestellten Punkt verdeutlicht)

$$m\ddot{s} = ml\ddot{\phi} = -mg\sin\phi.$$

Ist die Anfangsauslenkung $\phi_0$ gegeben und befindet sich der Massenpunkt zur Anfangszeit im Ruhezustand, so hat man also die Anfangswertaufgabe

$$(*) \qquad \ddot{\phi} + \frac{g}{l}\sin\phi = 0, \qquad \phi(0) = \phi_0, \quad \dot{\phi}(0) = 0.$$

[6] Siehe z. B.

H. Heuser (1989) *Gewöhnliche Differentialgleichungen*. B. G. Teubner, Stuttgart.

[7] Die Zeichnung ist mit xfig hergestellt worden, was (zumindestens für mich noch) relativ kompliziert ist und nicht das gewünschte Resultat lieferte.

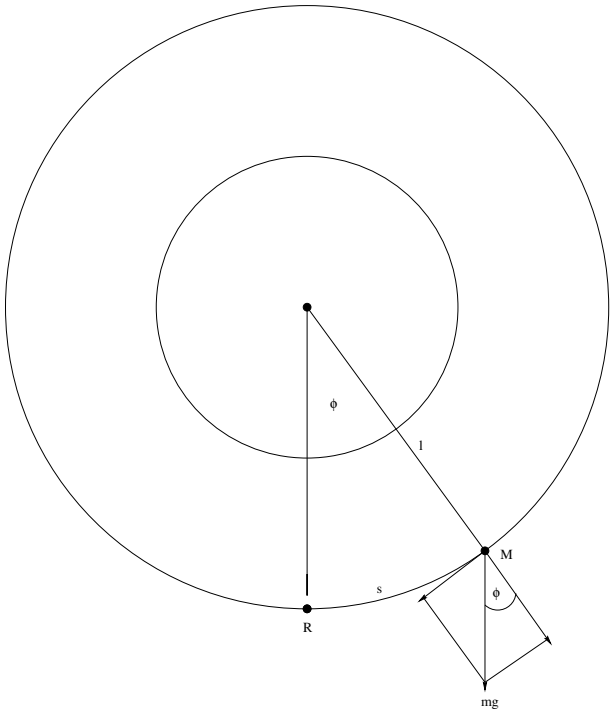


Abbildung 1.7: Das mathematische Pendel

Für kleine Winkel $\phi$ ist $\sin\phi \approx \phi$, die exakte Pendelgleichung $(*)$ geht also bei kleinen Ausschlägen über in die Näherungsgleichung (auch Gleichung des ungedämpften harmonischen Oszillators genannt)

$$(**) \qquad \ddot{\phi} + \frac{g}{l}\phi = 0, \qquad \phi(0) = \phi_0, \quad \dot{\phi}(0) = 0$$

mit der Lösung $\phi(t) = \phi_0 \cos\omega_0 t$, wobei

$$\omega_0 := \sqrt{\frac{g}{l}}$$

gesetzt wurde. Diese Näherungslösung hat die (von der Pendelmasse $m$ und der Anfangsauslenkung $\phi_0$ unabhängige) Periode $T_0 = 2\pi/\omega_0 = 2\pi\sqrt{l/g}$.

**Beispiel:** Für $\phi_0 := \pi/4$ (kein ganz kleiner Ausschlag mehr) und $\omega_0^2 = 4$ veranschaulichen wir die Lösungen von $(*)$ und $(**)$ in den Abbildungen 1.8 und 1.9. In 1.8 geben wir die Lösung $\phi$ und die zugehörige Bahn in der $(\phi, \dot{\phi})$-Phasenebene an. Dagegen findet man in Abbildung 1.9 die entsprechenden Plots für das linearisierte mathematische Pendel. Die Unterschiede sind offensichtlich nicht groß. Etwas überraschender (?) ist, dass auch das nichtlineare mathematische Pendel periodisch schwingt. In Kürze werden wir die Schwingungsdauer des mathematischen Pendels berechnen. Die beiden Abbildungen in 1.8 haben wir durch

```
eqn1:=diff(phi(t),t)=psi(t),diff(psi(t),t)=-4*sin(phi(t));
initial:=phi(0)=Pi/4;psi(0)=0;
```

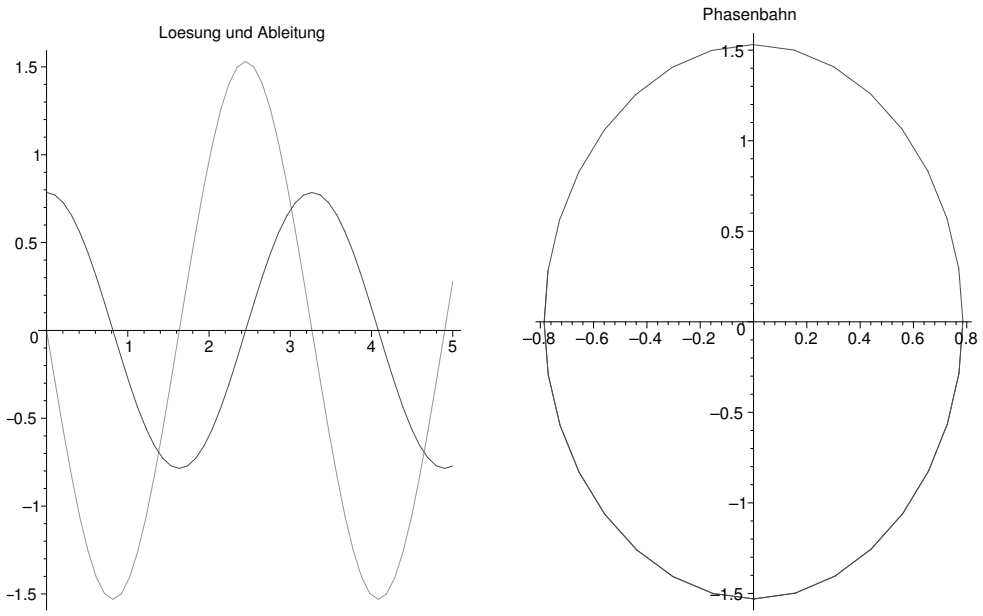


Abbildung 1.8: Mathematisches Pendel: Lösung und Bahn

```
sol1:=dsolve({eqn1,initial},{phi(t),psi(t)},type=numeric);
plots[odeplot](sol1,[[t,phi(t)],[t,psi(t)]],0..5,
          title="Loesung und Ableitung");
with(plots);
odeplot(sol1,[phi(t),psi(t)],0..5,title="Phasenbahn");
```

erhalten. Hierbei haben wir also die Anfangswertaufgabe zweiter Ordnung

$$(*) \qquad \ddot{\phi} + \omega_0^2 \sin\phi = 0, \qquad \phi(0) = \phi_0, \quad \dot{\phi}(0) = 0.$$

in eine Anfangswertaufgabe für ein System von Differentialgleichungen erster Ordnung umformuliert:

$$\begin{aligned} \dot{\phi} &= \psi, & \phi(0) &= \phi_0, \\ \dot{\psi} &= -\omega_0^2 \sin\phi, & \psi(0) &= 0. \end{aligned}$$

Entsprechendes gilt natürlich für die Abbildung 1.9. □

Definiert man

$$E(\phi, \dot{\phi}) := \frac{1}{2}\dot{\phi}^2 + \omega_0^2(1 - \cos\phi),$$

so stellt man mit einer Lösung $\phi$ von $(*)$ fest, dass

$$\frac{d}{dt}E(\phi(t), \dot{\phi}(t)) = \dot{\phi}(t)[\ddot{\phi}(t) + \omega_0^2 \sin\phi(t)] = 0.$$

Die Niveaulinien $E(\phi, \dot{\phi}) = C$ findet man in Abbildung 1.10. Diese wurden durch

```
f:=phi->4*(1-cos(phi));g:=psi->(1/2)*psi^2;
contourplot(f(phi)+g(psi),phi=-2*Pi..2*Pi,psi=-Pi..Pi,filled=true);
contourplot(f(phi)+g(psi),phi=-2*Pi..2*Pi,psi=-Pi..Pi);
```

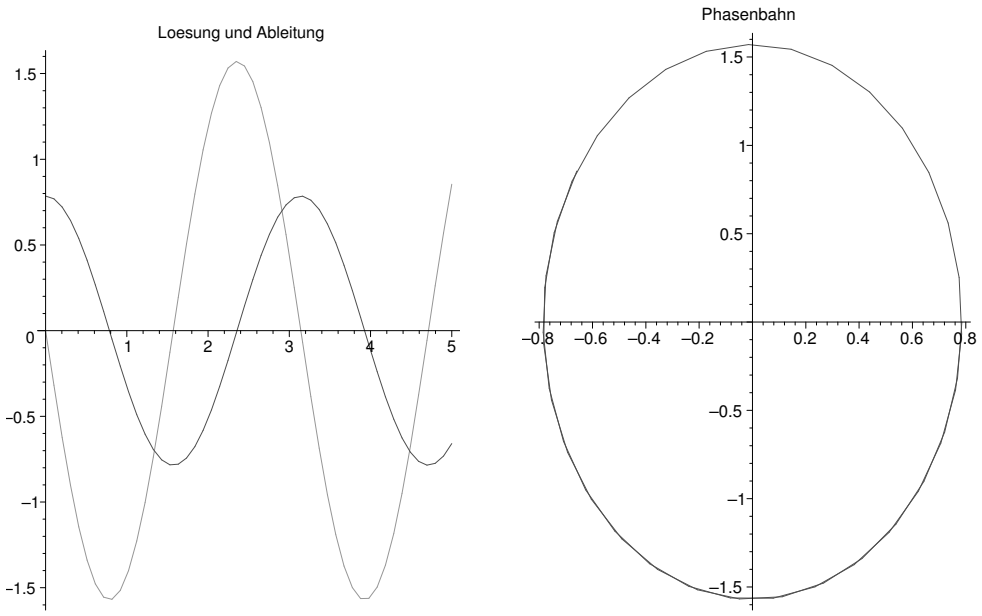


Abbildung 1.9: Linearisiertes mathematisches Pendel: Lösung und Bahn

hergestellt. Daher ist

$$\frac{1}{2}\dot{\phi}(t)^2 + \omega_0^2(1-\cos\phi(t)) = \omega_0^2(1-\cos\phi_0)$$

bzw.

$$\dot{\phi}(t)^2 = 2\omega_0^2[\cos\phi(t) - \cos\phi_0]$$

und daher

$$\dot{\phi}(t) = \pm\omega_0\sqrt{2(\cos\phi(t)-\cos\phi_0)}$$

für alle $t$. Die Pendelschwingungsdauer des mathematischen Pendels (welche von dem Anfangsausschlag $\phi_0$ abhängt) ist dann

$$T(\phi_0) = \frac{4}{\omega_0}\int_0^{\phi_0} \frac{d\phi}{\sqrt{2(\cos\phi-\cos\phi_0)}}.$$

Macht man hier die Substitution $\sin\frac{1}{2}\phi = k\sin\theta$ mit $k := \sin\frac{1}{2}\phi_0$, so erhält man unter Berücksichtigung von $\cos\phi = 1-2\sin^2\frac{1}{2}\phi$, dass

$$T(\phi_0) = \frac{4}{\omega_0}\int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \frac{4}{\omega_0}K(k),$$

wobei

$$K(k) := \int_0^{\pi/2} \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}$$

das *vollständige elliptische Integral erster Art* ist. Das Verhältnis zwischen der Schwingungsdauer des mathematischen Pendels und des linearisierten mathematischen Pendels ist also

$$\frac{T(\phi_0)}{T_0} = \frac{2K(\sin\frac{1}{2}\phi_0)}{\pi}.$$

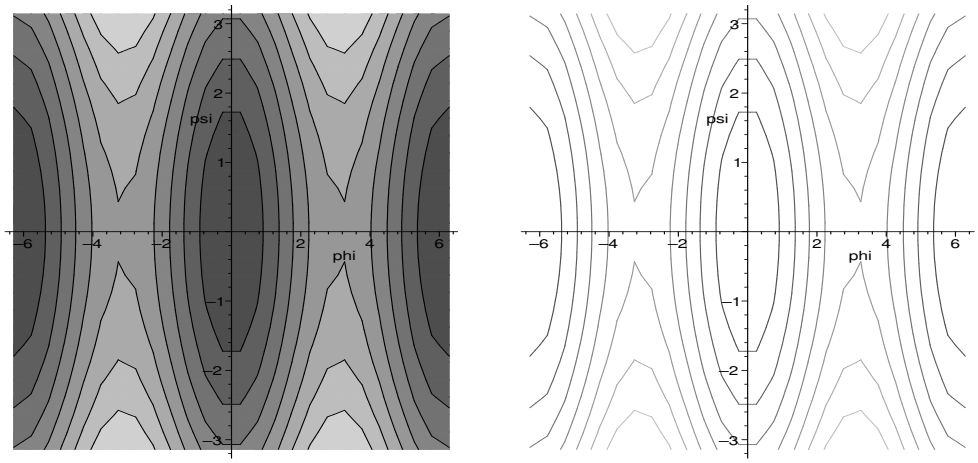


Abbildung 1.10: Niveaulinien zu $1 - \cos\phi + \frac{1}{2}\dot{\phi}^2 = C$

Dieses Verhältnis wird in Abbildung 1.11 als Funktion von $\phi_0$ aufgetragen. Den Plot

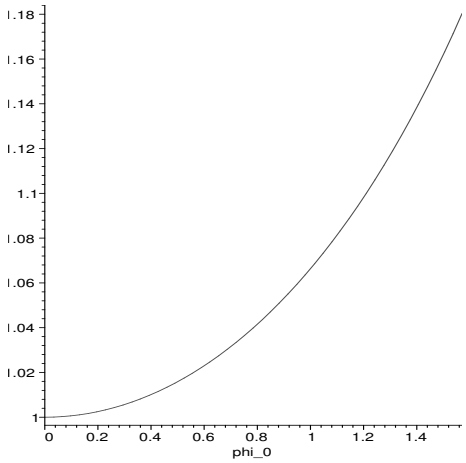


Abbildung 1.11: Das Verhältnis $T(\phi_0)/T_0$

haben wir hierbei durch

```
f:=phi_0->2*EllipticK(sin(0.5*phi_0))/Pi;
plot(f(phi_0),phi_0=0..Pi/2);
```

hergestellt.

Beim mathematischen Pendel bewegt sich der Massenpunkt am Ende der Stange (oder des Fadens) auf einem Kreisstück, dafür ist die Schwingungsdauer von dem Anfangsausschlag abhängig. Von C. Huygens (1629-1695) wurde gezeigt, dass die Schwingungsdauer eines Pendels, bei dem der Massenpunkt reibungsfrei allein unter dem Einfluss der Schwerkraft auf einem Zykloidenstück geführt wird, amplitudenunabhängig ist. In Abbildung 1.12 haben wir eine nach oben und eine nach unten geöffnete Zykloide gezeichnet:

```
plot([phi-sin(phi),1+cos(phi),phi=0..2*Pi],title="Nach oben
```

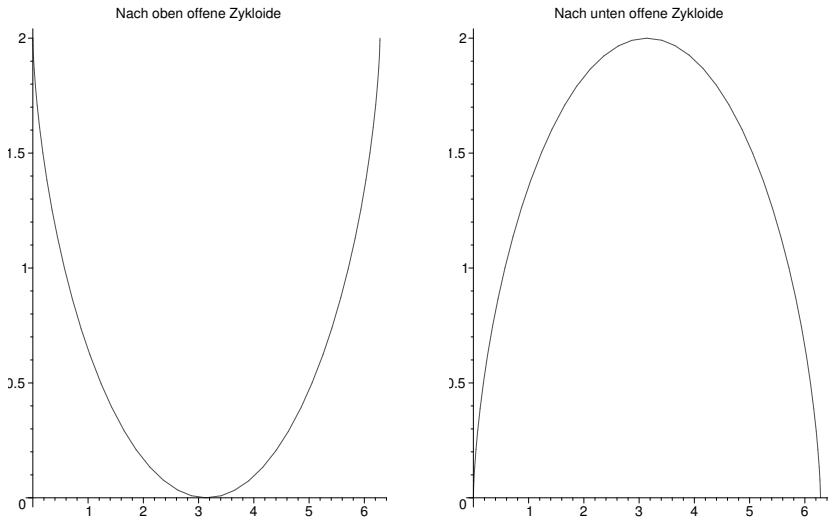


Abbildung 1.12: Zykloide: $x = r(\phi - \sin\phi)$, $y = r(1 \pm \cos\phi)$, $(0 \leq \phi \leq 2\pi)$

```
offene Zykloide");
plot([phi-sin(phi),1-cos(phi),phi=0..2*Pi],title="Nach unten
offene Zykloide");
```

Die Zykloide spielt auch bei dem klassischen Problem der Brachystochrone eine Rolle. Sind nämlich in einer vertikalen Ebene zwei Punkte $P_0$ und $P_1$ gegeben, so ist die $P_0$ und $P_1$ verbindende Zykloide unter allen $P_0$ und $P_1$ verbindenden Kurven diejenige, auf der ein nur der Schwerkraft unterworfener, reibungslos gleitender Massenpunkt in minimaler Zeit von $P_0$ nach $P_1$ gelangt.

Bei *Federschwingungen* stellt man sich im einfachsten Fall die Frage, wie sich ein an einer Feder befestigter Massenpunkt mit der Masse $m$ bewegt. Sei $x(t)$ die Auslenkung des Massenponktes zur Zeit $t$. Nimmt man an, dass die Feder dem Hookeschen Gesetz genügt, d. h. die Rückstellkraft proportional zur Auslenkung des Massenpunktes ist, so ist die auf den Massenpunkt wirkende Kraft durch $-cx(t)$ gegeben. Hierbei ist $c > 0$ die sogenannte Federkonstante. Das Minus-Zeichen tritt auf, da die Kraft der positiven $x$-Richtung entgegenwirkt. Als Bewegungsgleichung hat man also

$$m\ddot{x} = -cx,$$

die Gleichung des harmonischen Oszillators. Man kann leicht nachrechnen, dass für beliebige Konstanten $c_1, c_2$ die Funktion

$$x(t) := c_1 \cos\sqrt{\frac{c}{m}}t + c_2 \sin\sqrt{\frac{c}{m}}t$$

eine Lösung ist. Die Konstanten $c_1, c_2$ sind durch die Anfangsbedingungungen $x(0) = x_0$ und $\dot{x}(0) = \dot{x}_0$ mit vorgebenem $(x_0, \dot{x}_0)$ festgelegt:

```
> dsolve(m*diff(x(t),t,t)=-c*x(t),x(t));
```

$$\mathrm{x}(t) = \_C1\cos(\frac{\sqrt{c\,m}\,t}{m}) + \_C2\sin(\frac{\sqrt{c\,m}\,t}{m})$$

```
> dsolve({m*diff(x(t),t,t)=-c*x(t),x(0)=x_0,D(x)(0)=y_0},x(t));
```

$$\mathrm{x}(t) = x\_0\cos(\frac{\sqrt{c\,m}\,t}{m}) + \frac{y\_0\,m\sin(\frac{\sqrt{c\,m}\,t}{m})}{\sqrt{c\,m}}$$

Natürlich können die Verhältnisse komplizierter sein. Z. B. können mehrere Massenpunkte mit Federn verbunden sein, ferner können Reibungskräfte auftreten, die i. Allg. als proportional zur Geschwindigkeit angenommen werden.

Wir formulieren den folgenden Satz[8], dessen Beweis aus wesentlich allgemeineren Ergebnissen folgt, der aber hier schon als Übungsaufgabe gestellt wird.

**Satz 2.1** *Die Anfangswertaufgabe*

$$\ddot{x} + a\dot{x} + bx = 0, \qquad x(0) = x_0, \quad \dot{x}(0) = \dot{x}_0$$

*mit vorgegebenen reellen Zahlen $a, b$ sowie $x_0, \dot{x}_0$ besitzt genau eine (reelle) Lösung $x$. Diese kann je nach dem Vorzeichen der Diskriminante $\Delta := a^2 - 4b$ in einer der folgenden Formen dargestellt werden, wobei die Konstanten $c_1$ und $c_2$ durch die Anfangsbedingungen eindeutig festgelegt sind:*

1. $x(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$ *mit* $\lambda_{1,2} := (-a \pm \sqrt{\Delta})/2$, *falls* $\Delta > 0$,
2. $x(t) = (c_1 + c_2 t)e^{-(a/2)t}$, *falls* $\Delta = 0$,
3. $x(t) = e^{-(a/2)t}(c_1 \cos \beta t + c_2 \sin \beta t)$ *mit* $\beta := \sqrt{-\Delta}/2$, *falls* $\Delta < 0$.

**Bemerkung:** Sind in der Differentialgleichung $\ddot{x} + a\dot{x} + bx = 0$ der Reibungskoeffizient $a$ und die lineare Rückstellkraft (Federkonstante) $b$ positiv, so ist der erste Fall aperiodisch (Lösung der Anfangswertaufgabe strebt asymptotisch gegen 0 für $t \to \infty$), der zweite wird der aperodische Grenzfall genannt, während es sich im dritten Fall um eine gedämpfte Schwingung handelt. In den folgenden Abbildungen werden diese Fälle illustriert, wobei wir neben der Lösung auch noch die Bahn in der $(x, \dot{x})$-Phasenebene angeben. In Abbildung 1.13 findet man den aperiodischen Fall, in 1.14 den aperiodischen Grenzfall und schließlich in 1.15 eine gedämpfte Schwingung. □

### 1.2.2 Planetenbahnen

Das Newtonsche Gravitationsgesetz sagt aus, dass zwei (punktförmige) Objekte aufeinander eine Kraft ausüben, deren Länge (die Kraft ist ein Vektor) linear jeweils von ihrer Masse und umgekehrt proportional vom Quadrat ihrer Entfernung abhängt. Im folgenden sei das eine Objekt die Sonne mit der Masse $M$, das andere Objekt ein gegebener Planet mit der Masse $m$. Der Planet bewegt sich in einer Ebene, als dessen

[8] Siehe Satz 14.2 bei

H. Heuser (1989) *Gewöhnliche Differentialgleichungen*. B. G. Teubner, Stuttgart.

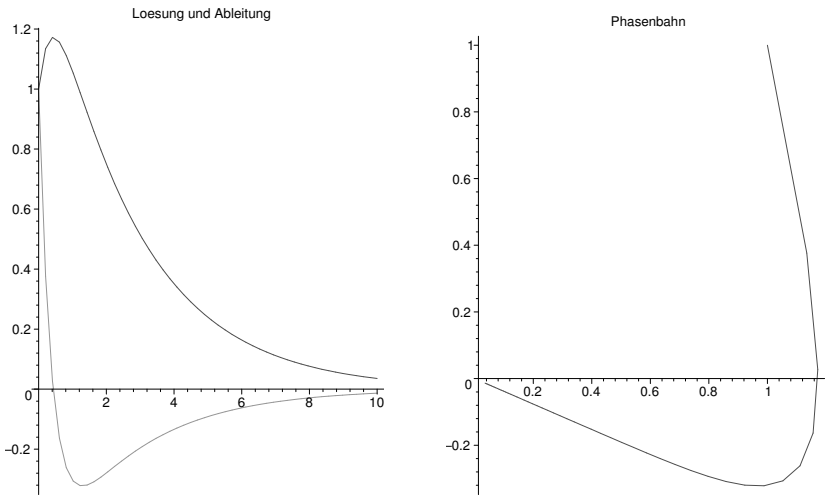


Abbildung 1.13: $\ddot{x} + 3\dot{x} + x = 0$, $x(0) = 1$, $\dot{x}(0) = 1$

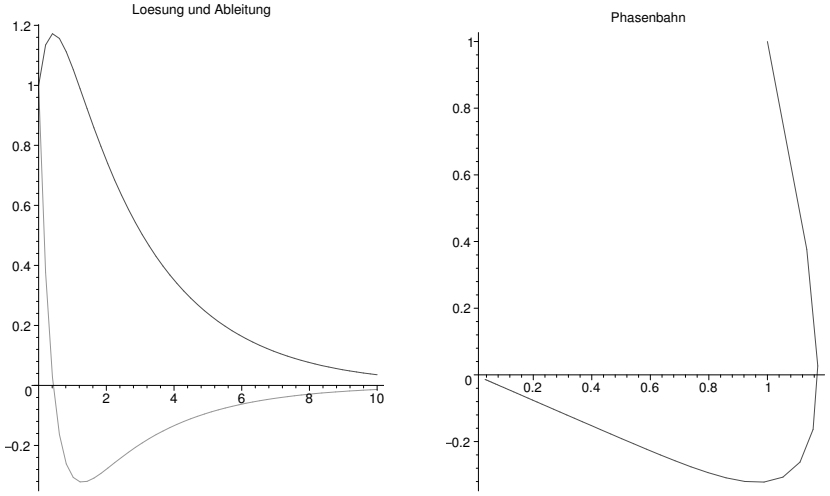


Abbildung 1.14: $\ddot{x} + \dot{x} + 0.25x = 0$, $x(0) = 0.5$, $\dot{x}(0) = 1.75$

Nullpunkt die (unbewegliche) Sonne genommen wird. Die Bewegungsgleichungen sind dann

$$(*) \qquad m\ddot{x} = -\frac{\gamma m M}{\|x\|^3} x.$$

Hierbei ist $x(t) = (x_1(t), x_2(t))$ der Ort des Planeten zur Zeit $t$ und $\gamma$ die Gravitationskonstante ist. Die Bewegungsgleichung $(*)$ ist also eigentlich ein System von zwei Differentialgleichungen zweiter Ordnung, nämlich

$$(*) \qquad \begin{aligned} \ddot{x}_1 &= -\frac{\gamma M}{(x_1^2 + x_2^2)^{3/2}} x_1, \\ \ddot{x}_2 &= -\frac{\gamma M}{(x_1^2 + x_2^2)^{3/2}} x_2. \end{aligned}$$

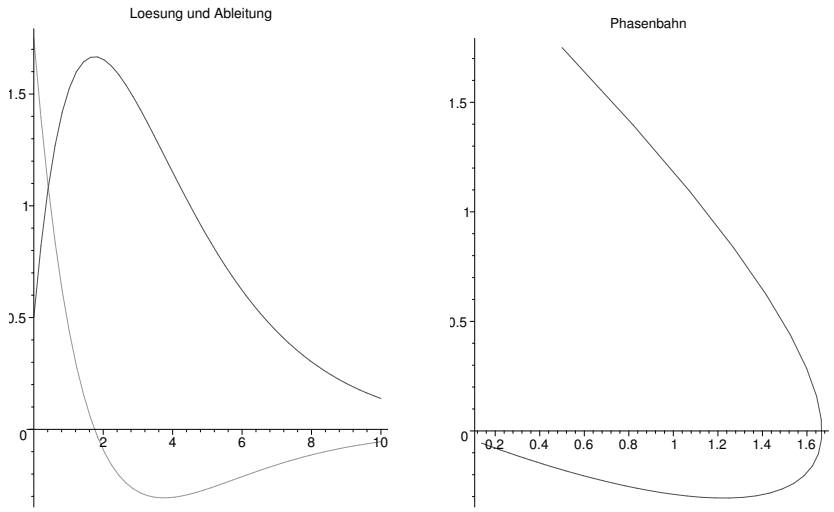


Abbildung 1.15: $\ddot{x} + 0.125\dot{x} + x = 0$, $x(0) = 2$, $\dot{x}(0) = 0$

Das Zweikörperproblem besteht darin, die Bahn des Planeten zu beschreiben. Die Lösung dieses Problems durch J. Kepler (1571-1630) gehört sicherlich zu den größten Leistungen in der Geschichte der Menschheit. Aus den Bewegungsgleichungen $(*)$ wollen wir die drei Keplerschen Gesetze[9] der Planetenbewegung ableiten[10]. Dies sind bekanntlich:

(K1) Die Bahnen der Planeten sind Ellipsen, in deren einem Brennpunkt die Sonne steht.

(K2) Der von der Sonne zu einem Planeten weisende Radiusvektor überstreicht in gleichen Zeiten gleiche Flächen.

(K3) Das Verhältnis zwischen dem Quadrat der Umlaufzeit und dem Kubus der großen Achse (der Bahnellipse) ist für alle Planeten des Sonnensystems konstant.

**Satz 2.2** *Für Planeten mit der Bewegungsgleichung* $(*)$ *gelten die drei Keplerschen Gesetze.*

**Beweis:** Wir suchen eine Lösung der Anfangswertaufgabe

$$\begin{aligned} \ddot{x}_1 &= -\frac{\gamma M}{(x_1^2+x_2^2)^{3/2}} x_1, & x_1(0) &= R, & \dot{x}_1(0) &= v_1, \\ \ddot{x}_2 &= -\frac{\gamma M}{(x_1^2+x_2^2)^{3/2}} x_2, & x_2(0) &= 0, & \dot{x}_2(0) &= v_2, \end{aligned}$$

[9]Es gibt im Internet einige Seiten, auf denen animierte Erläuterungen der Keplerschen Gesetze zu sehen sind. Man sehe sich z. B. an:

`http://www.cvc.org/science/kepler.htm`

`http://www.phy.syr.edu/courses/java/mc_html/kepler.html`

`http://sunsite.ubc.ca/LivingMathematics/V001N01/UBCExamples/Kepler/kepler.html`

[10]Wir folgen

W. Walter (1990) *Analysis II*. Springer-Verlag, Berlin-Heidelberg, New York.

die sich darstellen lässt in der Form

$$x_1(t) = r(t)\cos\phi(t), \qquad x_2(t) = r(t)\sin\phi(t),$$

wobei wir voraussetzen, dass $R > 0$ der Abstand des Planeten von der Sonne zur Zeit $t_0 = 0$ ist. Ferner wird $v_2 \neq 0$ vorausgesetzt (andernfalls wäre $x_2(t) = 0$, die Bewegung des Planeten würde auf einer Geraden erfolgen). Benutzt man die komplexe Schreibweise $z(t) = r(t)e^{i\phi(t)}$, so ist

$$\dot{z} = (\dot{r} + ir\dot{\phi})e^{i\phi}, \qquad \ddot{z} = (\ddot{r} + 2i\dot{r}\dot{\phi} + ir\ddot{\phi} - r\dot{\phi}^2)e^{i\phi}.$$

Die Bewegungsgleichungen sind also äquivalent zu

$$\ddot{r} + 2i\dot{r}\dot{\phi} + ir\ddot{\phi} - r\dot{\phi}^2 = -\frac{\gamma M}{r^2}.$$

Zerlegt man in Real- und Imaginärteil, so ergeben sich die beiden folgenden, mit $(*)$ äquivalenten Gleichungen:

$$\ddot{r} - r\dot{\phi}^2 + \frac{\gamma M}{r^2} = 0, \qquad 2\dot{r}\dot{\phi} + r\ddot{\phi} = 0. \tag{1}$$

Die Anfangsbedingungen sind gegeben durch

$$r(0) = R, \qquad \phi(0) = 0, \qquad \dot{r}(0) = v_1, \qquad \dot{\phi}(0) = \frac{v_2}{R}. \tag{2}$$

Also ist die gegebene Anfangswertaufgabe zu (1), (2) äquivalent.

Zum Nachweis des ersten Keplerschen Gesetzes gehen wir folgendermaßen vor. Zunächst definieren wir mit noch unbekannten Konstanten $\epsilon \geq 0$, $p > 0$ und $0 \leq \alpha < 2\pi$ die Funktion $f$ durch

$$f(\phi) := \frac{p}{1 + \epsilon\cos(\phi - \alpha)}.$$

Anschließend sei

$$t(\phi) := \frac{1}{A}\int_0^\phi f^2(\phi)\, d\phi \qquad \text{mit} \quad A := Rv_2,$$

also

$$\frac{dt(\phi)}{d\phi} = \frac{1}{A}f^2(\phi), \qquad t(0) = 0.$$

Mit $\phi = \phi(t)$ sei die Umkehrfunktion zu $t = t(\phi)$ bezeichnet und $r(t) := f(\phi(t))$ gesetzt. Aus

$$t = t(\phi(t)) = \frac{1}{A}\int_0^{\phi(t)} f^2(\phi)\, d\phi$$

erhält man durch Differentiation nach $t$, dass

$$\dot{\phi}(t) = \frac{A}{f^2(\phi(t))}.$$

Wir wollen uns überlegen, dass bei geeigneter Wahl der noch freien Konstanten $\epsilon, p, \alpha$ durch $(r, \phi)$ eine Lösung von (1) und (2) gegeben ist. Am einfachsten ist die zweite Gleichung in (1) einzusehen. Es ist nämlich

$$r^2(t)\dot{\phi}(t) = f^2(\phi(t))\frac{A}{f^2(\phi(t))} = A,$$

insbesondere also auch

$$0 = \frac{d}{dt}[r^2(t)\dot{\phi}(t)] = r(t)[2\dot{r}(t)\dot{\phi}(t) + r(t)\ddot{\phi}(t)].$$

Wegen $r(t) > 0$ ist die zweite Gleichung in (1) erfüllt. Von den Anfangsbedingungen in (2) ist $\phi(0) = 0$ schon erfüllt (wegen $t(0) = 0$). Nun kommen wir zu der ersten Gleichung in (1). Zunächst folgt aus $r(t) := f(\phi(t))$, dass

$$\begin{aligned}
\dot{r}(t) &= f'(\phi(t))\dot{\phi}(t) \\
&= f'(\phi(t))\frac{A}{f^2(\phi(t))} \\
&= \frac{p\epsilon\sin(\phi(t)-\alpha)}{(1+\epsilon\cos(\phi(t)-\alpha))^2} \cdot \frac{A(1+\epsilon\cos(\phi(t)-\alpha))^2}{p^2} \\
&= \frac{\epsilon A}{p}\sin(\phi(t)-\alpha),
\end{aligned}$$

anschließend

$$\ddot{r}(t) = \frac{\epsilon A}{p}\cos(\phi(t)-\alpha)\dot{\phi}(t).$$

Folglich ist

$$\begin{aligned}
r^2(t)\Big[\ddot{r}(t) - r(t)\dot{\phi}^2(t) + \frac{\gamma M}{r^2(t)}\Big] &= r^2(t)\Big[\frac{\epsilon A}{p}\cos(\phi(t)-\alpha)\dot{\phi}(t) - \frac{A^2}{r^3(t)} + \frac{\gamma M}{r^2(t)}\Big] \\
&= \frac{\epsilon A^2}{p}\cos(\phi(t)-\alpha) - \frac{A^2}{r(t)} + \gamma M \\
&= \frac{\epsilon A^2}{p}\cos(\phi(t)-\alpha) - \frac{A^2}{p}[1+\epsilon\cos(\phi(t)-\alpha)] + \gamma M \\
&= \gamma M - \frac{A^2}{p}.
\end{aligned}$$

Setzt man also

$$p := \frac{A^2}{\gamma M} = \frac{R^2 v_2^2}{\gamma M},$$

so erfüllt $(r, \phi)$ auch die erste Gleichung in (1). Die Konstanten $\epsilon$ und $\alpha$ werden durch die Anfangsbedingungen $r(0) = R$ und $\dot{r}(0) = v_1$ festgelegt. Dies führt auf die Gleichungen

$$R = \frac{p}{1+\epsilon\cos\alpha}, \qquad v_1 = -\frac{\epsilon A}{p}\sin\alpha$$

bzw. nach Einsetzen von $p$ auf

$$\epsilon \cos\alpha = \frac{Rv_2^2 - \gamma M}{\gamma M}, \qquad \epsilon \sin\alpha = -\frac{Rv_1 v_2}{\gamma M}.$$

Diese beiden Gleichungen sind lösbar, da $\epsilon \geq 0$ und $\alpha \in [0, 2\pi)$ als Polarkoordinaten des Punktes

$$Q := \frac{1}{\gamma M}(Rv_2^2 - \gamma M, -Rv_1 v_2)$$

bestimmt werden können. Für $Q \neq 0$ sind $\epsilon \geq 0$ und $\alpha \in [0, 2\pi)$ sogar eindeutig festgelegt. Die letzte Anfangsbedingung ist ebenfalls erfüllt:

$$\dot{\phi}(0) = \frac{A}{f^2(\phi(0))} = \frac{A}{r^2(0)} = \frac{Rv_2}{R^2} = \frac{v_2}{R}.$$

Damit ist gezeigt: Die Anfangswertaufgabe (1), (2) besitzt eine Lösung $(r, \phi)$ mit

$$r(t) = \frac{p}{1 + \epsilon\cos(\phi(t) - \alpha)},$$

wobei $p > 0$, $\epsilon \geq 0$ und $\alpha \in [0, 2\pi)$ geeignete Konstanten sind.

Durch

$$K := \{(f(\phi)\cos\phi, f(\phi)\sin\phi) : \phi \in [0, 2\pi]\}$$

mit

$$f(\phi) := \frac{p}{1 + \epsilon\cos(\phi - \alpha)}$$

ist ein Kegelschnitt gegeben und zwar eine Ellipse ($\epsilon < 1$), eine Parabel ($\epsilon = 1$) oder eine Hyperbel ($\epsilon > 1$). Wir gehen jetzt davon aus, dass die Anfangsdaten so vernünftig sind, dass es sich bei der Planetenbahn um eine Ellipse handelt, da sie ja geschlossen ist. Damit ist das erste Keplersche Gesetz bewiesen. Die Ellipse habe die Halbachsen $a \geq b$. Diese lassen sich aus $\epsilon$ und $p$ berechnen und man erhält (siehe Aufgabe 5)

$$a = \frac{p}{1 - \epsilon^2}, \qquad b = \frac{p}{\sqrt{1 - \epsilon^2}}.$$

Nun zum zweiten Keplerschen Gesetz. Man bezeichne mit $F(t_1, t_2)$ die Größe der vom Fahrstrahl für $t_1 \leq t \leq t_2$ überstrichenen Fläche, also den Flächeninhalt des von den Strahlen $\phi = \phi(t_1)$, $\phi = \phi(t_2)$ und der Kurve $f(\phi)e^{i\phi}$, $\phi(t_1) \leq \phi \leq \phi(t_2)$, begrenzten Gebietes

$$S_{1,2} := \{(r\cos\phi, r\sin\phi) : 0 \leq r \leq f(\phi),\ \phi(t_1) \leq \phi \leq \phi(t_2)\}.$$

Dann ist

$$F(t_1, t_2) = \frac{1}{2}\int_{\phi(t_1)}^{\phi(t_2)} f^2(\phi)\, d\phi = \frac{1}{2}\int_{t_1}^{t_2} f^2(\phi(t))\dot{\phi}(t)\, dt = \frac{1}{2}\int_{t_1}^{t_2} r^2(t)\dot{\phi}(t)\, dt = \frac{A}{2}(t_2 - t_1).$$

Hierbei erhält man die erste Gleichung (Leibnizsche Sektorformel[11]) z. B. aus der Transformationsformel für mehrfache Integrale. Damit ist auch das zweite Keplersche Gesetz bewiesen.

[11] Siehe

W. Walter (1990, S. 251) *Analysis II*. Springer-Verlag. Berlin-Heidelberg-New York.

Nun sei $T$ die Umlaufzeit des Planeten, also $\phi(T) = 2\pi$. Dann ist $F(0,T) = \frac{1}{2}AT$ die Fläche der Ellipse, die andererseits bekanntlich $\pi ab$ beträgt. Folglich ist $T = 2\pi ab/A$ und daher

$$T^2 = \frac{4\pi^2 a^2 b^2}{A^2} = \frac{4\pi^2 a^2 b^2}{p\gamma M} = \frac{4\pi^2 a^2 b^2}{(b^2/a)\gamma M} = \frac{4\pi^2}{\gamma M} a^3.$$

Damit ist auch das dritte Keplersche Gesetz bewiesen. □

**Beispiel:** Wir wollen einmal die Anfangswertaufgabe

$$\begin{aligned} \ddot{x}_1 &= -\frac{10}{(x_1^2+x_2^2)^{3/2}} x_1, & x_1(0) &= 3, & \dot{x}_1(0) &= 1, \\ \ddot{x}_2 &= -\frac{10}{(x_1^2+x_2^2)^{3/2}} x_2, & x_2(0) &= 0, & \dot{x}_2(0) &= 1 \end{aligned}$$

mit Maple lösen und die zugehörige Planetenbahn in einer $(x_1, x_2)$-Ebene plotten. Anschließend werden wir die Phasenbahn mit der im Beweis für die Keplerschen Gesetze erhaltenen vergleichen. Mit Hilfe der Eingabe

```
eqn1:=diff(x_1(t),t,t)=-10*x_1(t)/(x_1(t)^2+x_2(t)^2)^(3/2);
eqn2:=diff(x_2(t),t,t)=-10*x_2(t)/(x_1(t)^2+x_2(t)^2)^(3/2);
initial1:=x_1(0)=3,D(x_1)(0)=1;
initial2:=x_2(0)=0,D(x_2)(0)=1;
two_body:=dsolve({eqn1,eqn2,initial1,initial2},{x_1(t),x_2(t)},
      type=numeric);
with(plots);
odeplot(two_body,[x_1(t),x_2(t)],0..10,
     numpoints=200,title="Planetenbahn");
```

erhalten wir den in der folgenden Abbildung 1.16 links angegebenen Plot. Natürlich hätten wir das System von zwei Differentialgleichungen zweiter Ordnung auch als System von vier Differentialgleichungen erster Ordnung schreiben können. In der Darstellung der Ellipse

$$K := \{(f(\phi)\cos\phi, f(\phi)\sin\phi) : 0 \le \phi \le 2\pi\}, \qquad f(\phi) := \frac{p}{1+\epsilon\cos(\phi-\alpha)},$$

erhalten wir $p = 0.9$, während $\epsilon \ge 0$ und $\alpha \in [0, 2\pi)$ aus

$$\epsilon\cos\alpha = -0.7, \qquad \epsilon\sin\alpha = -0.3$$

zu berechnen sind. Dies ergibt mittels

```
fsolve({epsilon*cos(alpha)=-0.7,epsilon*sin(alpha)=-0.3},
     {epsilon,alpha},{epsilon=0..1,alpha=0..2*Pi});
```

die Werte

$$\epsilon = 0.7615773106, \qquad \alpha = 3.546484440.$$

Mit

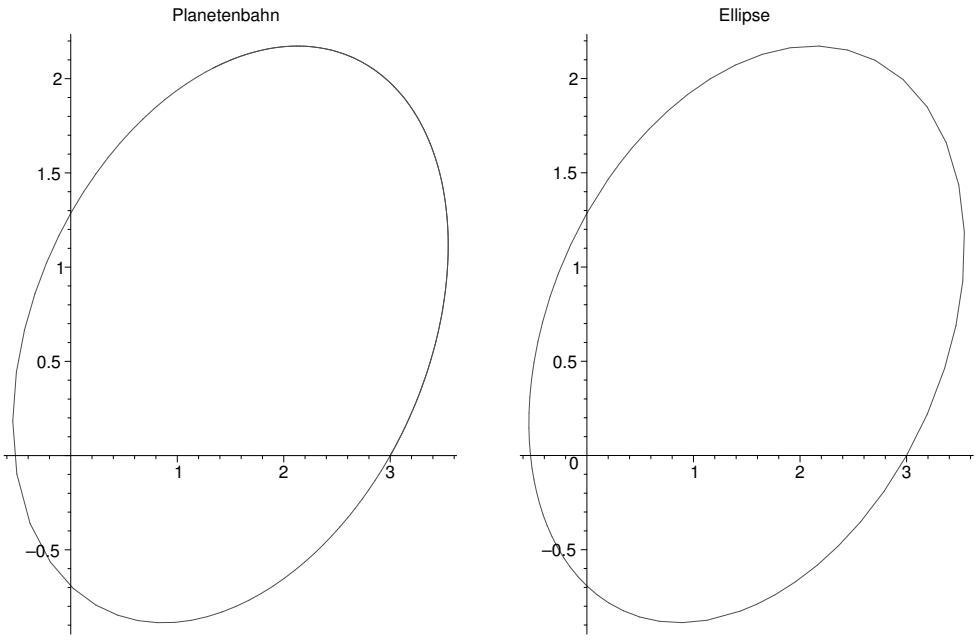


Abbildung 1.16: Eine auf zwei Arten berechnete Planetenbahn

```
p:=0.9;epsilon:=0.7615773106;alpha:=3.546484440;
f:=phi->p/(1+epsilon*cos(phi-alpha));
plot([f(phi)*cos(phi),f(phi)*sin(phi),phi=0..2*Pi],title="Ellipse");
```

erhalten wir die Ellipse in Abbildung 1.16 rechts. Einen Unterschied zum linken Pendant wird man nicht feststellen. □

Die Keplerschen Gesetze lösen das sogenannte Zweikörperproblem. Noch wesentlich interessanter und schwieriger ist das Dreikörperproblem[12]. Wir betrachten also die Bewegung von drei Massenpunkten mit Massen $m_1$ (diese sei die größte und stelle die Sonne dar), $m_2$ und $m_3$ unter dem Einfluss ihrer wechselseitigen Schwerkraftanziehung. Ist $x_i(t)$ die Position des $i$-ten Massenpunktes, $i = 1, 2, 3$, zur Zeit $t$ (im $\mathbb{R}^3$ im allgemeinen Fall, im $\mathbb{R}^2$ im sogenannten planaren Dreikörperproblem), so hat man als Bewegungsgleichungen ein System von drei Differentialgleichungen zweiter Ordnung, jeweils mit zwei (planarer Fall) oder drei Komponenten. Nach den Newtonschen Bewegungsgleichungen sind dies

$$\begin{aligned}
\ddot{x}_1 &= \frac{\gamma m_2}{\|x_2 - x_1\|^3}(x_2 - x_1) + \frac{\gamma m_3}{\|x_3 - x_1\|^3}(x_3 - x_1),\\
\ddot{x}_2 &= \frac{\gamma m_1}{\|x_1 - x_2\|^3}(x_1 - x_2) + \frac{\gamma m_3}{\|x_3 - x_2\|^3}(x_3 - x_2),\\
\ddot{x}_3 &= \frac{\gamma m_1}{\|x_1 - x_3\|^3}(x_1 - x_3) + \frac{\gamma m_2}{\|x_2 - x_3\|^3}(x_2 - x_3).
\end{aligned}$$

[12]Wir benutzen u. a.

W. Gander, J. Hřebiček (1993) *Solving Problems in Scientific Computing using Maple and MATLAB*. Springer-Verlag, Berlin-Heidelberg-New York.

R. Gass (1998) Mathematica *for Scientists and Engineers*. Prentice-Hall, Upper Saddle River.

Bei R. Gass (1998, S. 94 ff.) wird

$$M := m_1 + m_2 + m_3, \qquad \gamma := \frac{4\pi^2}{M}$$

gesetzt. Wir betrachten nur den planaren Fall (dies geschieht implizit bei Gass ebenfalls, da bei den Anfangsbedingungen stets die dritte Komponente als verschwindend vorausgesetzt wurde, so dass die Bewegung der Planeten in einer Ebene erfolgt). Als Anfangsbedingungen nehmen wir stets

$$x_1(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad x_1'(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(die Sonne befindet sich also zur Zeit $t = 0$ im Ursprung des Koordinatensystems), an den Anfangszuständen der beiden anderen Planeten kann "gedreht" werden. Es soll ein Maple-Programm zur Berechnung von $x_1(\cdot)$, $x_2(\cdot)$ und $x_3(\cdot)$ aufgestellt werden, anschließend sollen für gewisse Massenverhältnisse und Anfangsbedingungen die Planetenbahnen visualisiert werden.

Wir geben ziemlich genau wie bei R. Glass (1998, S. 91–96) ein Programm für Planetenbahnen an, bei dem die Massenverhältnisse etwa denen zwischen Sonne, Jupiter und Erde entsprechen. In Abbildung 1.17 werden die entsprechenden Bahnen von Jupiter und Erde angegeben.

```
m_1:=1; m_2:=0.001; m_3:=0.000001; M:=m_1+m_2+m_3; gamm:=4*Pi^2/M;
eqn1a:=diff(u_1(t),t,t)=gamm*m_2*(u_2(t)-u_1(t))/((u_2(t)-u_1(t))^2+(v_2(t)-v_1(t))^2)^(3/2)
+gamm*m_3*(u_3(t)-u_1(t))/((u_3(t)-u_1(t))^2+(v_3(t)-v_1(t))^2)^(3/2):
eqn1b:=diff(v_1(t),t,t)=gamm*m_2*(v_2(t)-v_1(t))/((u_2(t)-u_1(t))^2+(v_2(t)-v_1(t))^2)^(3/2)
+gamm*m_3*(v_3(t)-v_1(t))/((u_3(t)-u_1(t))^2+(v_3(t)-v_1(t))^2)^(3/2):
eqn2a:=diff(u_2(t),t,t)=gamm*m_1*(u_1(t)-u_2(t))/((u_1(t)-u_2(t))^2+(v_1(t)-v_2(t))^2)^(3/2)
+gamm*m_3*(u_3(t)-u_2(t))/((u_3(t)-u_2(t))^2+(v_3(t)-v_2(t))^2)^(3/2):
eqn2b:=diff(v_2(t),t,t)=gamm*m_1*(v_1(t)-v_2(t))/((u_1(t)-u_2(t))^2+(v_1(t)-v_2(t))^2)^(3/2)
+gamm*m_3*(v_3(t)-v_2(t))/((u_3(t)-u_2(t))^2+(v_3(t)-v_2(t))^2)^(3/2):
eqn3a:=diff(u_3(t),t,t)=gamm*m_1*(u_1(t)-u_3(t))/((u_1(t)-u_3(t))^2+(v_1(t)-v_3(t))^2)^(3/2)
+gamm*m_2*(u_2(t)-u_3(t))/((u_2(t)-u_3(t))^2+(v_2(t)-v_3(t))^2)^(3/2):
eqn3b:=diff(v_3(t),t,t)=gamm*m_1*(v_1(t)-v_3(t))/((u_1(t)-u_3(t))^2+(v_1(t)-v_3(t))^2)^(3/2)
+gamm*m_2*(v_2(t)-v_3(t))/((u_2(t)-u_3(t))^2+(v_2(t)-v_3(t))^2)^(3/2):
ini1a:=u_1(0)=0,D(u_1)(0)=0:
ini1b:=v_1(0)=0,D(v_1)(0)=0:
ini2a:=u_2(0)=1,D(u_2)(0)=0:
ini2b:=v_2(0)=0,D(v_2)(0)=sqrt(gamm*M):
ini3a:=u_3(0)=1.5,D(u_3)(0)=0:
ini3b:=v_3(0)=0,D(v_3)(0)=-sqrt(gamm*M/1.5):
sol:=dsolve({eqn1a,eqn1b,eqn2a,eqn2b,eqn3a,eqn3b,ini1a,ini1b,ini2a,ini2b,ini3a,ini3b},
 {u_1(t),v_1(t),u_2(t),v_2(t),u_3(t),v_3(t)},type=numeric):
with(plots):
plot1:=odeplot(sol,[u_1(t),v_1(t)],0..10,numpoints=200,color=red):
plot2:=odeplot(sol,[u_2(t),v_2(t)],0..10,numpoints=200,color=green):
plot3:=odeplot(sol,[u_3(t),v_3(t)],0..10,numpoints=200,color=blue):
display(plot1,plot2,plot3);
```

In Abbildung 1.17 links geben wir die Bewegungen der Sonne, des Jupiter und der Erde bezüglich eines festen Koordinatensystems an, wobei man die Bewegung der Sonne nicht sieht, da sie verschwindend gering ist verglichen mit der der Planeten. Daher wollen wir in Abbildung 1.17 die Bewegung von Jupiter und Erde auch noch in einem heliozentrischen Koordinatensystem angeben. Dies geschieht durch

```
plot4:=odeplot(sol,[u_2(t)-u_1(t),v_2(t)-v_1(t)],0..10,numpoints=200,color=green):
plot5:=odeplot(sol,[u_3(t)-u_1(t),v_3(t)-v_1(t)],0..10,numpoints=200,color=blue):
display(plot4,plot5);
```

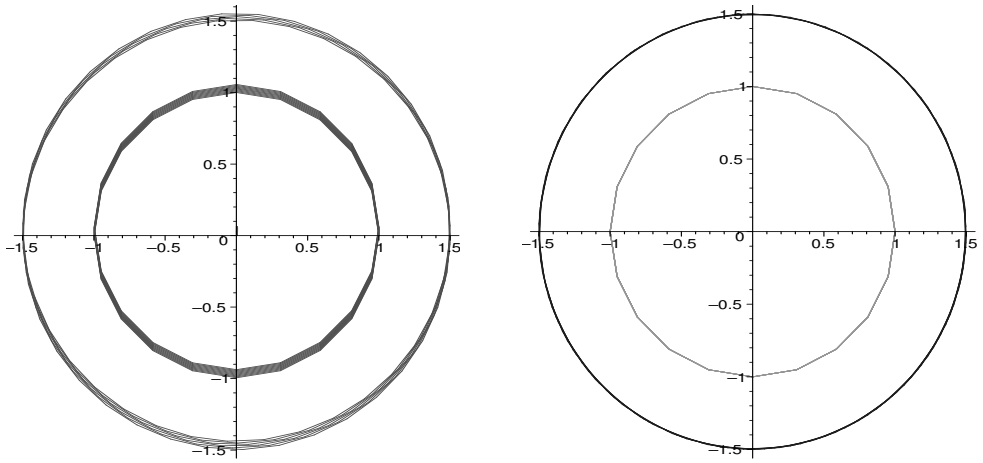

Abbildung 1.17: Bahnen von Jupiter und Erde um die Sonne

Man erkennt, dass keine Schwankungen mehr auftreten. Bei anderen Daten erhält man völlig andere Bilder. Setzt man z. B. (wie bei R. Gass (1998, S. 98)) $m_1 = 1$, $m_2 = 0.02$, $m_3 = 0.05$ und nimmt als Anfangsbedingungen (abweichend vom obigen Programm) $x_2'(0) = (\sqrt{\gamma M/2}, -\sqrt{\gamma M/2})$, $x_3(0) = (2, 0)$ und $x_3'(0) = (0, \sqrt{\gamma M/4})$ und integriert man nur über das Zeitintervall $[0, 5]$, so erhält man die in Abbildung 1.18 angebenen Bahnen, wobei die rechte wieder heliozentrisch zu verstehen ist.

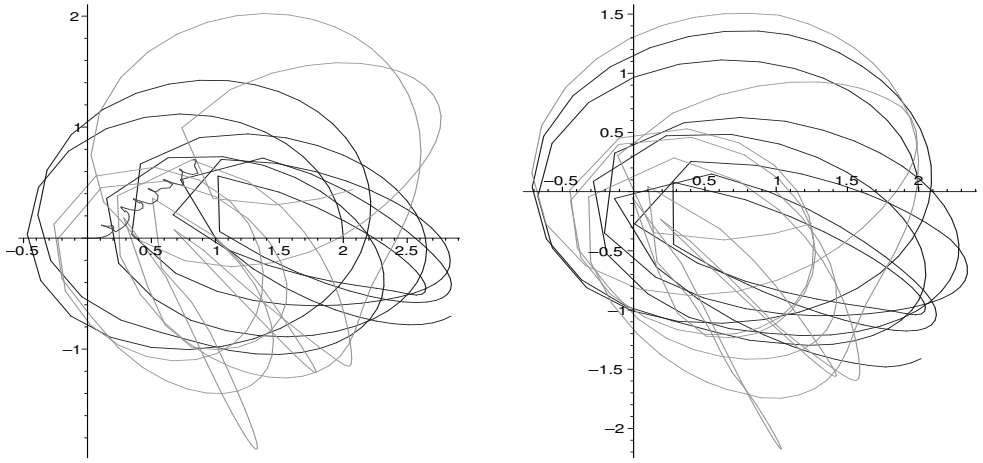

Abbildung 1.18: Bahnen beim Dreikörperproblem

## 1.2.3 Aufgaben

1. Das vollständige elliptische Integral erster Ordnung ist mit dem Gaußschen arithmetisch-geometrischen Mittel (AGM) verwandt[13]. Insbesondere zeige man:

[13] Hierüber kann man sich sehr gut auf den ersten Seiten von

J. M. Borwein, P. B. Borwein (1987) *Pi and the AGM*. J. Wiley, New York

informieren. Zu recht bezeichnen sie das arithmetisch-geometrische Mittel als eine der Juwelen der klassischen Analysis.

(a) Gegeben seien Zahlen $a, b$ mit $0 < b \leq a$. Auf die folgende Weise erzeuge man Folgen $\{a_k\}$, $\{b_k\}$.

- Setze $a_0 := a$, $b_0 := b$.
- Für $k = 0, 1, \ldots$:
  - Berechne $a_{k+1} := \frac{1}{2}(a_k + b_k)$.
  - Berechne $b_{k+1} := \sqrt{a_k b_k}$.

Man zeige: Die Folgen $\{a_k\}$ und $\{b_k\}$ konvergieren monoton nicht wachsend bzw. monoton nicht fallend gegen einen gemeinsamen Grenzwert $M(a, b)$, das sogenannte arithmetisch-geometrische Mittel von $a$ und $b$.

(b) Für $0 < b \leq a$ und $\lambda > 0$ ist $M(\lambda a, \lambda b) = \lambda M(a, b)$.

(c) Für $0 < b \leq a$ ist

$$M(a, b) = M\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

(d) Für $0 < b \leq 1$ ist

$$M(1, b) = \frac{1+b}{2} M\left(1, \frac{2\sqrt{b}}{1+b}\right).$$

(e) Für $0 < x \leq 1$ ist

$$\frac{1}{M(1, x)} = \frac{2}{\pi} K(\sqrt{1 - x^2}),$$

wobei mit

$$K(k) := \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}$$

das vollständige elliptische Integral erster Art bezeichnet wird[14].

(f) Man zeige[15], dass

$$\frac{1}{M(\sqrt{2}, 1)} = \frac{2}{\pi} \int_0^1 \frac{dt}{\sqrt{1 - t^4}}.$$

[14] Für einen Beweis kann man J. M. Borwein, P. B. Borwein (1987, S. 5) oder auch J. Todd (1979, S. 18) *Basic Numerical Mathematics*, vol. 1. Birkhäuser Verlag, Basel-Stuttgart konsultieren. Aber selbst dann wird der Beweis nicht ganz einfach sein

[15] Bei J. Todd (1979, S. 17) kann man nachlesen:

As a teenager in 1791 Gauss, without computers, made extensive calculations of arithmetic-geometric means. In particular he found that

$$M(\sqrt{2}, 1) = 1.19814\,02347\,35592\,20744.$$

It seems clear that he was searching for a formula for $M(a, b)$.... It was not until 1799 that he made progress. At that time he computed the definite integral

$$A = \int_0^1 \frac{dt}{\sqrt{1 - t^4}}.$$

He then recalled his value of $M(\sqrt{2}, 1)$ given above and observed that the product $AM(\sqrt{2}, 1)$ coincided to many decimal places with $\frac{1}{2}\pi$. In his diary, on 30 May 1799, Gauss wrote that if one could prove rigorously that $AM(\sqrt{2}, 1) = \frac{1}{2}\pi$, then new fields of mathematics would open. In his diary, on 23 December 1799, Gauss noted that he had proved this result, and more; in later years his prophesy was fulfilled.

2. Man schreibe ein Maple-Programm, mit dem k Schritte des Gauß-Verfahrens zur Berechnung von $M(a,b)$ durchgeführt werden. Insbesondere berechne man $M(\sqrt{2},1)$ und prüfe numerisch die von Gauß gefundene Identität

$$M(\sqrt{2},1)\int_0^1 \frac{dt}{\sqrt{1-t^4}} = \frac{\pi}{2}$$

nach.

3. Für einige Jahre[16] entledigte man sich in den USA eines Teils des radioaktiven Mülls, indem dieser in Fässer kam, die in die See geworfen wurden. Es wurde davon ausgegangen (nach hoffentlich sorgfältigen Tests), dass die Fässer so dicht sind, dass eine Lagerung unbedenklich ist. Es stellte sich aber die Frage, ob eine zu hohe Aufprallgeschwindigkeit zu einem Leck führen könnte. Nach Tests ergab sich, dass die Fässer ab einer Aufprallgeschwindigkeit von 12.2 m/sec platzen konnten, so dass die Aufgabe darin besteht, die Aufprallgeschwindigkeit zu ermitteln.

   Ein Fass wiege $m := 240\,\mathrm{kg}$, das Volumen sei $V := 0.21\,\mathrm{m}^3$. Der Wasserwiderstand $D$ sei proportional zur Geschwindigkeit $v$ des Fasses: $D = cv$, wobei durch Experimente $c = 0.12\,\mathrm{kg}\cdot\mathrm{sec/m}$ festgestellt wurde. Durch den Auftrieb erleidet das Fass einen Gewichtsverlust $B$, der gleich dem Gewicht des verdrängten Salzwassers ist (Prinzip des Archimedes). Daher ist $B$ das Produkt aus Volumen $V = 0.21\,\mathrm{m}^3$ des Fasses und der Dichte $1025\,\mathrm{kg/m}^3$ von Salzwasser, also ist $B = 215.25\,\mathrm{kg}$. Bezeichnet man mit $x(t)$ die Tiefe des Fasses zur Zeit $t$ ($x = 0$ sei die Meeresoberfläche), so lautet die Newtonsche Bewegungsgleichung daher

$$m\ddot{x} = g(m - B - cv) = g(m - B - c\dot{x}),$$

   wobei $g = 9.81\,\mathrm{m/sec}^2$. Ferner sind die Anfangsbedingungen

$$x(0) = 0, \qquad \dot{x}(0) = 0$$

   gegeben.

   Bei welcher Wassertiefe übersteigt die Geschwindigkeit $v$ die kritische Aufprallgeschwindigkeit von 12.2 m/sec?

4. Bei W. Walter (1993, S. 5)[17] findet man ein System von zwei Differentialgleichungen zweiter Ordnung, durch das die Bewegung eines Satelliten im Gravitationsfeld zweier Körper (z. B. Erde und Mond) modelliert wird, nämlich

$$\begin{aligned}
\ddot{x} &= x + 2\dot{y} - \mu'\frac{x+\mu}{[(x+\mu)^2+y^2]^{3/2}} - \mu\frac{x-\mu'}{[(x-\mu')^2+y^2]^{3/2}},\\
\ddot{y} &= y - 2\dot{x} - \mu'\frac{y}{[(x+\mu)^2+y^2]^{3/2}} - \mu\frac{y}{[(x-\mu')^2+y^2]^{3/2}}.
\end{aligned}$$

   Hierbei ist $\mu$ eine gegebene Konstante und $\mu' := 1-\mu$. Für $\mu := 0.01213$ und die Anfangsbedingungen

$$x(0) = 1.2, \quad \dot{x}(0) = 0, \quad y(0) = 0, \quad \dot{y}(0) = -1.04936$$

   plotte man die Bahn $\{(x(t), y(t)) : 0 \le t \le 10\}$.

[16] Siehe M. Braun (1975, S. 68 ff.).

[17] W. Walter (1993) *Gewöhnliche Differentialgleichungen. 5. Auflage*. Springer-Verlag, Berlin-Heidelberg-New York.

5. Mit $p > 0$, $\epsilon \in [0, 1)$ und $\alpha \in [0, 2\pi)$ sei die Ellipse $K_\alpha$ durch

$$K_\alpha := \left\{ \frac{p}{1 + \epsilon \cos(\phi - \alpha)} (\cos\phi, \sin\phi) : \phi \in [0, 2\pi] \right\}$$

gegeben. Man zeige:

(a) Dreht man die Ellipse $K_\alpha$ um den Winkel $\alpha$ im Uhrzeigersinn, so erhält man $K_0$.

(b) Die Ellipse $K_0$ (und damit auch $K_\alpha$) hat die Halbachsen

$$a = \frac{p}{1 - \epsilon^2}, \qquad b = \frac{p}{\sqrt{1 - \epsilon^2}}$$

und die Brennpunkte $(0, 0)$ und $(-2p\epsilon/(1 - \epsilon^2), 0)$.

(c) Die Ellipse $K_0$ lässt sich in der Form

$$K_0 = \left\{ (x, y) : \frac{(x - x_0)^2}{a^2} + \frac{y^2}{b^2} = 1 \right\}$$

darstellen, wobei

$$x_0 := -\frac{p\epsilon}{1 - \epsilon^2}.$$

6. Wir betrachten den Fall eines Körpers der Masse $m$, der unter dem Einfluss der Schwerkraft sich senkrecht nach unten bewegt, wobei der zur Geschwindigkeit proportionale Luftwiderstand berücksichtigt werde. Mit einer Konstanten $\rho > 0$ und (konstantem) $g = 9.81\,\mathrm{m/sec^2}$ hat man die Anfangswertaufgabe

$$m\ddot{x} = mg - \rho\dot{x}, \qquad x(0) = 0, \quad \dot{x}(0) = v_0$$

zu lösen. Man zeige, dass $\lim_{t\to\infty} \dot{x}(t)$ existiert, der Körper also eine endliche Endgeschwindigkeit erreicht[18].

## 1.3 Elementar lösbare Differentialgleichungen

Die wenigsten Differentialgleichungen können geschlossen gelöst oder, wie man auch sagt, elementar oder exakt integriert werden, wobei wir nicht versuchen wollen, genau zu definieren, was eine "geschlossene Lösung" ist. Das ist mit ein Grund, weshalb wir auch auf die numerische Behandlung von Differentialgleichungen eingehen werden und die exakt lösbaren Differentialgleichungen nur kurz streifen werden. Eine Standardsammlung elementar integrierbarer Differentialgleichungen findet man bei E. Kamke (1967). In diesem Abschnitt gehen wir oft ziemlich skrupellos vor. Es werden i. Allg. keine Voraussetzungen angegeben, unter denen gewisse Umformungen erlaubt sind. Im Einzelfall wird man sich nachträglich durch Einsetzen vergewissern müssen, dass ein Lösungskandidat wirklich eine Lösung ist.

Wir werden lediglich auf explizite Differentialgleichungen (erster Ordnung) und nicht auf implizite Differentialgleichungen eingehen. Erstere, also etwa die Gleichung $x' =$

[18] Bei H. Heuser (1989, S. 30) findet man hierzu die Bemerkung: Von dieser Tatsache profitiert der Fallschirmspringer immer dann, wenn sein Schirm überhaupt aufgeht.

$f(t,x)$, hat den Vorteil, dass sie eine einfache geometrische Interpretation in der $(t,x)$-Ebene erlaubt. Ist nämlich $x(\cdot)$ eine Lösung dieser Differentialgleichung mit $x(t_0) = x_0$, so ist $x'(t_0) = f(t_0, x(t_0)) = f(t_0, x_0)$, durch $f(t_0, x_0)$ ist also die Steigung der durch $(t_0, x_0)$ gehenden Lösungskurve gegeben. Ein Tripel $(t, x, f(t,x))$ (hierbei denke man sich $f(t,x)$ über $\tan\alpha = f(t,x)$ als Steigung interpretiert) heißt *Linienelelement*, die Gesamtheit der Linienelemente heißt das zu der Differentialgleichung gehörende *Richtungsfeld*. Das Richtungsfeld kann man sich dadurch graphisch veranschaulichen, dass man in den (zulässigen) Punkten $(t,x)$ ein kleines Geradenstück mit der Steigung $f(t,x)$ anträgt, wodurch man sich einen Überblick über den möglichen Verlauf von Lösungskurven verschaffen kann. In Abbildung 1.19 links haben wir das Richtungsfeld

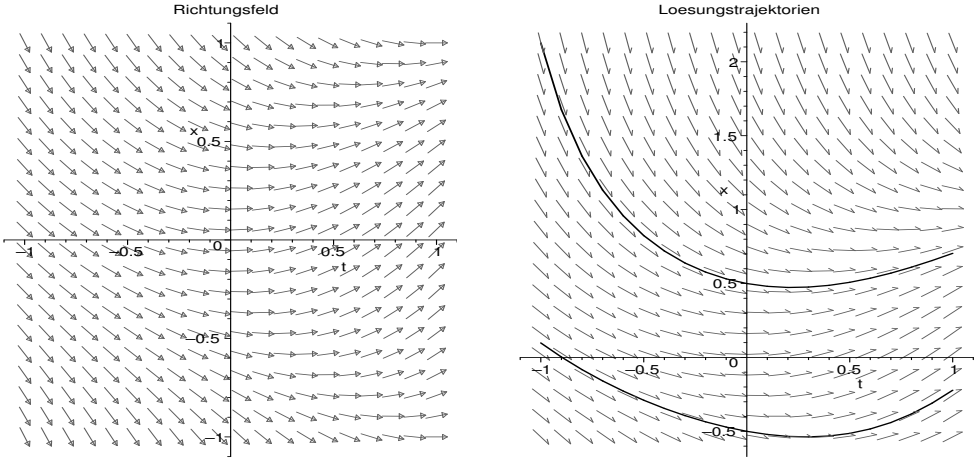


Abbildung 1.19: Richtungsfeld und Lösungstrajektorie zu $x' = t - x^2$

zu $x' = t - x^2$ eingetragen, rechts findet man neben dem Richtungsfeld die Lösungstrajektorien durch $(0, -0.5)$ und $(0, 0.5)$. Diese Abbildungen haben wir mittels

```
with(DEtools):
eqn:=diff(x(t),t)=t-x(t)^2;
dfieldplot(eqn,x(t),t=-1..1,x=-1..1,arrows=medium,
    title="Richtungsfeld",scene=[t,x]);
phaseportrait(eqn,x(t),t=-1..1,[[x(0)=-0.5],[x(0)=0.5]],
    title="Loesungstrajektorien",linecolor=black,scene=[t,x]);
```

erhalten. Zum Zeichnen des Richtungsfeldes sucht man zweckmäßig nicht die Steigungen einzelner Punkte auf, sondern zeichnet die Kurven gleicher Steigung, die *Isoklinen*[19]. Für $f(t,x) := t - x^2$ zeichnen wir die Isoklinen in Abbildung 1.20 (einmal mit `filled=true`, einmal ohne.

[19]Wörtlich

L. Collatz (1967) *Differentialgleichungen. Eine Einführung unter besonderer Berücksichtigung der Anwendungen*. B. G. Teubner, Stuttgart

entnommen.

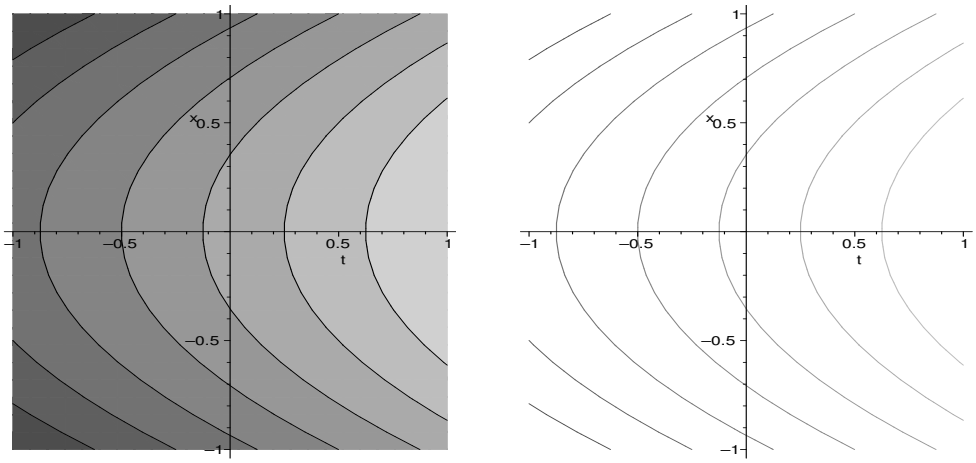


Abbildung 1.20: Isoklinen zu $x' = t - x^2$

### 1.3.1 Differentialgleichung mit getrennten Veränderlichen

Zu lösen sei die Differentialgleichung

$$x' = g(t)h(x),$$

die man aus naheliegenden Gründen Differentialgleichung mit *getrennten Veränderlichen* nennt. Sei $x(\cdot)$ eine Lösung und

$$H(x) := \int^x \frac{1}{h(\xi)}\, d\xi$$

eine Stammfunktion von $1/h$. Dann ist

$$\frac{d}{dt} H(x(t)) = \frac{x'(t)}{h(x(t))} = g(t)$$

und daher

$$H(x(t)) = \int^t g(\tau)\, d\tau + \text{const.}$$

Ist der Anfangswert $x(t_0) = x_0$ vorgeschrieben, so erhält man

$$H(x(t)) = \int_{t_0}^t g(\tau)\, d\tau + H(x_0),$$

woraus man (unter günstigen Umständen) $x(\cdot)$ berechnen kann.

**Beispiel:** Will man die allgemeine Lösung von $x' = t^2/x^2$ bestimmen, so ist also $g(t) = t^2$ und $h(x) = 1/x^2$. Obiges Vorgehen liefert

$$\frac{x(t)^3}{3} = \frac{t^3}{3} + \text{const} \qquad \text{bzw.} \qquad x(t) = (t^3 + c)^{1/3}$$

mit einer Konstanten $c$. □

**Beispiel:** Man löse die Anfangswertaufgabe

$$x' = 1 + x^2, \qquad x(0) = 0.$$

Wir erhalten $\arctan x(t) = t$ bzw. $x(t) = \tan t$. Wir beachten: Obwohl die Differentialgleichung relativ harmlos aussieht, existiert die Lösung der gestellten Anfangswertaufgabe nur auf dem Intervall $(-\pi/2, \pi/2)$. Man merke sich also: Lösungen von Anfangswertaufgaben existieren i. Allg. nur auf einem endlichen Intervall. □

**Beispiel:** Zu lösen sei die Anfangswertaufgabe

$$x' = \frac{\cos t}{1 + e^x}, \qquad x(\pi/2) = 3.$$

Aus

$$\int_3^{x(t)} (1 + e^\xi)\, d\xi = \int_{\pi/2}^{t} \cos\tau\, d\tau$$

erhält man

$$x(t) + e^{x(t)} = 2 + e^3 + \sin t.$$

Man erhält also keine explizite Darstellung der gesuchten Lösung $x(\cdot)$ als Funktion von $t$, sondern eine implizite, da man zur Berechnung von $x(t)$ noch die nichtlineare Gleichung

$$x + e^x = 2 + e^3 + \sin t$$

zu lösen hat. Diese hat offenbar eine eindeutige positive Lösung, die etwa mit Hilfe des Newton-Verfahrens berechnet werden kann. Mit Maple erhält man z. B. folgendes:

```
> dsolve(diff(x(t),t)=cos(t)/(1+exp(x(t))),x(t));
```

$$\mathrm{x}(t) = -\mathrm{LambertW}(e^{(\sin(t)+\_C1)}) + \sin(t) + \_C1$$

```
> dsolve(diff(x(t),t)=cos(t)/(1+exp(x(t))),x(t),implicit);
```

$$\sin(t) - \mathrm{x}(t) - e^{\mathrm{x}(t)} + \_C1 = 0$$

```
> dsolve({diff(x(t),t)=cos(t)/(1+exp(x(t))),x(Pi/2)=3},x(t));
```

$$\mathrm{x}(t) = -\mathrm{LambertW}(e^{(\sin(t)+e^3+2)}) + \sin(t) + e^3 + 2$$

```
> dsolve({diff(x(t),t)=cos(t)/(1+exp(x(t))),x(Pi/2)=3
> },x(t),implicit);

Error, (in dsolve/IC) The 'implicit' option is not available when
giving Initial Conditions.
```

Durch `?LambertW` kann man sich über die LambertW-Funktion informieren, wir verzichten hier auf eine entsprechende Information. □

Durch geeignete Substitutionen oder Transformationen kann man gelegentlich eine gegebene Differentialgleichung auf eine mit getrennten Veränderlichen zurückführen. Ist etwa die Differentialgleichung

$$x' = f(at + bx + c)$$

gegeben, wobei o. B. d. A. $b \neq 0$ angenommen werden kann, und ist $x(\cdot)$ eine Lösung, so setze man $y(t) := at + bx(t) + c$. Dann ist

$$y'(t) = a + bx'(t) = a + bf(at + bx(t) + c) = a + bf(y(t)),$$

also $y(\cdot)$ Lösung der Differentialgleichung $y' = a + bf(y)$. Ist umgekehrt $y(\cdot)$ Lösung von $y' = a + bf(y)$ und setzt man dann

$$x(t) := \frac{y(t) - at - c}{b},$$

so ist $x(\cdot)$ Lösung von $x' = f(at + bx + c)$.

**Beispiel:** Gegeben sei die Anfangswertaufgabe (siehe W. Walter (1993, S. 18))

$$x' = (t + x)^2, \qquad x(0) = 0.$$

Nach obiger Vorgehensweise hat man zunächst

$$y' = 1 + y^2, \qquad y(0) = 0$$

zu lösen, erhält $y(t) = \tan t$, und anschließend die gesuchte Lösung $x(t) = \tan t - t$. Dasselbe Ergebnis erhält man natürlich auch nach

```
dsolve({diff(x(t),t)=(t+x(t))^2,x(0)=0},x(t));
```

□

Ähnlich ist es, wenn die *homogene Differentialgleichung*

$$x' = f(x/t)$$

gegeben ist. Ist $x(\cdot)$ eine Lösung, so setze $y(t) := x(t)/t$ (es sei $t \neq 0$). Dann ist

$$y'(t) = \frac{x'(t)t - x(t)}{t^2} = \frac{f(y(t)) - y(t)}{t},$$

so dass $y$ der Differentialgleichung mit getrennten Veränderlichen

$$y' = \frac{f(y) - y}{t}$$

genügt. Ist umgekehrt $y(\cdot)$ eine Lösung dieser Differentialgleichung, so ist $x(t) := ty(t)$ Lösung der Ausgangsgleichung.

**Beispiel:** Zu lösen sei die Anfangswertaufgabe (siehe W. Walter (1993, S. 19))

$$x' = \frac{x}{t} - \frac{t^2}{x^2}, \qquad x(1) = 1.$$

Wie oben erklärt (jetzt ist $f(y) := y - 1/y^2$) hat man zunächst

$$y' = \frac{f(y) - y}{t} = -\frac{1}{ty^2}, \qquad y(1) = 1$$

zu lösen. Dies geschieht wie oben beschrieben. Man erhält

$$\frac{y(t)^3 - 1}{3} = -\ln t,$$

hieraus $y(t) = (1 - 3\ln t)^{1/3}$ und anschließend

$$x(t) = ty(t) = t(1 - 3\ln t)^{1/3},$$

was natürlich nur für $0 < t < e^{1/3}$ Sinn macht. Dasselbe Ergebnis erhält man durch

```
dsolve({diff(x(t),t)=x(t)/t-t^2/x(t)^2,x(1)=1},x(t));
```

□

### 1.3.2 Lineare Differentialgleichungen erster Ordnung

Die Gleichung

$$x' + g(t)x = h(t)$$

heißt *lineare Differentialgleichung erster Ordnung*. Sie heißt *homogen*, falls $h \equiv 0$, andernfalls *inhomogen*. Die Menge der Lösungen der inhomogenen Aufgabe ist nun bekanntlich, etwas lax formuliert, die Menge der Lösungen der homogenen Aufgabe plus einer speziellen Lösung der inhomogenen Aufgabe. Genauer:

- *Seien $X$ und $Y$ lineare Räume, $A\colon X \longrightarrow Y$ eine lineare Abbildung und $b \in Y$. Ist dann $x_0 \in X$ ein Element mit $Ax_0 = b$ (spezielle Lösung der inhomogenen Gleichung), so ist*

$$\{x \in X : Ax = b\} = \{z + x_0 : Az = 0\}.$$

Wir berechnen daher zunächst die allgemeine Lösung der homogenen Aufgabe

$$x' + g(t)x = 0 \qquad \text{bzw.} \qquad x' = -g(t)x.$$

Dies ist eine Differentialgleichung mit getrennten Veränderlichen. Man erhält

$$\ln x(t) = -\int_{t_0}^{t} g(\tau)\,d\tau + \ln x(t_0)$$

und hieraus

$$x(t) = x_0 \exp\Big(-\int_{t_0}^{t} g(\tau)\,d\tau\Big).$$

Ist etwa $g(\cdot)$ auf einem Intervall $I \subset \mathbb{R}$ stetig und $t_0 \in I$ fest, so ist

$$x(t; x_0) := x_0 \exp\Big(-\int_{t_0}^{t} g(\tau)\,d\tau\Big)$$

für jedes $x_0 \in \mathbb{R}$ eine Lösung der gegebenen homogenen Differentialgleichung erster Ordnung. Um sich zu überlegen, dass hierdurch *alle* Lösungen gegeben sind, nehmen wir an, $z(\cdot)$ sei eine weitere. Anschließend definieren wir

$$y(t) := \exp\Big(\int_{t_0}^{t} g(\tau)\,d\tau\Big)\, z(t).$$

Differenzieren liefert

$$y'(t) = \exp\Big(\int_{t_0}^{t} g(\tau)\,d\tau\Big)\, \underbrace{(g(t)z(t) + z'(t))}_{=0} = 0.$$

Daher ist $y(t) = y_0$ konstant, woraus wir die gewünschte Aussage erhalten.

Zur Berechnung einer speziellen Lösung der inhomogenen Differentialgleichung

$$x' + g(t)x = h(t)$$

machen wir einen Ansatz, den man *Variation der Konstanten* nennt. Und zwar machen wir den Ansatz

$$x(t) = x_0(t)\exp(-G(t)) \qquad \text{mit} \quad G(t) := \int_{t_0}^{t} g(\tau)\,d\tau.$$

Ist hier $x_0(t) = x_0$ konstant, so ist $x$ Lösung der homogenenen Gleichung. Durch "Variieren" versucht man zu einer Lösung der inhomogenen Gleichung zu kommen. Zur Bestimmung von $x_0(\cdot)$ berechnen wir

$$x'(t) + g(t)x(t) = [x_0'(t) - g(t)x_0(t) + g(t)x_0(t)]\,\exp(-G(t)) = x_0'(t)\exp(-G(t)).$$

Daher ist $x_0(\cdot)$ so zu bestimmen, dass $x_0' = \exp(G(t))\,h(t)$, was auf

$$x_0(t) = \int_{t_0}^{t} \exp(G(\tau))h(\tau)\,d\tau + x_0$$

mit einer Konstanten $x_0$ führt. Wählen wir hier speziell $x_0 = 0$, so erhalten wir, dass

$$x(t) := \exp(-G(t)) \int_{t_0}^{t} \exp(G(\tau))h(\tau)\,d\tau \qquad \text{mit} \quad G(t) := \int_{t_0}^{t} g(\tau)\,d\tau$$

eine spezielle Lösung der inhomogenen Gleichung ist. Daher ist

$$x(t; x_0) := \exp(-G(t))\,\Big[x_0 + \int_{t_0}^{t} \exp(G(\tau))h(\tau)\,d\tau\Big] \qquad \text{mit} \quad x_0 \in \mathbb{R}$$

die allgemeine Lösung der inhomogenen Gleichung. Sie existiert auf ganz $I$. Sind $g$ und $h$ auf *ganz* $\mathbb{R}$ stetig, so existiert sie also auf *ganz* $\mathbb{R}$. Ferner ist $x(\cdot; x_0)$ die *eindeutige* Lösung der Anfangswertaufgabe

$$x' + g(t)x = h(t), \qquad x(t_0) = x_0.$$

Denn ist $\bar{x}(\cdot)$ eine weitere Lösung, so ist die Differenz $w(t) := x(t; x_0) - \bar{x}(t)$ Lösung der homogenen Differentialgleichung. Nach obiger Überlegung existiert ein $w_0 \in \mathbb{R}$ mit $w(t) = w_0 \exp(-G(t))$. Da außerdem $w(t_0) = 0$ ist $w_0 = 0$, was die gewünschte Aussage nach sich zieht.

**Beispiel:** An einem Stromkreis mit dem (positiven) Ohmschen Widerstand $W$ und dem (positiven) Selbstinduktionskoeffizienten $L$ sei eine mit der Zeit $t$ veränderliche Spannung $E(t) := E_0 \sin \omega t$ mit $\omega > 0$ angelegt. Die Stromstärke $I$ befolgt das Gesetz

$$L \frac{dI}{dt} + W\, I = E(t),$$

d. h. $I$ ist Lösung der linearen Differentialgleichung erster Ordnung

$$\frac{dI}{dt} + \frac{W}{L}\, I = \frac{E_0}{L} \sin \omega t.$$

Nach obiger Überlegung ist

$$I(t) = e^{-Wt/l} \Big[ I_0 + \frac{E_0}{L} \int_0^t e^{W\tau/L} \sin \omega\tau \, d\tau \Big].$$

Es ist nicht schwer zu zeigen (siehe E. Kamke (1969, S. 34)): Ist $\gamma \in (0, \pi/2)$ so bestimmt, dass $\tan \gamma = \omega L/W$, so läßt sich die Lösung $I$ der Anfangswertaufgabe

$$\frac{dI}{dt} + \frac{W}{L}\, I = \frac{E_0}{L} \sin \omega t, \qquad I(0) = I_0$$

darstellen durch

$$I(t) = e^{-Wt/L} \Big( I_0 + \frac{\omega L E_0}{W^2 + \omega^2 L^2} \Big) + \frac{E_0}{\sqrt{W^2 + \omega^2 L^2}} \sin(\omega t - \gamma).$$

Die Stromstärke setzt sich hiernach zusammen aus einem durch das zweite Glied dargestellten rein periodischen Teil und einem wegen des Faktors $e^{-Wt/L}$ abklingenden aperiodischen Teils. □

**Beispiel:** Man löse die Anfangswertaufgabe

$$x' + \frac{4t}{1+t^2}\, x = \frac{t}{1+t^2}, \qquad x(2) = 1.$$

Hierzu berechnen wir zunächst

$$G(t) = \int_2^t \frac{4\tau}{1+\tau^2}\, d\tau = \ln(1+t^2)^2 - \ln 25$$

und anschließend

$$\begin{aligned} x(t) &= \frac{25}{(1+t^2)^2} \Big( 1 + \int_2^t \frac{1}{25} (1+\tau^2)^2 \, \frac{\tau}{1+\tau^2}\, d\tau \Big) \\ &= \frac{1}{(1+t^2)^2} \Big( 25 + \frac{t^2}{2} + \frac{t^4}{4} - 6 \Big) \\ &= \frac{1}{(1+t^2)^2} \Big( \frac{t^4}{4} + \frac{t^2}{2} + 19 \Big), \end{aligned}$$

ein Ergebnis, das man auch sofort durch

```
> infolevel[dsolve]:=3;
```

$$infolevel_{dsolve} := 3$$

```
> dsolve({diff(x(t),t)+4*t/(1+t^2)*x(t)=t/(1+t^2),x(2)=1},x(t));

Methods for first order ODEs:

Trying to isolate the derivative dx/dt...

Successful isolation of dx/dt

-> Trying classification methods

trying a quadrature

trying 1st order linear

1st order linear successful
```

$$\mathrm{x}(t) = \frac{\frac{1}{4}\,t^4 + \frac{1}{2}\,t^2 + 19}{(1+t^2)^2}$$

erhält. Hierbei haben wir durch `infolevel[dsolve]:=3;` dafür gesorgt, dass Maple uns etwas mehr darüber mitteilt, wie es zum Ergebnis kommt. □

### 1.3.3 Bernoullische Differentialgleichung

Ebenso wie man durch geeignete Transformationen eine gegebene Differentialgleichung unter (glücklichen) Umständen auf eine Differentialgleichung mit getrennten Veränderlichen zurückführen kann, so kann man gewisse nichtlineare Differentialgleichungen auf lineare zurückführen.

Die Differentialgleichung

$$x' + g(t)x + h(t)x^\alpha = 0$$

mit $\alpha \neq 1$ (besonders einfach ist der Fall $\alpha = 2$) heißt *Bernoullische Differentialgleichung*. Multipliziert man sie mit $(1-\alpha)x^{-\alpha}$, so erhält man

$$(x^{1-\alpha})' + (1-\alpha)g(t)x^{1-\alpha} + (1-\alpha)h(t) = 0,$$

so dass sich für $z = x^{1-\alpha}$ die lineare Differentialgleichung

$$z' + (1-\alpha)g(t)z + (1-\alpha)h(t) = 0$$

ergibt. Aus einer Lösung $z$ dieser Differentialgleichung erhält man durch

$$x(t) := z(t)^{1/(1-\alpha)}$$

eine Lösung der Bernoullischen Differentialgleichung, wobei zu beachten ist, dass bei der Transformation die Lösung $x = 0$ verloren gehen kann.

**Beispiel:** Zu bestimmen sei die allgemeine Lösung der Differentialgleichung

$$x' + \frac{x}{t} = t^2 x^2.$$

Für $z = 1/x$ erhält man also die lineare Differentialgleichung

$$z' - \frac{1}{t}\, z = -t^2.$$

Diese hat, wie man leicht nachrechnet, die allgemeine Lösung

$$z(t) = C\, t - \frac{t^3}{2},$$

so dass man als allgemeine Lösung der gegebenen Differentialgleichung

$$x(t) = \frac{1}{C\, t - t^3/2}$$

erhält. Dasselbe Ergebnis (in geringfügig anderer Form) findet auch Maple. □

### 1.3.4 Riccatische Differentialgleichung

Die Differentialgleichung

$$x' + g(t)x + h(t)x^2 = k(t)$$

heißt (allgemeine) *Riccatische Differentialgleichung*. Die Lösungen lassen sich i. Allg. nicht in geschlossener Form darstellen. Kennt man allerdings *eine* spezielle Lösung, so lassen sich *alle* angeben. Denn sei $\phi$ eine spezielle Lösung. Ist $x$ ebenfalls eine Lösung, so genügt die Differenz $z := x - \phi$ der Differentialgleichung

$$z' + g(t)z + h(t)\,[x^2 - \phi^2] = 0.$$

Wegen $x^2 - \phi^2 = (x - \phi)(x + \phi) = z(z + 2\phi)$ erhält man für $z$ die Differentialgleichung

$$z' + [g(t) + 2\phi(t)h(t)]z + h(t)z^2 = 0.$$

Dies ist eine Bernoullische Differentialgleichung, die sich mittels der Transformation $y = 1/z$ in die lineare Differentialgleichung

$$y' - [g(t) + 2\phi(t)h(t)]y - h(t) = 0$$

überführen lässt. Von dieser kann man (wenigstens im Prinzip) die allgemeine Lösung bestimmen und erhält daher insgesamt die allgemeine Lösung der Riccatischen Differentialgleichung.

**Beispiel:** Es sei die allgemeine Lösung von

$$x' - tx + x^2 = 1$$

zu bestimmen. Eine spezielle Lösung ist $\phi(t) := t$. Die allgemeine Lösung $x$ hat dann die Form $x = \phi + z$, wobei $z$ die allgemeine Lösung von

$$z' + tz + z^2 = 0$$

ist. Nach der Transformation $y = 1/z$ ist also die allgemeine Lösung $y$ der linearen Differentialgleichung

$$y' - ty - 1 = 0$$

zu bestimmen, was nach dem oben gesagten auf

$$y(t) = e^{t^2/2}\,[c + \int_0^t e^{-\tau^2/2}\,d\tau]$$

mit einer beliebigen Konstanten $c \in \mathbb{R}$ führt. Die allgemeine Lösung der gegebenen Riccattischen Differentialgleichung ist daher

$$x(t) = t + \frac{e^{-t^2/2}}{c + \int_0^t e^{-\tau^2/2}\,d\tau}.$$

Mit Maple erhält man (auch nach `simplify(%);`)

$$x(t) = \frac{t\sqrt{\pi}\sqrt{2}\mathrm{erf}(\frac{1}{2}\sqrt{2}t) + 2tC + 2e^{-t^2/2}}{\sqrt{\pi}\sqrt{2}\mathrm{erf}(\frac{1}{2}\sqrt{2}t) + 2C},$$

wobei

$$\mathrm{erf}(t) := \frac{2}{\sqrt{\pi}} \int_0^t e^{-\tau^2}\,d\tau$$

die Fehlerfunktion ist. Man stellt leicht fest, dass dies genau dasselbe Ergebnis ist. □

### 1.3.5 Aufgaben

1. Man löse[20], "per hand" und mit Maple, die Anfangswertaufgabe

   $$x' - x + e^{-2t}x^2 = 0, \qquad x(0) = 1.$$

   Wo ist die Lösung erklärt?

2. Man bestimme, "per hand" und mit Maple, die allgemeine Lösung der Differentialgleichungen

   (a) $x' = e^x \sin t$,

   (b) $x' = (t - x + 3)^2$,

   (c) $(1 + t^2)x' + tx = t\sqrt{1 + t^2}$.

[20] Diese Aufgabe wurde in der Staatsexamensklausur September 2001 gestellt.

3. Man gebe für die Differentialgleichung

$$x' = x^2 + 1 - t^2$$

das Richtungsfeld an und plotte Isoklinen. Ferner bestimme man sämtliche Lösungen (eine Lösung ist aus dem Richtungsfeld ersichtlich).

4. Man bestimme, "per hand" und mit Maple, die Lösung der Anfangswertaufgaben

(a) $x' = t^2\sqrt{1-x^2}$, $x(1) = 0$,

(b) $x' = t^2/(e^x + \cos x)$, $x(-1) = 0$,

(c) $x' = (2/t)x + t^4$, $x(1) = -6$,

(d) $x' = 2(x/t)^3 + x/t$, $x(1) = 2$.

5. Sei $F$ eine auf einem Gebiet $D$ der $(t,x)$-Ebene definierte reellwertige Funktion, die dort stetig partiell differenzierbar sei. Ist $x(\cdot)$ eine auf einem Intervall $I$ stetig differenzierbare Funktion mit $(t, x(t)) \in D$ und $F(t, x(t)) = \text{const}$ für alle $t \in I$, so genügt $x$ auf $I$ der Differentialgleichung $F_t(t,x) + F_x(t,x)x' = 0$. Umgekehrt nennen wir eine Differentialgleichung $g(t,x) + h(t,x)x' = 0$ *exakt*, wenn ein (hinreichend glattes) $F$ mit $F_t(t,x) = g(t,x)$, $F_x(t,x) = h(t,x)$ in $D$ existiert. Notwendig und hinreichend hierfür ist $g_x(t,x) = h_t(t,x)$, $F$ nennt man die zugehörige Stammfunktion.

Man löse die folgenden Anfangswertaufgaben für eine exakte Differentialgleichung. Auch die Möglichkeiten von Maple können getestet werden.

(a) $2tx + t^2x' = 0$, $x(2) = -3$,

(b) $(x^2 + \cos t) + 2txx' = 0$, $x(\pi) = -3$,

(c) $x + (t - \sin x)x' = 0$, $x(0) = \pi/2$.

6. Gegeben sei die Differentialgleichung $g(t,x) + h(t,x)x' = 0$, wobei $g$ und $h$ "hinreichend glatt" sind. Man nennt eine nicht verschwindende Funktion $k$ einen *integrierenden Faktor* für diese Differentialgleichung, wenn $k(t,x)g(t,x) + k(t,x)h(t,x)x'$ eine exakte Differentialgleichung ist. Es liegt nahe, als integrienden Faktor eine Funktion $k$ anzusetzen, die alleine von $t$ oder $x$ abhängt. Z. B. ist eine nicht verschwindende Funktion $k = k(t)$ genau dann ein integrierender Faktor für obige Differentialgleichung, wenn

$$\frac{g_x(t,x) - h_t(t,x)}{h(t,x)} = \frac{k'(t)}{k(t)} = \frac{d}{dt}\ln k(t),$$

insbesondere muss hier die linke Seite von $x$ unabhängig sein.

Durch die Bestimmung eines integrierenden Faktors löse man die folgenden Anfangswertaufgaben. Ferner plotte man die Lösungen auf einem geeigneten Intervall.

(a) $\cos x + (t \sin x)x' = 1$, $x(-1) = \pi/2$,

(b) $(2t^2 + 2tx^2 + 1)x + (3x^2 + t)x' = 0$, $x(0) = 1$.

7. Eine implizite Differentialgleichung der Form $x = tx' + g(x')$ nennt man eine *Clairautsche Differentialgleichung*. Man zeige:

(a) Ist $g(\cdot)$ an der Stelle $a$ definiert, so ist $x(t) := at + g(a)$ eine Lösung der Differentialgleichung.

(b) Ist $g$ stetig differenzierbar und $g'(t) \neq 0$ auf einem Intervall $I$, so ist eine Lösung in Parameterdarstellung durch

$$t = -\dot{g}(s), \qquad x = -s\dot{g}(s) + g(s)$$

gegeben.

Schließlich bestimme man Lösungen der Clairautschen Differentialgleichung $x = tx' + \exp(x')$ und plotte sie.

8. Man bestimme für die Clairautschen Differentialgleichungen

   (a) $x = tx' - \sqrt{x' - 1}$,

   (b) $x = tx' + x'^2$

   Lösungen in expliziter Form.

9. Zur Lösung von $x' = t^2 + x^2$, $x(0) = 1$, mache man einen Potenzreihenansatz $x(t) = \sum_{k=0}^{\infty} a_k t^k$. Man stelle eine Rekursionsformel für die Koeffizienten auf und zeige, dass $a_k \geq 1$, $k = 0, 1, \ldots$. Man berechne die ersten 15 Koeffizienten mit Maple.

10. Sei $x \in C^1(0, a)$ auf $(0, a)$ positiv und $\lim_{t \to a} x(t) = 0$. Für jedes $t \in (0, a)$ sei der Abstand des Punktes $P := (t, x(t))$ vom Schnittpunkt $T$ der Tangente an $x(t)$ in $P$ mit der $x$-Achse gleich dem Abstand von $T$ zum Nullpunkt. Man zeige[21], dass $x$ einer Differentialgleichung erster Ordnung genügt und löse diese.

## 1.4 Funktionalanalytische Grundlagen

Einige der klassischen Ergebnisse über gewöhnliche Differentialgleichungen, insbesondere Existenz- und Eindeutigkeitsaussagen bei Anfangs- oder auch Randwertaufgaben, lassen sich mit funktionalanalytischen Methoden und Hilfsmitteln besonders übersichtlich herleiten. Das Schwergewicht werden wir hier auf *Fixpunktsätze* legen. So wird z. B. die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t, x), \qquad x(t_0) = x_0,$$

wobei die Anfangszeit $t_0 \in \mathbb{R}$, der Anfangszustand $x_0 \in \mathbb{R}^n$ und die auf einer Umgebung $U$ von $(t_0, x_0)$ definierte und dort zumindestens stetige Abbildung $f\colon U \longrightarrow \mathbb{R}^n$ gegeben sind, in die äquivalente *Fixpunktaufgabe*

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds$$

umformuliert und auf diese Fixpunktaufgabe der Kontraktionssatz (dies führt auf den Satz von Picard-Lindelöf) bzw. der Schaudersche Fixpunktsatz (dies liefert den Existenzsatz von Peano) angewendet. Man beachte, dass man eine Anfangswertaufgabe für eine (explizite) Differentialgleichung $n$-ter Ordnung, etwa

$$x^{(n)} = f(t, x, x', \ldots, x^{(n-1)}), \qquad x(t_0) = x_0,\ x'(t_0) = x_0',\ \ldots, x^{(n-1)}(t_0) = x_0^{(n-1)},$$

[21] Diese Aufgabe wurde, mit geringfügig anderer Notation, in der Staatsexamensklausur September 2001 gestellt.

als eine Anfangswertaufgabe für ein System von $n$ Differentialgleichungen erster Ordnung schreiben kann:

$$\begin{array}{rcl@{\qquad}rcl} x_1' &=& x_2, & x_1(t_0) &=& x_0, \\ x_2' &=& x_2, & x_2(t_0) &=& x_0', \\ & \vdots & & & \vdots & \\ x_{n-1}' &=& x_n, & x_{n-1}(t_0) &=& x_0^{(n-2)} \\ x_n' &=& f(t, x_1, x_2, \ldots, x_n), & x_n(t_0) &=& x_0^{(n-1)}. \end{array}$$

### 1.4.1 Der Fixpunktsatz für kontrahierende Abbildungen

Der Fixpunktsatz für kontrahierende Abbildungen, gelegentlich auch *Kontraktionssatz* oder *Banachscher Fixpunktsatz* genannt, ist vielen schon aus der Analysis oder einer Vorlesung über Numerische Mathematik bekannt. Er wird unterschiedlich allgemein formuliert. Wir wählen hier einen Mittelweg. Hierbei nehmen wir an, dass die folgenden Begriffe aus der Analysis bekannt sind:

- Linearer[22] normierter Raum, Norm, Konvergenz und Cauchy-Folge in einem linearen normierten Raum , Banach-Raum,
- Abgeschlossene, kompakte, relativ kompakte[23] Teilmengen eines linearen normierten Raumes,
- Stetige Abbildungen zwischen linearen normierten Räumen.

**Beispiele:** Der $\mathbb{R}^n$, versehen mit einer beliebigen Vektornorm $\|\cdot\|$, ist bekanntlich ein Banach-Raum. Die wichtigsten Normen im $\mathbb{R}^n$ sind die euklidische Vektornorm, die Betragssummennorm und die Maximumnorm, welche für einen Vektor $x = (x_j) \in \mathbb{R}^n$ durch

$$\|x\|_2 := \Big(\sum_{j=1}^{n} x_j^2\Big)^{1/2}, \qquad \|x\|_1 := \sum_{j=1}^{n} |x_j|, \qquad \|x\|_\infty := \max_{j=1,\ldots,n} |x_j|$$

gegeben sind.

Mit $C_n[a,b]$ bezeichnen wir den linearen Raum der stetigen Abbildungen $x\colon [a,b] \longrightarrow \mathbb{R}^n$, wobei bei dieser Schreibweise $[a,b]$ stets als ein kompaktes Intervall in $\mathbb{R}$ verstanden wird und $C[a,b]$ statt $C_1[a,b]$ geschrieben wird. Elemente von $C_n[a,b]$ sind also vektorwertige, stetige Funktionen auf $[a,b]$. (Entsprechend ist $A \in C_{n\times n}[a,b]$ eine stetige Abbildung von $[a,b]$ nach $\mathbb{R}^{n\times n}$.) Definiert man auf $C_n[a,b]$ eine Norm durch

$$\|x\| := \max_{t\in[a,b]} \|x(t)\|,$$

wobei auf der rechten Seite $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist, so ist $C_n[a,b]$ versehen mit dieser Norm ein Banach-Raum (siehe Aufgabe 2). Man beachte, dass aus

[22] Der zugrunde gelegte Skalarkörper ist bei uns grundsätzlich der Körper $\mathbb{R}$ der reellen Zahlen.

[23] Hierauf gehen wir allerdings im Zusammenhang mit dem Schauderschen Fixpunktsatz noch kurz ein.

dem Zusammenhang immer hervorgeht, ob es sich bei $\|\cdot\|$ um eine Norm auf dem $\mathbb{R}^n$ oder auf $C_n[a,b]$ handelt. □

Ein häufig benutztes Hilfsmittel ist in dem folgenden Lemma angegeben, siehe auch W. Walter (1993, S. 98).

**Lemma 4.1** *Ist $x \in C_n[a,b]$ und $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$, so ist*

$$\Big\| \int_a^b x(t)\,dt \Big\| \le \int_a^b \|x(t)\|\,dt.$$

**Beweis:** Als Komposition zweier stetiger Abbildungen ist $\|x(\cdot)\| \in C[a,b]$. Zu vorgegebenem $\epsilon > 0$ existiert eine hinreichend feine Zerlegung

$$a = t_0 < t_1 < \cdots < t_{p-1} < t_p = b$$

mit

$$\Big\| \int_a^b x(t)\,dt - \sum_{i=1}^{p} x(t_i)(t_i - t_{i-1}) \Big\| \le \frac{\epsilon}{2}, \qquad \Big| \int_a^b \|x(t)\|\,dt - \sum_{i=1}^{p} \|x(t_i)\|(t_i - t_{i-1}) \Big| \le \frac{\epsilon}{2}.$$

Daher ist

$$\begin{aligned} \Big\| \int_a^b x(t)\,dt \Big\| &\le \Big\| \sum_{i=1}^{p} x(t_i)(t_i - t_{i-1}) \Big\| + \frac{\epsilon}{2} \\ &\le \sum_{i=1}^{p} \|x(t_i)\|(t_i - t_{i-1}) + \frac{\epsilon}{2} \\ &\le \int_a^b \|x(t)\|\,dt + \epsilon. \end{aligned}$$

Mit $\epsilon \to 0$ folgt die Behauptung. □

Jetzt folgt der bekannte Fixpunktsatz für kontrahierende Abbildungen bzw. der Kontraktionssatz (auch Banachscher Fixpunktsatz genannt), den wir nur der Vollständigkeit halber beweisen.

**Satz 4.2 (Banach)** *Sei $(X, \|\cdot\|)$ ein Banach-Raum, $K \subset X$ abgeschlossen und $F$ eine Abbildung mit $F(K) \subset K$, die also $K$ in sich abbildet, und die auf $K$ kontrahierend ist, zu der also eine Konstante $q \in (0,1)$ mit*

$$\|F(x) - F(y)\| \le q\,\|x - y\| \qquad \text{für alle } x, y \in K$$

*existiert. Dann besitzt $F$ genau einen Fixpunkt $x^*$ in $K$ und es gilt die Fehlerabschätzung*

$$\|x^* - F(x)\| \le \frac{q}{1-q}\,\|x - F(x)\| \qquad \text{für alle } x \in K.$$

*Genauer gilt: Für jedes $x_0 \in K$ und $x_{k+1} := F(x_k)$ konvergiert die Folge $\{x_k\}$ gegen den einzigen Fixpunkt $x^*$ von $F$ in $K$ und es gilt die a priori Fehlerabschätzung*

$$\|x_k - x^*\| \le \frac{q^k}{1-q}\|x_1 - x_0\|, \qquad k = 0, 1, \ldots$$

*sowie die a posteriori Fehlerabschätzung*

$$\|x_k - x^*\| \le \frac{q}{1-q}\|x_k - x_{k-1}\|, \qquad k = 1, 2, \ldots.$$

**Beweis:** Sei $x \in K$ beliebig. Man definiere eine Folge $\{x_k\} \subset K$ durch

$$x_0 := x, \qquad x_{k+1} := F(x_k) \qquad (k = 0, 1, \ldots).$$

Durch vollständige Induktion nach $k$ zeigt man, dass

$$\|x_{k+1} - x_k\| \le q^k \, \|x_1 - x_0\|, \qquad k = 0, 1, \ldots,$$

anschließend folgt mit Hilfe der Dreiecksungleichung

$$(*) \qquad \|x_{k+p} - x_k\| \le \frac{q^k}{1-q} \, \|x_1 - x_0\| \qquad (k = 0, 1, \ldots, \; p \in \mathbb{N}).$$

Daher ist $\{x_k\}$ eine Cauchy-Folge, also konvergent gegen ein $x^* \in X$, welches wegen der Abgeschlossenheit von $K$ sogar in $K$ liegt. Da $F$ insbesondere stetig ist und $x_{k+1} = F(x_k)$ gilt, ist $x^* = F(x^*)$, also $x^* \in K$ ein Fixpunkt von $F$. Da $F$ auf $K$ kontrahiert, ist $x^*$ einziger Fixpunkt von $F$ in $K$. Mit $p \to \infty$ folgt aus $(*)$, dass

$$\|x^* - x_k\| \le \frac{q^k}{1-q} \, \|x_1 - x_0\|, \qquad k = 0, 1, \ldots.$$

Mit $k = 1$ folgt hieraus die behauptete Fehlerabschätzung. □

### 1.4.2 Der Brouwersche und der Schaudersche Fixpunktsatz

Einer der schönsten und berühmtesten Sätze der Analysis ist wohl der folgende, auf L. E. J. Brouwer (1910) zurückgehende Satz

**Brouwerscher Fixpunktsatz:** *Sei $K \subset \mathbb{R}^n$ nichtleer, kompakt und konvex. Ist dann $F\colon K \longrightarrow \mathbb{R}^n$ stetig und $F(K) \subset K$, so besitzt $F$ mindestens einen Fixpunkt in $K$, d. h. es existiert mindestens ein $x \in K$ mit $x = F(x)$.*

Ein Beweis dieses Satzes würde die Vorlesung sprengen. Gewöhnlich wird er zunächst (z. B. in der algebraischen Topologie) für den Spezialfall der euklidischen Einheitskugel $B^n := \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$ bewiesen. Wie die obige Formulierung dann aus diesem Spezialfall folgt, kann verhältnismäßig einfach gezeigt werden. Der erste "elementare" analytische Beweis des Brouwerschen Fixpunktsatzes stammt von E. Heinz (1959), siehe auch J. M. Ortega, W. C. Rheinboldt (1970, S. 161)[24]. Ein außerordentlich eleganter analytischer Beweis des Brouwerschen Fixpuntsatzes stammt von J. Milnor (1978)[25], siehe auch J. Franklin (1980)[26]. Dort kann man auf S. 232 nachlesen:

[24] Ortega, J. M. and W. C. Rheiboldt (1970) *Iterative Solution of Nonlinear Equations in Several Variables.* Academic Press, New York-London.

[25] Milnor, J. (1978) "Analytic proofs of the "hairy ball theorem" and the Brouwer fixed point theorem". American Math. Monthly 85, 521–524.

[26] Franklin, J. (1980) *Methods of Mathematical Economics. Linear and Nonlinear Programming. Fixed Point Theorems.* Springer-Verlag, New York-Heidelberg-Berlin.

- This is what the Brouwer theorem says in everyday terms: Sit down with a cup of coffee. Gently and continuously swirl the coffee about in the cup. Put the cup down, and let the motion subside. When the coffee is still, Brouwer says there is at least one point in the coffee that has returned to the exact spot where it was when you first sat down.

Wir kommen nun zum *Schauderschen Fixpunktsatz*, den man als eine Verallgemeinerung des Brouwerschen Fixpunktsatzes auffassen kann.

**Satz 4.3 (Schauderscher Fixpunktsatz)** *Sei $(X, \|\cdot\|)$ ein linearer normierter Raum und $K \subset X$ nichtleer, abgeschlossen und konvex. Es sei $F\colon K \longrightarrow X$ stetig, $F(K) \subset K$ und $F(K)$ relativ kompakt. Dann besitzt $F$ mindestens einen Fixpunkt in $K$.*

Bevor wir den Schauderschen Fixpunktsatz mit Hilfe des Brouwerschen Fixpunktsatzes beweisen, müssen wir eventuell noch nicht bekannte Vokabeln in der obigen Formulierung erklären. Der Begriff der *Kompaktheit* ist einer der wichtigsten in der Analysis, der Funktionalanalysis und der Topologie. In linearen normierten Räumen ist der Begriff der "Überdeckungs-Kompaktheit" äquivalent dem der "Folgen-Kompaktheit", wie in der Funktionalanalysis bewiesen wird. Daher können wir für unsere Zwecke definieren:

- Sei $(X, \|\cdot\|)$ ein linearer normierter Raum. Eine Menge $K \subset X$ heißt *kompakt*, wenn es zu jeder Folge $\{x_k\} \subset K$ eine konvergente Teilfolge gibt, deren Limes in $K$ liegt. Dagegen heißt eine Menge $K \subset X$ *relativ kompakt*, wenn aus jeder Folge $\{x_k\} \subset K$ eine konvergente Teilfolge ausgewählt werden kann (deren Limes nicht notwendig in $K$ liegt).

Einige wenige Bemerkungen hierzu sind nützlich. Nach wie vor sei $(X, \|\cdot\|)$ ein linearer normierter Raum.

- Ist $K \subset X$ relativ kompakt, so ist $K$ *beschränkt*, d. h. es existiert eine Konstante $r > 0$ mit $\|x\| \leq r$ für alle $x \in K$.

Denn: Wäre $K$ nicht beschränkt, so existierte eine Folge $\{x_k\} \subset K$ mit $\|x_k\| \to \infty$. Dann wäre aus $\{x_k\}$ aber keine konvergente Teilfolge auswählbar. Im $\mathbb{R}^n$ gilt bekanntlich auch die Umkehrung, dass eine beschränkte Menge relativ kompakt ist, was in unendlichdimensionalen linearen normierten Räumen aber i. Allg. falsch ist.

- Ist $K \subset X$ kompakt, so ist $K$ abgeschlossen.

Denn: Ist $\{x_k\} \subset K$ eine Folge mit dem Limes $x \in X$, so ist notwendig $x \in K$ wegen der Kompaktheit von $K$. Folglich gilt die folgende Aussage, von der im $\mathbb{R}^n$ (und in jedem endlichdimensionalen linearen normierten Raum) auch die Umkehrung richtig ist:

- Ist $K \subset X$ kompakt, so ist $K$ beschränkt und abgeschlossen.

**Beweis von Satz 4.3:** Wir bezeichnen mit $B(y; \epsilon)$ die offene Kugel um $y \in X$ mit dem Radius $\epsilon > 0$. Zunächst zeigen wir:

- Ist $Y \subset X$ relativ kompakt, so existieren zu vorgegebenem $\epsilon > 0$ endlich viele $y_1, \ldots, y_N \in Y$ mit $Y \subset \bigcup_{i=1}^{N} B(y_i; \epsilon)$.

Denn: Angenommen, dies sei nicht wahr. Man wähle $z_1 \in Y$ beliebig, anschließend $z_2 \in Y$ mit $\|z_1 - z_2\| \geq \epsilon$. Dies ist möglich, da $Y \not\subset B(z_1, \epsilon)$. Angenommen, $z_1, \ldots, z_{k-1} \in Y$ seien schon so bestimmt, daß $\|z_i - z_j\| \geq \epsilon$ für $1 \leq i < j \leq k-1$. Da $Y \not\subset \bigcup_{i=1}^{k-1} B(z_i; \epsilon)$, existiert ein $z_k \in Y$ mit $\|z_k - z_j\| \geq \epsilon$ für $j = 1, \ldots, k-1$. Insgesamt gewinnt man eine Folge $\{z_k\} \subset Y$ mit $\|z_k - z_j\| \geq \epsilon$ für $j \neq k$. Aus $\{z_k\} \subset Y$ ist aber keine konvergente Teilfolge auswählbar, ein Widerspruch dazu, daß $Y$ relativ kompakt.

Nun sei $Y := F(K)$, ferner sei $\epsilon > 0$ fest. Wegen der gerade eben bewiesenen Aussage existieren $y_1, \ldots, y_N \in Y$ mit $Y \subset \bigcup_{i=1}^{N} B(y_i; \epsilon)$. Wir definieren die Menge

$$K_\epsilon := \Big\{ \sum_{i=1}^{N} \lambda_i y_i : \lambda_i \geq 0 \;\; (i = 1, \ldots, N),\; \sum_{i=1}^{N} \lambda_i = 1 \Big\},$$

die *konvexe Hülle* der Punkte $y_1, \ldots, y_N$, also die kleinste konvexe Teilmenge in $X$, welche $\{y_1, \ldots, y_N\}$ enthält. Wegen $Y := F(K) \subset K$ und der Konvexität von $K$ ist $K_\epsilon \subset K$.

- Es existiert eine stetige Abbildung $p_\epsilon \colon Y \longrightarrow K_\epsilon$ mit $\|p_\epsilon(y) - y\| < \epsilon$ für alle $y \in Y$.

Denn: Für $i = 1, \ldots, N$ definiere man $\phi_i \colon Y \longrightarrow \mathbb{R}$ durch

$$\phi_i(y) := \begin{cases} 0 & \text{falls} \quad \|y_i - y\| \geq \epsilon, \\ \epsilon - \|y_i - y\| & \text{sonst.} \end{cases}$$

Dann sind die $\phi_i$ stetig. Wegen $Y \subset \bigcup_{i=1}^{N} B(y_i; \epsilon)$ existiert zu jedem $y \in Y$ ein $i \in \{1, \ldots, N\}$ mit $\phi_i(y) > 0$. Jetzt definiere man $p_\epsilon \colon Y \longrightarrow K_\epsilon$ durch

$$p_\epsilon(y) := \sum_{i=1}^{N} \lambda_i(y) y_i \qquad \text{mit} \qquad \lambda_i(y) := \phi_i(y) \Big/ \sum_{i=1}^{N} \phi_i(y) \quad (i = 1, \ldots, N).$$

Offensichtlich ist $p_\epsilon \colon Y \longrightarrow K_\epsilon \subset X$ stetig. Für beliebiges $y \in Y$ ist ferner

$$\begin{aligned} \|p_\epsilon(y) - y\| &= \Big\| \sum_{i=1}^{N} \lambda_i(y)(y_i - y) \Big\| \\ &\leq \sum_{i : \|y_i - y\| < \epsilon} \lambda_i(y) \, \|y_i - y\| \\ &< \epsilon. \end{aligned}$$

- Es existiert ein $x^* \in K$ mit $x^* = F(x^*)$.

Denn: Man definiere $F_\epsilon := p_\epsilon \circ F$. Offensichtlich ist $F_\epsilon \colon K_\epsilon \longrightarrow K_\epsilon$ stetig. Der Brouwersche Fixpunktsatz (beachte: $K_\epsilon$ ist eine nichtleere, beschränkte, abgeschlossene und konvexe Menge in dem *endlich*dimensionalen linearen Raum $\operatorname{span}\{y_1, \ldots, y_N\}$) liefert die Existenz eines $x_\epsilon \in K_\epsilon$ mit $x_\epsilon = F_\epsilon(x_\epsilon)$. Nun sei $\{\epsilon_k\} \subset \mathbb{R}_+$ eine Nullfolge, man setze $x_k := x_{\epsilon_k}$, $y_k := F(x_k)$ (eine Verwechslung mit obigen $y_i$ ist nicht mehr möglich, diese haben ihre Schuldigkeit getan). Dann ist $\{x_k\} \subset K$ und $\{y_k\} \subset F(K) \subset K$. Nach

Voraussetzung ist $F(K)$ relativ kompakt, so daß aus $\{y_k\}$ eine gegen ein $x^*$ konvergente Teilfolge ausgewählt werden kann. O. B. d. A. konvergiere $\{y_k\}$ schon selber gegen $x^*$ (dadurch ersparen wir uns nur Subindizes). Da $K$ abgeschlossen und $\{y_k\} \subset K$, ist $x^* \in K$. Wir zeigen, daß auch die Folge $\{x_k\}$ gegen $x^*$ konvergiert. Denn es ist

$$\begin{aligned} \|x_k - x^*\| &= \|F_{\epsilon_k}(x_k) - x^*\| \\ &= \|p_{\epsilon_k}(y_k) - x^*\| \\ &\leq \|p_{\epsilon_k}(y_k) - y_k\| + \|y_k - x^*\| \\ &< \underbrace{\epsilon_k}_{\to 0} + \underbrace{\|y_k - x^*\|}_{\to 0} \\ &\to 0. \end{aligned}$$

Wegen der vorausgesetzten Stetigkeit von $F$ konvergiert die Folge $\{y_k\} = \{F(x_k)\}$ also einerseits gegen $x^*$, andererseits gegen $F(x^*)$, so daß notwendig $x^* = F(x^*)$. Damit ist der Schaudersche Fixpunktsatz bewiesen.

### 1.4.3 Der Satz von Arzela-Ascoli

Der Satz von Arzela-Ascoli dient dazu, die relative Kompaktheit einer Teilmenge des Banach-Raumes $C_n[a,b]$ (wie stets versehen mit einer Norm $\|x\| := \max_{t\in[a,b]} \|x(t)\|$) nachzuweisen.

**Satz 4.4 (Arzela-Ascoli)** *Die Menge $K \subset C_n[a,b]$ habe die folgenden beiden Eigenschaften:*

1. *Die Menge $K$ ist beschränkt, d. h. es existiert eine Konstante $C > 0$ mit $\|x\| \leq C$ für alle $x \in K$.*

2. *Die Menge $K$ ist gleichgradig stetig, d. h. zu jedem $\epsilon > 0$ gibt es ein $\delta = \delta(\epsilon) > 0$ mit*
$$t, s \in [a,b], \quad |t-s| \leq \delta, \quad x \in K \Longrightarrow \|x(t) - x(s)\| \leq \epsilon.$$

*Dann ist $K$ relativ kompakt in $C_n[a,b]$.*

**Beweis:** Sei $\{x_k\} \subset K$ eine Folge in $K$. Wir haben zu zeigen, dass aus $\{x_k\}$ eine (gleichmäßig) konvergente Teilfolge ausgewählt werden kann. Seien $r_1, r_2, \ldots$ die (abzählbar vielen) rationalen Zahlen in $[a,b]$. Da $\{x_k(r_1)\} \subset \mathbb{R}^n$ beschränkt ist (und beschränkte Teilmengen des $\mathbb{R}^n$ relativ kompakt sind), existiert eine unendliche Teilmenge $K_1 = \{k_{11}, k_{12}, \ldots\}$ derart, dass $\{x_k(r_1)\}_{k\in K_1}$ konvergent ist. Entsprechend kann aus $\{x_k(r_2)\}_{k\in K_1}$ eine konvergente Teilfolge ausgewählt werden, es existiert also $K_2 = \{k_{21}, k_{22}, \ldots\} \subset K_1$ derart, dass $\{x_k(r_2)\}_{k\in K_2}$ konvergiert. Im allgemeinen Schritt wähle man $K_j = \{k_{j1}, k_{j2}, \ldots\} \subset K_{j-1}$ derart, dass $\{x_k(r_j)\}_{k\in K_j}$ konvergiert. Nun definiere man $D := \{k_{11}, k_{22}, \ldots\}$ und zeige von der Folge $\{x_k\}_{k\in D}$, dass sie eine Cauchy-Folge und folglich konvergent ist. Wegen der gleichgradigen Stetigkeit von $\{x_k\}_{k\in D}$ existiert zu vorgegebenem $\epsilon > 0$ ein $\delta(\epsilon) > 0$ mit

$$t, s \in [a,b], \quad |t-s| \leq \delta, \quad k \in D \Longrightarrow \|x_k(t) - x_k(s)\| \leq \frac{\epsilon}{3}.$$

Wegen der Kompaktheit des Intervalls $[a,b]$ können wir eine *endliche* Menge $R \subset \{r_1, r_2, \ldots\}$ finden mit der Eigenschaft, dass es zu jedem $t \in [a,b]$ ein $r \in R$ mit $|t-r| \leq \delta(\epsilon)$ gibt. Für jedes $r \in R$ ist $\{x_k(r)\}_{k\in D}$ konvergent, folglich eine Cauchy-Folge, so dass ein $N_r(\epsilon)$ mit $\|x_k(r) - x_l(r)\| \leq \epsilon/3$ für alle $k,l \in D$ mit $k,l \geq N_r(\epsilon)$ existiert. Nun setze man $N(\epsilon) := \max_{r\in R} N_r(\epsilon)$ (beachte: $R$ ist endlich). Um zu zeigen, dass $\{x_k\}_{k\in D}$ eine Cauchy-Folge in $C_n[a,b]$ ist, gebe man sich $k,l \in D$ mit $k,l \geq N(\epsilon)$ und ein beliebiges $t \in [a,b]$ vor. Nach Konstruktion von $R$ existiert $r \in R$ mit $|t-r| \leq \delta(\epsilon)$. Folglich ist

$$\|x_k(t) - x_l(t)\| \leq \underbrace{\|x_k(t) - x_k(r)\|}_{\leq \epsilon/3} + \underbrace{\|x_k(r) - x_l(r)\|}_{\leq \epsilon/3} + \underbrace{\|x_l(r) - x_l(t)\|}_{\leq \epsilon/3} \leq \epsilon$$

und folglich

$$\|x_k - x_l\| \leq \epsilon \qquad \text{für alle } k,l \in D \text{ mit } k,l \geq N(\epsilon).$$

Folglich ist $\{x_k\}_{k\in D}$ eine Cauchy-Folge, der Satz von Arzela-Ascoli ist bewiesen. □

**Bemerkung:** In Aufgabe 17 kann bewiesen werden, dass im Satz von Arzela-Ascoli auch die Umkehrung richtig ist, dass also eine relativ kompakte Teilmenge von $C_n[a,b]$ (versehen mit der Maximumnorm) beschränkt und gleichgradig stetig ist. □

### 1.4.4 Aufgaben

1. Auf $C[a,b]$ ist durch

$$\|x\|_\infty := \max_{t\in[a,b]} |x(t)|, \qquad \|x\|_2 := \left(\int_a^b x(t)^2\,dt\right)^{1/2}$$

jeweils eine Norm gegeben. Man zeige, dass *keine* Konstante $C > 0$ mit $\|x\|_\infty \leq C\,\|x\|_2$ für alle $x \in C[a,b]$ existiert, d. h. in $C[a,b]$ sind nicht je zwei Normen äquivalent.

Hinweis: Man konstruiere eine Folge $\{x_k\} \subset C[a,b]$ mit $\|x_k\|_\infty = 1$ für alle $k$ und $\lim_{k\to\infty} \|x_k\|_2 = 0$.

2. Der lineare Raum $C_n[a,b]$ aller stetigen Abbildungen $x\colon [a,b] \longrightarrow \mathbb{R}^n$, versehen mit der Norm

$$\|x\| := \max_{t\in[a,b]} \|x(t)\|,$$

wobei auf der rechten Seite $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist, ist ein Banach-Raum.

3. Der lineare Raum $C_n^1[a,b]$ aller stetig differenzierbaren Abbildungen $x\colon [a,b] \longrightarrow \mathbb{R}^n$, versehen mit der Norm

$$\|x\| := \max(\max_{t\in[a,b]} \|x(t)\|, \max_{t\in[a,b]} \|x'(t)\|),$$

wobei auf der rechten Seite $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist, ist ein Banach-Raum.

4. Auf $C[0,1]$, dem linearen Raum der auf dem Intervall $[0,1]$ stetigen, reellwertigen Funktionen definiere man die reellwertige Abbildung $\|\cdot\|$ durch $\|x\| := \max_{t\in[0,1]} t^2|x(t)|$. Man zeige[27], dass $(C[0,1], \|\cdot\|)$ ein linearer normierter Raum, aber kein Banach-Raum ist.

   Hinweis: Man betrachte die Folge $\{x_k\} \subset C[0,1]$, die durch

   $$x_k(t) := \begin{cases} k, & t \in [0, 1/k], \\ 1/t, & t \in [1/k, 1] \end{cases}$$

   definiert ist.

5. Man beweise den Brouwerschen Fixpunktsatz im eindimensionalen Fall. Man zeige also: Sei $[a,b] \subset \mathbb{R}$ ein kompaktes Intervall, $F\colon [a,b] \longrightarrow \mathbb{R}$ eine stetige Funktion mit $F([a,b]) \subset [a,b]$. Dann existiert ein $x \in [a,b]$ mit $F(x) = x$.

6. Sei $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ lipschitzstetig und gleichmäßig monoton auf dem $\mathbb{R}^n$, d. h. es existieren positive Konstanten $L$ und $c$ mit

   $$\begin{aligned} \|f(x) - f(y)\| &\leq L\,\|x - y\| \\ (f(x) - f(y))^T(x - y) &\geq c\,\|x - y\|^2 \end{aligned} \qquad \text{für alle } x, y \in \mathbb{R}^n.$$

   Hierbei sei $\|\cdot\|$ die euklidische Norm im $\mathbb{R}^n$. Dann besitzt die Gleichung $f(x) = u$ für jedes $u \in \mathbb{R}^n$ genau eine Lösung.

7. Man beweise die folgende Variante zum Kontraktionssatz: Sei $X$ ein Banach-Raum (mit einer Norm $\|\cdot\|$) und $F\colon X \longrightarrow X$ eine Abbildung.. Es existiere ein $x_0 \in X$ und ein $r > 0$ derart, dass $F$ auf der Kugel

   $$B[x_0; r] := \{x \in X : \|x - x_0\| \leq r\}$$

   kontrahierend mit einer Lipschitzkonstanten $q < 1$ ist. Ferner sei $\|F(x_0) - x_0\| \leq (1-q)r$. Dann besitzt $F$ in $B[x_0; r]$ genau einen Fixpunkt.

8. Man zeige:

   (a) Die durch die Iterationsvorschrift $x_{k+1} := \exp(-x_k)$ gewonnene Folge $\{x_k\}$ konvergiert für jedes $x_0 \in \mathbb{R}$ gegen die eindeutige Lösung von $x = e^{-x}$.

   (b) Die durch die Iterationsvorschrift

   $$x_{k+1} := \frac{\exp(-x_k)(1 + x_k)}{1 + \exp(-x_k)}$$

   gewonnene Folge $\{x_k\}$ konvergiert für jedes $x_0 \in [0,1]$ gegen die eindeutige Lösung von $x = e^{-x}$.

   Ausgehend von $x_0 := 0.3$ berechne man mit Maple für beide Iterationsvorschriften $x_1, \ldots, x_{10}$ und vergleiche die Ergebnisse.

[27] Diese Aufgabe haben wir W. Walter (1993, S. 55) entnommen.

9. Wir[28] definieren in $C(I)$, $I := [0, a]$, drei Normen, die Maximumnorm

$$\|x\|_0 := \max_{t \in I} |x(t)|$$

sowie die Normen

$$\|x\|_1 := \max_{t \in I} |x(t)| e^{-at}, \qquad \|x\|_2 := \max_{t \in I} |x(t)| e^{-t^2}.$$

Man berechne für den durch

$$T(x)(t) := \int_0^t \tau x(\tau)\, d\tau$$

definierten Operator $T: C(I) \longrightarrow C(I)$ die entsprechenden Operatornormen $\|T\|_0$, $\|T\|_1$ und $\|T\|_2$. Hierbei ist die Operatornorm $\|T\|_j$, $j = 0, 1, 2$, gegeben durch

$$\|T\|_j := \sup_{x \in C(I) \setminus \{0\}} \frac{\|T(x)\|_j}{\|x\|_j}.$$

10. Man zeige[29], dass die Integralgleichung

$$x(t) = \frac{1}{2} t^2 + \int_0^t \tau x(\tau)\, d\tau, \qquad t \in I := [0, a]$$

genau eine Lösung besitzt und bestimme diese durch Zurückführung auf ein Anfangswertproblem bzw. durch explizite Berechnung der sukzessiven Approximationen unter Benutzung der Aufgabe 9, beginnend etwa mit $x_0 := 0$.

11. Man zeige: Ist $A \in \mathbb{R}^{n \times n}$ eine positive Matrix, also alle Einträge von $A$ positiv, so besitzt $A$ einen positiven Eigenwert $\lambda^*$ mit zugehörigem positiven Eigenvektor $x^*$, d. h. alle Komponenten von $x^*$ sind positiv. Ferner ist $|\lambda| \leq \lambda^*$ für alle Eigenwerte $\lambda$ von $A$, d. h. $\lambda^*$ ist der *Spektralradius* von $A$.

Hinweis: Für den ersten Teil setze man $K := \{x \in \mathbb{R}^n : x \geq 0,\ e^T x = 1\}$ (hierbei ist $e \in \mathbb{R}^n$ der Vektor, dessen Komponenten alle gleich 1 sind), definiere $F: K \longrightarrow \mathbb{R}$ durch $F(x) := Ax / e^T Ax$ und wende den Brouwerschen Fixpunktsatz an. Für den zweiten Teil kann man benutzen, dass jeder Eigenwert von $A$ auch Eigenwert von $A^T$ ist.

12. Man zeige: Sei $F: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ stetig. Es mögen Konstanten $\alpha \in (0, 1)$, $\beta > 0$ existieren derart, dass $\|F(x)\| \leq \alpha \|x\| + \beta$ für alle $x \in \mathbb{R}^n$. Hierbei sei $\|\cdot\|$ eine beliebige Norm im $\mathbb{R}^n$. Dann besitzt $F$ mindestens einen Fixpunkt.

13. Man zeige[30]: Sei $K := \{x \in \mathbb{R}^n : \|x\| \leq r\}$ mit $r > 0$ die abgeschlossene Kugel um den Nullpunkt mit dem Radius $r$, wobei $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist. Sei $F: K \longrightarrow \mathbb{R}^n$ stetig. Gilt dann die Implikation

$$\lambda > 1,\ \|x\| = r \Longrightarrow F(x) \neq \lambda x,$$

so besitzt $F$ einen Fixpunkt in $K$.

Hinweis: Man mache einen Widerspruchsbeweis und wende den Brouwerschen Fixpunktsatz auf die durch $G(x) := r[F(x) - x] / \|F(x) - x\|$ definierte Abbildung an.

[28] Diese Aufgabe haben wir W. Walter (1993, S. 56) entnommen.
[29] Diese Aufgabe haben wir W. Walter (1993, S. 56) entnommen.
[30] Siehe J. M. Ortega, W. C. Rheinboldt (1970, S. 163).

14. Man beweise die folgende direkte Verallgemeinerung des Brouwerschen Fixpunktsatzes.

    Sei $(X, \|\cdot\|)$ ein linearer normierter Raum und $K \subset X$ nichtleer, konvex und kompakt. Die stetige Abbildung $F\colon K \longrightarrow X$ bilde $K$ in sich ab, d. h. es sei $F(K) \subset K$. Dann besitzt $F$ mindestens einen Fixpunkt in $K$.

15. Durch das folgende Gegenbeispiel[31] zeige man, dass die Aussage in Aufgabe 14 falsch wird, wenn man "kompakt" durch "abgeschlossen und beschränkt" ersetzt. Sei $X := l^2$ der Hilbertsche Folgenraum aller Folgen $x := \{x_j\}$ reeller Zahlen mit $\sum_{j=1}^{\infty} x_j^2 < \infty$, versehen mit der Norm $\|x\| := (\sum_{j=1}^{\infty} x_j^2)^{1/2}$. Ferner sei $K := \{x \in l^2 : \|x\| \le 1\}$ die abgeschlossene Einheitskugel und $F\colon K \longrightarrow l^2$ definiert durch $y = F(x)$ mit

$$y_1 := (1 - \|x\|^2)^{1/2}, \quad y_j := x_{j-1} \quad (j = 2, 3, \ldots).$$

    Man zeige:

    (a) $(l^2, \|\cdot\|)$ ist ein Banach-Raum.

    (b) Die Menge $K \subset l^2$ ist nichtleer, abgeschlossen, beschränkt und konvex.

    (c) Die Abbildung $F$ bildet $K$ stetig in sich ab.

    (d) Die Abbildung $F$ besitzt keinen Fixpunkt in $K$.

16. Mit positiven Konstanten $c_0, c_1$ sei

$$K := \{x \in C^1[a,b] : \|x\|_\infty \le c_0,\ \|x'\|_\infty \le c_1\}.$$

    Man zeige, dass aus jeder Folge $\{x_k\} \subset K$ eine gleichmäßig konvergente Teilfolge ausgewählt werden kann.

    Hinweis: Man wende den Satz von Arzela-Ascoli an.

17. Man zeige die Umkehrung im Satz von Arzela-Ascoli, also:

    Sei $C_n[a,b]$ versehen mit der Maximumnorm $\|x\| := \max_{t\in[a,b]} \|x(t)\|$ für $x \in C_n[a,b]$, wobei $\|\cdot\|$ rechts eine beliebige Norm auf dem $\mathbb{R}^n$ ist. Eine relativ kompakte Menge $K \subset C_n[a,b]$ ist beschränkt und gleichgradig stetig.

    Hinweis: Zum Nachweis der gleichgradigen Stetigkeit von $K$ zeige man zunächst, dass es zu vorgegebenem $\epsilon > 0$ endlich viele $\{z_1, \ldots, z_p\} \subset K$ mit $\min_{i=1,\ldots,p} \|x - z_i\| \le \epsilon/3$ für alle $x \in K$ gibt. Mit Hilfe der gleichmäßigen Stetigkeit der $z_i$, $i = 1, \ldots, p$, schließe man auf die gleichgradige Stetigkeit von $K$.

[31] Siehe z. B. J. Franklin (1980, S. 275).

# Kapitel 2

# Die Theorie gewöhnlicher Anfangswertaufgaben

In diesem Kapitel betrachten wir die Anfangswertaufgabe für ein (explizites) System von $n$ Differentialgleichungen erster Ordnung, also die Aufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

mit der Abbildung $f\colon D \subset \mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$. Die für uns interessanten theoretischen Fragen, auf die wir in diesem Kapitel Antworten geben wollen, sind:

- Unter welchen Voraussetzungen besitzt die Anfangswertaufgabe (P) eine Lösung? Auf welches Intervall ist diese fortsetzbar?

- Unter welchen Voraussetzungen ist eine Lösung von (P) eindeutig?

- Wenn die rechte Seite $f$ stetig oder differenzierbar von gewissen Parametern abhängt, so würde man erwarten, dass das entsprechende auch für die Lösung gilt. Ist das richtig?

- Was für Stabilitätsaussagen können bewiesen werden? Hierbei wird man, grob gesagt, eine Lösung von (P) stabil nennen, wenn kleine Änderungen im Anfangszustand für alle $t \geq t_0$, also auch in der fernen Zukunft, nur kleine Änderungen in der Lösung nach sich ziehen.

- Sei $f(t,x) = A(t)x + b(t)$ mit $A \in C_{n\times n}(I)$, $b \in C(I;\mathbb{R}^n)$ (hierbei sei $I$ ein kompaktes Intervall, welches den Anfangszeitpunkt $t_0$ enthält), also (P) eine Anfangswertaufgabe für ein lineares System von $n$ Differentialgleichungen erster Ordnung. Kann die Lösung (wir werden sehen, dass diese in der Tat eindeutig existiert) geschlossen angegeben werden? Was kann ausgesagt werden, wenn $A(\cdot)$ konstant ist? Welche notwendigen und hinreichenden Stabilitätsbedingungen können in diesem Fall aufgestellt werden?

# 2.1 Existenz- und Eindeutigkeitsaussagen

## 2.1.1 Der Satz von Picard-Lindelöf

Wir beginnen mit dem Satz von Picard-Lindelöf, in dem im wesentlichen ausgesagt wird, dass unter der Voraussetzung der lokalen Lipschitzstetigkeit an die rechte Seite der gegebenen Anfangswertaufgabe die lokale eindeutige Lösbarkeit folgt.

**Satz 1.1 (Picard-Lindelöf)** *Sei $D \subset \mathbb{R}^{n+1}$ offen und $f\colon D \longrightarrow \mathbb{R}^n$ stetig und bezüglich der letzten $n$ Variablen lokal lischitzstetig, d. h. zu jedem $(t_0, x_0) \in D$ existiert eine Umgebung $U = U(t_0, x_0)$ und eine Konstante $L = L(t_0, x_0)$ derart, dass*

$$\|f(t,x) - f(t,y)\| \le L\,\|x - y\| \qquad \text{für alle } (t,x),\ (t,y) \in D \cap U.$$

*Dann existiert zu jedem $(t_0, x_0) \in D$ ein $\alpha^* > 0$ derart, dass die Anfangswertaufgabe*

$$x' = f(t,x), \qquad x(t_0) = x_0$$

*auf $I^* := \{t \in \mathbb{R} : |t - t_0| \le \alpha^*\}$ eindeutig lösbar ist.*

**Beweis:** Sei $(t_0, x_0) \in D$ gegeben. Man wähle zunächst positive Zahlen $\alpha, \beta$ derart, dass $I_\alpha \times B_\beta \subset D \cap U$, wobei

$$I_\alpha := \{t \in \mathbb{R} : |t - t_0| \le \alpha\}, \qquad B_\beta := \{x \in \mathbb{R}^n : \|x - x_0\| \le \beta\}.$$

Dann definiere man

$$M := \max_{(t,x) \in I_\alpha \times B_\beta} \|f(t,x)\|$$

und bestimme $\alpha^* \in (0, \alpha]$, $\beta^* \in (0, \beta]$ mit $M\alpha^* \le \beta^*$. Z. B. setze man $\beta^* := \beta$ und $\alpha^* := \min(\alpha, \beta/M)$. Anschließend definiere man $I^* := \{t \in \mathbb{R} : |t - t_0| \le \alpha^*\}$ und die nichtleere Menge

$$K := \Big\{x \in C_n(I^*) : \max_{t \in I^*} \|x(t) - x_0\| \le \beta^*\Big\}.$$

Dann bildet die Abbildung $F\colon C_n(I^*) \longrightarrow C_n(I^*)$, definiert durch

$$F(x)(t) := x_0 + \int_{t_0}^{t} f(s, x(s))\,ds$$

die Menge $K$ in sich ab. Denn ist $x \in K$ und $t \in I^*$, so ist

$$\|F(x)(t) - x_0\| = \Big\|\int_{t_0}^{t} f(s, x(s))\,ds\Big\| \le |t - t_0|\,M \le \alpha^* M \le \beta^*$$

und daher auch $F(x) \in K$. Definiert man auf $C_n(I^*)$ die gewichtete Maximumnorm

$$\|x\|_* := \max_{t \in I^*} e^{-2L|t - t_0|}\,\|x(t)\|,$$

wobei $L$ die Lipschitzkonstante von $f$ bezüglich des zweiten Argumentes auf $D \cap U$ ist, so ist

$$\|F(x) - F(y)\|_* \leq \frac{1}{2} \|x - y\|_* \quad \text{für alle } x, y \in K,$$

insbesondere bildet $F$ also die Menge $K$ bezüglich der Norm $\|\cdot\|_*$ kontrahierend in sich ab. Denn sind $x, y \in K$ und $t \in I^*$, so ist

$$\begin{aligned}
e^{-2L|t-t_0|}\|F(x)(t) - F(y)(t)\| &= e^{-2L|t-t_0|}\left\| \int_{t_0}^{t} [f(s, x(s)) - f(s, y(s))]\, ds \right\| \\
&\leq e^{-2L|t-t_0|}\text{sign}\,(t - t_0) \int_{t_0}^{t} \|f(s, x(s)) - f(s, y(s))\|\, ds \\
&\leq Le^{-2L|t-t_0|}\text{sign}\,(t - t_0) \int_{t_0}^{t} \|x(s) - y(s)\|\, ds \\
&\leq Le^{-2L|t-t_0|}\text{sign}\,(t - t_0) \int_{t_0}^{t} e^{2L|s-t_0|}\, ds\, \|x - y\|_* \\
&= Le^{-2L|t-t_0|} \frac{1}{2L}(e^{2L|t-t_0|} - 1)\, \|x - y\|_* \\
&\leq \frac{1}{2}\|x - y\|_*.
\end{aligned}$$

Der Kontraktionssatz (angewandt auf den mit der Norm $\|\cdot\|_*$ versehenen Banach-Raum $C_n(I^*)$ und die Abbildung $F$, welche die abgeschlossene Menge $K$ kontrahierend in sich abbildet) liefert, dass $F$ in $K$ genau einen Fixpunkt $x$ besitzt bzw. die Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, auf dem Intervall $I^*$ genau eine Lösung $x$ in $K$ besitzt. Um die Eindeutigkeit einer Lösung in $I^*$ zu beweisen, nehmen wir an, $y$ sei eine weitere Lösung der Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, auf dem Intervall $I^*$. Sei

$$\hat{\alpha} := \sup\{\alpha \in [0, \alpha^*] : \|y(t) - x_0\| \leq \beta^* \text{ für alle } t \in I_\alpha\}.$$

Ist $\hat{\alpha} = \alpha^*$, so ist auch $y \in K$ und folglich $x = y$. Wir nehmen daher an, es sei $\hat{\alpha} < \alpha^*$ und führen dies zum Widerspruch. Für $t \in I_{\hat{\alpha}}$ ist

$$\|y(t) - x_0\| = \|F(y)(t) - x_0\| \leq M\,|t - t_0| \leq M\hat{\alpha} < M\alpha^* \leq \beta^*,$$

was ein Widerspruch zur Definition von $\hat{\alpha}$ ist. Insgesamt ist der Satz von Picard-Lindelöf bewiesen.

□

**Bemerkung:** Entscheidendes Hilfmittel im Beweis des Satzes von Picard-Lindelöf ist der Fixpunktsatz für kontrahierende Abbildungen und damit ein konstruktiver Fixpunktsatz. Daher ist sogar gezeigt worden, dass mit $x_0(t) := x_0$ und

$$x_{k+1}(t) := x_0 + \int_{t_0}^{t} f(s, x_k(s))\, ds$$

die Folge $\{x_k\}$ auf einem Interval $I^* = [t_0 - \alpha^*, t_0 + \alpha^*]$ gleichmäßig gegen die auf diesem Intervall $I^*$ eindeutig existierende Lösung der gegebenen Anfangswertaufgabe konvergiert. □

**Bemerkung:** Die lokale Lipschitzstetigkeit von $f$ bezüglich der letzten $n$ Komponenten ist eine schwache Voraussetzung. Ist z. B. $D \subset \mathbb{R}^n$ konvex und sind in $D$ die Ableitungen $\partial f_i/\partial x_j$ stetig und beschränkt $(i, j = 1, \ldots, n)$, so genügt $f$ in $D$ einer Lipschitzbedingung. Denn wegen des Mittelwertsatzes existiert zu $(t, x)$, $(t, y) \in D$ und $i \in \{1, \ldots, n\}$ ein $z \in \mathbb{R}^n$ mit $(t, z) \in D$ (nämlich auf der Verbindungsstrecke zwischen $(t, x)$ und $(t, y)$) und

$$f_i(t, x) - f_i(t, y) = \sum_{j=1}^{n} \frac{\partial f_i(t, z)}{\partial x_j}(x_j - y_j),$$

woraus die Behauptung sofort folgt. □

Etliche Varianten zum obigen Satz von Picard-Lindelöf sind denkbar.

**Korollar 1.2** *Sei $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$. Mit positiven Zahlen $\alpha^*, \beta^*$ sei*

$$I^* := \{t \in \mathbb{R} : |t - t_0| \le \alpha^*\}, \qquad B^* := \{x \in \mathbb{R}^n : \|x - x_0\| \le \beta^*\}.$$

*Die Funktion $f\colon I^* \times B^* \longrightarrow \mathbb{R}^n$ sei stetig und bezüglich der zweiten Variablen lipschitzstetig auf $I^* \times B^*$. Sei $M := \max_{(t,x)\in I^*\times B^*} \|f(t, x)\|$ und $M\alpha^* \le \beta^*$. Dann besitzt die Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, genau eine Lösung $x$ auf $I^*$ mit $x(t) \in B^*$ für alle $t \in I^*$.*

Die Bedingung $M\alpha^* \le \beta^*$ dient jeweils dazu, zu sichern, dass die Abbildung $F$ die abgeschlossene Kugel $K \subset C_n(I^*)$ in sich abbildet. Ist $\beta^* = \infty$ im obigen Korollar, die rechte Seite $f$ der Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, also bezüglich des zweiten Argumentes global lipschitzstetig, so kann man auf diese Bedingung verzichten und erhält:

**Korollar 1.3** *Sei $I \subset \mathbb{R}$ ein kompaktes Intervall. Die Funktion $f\colon I \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sei stetig und bezüglich der zweiten Variablen lipschitzstetig auf $I \times \mathbb{R}^n$. Dann besitzt die Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, für jedes $(t_0, x_0) \in I \times \mathbb{R}^n$ genau eine Lösung auf dem Intervall $I$.*

**Beispiel:** Der wichtigste Spezialfall des letzten Korollars besteht sicherlich darin, dass $f(t, x) = A(t)x + b(t)$, wobei $A \in C_{n\times n}(I)$ und $b \in C_n(I)$, also $A\colon I \longrightarrow \mathbb{R}^{n\times n}$ und $b\colon I \longrightarrow \mathbb{R}^n$ stetig sind. Denn für beliebige $(t, x), (t, y) \in I \times \mathbb{R}^n$ ist dann

$$\|f(t, x) - f(t, y)\| = \|A(t)(x - y)\| \le \|A(t)\|\,\|x - y\|.$$

Hierbei bezeichnen wir die der gegebenen Vektornorm $\|\cdot\|$ zugeordnete Matrixnorm ebenfalls mit $\|\cdot\|$. Genauer:

- Ist $\|\cdot\|$ eine Norm auf $\mathbb{R}^n$, so wird auf $\mathbb{R}^{n\times n}$ durch

$$\|A\| := \max_{\|x\|=1} \|Ax\|, \qquad A \in \mathbb{R}^{n\times n},$$

die sogenannte *zugeordnete Matrixnorm* definiert.

Die Abbildung $t \mapsto \|A(t)\|$ ist stetig. Ist also $I$ ein kompaktes Intervall, so ist $f$ global lipschitzstetig mit der Lipschitzkonstanten $L := \max_{t\in I} \|A(t)\|$. □

**Beispiel:** Eine direkte Anwendung des Kontraktionssatzes auf spezielle Anfangswertaufgaben kann auch numerisch befriedigende Ergebnisse liefern. Dies wollen wir an einem Beispiel demonstrieren.

Gesucht sei die Lösung der Anfangswertaufgabe

$$x' = t^2 + x^2, \qquad x(0) = 0$$

auf dem Intervall $I := [-\frac{1}{2}, \frac{1}{2}]$. Als Näherung nehmen wir $x_0(t) := \frac{1}{3}t^3$. Mit

$$F(x)(t) := \int_0^t [s^2 + x(s)^2]\, ds$$

wird

$$x_1(t) := F(x_0)(t) = \frac{1}{3}t^3 + \frac{1}{63}t^7.$$

Wenn wir hier, was natürlich nicht unbedingt nötig ist, Maple anwenden wollen, so würden wir dies als Ergebnis von

```
int(s^2+(1/3*s^3)^2,s=0..t);
```

erhalten. Will (oder muss) man aber weitere Iterationen durchführen, so lernt man Maple sehr schnell schätzen. So erhält man

$$x_2(t) = \frac{t^3}{3} + \frac{t^7}{63} + \frac{2t^{11}}{2079} + \frac{t^{15}}{59535}$$

und (das wird keiner mehr zu Fuß ausrechnen wollen)

$$\begin{aligned} x_3(t) \;=\; & \frac{t^3}{3} + \frac{t^7}{63} + \frac{2t^{11}}{2079} + \frac{13t^{15}}{218295} + \frac{82t^{19}}{37328445} + \frac{662t^{23}}{10438212015} \\ & + \frac{4t^{27}}{3341878155} + \frac{t^{31}}{109876902975}. \end{aligned}$$

Übrigens gibt es in Maple bei der Lösung von Anfangswertaufgaben mit `dsolve` die Option `series`, was wir durch den folgenden Ausschnitt illustrieren:

```
> restart;
> Order:=10;dsolve({diff(x(t),t)=t^2+x(t)^2,x(0)=0},x(t),series);
```

$$Order := 10$$

$$\mathrm{x}(t) = \frac{1}{3}\,t^3 + \frac{1}{63}\,t^7 + \mathrm{O}(t^{10})$$

```
> x_1:=convert(rhs(%),polynom);
```

$$x\_1 := \frac{1}{3}\,t^3 + \frac{1}{63}\,t^7$$

Als Norm auf $C(I)$ wählen wir wie üblich die Maximumnorm, also $\|x\| := \max_{t\in I} |x(t)|$. Als Menge $K$ wählen wir $K := \{x \in C(I) : \|x - x_0\| \le \beta\}$ und bestimmen $\beta > 0$ so,

dass einerseits $F(K) \subset K$ und andererseits $F$ auf $K$ kontrahiert. Ersteres ist erfüllt, wenn $\frac{1}{2}(\frac{1}{24}+\beta)^2 \le \beta$. Andererseits ist für $x, y \in K$ und $t \in I$ offenbar

$$\begin{aligned} |F(x)(t) - F(y)(t)| &= \left| \int_0^t [x(s)+y(s)]\,[x(s)-y(s)]\,ds \right| \\ &\le \frac{1}{2} 2 \Big( \frac{1}{24} + \beta \Big) \, \|x-y\| \end{aligned}$$

und daher $\|F(x) - F(y)\| \le (\frac{1}{24}+\beta)\,\|x-y\|$. Daher kontrahiert die Abbildung $F$ auf $K$, wenn $\frac{1}{24}+\beta < 1$. Z. B. kann man $\beta = 0.001$ wählen und hat als Lipschitzkonstante $q := \frac{1}{24} + 0.001 \le 0.042667$. Wegen $\|x_1 - x_0\| = \frac{1}{63}(\frac{1}{2})^7 \le 0.000125$ erhält man aus dem Kontraktionssatz für die Lösung $x$ die Fehlerabschätzung

$$\|x - x_1\| \le \frac{q}{1-q}\,\|x_1 - x_0\| \le 6 \cdot 10^{-6},$$

was schon ganz befriedigend ist.

Der Versuch, die gegebene Anfangswertaufgabe geschlossen zu lösen, führt zu einem gewissen Erfolg:

```
> restart;
> dsolve({diff(x(t),t)=t^2+x(t)^2,x(0)=0},x(t));
```

$$\mathrm{x}(t) = -\frac{t\,(-\mathrm{BesselJ}(\frac{-3}{4},\,\frac{1}{2}\,t^2) + \mathrm{BesselY}(\frac{-3}{4},\,\frac{1}{2}\,t^2))}{-\mathrm{BesselJ}(\frac{1}{4},\,\frac{1}{2}\,t^2) + \mathrm{BesselY}(\frac{1}{4},\,\frac{1}{2}\,t^2)}$$

Hierbei sind BesselJ und BesselY Bessel-Funktionen erster bzw. zweiter Art. □

Ohne die lokale Lipschitzbedingung an die rechte Seite der gegebenen Differentialgleichung ist die Eindeutigkeit einer Lösung i. Allg. nicht gegeben, wie das folgende Beispiel zeigt.

**Beispiel:** Gegeben sei die Anfangswertaufgabe

$$x' = \sqrt{|x|}, \qquad x(0) = 0.$$

Natürlich ist $x := 0$ eine Lösung. Weitere Lösungen erhält man, wenn man mit beliebigem $a \le 0$ definiert:

$$x(t) := \frac{1}{4} \begin{cases} t^2, & \text{für} \quad t > 0, \\ 0, & \text{für} \quad a \le t \le 0, \\ -(t-a)^2, & \text{für} \quad t < a. \end{cases}$$

Dies liegt natürlich daran, dass für $f(t,x) := \sqrt{|t|}$ die lokale Lipschitzbedingung auf einer Umgebung von $(0,0)$ nicht erfüllt ist. □

Der Satz von Picard-Lindelöf liefert eine lokale Existenz- und Eindeutigkeitsaussage für die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0,$$

also für offenes $D \subset \mathbb{R}^{n+1}$, stetiges $f\colon D \longrightarrow \mathbb{R}^n$, welches bezüglich der letzten Variablen $x$ lokal lipschitzstetig ist, die Existenz einer Lösung $x$ von (P) auf einem Intervall $I = [t_0 - \alpha, t_0 + \alpha]$. Im folgenden Satz wollen wir uns überlegen, dass diese durch den Existenzsatz von Picard-Lindelöf gewonnene Lösung auf ein *maximales Intervall* $(t_{\min}, t_{\max})$ *fortgesetzt* werden kann. Hierbei sollte es sich von alleine verstehen, was wir unter einer *Fortsetzung* auf ein *maximales* Intervall verstehen[1].

**Satz 1.4** *Sei $D \subset \mathbb{R}^{n+1}$ offen, $f\colon D \longrightarrow \mathbb{R}^n$ stetig und bezüglich der letzten Variablen $x$ lokal lipschitzstetig. Dann besitzt die Anfangswertaufgabe*

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0,$$

*für jedes $(t_0, x_0) \in D$ eine eindeutig bestimmte Lösung $x$ auf einem maximalen Intervall $(t_{\min}, t_{\max})$. D. h. $x$ ist nicht auf ein größeres Intervall fortsetzbar und jede Lösung von (P) ist Restriktion von $x$. Ferner kommt $x$ nach links und rechts dem Rand von $D$ beliebig nahe. Dies soll heißen:*

- *Ist $t_{\max} < +\infty$, so ist*

$$\limsup_{t \to t_{\max}-} \|x(t)\| = \infty$$

*(Lösung "explodiert") oder*

$$\liminf_{t \to t_{\max}-} \operatorname{dist}((t,x(t)), \partial D) = 0$$

*(Lösung "kollabiert")*

*und entsprechend für den linken Endpunkt des maximalen Existenzintervalles:*

- *Ist $t_{\min} > -\infty$, so ist*

$$\limsup_{t \to t_{\min}+} \|x(t)\| = \infty$$

*oder*

$$\liminf_{t \to t_{\min}+} \operatorname{dist}((t,x(t)), \partial D) = 0.$$

Bevor wir den Satz beweisen, wollen wir seine Aussage durch einfache Beispiele illustrieren. Hierbei ist wegen des lokalen Existenz- und Eindeutigkeitssatzes klar, dass das maximale Intervall ein *offenes* Intervall ist.

**Beispiele:** Bei der linearen Anfangswertaufgabe $x' = tx$, $x(0) = 1$ ist die rechte Seite $f(t,x) = tx$ stetig auf $D := \mathbb{R}^2$ und dort lokal lipschitzstetig. Die Lösung $x(t) = e^{t^2/2}$ existiert auf dem maximalen Intervall $(-\infty, \infty)$.

Für das "Explodieren" einer Lösung geben wir zwei Beispiele. Die Anfangswertaufgabe $x' = -x^2$, $x(0) = 1$, besitzt die Lösung $x(t) = 1/(t+1)$ auf $(-1, \infty)$. Die Anfangswertaufgabe $x' = tx^2$, $x(0) = 1$, besitzt die Lösung $x(t) = 2/(2 - t^2)$ auf dem Intervall $(-\sqrt{2}, \sqrt{2})$. In beiden Fällen ist die rechte Seite auf $D := \mathbb{R}^2$ stetig und lokal lipschitzstetig.

[1] Wir halten uns hier an W. Walter (1993, S. 63 ff.).

Nun noch ein Beispiel zum "Kollabieren" einer Lösung. Gegeben sei die Anfangswertaufgabe $x' = -1/\sqrt{x}$, $x(0) = 1$. Hier ist die rechte Seite $f(t,x) := -1/\sqrt{x}$ auf $D := \mathbb{R}\times\mathbb{R}_+$ stetig und bezüglich $x$ lokal lipschitzstetig. Die Lösung $x(t) = (1-3t/2)^{2/3}$ existiert auf dem Intervall $(-\infty, 2/3)$. Hier ist $\lim_{t\to 2/3-}(t,x(t)) = (2/3,0) \in \partial D$. □

**Beweis von Satz 1.4:** Für die *Eindeutigkeit* beweisen wir die Aussage:

- Sind $x$ und $y$ zwei Lösungen des Anfangswertproblems (P) und ist $I$ ein gemeinsames Existenzintervall beider Lösungen mit $t_0 \in I$, so ist $x = y$ in $I$.

Denn: Dies ist eine direkte Folgerung aus der lokalen eindeutigen Lösbarkeit.

Nun zur *Existenz* einer Lösung auf einem maximalen Intervall. Zunächst existiert wegen des Satzes von Picard-Lindelöf eine lokale Lösung von (P). Diese lässt sich zu einer Lösung $x$ auf einem maximalen Intervall $(t_{\min}, t_{\max})$ fortsetzen. Wir nehmen an, es sei $t_{\max} < +\infty$. Angenommen, es ist $\limsup_{t\to t_{\max}-} \|x(t)\| < \infty$ und $\liminf_{t\to t_{\max}-} \operatorname{dist}((t,x(t)), \partial D) > 0$. Dann existiert eine kompakte Menge $A \subset D$ mit $(t,x(t)) \in A$ für $t \in [t_0, t_{\max})$. Hieraus schließen wir, dass sich $x$ auf das abgeschlossene Intervall $[t_0, t_{\max}]$ als Lösung fortsetzen lässt[2]. Wegen $(t_{\max}, x(t_{\max})) \in D$ kann man wegen Picard-Lindelöf eine lokale Lösung durch diesen Punkt erhalten und man hätte einen Widerspruch dazu, dass sich eine Lösung nicht über $t_{\max}$ hinaus fortsetzen lässt. Für den linken Endpunkt des maximalen Intervalls kann natürlich genau so argumentiert werden. Damit ist der Satz schließlich bewiesen. □

### 2.1.2 Der Satz von Peano

Grob sagt der Satz von Peano aus: Ist $D \subset \mathbb{R}^{n+1}$ offen, und $f\colon D \longrightarrow \mathbb{R}^n$ stetig, so "geht durch jeden Punkt $(t_0, x_0) \in D$ eine Lösung von $x' = f(t,x)$". Wichtig ist, dass es sich hier um einen Existenzsatz aber keine Eindeutigkeitsaussage handelt.

**Satz 1.5 (Peano)** *Sei $D \subset \mathbb{R}^{n+1}$ offen und $f\colon D \longrightarrow \mathbb{R}^n$ stetig. Dann existieren zu jedem $(t_0, x_0) \in D$ positive Zahlen $\alpha^*, \beta^*$ derart, dass mit*

$$I^* := \{t \in \mathbb{R} : |t - t_0| \le \alpha^*\}, \qquad B^* := \{x \in \mathbb{R}^n : \|x - x_0\| \le \beta^*\}$$

[2]Da $A \subset D$ kompakt, $f$ auf $D$ und daher auch auf $A$ stetig ist und $(t,x(t)) \in A$ für alle $t \in [t_0, t_{\max})$, existiert eine Konstante $C > 0$ mit $\|f(t,x(t))\| \le C$ für alle $t \in [t_0, t_{\max})$. Für $s, t \in [t_0, t_{\max})$ ist dann

$$\|x(t) - x(s)\| \le \left\| \int_s^t f(\tau, x(\tau))\, d\tau \right\| \le C\,|t - s|.$$

Hieraus folgt, dass $\lim_{t\to t_{\max}-} x(t)$ existiert. Denn ist $\{t_k\} \subset [t_0, t_{\max})$ eine Folge mit $\lim_{k\to\infty} t_k = t_{\max}$, so ist $\{x(t_k)\}$ eine Cauchy-Folge und folglich konvergent. Wiederum wegen der Lipschitzstetigkeit von $x(\cdot)$ auf $[t_0, t_{\max})$ ist der Limes von der Wahl der Folge $\{t_k\}$ unabhängig. Daher existiert $\lim_{t\to t_{\max}-} x(t)$ und wir setzen daher $x(t_{\max}) := \lim_{t\to t_{\max}-} x(t)$. Da $A$ abgeschlossen ist, ist $(t,x(t)) \in A$ für alle $t \in [t_0, t_{\max}]$. Da $x$ auf $[t_0, t_{\max})$ eine Lösung der Differentialgleichung $x' = f(t,x)$ ist, gilt

$$x(t) = x(t_0) + \int_{t_0}^t f(s, x(s))\, ds$$

für alle $t \in [t_0, t_{\max})$. Mit $t \to t_{\max}-$ folgt, dass diese Gleichung auch für $t = t_{\max}$, insgesamt also für alle $t \in [t_0, t_{\max}]$ besteht. Damit ist $x$ an der Stelle $t_{\max}$ (linksseitig) differenzierbar und $x'(t_{\max}-) = f(t_{\max}, x(t_{\max}))$.

*gilt: Es existiert mindestens ein $x \in C_n(I^*)$ mit*

*(a) Es ist $x(t) \in B^*$ für alle $t \in I^*$.*

*(b) Auf $I^*$ ist $x$ eine Lösung der Anfangswertaufgabe $x' = f(t,x)$, $x(t_0) = x_0$.*

**Beweis:** Sei $(t_0, x_0) \in D$. Man wähle positive Zahlen $\alpha, \beta$ derart, dass $I_\alpha \times B_\beta \subset D$ mit

$$I_\alpha := \{t \in \mathbb{R} : |t - t_0| \le \alpha\}, \qquad B_\beta := \{x \in \mathbb{R}^n : \|x - x_0\| \le \beta\},$$

was wegen der vorausgesetzten Offenheit von $D$ möglich ist. Anschließend definiere man

$$M := \max_{(t,x) \in I_\alpha \times B_\beta} \|f(t,x)\|,$$

bestimme $\alpha^* \in (0, \alpha]$ und $\beta^* \in (0, \beta]$ mit $M\alpha^* \le \beta^*$ und setze $I^*$ und $B^*$ wie oben angegeben. Nun wende man den Schauderschen Fixpunktsatz mit folgenden Daten an: Als linearen normierten Raum $(X, \|\cdot\|)$ nehme man $X := C_n(I^*)$ mit der Norm $\|x\| := \max_{t \in I^*} \|x(t)\|$ (wobei rechts $\|\cdot\|$ eine gegebene Norm auf dem $\mathbb{R}^n$ ist), die nichtleere, abgeschlossene und konvexe Menge $K$ sei durch

$$K := \{x \in C_n(I^*) : x(t) \in B^* \text{ für alle } t \in I^*\}$$

gegeben und schließlich sei die Abbildung $F\colon C_n(I^*) \longrightarrow C_n(I^*)$ definiert durch

$$F(x)(t) := x_0 + \int_{t_0}^{t} f(s, x(s))\, ds.$$

Die Voraussetzungen des Schauderschen Fixpunktsatzes sind leicht nachgeprüft:

1. Die Abbildung $F\colon K \longrightarrow C_n(I^*)$ ist stetig.

   Denn: Sei $\epsilon > 0$ beliebig vorgegeben. Die Menge $I^* \times B^*$ ist kompakt und daher $f$ auf $I^* \times B^*$ gleichmäßig stetig. Folglich existiert ein $\delta = \delta(\epsilon) > 0$ mit

   $$(t,x),\ (t,y) \in I^* \times B^*, \quad \|x - y\| \le \delta \Longrightarrow \|f(t,x) - f(t,y)\| \le \frac{\epsilon}{\alpha^*}.$$

   Sind nun $x, y \in K$ mit $\|x - y\| \le \delta$ gegeben, so ist $(s, x(s)),\ (s, y(s)) \in I^* \times B^*$ für alle $s \in I^*$ und daher

   $$\begin{aligned} \|F(x)(t) - F(y)(t)\| &= \left\| \int_{t_0}^{t} [f(s,x(s)) - f(s,y(s))]\, ds \right\| \\ &\le \operatorname{sign}(t - t_0) \int_{t_0}^{t} \|f(s,x(s)) - f(s,y(s))\|\, ds \\ &\le |t - t_0| \frac{\epsilon}{\alpha^*} \\ &\le \epsilon. \end{aligned}$$

   Also gilt die Implikation

   $$x, y \in K, \quad \|x - y\| \le \delta \Longrightarrow \|F(x) - F(y)\| \le \epsilon,$$

   womit die Stetigkeit von $F$ auf $K$ bewiesen ist.

2. Es ist $F(K) \subset K$.

   Sind $x \in K$ und $t \in I^*$ beliebig, so ist

   $$\|F(x)(t) - x_0\| \leq M\,|t - t_0| \leq M\alpha^* \leq \beta^*$$

   und damit $F(x) \in K$.

3. $K$ ist nichtleer, abgeschlossen und konvex, $F(K)$ ist relativ kompakt.

   Die erste Aussage ist trivial. Die relative Kompaktheit von $F(K)$ folgt aus dem Satz von Arzela-Ascoli, indem man zeigt, dass $F(K)$ beschränkt und gleichgradig stetig ist.

   (a) $F(K)$ ist beschränkt.
   Denn: Sind $x \in K$ und $t \in I^*$ beliebig, so ist

   $$\|F(x)(t)\| \leq \|x_0\| + \operatorname{sign}(t - t_0) \int_{t_0}^{t} \|f(s, x(s))\|\, ds \leq \|x_0\| + M\alpha^*,$$

   also $\|F(x)\| \leq C := \|x_0\| + M\alpha^*$ für alle $x \in K$, womit die Beschränktheit von $F(K)$ nachgewiesen ist.

   (b) $F(K)$ ist gleichgradig stetig.
   Denn: Sei $x \in K$ beliebig vorgegeben. Für beliebige $s, t \in I^*$ ist

   $$\|F(x)(s) - F(x)(t)\| = \left\| \int_t^s f(\tau, x(\tau))\, d\tau \right\| \leq M\,|s - t|,$$

   woraus man die gleichgradige Stetigkeit von $F(K)$ abliest. Denn zu vorgegebenem $\epsilon > 0$ setze man $\delta := \epsilon/M$ und erhält die Implikation

   $$s, t \in I^*, \quad |s - t| \leq \delta, \quad x \in K \Longrightarrow \|F(x)(s) - F(x)(t)\| \leq \epsilon,$$

   was die gleichgradige Stetigkeit von $F(K)$ nach sich zieht.

Der Schaudersche Fixpunktsatz liefert die Behauptung, nämlich die Existenz eines $x \in K$ mit $F(x) = x$. Offensichtlich ist $x \in C_n^1(I^*)$ auf $I^*$ eine Lösung der gegebenen Anfangswertaufgabe. □

Wieder sind etliche Varianten möglich. Im folgenden Korollar beschränken wir uns darauf, eine Existenzaussage "in die Zukunft zu machen".

**Korollar 1.6** *Gegeben sei die Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$. Mit vorgegebenen positiven $\alpha, \beta$ sei $f\colon D \longrightarrow \mathbb{R}^n$ stetig, wobei*

$$D := \{(t, x) \in \mathbb{R}^{n+1} : t_0 \leq t \leq t_0 + \alpha,\ \|x - x_0\| \leq \beta\}.$$

*Mit*

$$M := \max_{(t,x)\in D} \|f(t, x)\|, \qquad \alpha^* := \min(\alpha, \beta/M)$$

*besitzt dann die gegebene Anfangswertaufgabe eine Lösung auf $[t_0, t_0 + \alpha^*]$.*

**Beispiel:** Die Anfangswertaufgabe $x' = -x^2$, $x(1) = 1$ hat die Lösung $x(t) = 1/t$. Das maximale Existenzintervall der Lösung ist also $(0, \infty)$. Zum Vergleich rechnen wir uns das Existenzintervall aus, das uns der Satz von Peano bzw. dessen Beweis liefert. Hierbei gibt man sich positive $\alpha, \beta$ beliebig vor, berechnet $M := \max_{|x-1|\le\beta} x^2 = (1+\beta)^2$ und anschließend

$$\alpha^* := \min\Big(\alpha, \frac{\beta}{M}\Big) = \min\Big(\alpha, \frac{\beta}{(1+\beta)^2}\Big),$$

womit die Existenz einer Lösung auf $[1-\alpha^*, 1+\alpha^*]$ gesichert ist. Wegen $\beta/(1+\beta)^2 \le \frac{1}{4}$ ist die Existenz einer Lösung daher lediglich auf dem Intervall $[\frac{3}{4}, \frac{5}{4}]$ gesichert. □

**Bemerkung:** Die Voraussetzungen des Satzes von Peano seien erfüllt. Dann kann jede Lösung der gegebenen Anfangswertaufgabe $x' = f(t,x)$, $x(t_0) = x_0$, (wobei $D \subset \mathbb{R}^{n+1}$ offen, $f\colon D \longrightarrow \mathbb{R}^n$ stetig und $(t_0, x_0) \in D$) nach rechts und links bis zum Rande von $D$ fortgesetzt werden bzw. dem Rand von $D$ beliebig nahe kommen. In Satz 1.4 ist dies näher erklärt. Der Beweis ist ohne lokale Eindeutigkeit schwieriger, wir verweisen nur auf W. Walter (1993, S. 67 ff.). □

### 2.1.3 Das Lemma von Gronwall

Eine wichtige Ungleichung zur Untersuchung gewöhnlicher Differentialgleichungen ist die *Gronwallsche Ungleichung*. Es gibt hierzu etliche Varianten, wir beginnen mit einer verhältnismäßig allgemeinen Aussage und spezialisieren diese später.

**Lemma 1.7 (Gronwall)** *Sei $I := [t_0, T]$ ein Intervall, ferner seien $\phi, \alpha, \beta \in C(I)$ Funktionen mit*

$$\beta(t) \ge 0, \qquad \phi(t) \le \alpha(t) + \int_{t_0}^{t} \beta(s)\phi(s)\,ds \qquad \text{für alle } t \in I.$$

*Dann ist*

$$\phi(t) \le \alpha(t) + \int_{t_0}^{t} \alpha(s)\beta(s) \exp\Big(\int_s^t \beta(\tau)\,d\tau\Big)\,ds \qquad \text{für alle } t \in I.$$

**Beweis:** Man definiere $\psi \in C^1(I)$ durch

$$\psi(t) := \int_{t_0}^{t} \beta(s)\phi(s)\,ds.$$

Dann ist $\phi(t) \le \alpha(t) + \psi(t)$ und wegen $\beta(t) \ge 0$ offenbar

$$\psi'(t) = \beta(t)\phi(t) \le \beta(t)\,[\alpha(t) + \psi(t)]$$

bzw.

$$\psi'(t) - \beta(t)\psi(t) \le \alpha(t)\beta(t).$$

Mit einer nichtnegativen Funktion $r \in C(I)$ (nämlich $r := \alpha\beta - (\psi' - \beta\psi)$) ist also $\psi$ Lösung der linearen Anfangswertaufgabe erster Ordnung

$$\psi' - \beta(t)\psi = \alpha(t)\beta(t) - r(t), \qquad \psi(t_0) = 0.$$

Daher ist

$$\begin{aligned}\psi(t) &= \exp\Big(\int_{t_0}^t \beta(\tau)\,d\tau\Big)\int_{t_0}^t \exp\Big(-\int_{t_0}^s \beta(\tau)\,d\tau\Big)[\alpha(s)\beta(s)-r(s)]\,ds\\ &\le \exp\Big(\int_{t_0}^t \beta(\tau)\,d\tau\Big)\int_{t_0}^t \exp\Big(-\int_{t_0}^s \beta(\tau)\,d\tau\Big)\alpha(s)\beta(s)\,ds\\ &= \int_{t_0}^t \alpha(s)\beta(s)\exp\Big(\int_s^t \beta(\tau)\,d\tau\Big)\,ds,\end{aligned}$$

was wegen $\phi(t) \le \alpha(t) + \psi(t)$ zu zeigen war. □

Ist speziell $\alpha(t) \equiv \alpha$ konstant, so erhalten wir die sogenannte *spezielle Gronwallsche Ungleichung*.

**Korollar 1.8** *Sei $I := [t_0, T]$ ein Intervall, ferner seien $\alpha \in \mathbb{R}$, $\beta, \phi \in C(I)$ mit*

$$\beta(t) \ge 0, \qquad \phi(t) \le \alpha + \int_{t_0}^t \beta(s)\phi(s)\,ds \qquad \text{für alle } t \in I.$$

*Dann ist*

$$\phi(t) \le \alpha\,\exp\Big(\int_{t_0}^t \beta(s)\,ds\Big) \qquad \text{für alle } t \in I.$$

**Beweis:** Eine direkte Anwendung von Lemma 1.7 liefert

$$\phi(t) \le \alpha\underbrace{\Big[1 + \int_{t_0}^t \beta(s)\exp\Big(\int_s^t \beta(\tau)\,d\tau\Big)\,ds\Big]}_{=\gamma(t)}.$$

Wie angegeben definiere man $\gamma \in C^1(I)$ durch

$$\begin{aligned}\gamma(t) &:= 1 + \int_{t_0}^t \beta(s)\exp\Big(\int_s^t \beta(\tau)\,d\tau\Big)\,ds\\ &= 1 + \exp\Big(\int_{t_0}^t \beta(\tau)\,d\tau\Big)\int_{t_0}^t \exp\Big(-\int_{t_0}^s \beta(\tau)\,d\tau\Big)\beta(s)\,ds.\end{aligned}$$

Dann ist $\gamma(t_0) = 1$ und

$$\gamma'(t) = \beta(t)[\gamma(t) - 1] + \beta(t) = \beta(t)\gamma(t)$$

und folglich $\gamma(t) = \exp(\int_{t_0}^t \beta(s)\,ds)$. Damit ist die spezielle Gronwallsche Ungleichung bewiesen. □

**Bemerkung:** Mit Hilfe der speziellen Gronwallschen Ungleichung kann eine Aussage über die stetige Abhängigkeit der Lösung einer Anfangsaufgabe vom Anfangswert bewiesen werden. Das wollen wir jetzt skizzieren.

Sei $I = [t_0, T]$ ein Intervall und $x, y \in C^1(I; \mathbb{R}^n)$ zwei Funktionen mit

$$x'(t) = f(t, x(t)), \qquad x(t_0) = x_0$$

und

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

für alle $t \in I$. Dann gilt für alle $t \in I$, wenn wir noch voraussetzen, dass $f$ bezüglich des zweiten Arguments global lipschitzstetig mit der Lipschitzkonstanten $L$ ist:

$$\begin{aligned} \|x(t) - y(t)\| &= \left\| x_0 + \int_{t_0}^{t} x'(s)\,ds - y_0 - \int_{t_0}^{t} y'(s) \right\| \\ &\le \|x_0 - y_0\| + \int_{t_0}^{t} \|f(s, x(s)) - f(s, y(s))\|\,ds \\ &\le \|x_0 - y_0\| + \int_{t_0}^{t} L\,\|x(s) - y(s)\|\,ds. \end{aligned}$$

Aus der speziellen Gronwallschen Ungleichung erhält man

$$\|x(t) - y(t)\| \le \|x_0 - y_0\|\, e^{L(t-t_0)} \qquad \text{für alle } t \in I.$$

Hieraus folgt nicht nur eine nicht überraschende Eindeutigkeitsaussage (wenn $x_0 = y_0$), sondern auch eine Aussage über die (lipschitz-) stetige Abhängigkeit der Lösung vom Anfangswert. In Aufgabe 8 wird diese Aussage noch etwas verallgemeinert. □

### 2.1.4 Aufgaben

1. Die Anfangswertaufgabe

   $$x'' = f(t, x), \qquad x(t_0) = x_0,\ x'(t_0) = x_0'$$

   für eine Differentialgleichung zweiter Ordnung kann natürlich als eine Anfangswertaufgabe für zwei Differentialgleichungen erster Ordnung geschieben werden. Man zeige, dass sie auch äquivalent ist zu der Integralgleichung

   $$x(t) = x_0 + x_0'(t - t_0) + \int_{t_0}^{t} (t - s) f(s, x(s))\,ds.$$

   Genauer: Sei $I$ ein Intervall mit $t_0 \in I$. Ist $x \in C^2(I)$ eine Lösung der Anfangswertaufgabe, so ist $x$ auch eine Lösung der Integralgleichung. Ist umgekehrt $x \in C(I)$ eine Lösung der Integralgleichung, so ist sogar $x \in C^2(I)$ und $x$ ist eine Lösung der Anfangswertaufgabe.

2. Bei gegebenem $a > 0$ sei die Funktion $f = f(t, s, x)$ auf $D := \{(t, s, x) \in \mathbb{R}^3 : 0 \le s \le t \le a\}$ stetig und dort bezüglich der letzten Variablen $x$ global lipschitzstetig, d. h. es existiere $L > 0$ mit

   $$|f(t, s, x) - f(t, s, y)| \le L\,|x - y| \qquad \text{für alle } (t, s, x),\ (t, s, y) \in D.$$

   Dann besitzt die Volterrasche Integralgleichung

   $$x(t) = g(t) + \int_{0}^{t} f(t, s, x(s))\,ds$$

   für jedes $g \in C[0, a]$ genau eine auf $[0, a]$ stetige Lösung.

3. Gegeben sei die Anfangswertaufgabe

$$x' = tx^2, \qquad x(0) = 1.$$

Mit $x_0 := 1$ und

$$x_{k+1}(t) := 1 + \int_0^t s x_k(s)^2 \, ds$$

berechne man $x_1, x_2, x_3$. Man bestimme ein Intervall $[0, \alpha]$ mit $\alpha > 0$, auf dem eine Lösung eindeutig existiert und mache eine Fehlerabschätzung.

4. Die lineare Anfangswertaufgabe erster Ordnung

$$x' = 2tx + t, \qquad x(0) = x_0$$

besitzt die Lösung

$$x(t) = x_0 e^{t^2} + \frac{1}{2}(e^{t^2} - 1),$$

wie man durch `dsolve({diff(x(t),t)=2*t*x(t)+t,x(0)=x_0},x(t));` oder eigene Rechnung feststellt. Mit $x_0(t) := x_0$ sei

$$x_{k+1}(t) := x_0 + \int_0^t (2s x_k(s) + s) \, ds.$$

Man stelle die Iterierten $x_k$ geschlossen dar und begründe, weshalb die Folge $\{x_k\}$ auf jedem kompakten Intervall in $\mathbb{R}$ gleichmäßig gegen die Lösung der gegebenen Anfangswertaufgabe konvergiert.

5. Man zeige, dass die Anfangswertaufgabe für das mathematische Pendel, also

$$x'' + \omega_0^2 \sin x = 0, \quad x(0) = x_0, \quad x'(0) = 0,$$

für beliebige $\omega_0$ und $x_0$ genau eine Lösung besitzt. Diese existiert auf ganz $\mathbb{R}$ und ist gerade, also $x(t) = x(-t)$ für alle $t$. Für $\omega_0 := 2$ und $x_0 := 1$ berechne man mit Hilfe des Gaußschen Verfahrens vom arithmetisch-geometrischen Mittel die Periodenlänge $T = (4/\omega_0) K(\sin \frac{1}{2} x_0)$. Schließlich plotte man die Lösung auf $[0, 2T]$.

6. Gegeben sei die Anfangswertaufgabe

(P) $$x' = t + \sin x, \qquad x(0) = 0.$$

   (a) Man zeige, dass (P) auf jedem kompakten Teilintervall $I$ von $\mathbb{R}$ mit $0 \in I$ genau eine Lösung besitzt.

   (b) Mit Hilfe von Maple-Befehlen plotte man die Lösung von (P) auf dem Intervall $[-1, 1]$.

   (c) Man zeige, dass die Lösung $x$ von (P) nichtnegativ ist.

7. Man beweise[3], dass die Volterra-Integralgleichung

$$x(t) = g(t) + \int_0^t k(t, s, x(s)) \, ds$$

[3]Diese Aufgabe ist dem Buch von W. Walter über Gewöhnliche Differentialgleichungen entnommen.

mindestens eine in $[0,a]$ stetige Lösung besitzt, wenn $g \in C[0,a]$ und der „Kern“ $k(t,s,z)$ für $0 \le s \le t \le a$, $z \in \mathbb{R}$, stetig ist und einer Wachstumsbedingung $|k(t,s,z)| \le L(1+|z|)$ mit eine Konstanten $L > 0$ genügt.

Hinweis: Man wende den Schauderschen Fixpunktsatz an mit den folgenden Daten: Sei $X := C[0,a]$ der mit der Maximumnorm $\|x\| := \max_{t\in[0,a]} |x(t)|$ versehene Banach-Raum. Sei

$$K := \{x \in C[0,a] : |x(t)| \le \rho(t) \text{ für alle } t \in [0,a]\},$$

wobei $\rho(\cdot)$ die Lösung der Anfangswertaufgabe

$$\rho' = L(1+\rho), \qquad \rho(0) = \|g\|$$

ist und $F\colon K \longrightarrow C[0,a]$ durch

$$F(x)(t) := g(t) + \int_0^t k(t,s,x(s))\,ds$$

definiert ist. Man zeige also, dass mit diesen Daten die Voraussetzungen des Schauderschen Fixpunktsatzes erfüllt sind.

8. Man beweise die folgende Aussage:

   Sei $D \subset \mathbb{R}^{n+1}$ offen, $f\colon D \longrightarrow \mathbb{R}^n$ stetig und bezüglich des zweiten Arguments (global) lipschitzstetig mit einer Lipschitzkonstanten $L$. Sei $(t_0,x_0) \in D$ und $x$ eine Lösung von $x' = f(t,x)$, $x(t_0) = x_0$, auf $I := \{t \in \mathbb{R} : |t-t_0| \le \alpha\}$ mit $(t,x(t)) \in D$ für alle $t \in I$. Entsprechend sei auch $(\hat{t}_0,\hat{x}_0) \in D$ mit $\hat{t}_0 \in I$ und $\hat{x}$ eine Lösung von $x' = f(t,x)$, $x(\hat{t}_0) = \hat{x}_0$, auf $\hat{I} := \{t \in \mathbb{R} : |t-t_0| \le \hat{\alpha}\}$ mit $(t,\hat{x}(t)) \in D$ für alle $t \in \hat{I}$. Dann ist

   $$\|x(t) - \hat{x}(t)\| \le (M\,|t_0 - \hat{t}_0| + \|x_0 - \hat{x}_0\|)\,e^{L\,|t-\hat{t}_0|} \qquad \text{für alle } t \in I \cap \hat{I},$$

   wobei $M := \max_{t\in I} \|f(t,x(t))\|$.

9. Die Abbildung $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sei stetig und genüge einer einseitigen Lipschitzbedingung, d. h. es existiere ein $l \in \mathbb{R}$ (welches auch negativ sein kann) mit

   $$[f(x) - f(y)]^T (x-y) \le l\,\|x-y\|^2 \qquad \text{für alle } x,y \in \mathbb{R}^n,$$

   wobei hier $\|\cdot\|$ die euklidische Norm bedeute. Man zeige, dass die Anfangswertaufgabe

   $$x' = f(x), \qquad x(0) = x_0$$

   für jedes $x_0 \in \mathbb{R}^n$ höchstens eine[4] Lösung auf $[0,\infty)$ besitzt.

10. Mit Hilfe von Maple[5] bestimme man das maximale Existenzintervall für die Anfangswertaufgaben:

    (a) $x' = (1-2t)/\cos x$, $x(1) = 2$,
    (b) $x' = x/t + 4t^2x^2$, $x(1) = 1/15$,
    (c) $x' = x(1-x)$, $x(0) = 2$.

[4] Es kann auch die Existenz einer Lösung nachgewiesen werden, siehe Theorem 1.4.1 bei

K. Strehmel, R. Weiner (1992) *Linear-implizite Runge-Kutta-Methoden und ihre Anwendung*. B. G. Teubner, Stuttgart-Leipzig.

[5] Die Aufgabe haben wir dem Buch

D. Betounes (2001) *Differential Equations. Theory and Applications with Maple*. Springe-Verlag, Berlin-New York-Heidelberg

entnommen.

## 2.2 Lineare Differentialgleichungssysteme

Dass lineare Probleme bei Differentialgleichungen eine ganz besondere Rolle spielen und dass man nicht hoffen kann, nichtlineare Aufgaben adäquat untersuchen zu können, wenn man dazu bei linearen Problemen nicht in der Lage ist, versteht sich fast von selbst.

### 2.2.1 Lineare Systeme mit variablen Koeffizienten

In diesem Unterabschnitt betrachten wir lineare Differentialgleichungssysteme der Form

$$(*) \qquad x' = A(t)x + b(t).$$

Es werde vorausgesetzt, dass $A\colon \mathbb{R} \longrightarrow \mathbb{R}^{n\times n}$ und $b\colon \mathbb{R} \longrightarrow \mathbb{R}^n$ stetig sind. Die Modifikationen für den Fall, dass $A$ und $b$ nicht auf ganz $\mathbb{R}$, sondern nur einem Teilintervall $I \subset \mathbb{R}$ erklärt und stetig sind, werden offensichtlich sein. Alle Lösungen von $(*)$ erhält man natürlich dadurch, dass man zu einer speziellen Lösung von $(*)$ alle Lösungen der zugehörigen homogenen Aufgabe $x' = A(t)x$ addiert.

Bei gegebenem $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$ ist die Anfangswertaufgabe

$$x' = A(t)x, \qquad x(t_0) = x_0$$

eindeutig lösbar, ferner existiert die Lösung, die wir mit $x(\cdot; t_0, x_0)$ bezeichnen wollen, auf ganz $\mathbb{R}$. Dies hatten wir uns im Anschluss an den Existenz- und Eindeutigkeitssatz von Picard-Lindelöf überlegt, siehe Korollar 1.3. Die Abbildung $x_0 \mapsto x(\cdot; t_0, x_0)$ vom $\mathbb{R}^n$ in den Lösungsraum der homogenen Aufgabe $x' = A(t)x$ ist linear und bijektiv. Daher hat der Lösungsraum der homogenen Aufgabe die Dimension $n$.

Seien $x_1, \dots, x_n$ (nicht notwendig linear unabhängige) Lösungen von $x' = A(t)x$ und $X(t) = (\ x_1(t)\ \cdots\ x_n(t)\ )$ diejenige Matrix, die $x_1(t), \dots, x_n(t)$ als Spalten besitzt. Dann ist $X'(t) = A(t)X(t)$ und es gilt:

$$\det X(\tau) \neq 0 \text{ für } \textit{ein } \tau \in \mathbb{R} \Longleftrightarrow \det X(t) \neq 0 \text{ für } \textit{alle } t \in \mathbb{R}.$$

Denn: Angenommen, es ist $\det X(\tau) = 0$, also $X(\tau)$ singulär. Dann existiert ein $c \neq 0$ mit $X(\tau)c = 0$. Dann ist aber $X(\cdot)c = \sum_{j=1}^n c_j x_j(\cdot)$ eine Lösung von $x' = A(t)x$, die zur Zeit $\tau$ verschwindet. Da die triviale Lösung $x \equiv 0$ eine ebensolche ist und eine Lösung linearer Anfangswertaufgaben eindeutig bestimmt ist, ist $X(t)c = 0$ für alle $t$ und daher $\det X(t) = 0$ für alle $t \in \mathbb{R}$.

Ist $X'(t) = A(t)X(t)$ für alle $t$ und $\det X(t) \neq 0$ (für ein oder für alle $t$, das bleibt sich gleich), so nennt man $X$ ein *Fundamentalsystem* von $x' = A(t)x$. Ist $X$ ein Fundamentalsystem, so erhält man *sämtliche* Lösungen des homogenen Systems in der Form $x(t) = X(t)c$ mit $c \in \mathbb{R}^n$.

Ist $X$ ein beliebiges Fundamentalsystem von $x' = A(t)x$, so ist durch $x(t) := X(t) \int_{t_0}^t X^{-1}(s)b(s)\,ds$ eine spezielle Lösung von $(*)$ gegeben, wie man sofort nachrechnet. Daher ist

$$x(t) := X(t)c + X(t) \int_{t_0}^t X^{-1}(s)b(s)\,ds$$

mit beliebigem $c \in \mathbb{R}^n$ die allgemeine Lösung von $(*)$, während durch

$$x(t) := X(t)\left[X^{-1}(t_0)x_0 + \int_{t_0}^{t} X^{-1}(s)b(s)\,ds\right]$$

die (eindeutige) Lösung der Anfangswertaufgabe

$$x' = A(t)x + b(t), \qquad x(t_0) = x_0$$

gegeben ist.

**Beispiel:** Nur in Ausnahmefällen kann man bei einem linearen Differentialgleichungssystem mit variablen Koeffizienten ein Fundamentalsystem geschlossen angeben. Wir reproduzieren ein Beispiel bei W. Walter (1996, S. 144), in welchem ein Fundamentalsystem angegeben werden kann. Und zwar betrachten wir die inhomogene Aufgabe

$$(*) \qquad \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} \frac{1}{t} & -1 \\ \frac{1}{t^2} & \frac{2}{t} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} t \\ -t^2 \end{pmatrix}, \qquad \begin{pmatrix} x_1(1) \\ x_2(1) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Mit Hilfe von

```
dsolve({diff(x_1(t),t)=(1/t)*x_1(t)-x_2(t),
     diff(x_2(t),t)=(1/t^2)* x_1(t)+(2/t)*x_2(t),
     x_1(1)=1, x_2(1)=0},{x_1(t),x_2(t)});
```

und

```
dsolve({diff(x_1(t),t)=(1/t)*x_1(t)-x_2(t),
     diff(x_2(t),t)=(1/t^2)*x_1(t)+(2/t)*x_2(t),
     x_1(1)=0, x_2(1)=1},{x_1(t),x_2(t)});
```

erhält man das Fundamentalsystem

$$X(t) = \begin{pmatrix} (-\ln t + 1)t^2 & -t^2 \ln t \\ t \ln t & t(1 + \ln t) \end{pmatrix}.$$

Die Berechnung der Lösung von $(*)$ mittels der Formel

$$x(t) = X(t)\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_1^t X(s)^{-1} \begin{pmatrix} s \\ -s^2 \end{pmatrix} ds\right)$$

scheint (wer kennt eine bessere Lösung?) komplizierter zu sein als z. B. in *Mathematica*. Dies liegt daran, dass `int` nicht auf Vektoren komponentenweise angewandt werden kann. Durch

```
with(LinearAlgebra):
b:=t-><t,-t^2>:
X:=t-><<(-ln(t)+1)*t^2|-t^2*ln(t)>,<t*ln(t)|t*(1+ln(t))>>:
x:=t->X(t).(<1,0>+<seq(int((MatrixInverse(X(s)).b(s))[i],s=1..t),i=1..2)>):
<seq(expand(x(t)[i]),i=1..2)>;
```

erhält man mit

$$\begin{pmatrix} -\frac{1}{2}\,t^2\ln(t) - \frac{1}{2}\,t^2\ln(t)^2 + \frac{3}{4}\,t^2 + \frac{1}{4}\,t^4 \\ \frac{3}{2}\,t\ln(t) + \frac{1}{2}\,t\ln(t)^2 - \frac{3}{4}\,t^3 + \frac{3}{4}\,t \end{pmatrix}$$

die gesuchte Lösung. Diese kann man natürlich auch versuchen, direkt durch

```
dsolve({diff(x_1(t),t)=(1/t)*x_1(t)-x_2(t)+t,
diff(x_2(t),t)=(1/t^2)*x_1(t)+(2/t)*x_2(t)-t^2,
x_1(1)=1,x_2(1)=0},{x_1(t),x_2(t)});
```

zu berechnen. Genau we bei *Mathematica* ist man hiermit nicht erfolgreich. □

Ein Fundamentalsystem mit $X(t_0) = I$ nennt man ein *normiertes Fundamentalsystem*. Die $j$-te Spalte $x_j$ eines normierten Fundamentalsystems erhält man als Lösung der Anfangswertaufgabe

$$x' = A(t)x, \qquad x(t_0) = e_j,$$

wobei $e_j$ den $j$-ten Einheitsvektor im $\mathbb{R}^n$ bedeutet.

Ist $X'(t) = A(t)X(t)$, also $X(t) = (\ x_1(t)\ \cdots\ x_n(t)\ )$ ein "Lösungssystem", so heißt $\det X(t)$ die *Wronski-Determinante* von $X(t)$. Über diese wird im folgenden Lemma eine Aussage gemacht.

**Lemma 2.1** *Sei $X'(t) = A(t)X(t)$ für alle $t$. Dann ist*

$$\det X(t) = \det X(t_0)\,\exp\Big(\int_{t_0}^{t} \operatorname{tr} A(s)\,ds\Big).$$

*Hierbei bedeutet* $\operatorname{tr} A(s) := \sum_{i=1}^{n} a_{ii}(s)$ *die Spur von* $A(s)$.

**Beweis:** Offenbar können wir annehmen, dass $\det X(t) \neq 0$ für alle $t$. Bei festem, beliebigen $\tau \in \mathbb{R}$ definiere man $Y(t) := X(t)X^{-1}(\tau)$.Wegen

$$Y'(t) = X'(t)X^{-1}(\tau) = A(t)X(t)X^{-1}(\tau) = A(t)Y(t)$$

und $Y(\tau) = I$ ist $Y$ ein normiertes Fundamenalsystem.

Sei $Y(t) = (\ y_1(t)\ \cdots\ y_n(t)\ )$. Bekanntlich ist die Determinante einer $n\times n$-Matrix $A = (a_{ij})$ gegeben durch

$$\det A = \sum_{p} (-1)^{v(p)} a_{1p_1} \cdots a_{np_n},$$

wobei die Summe über alle Permutionen $p = (p_1, \ldots, p_n)$ der Zahlen $\{1, \ldots, n\}$ zu nehmen ist und $v(p)$ die Anzahl der Inversionen von $p$ bedeutet. Hieraus erhält man

$$\begin{aligned} \frac{d}{dt}\det Y(t) &= \sum_{j=1}^{n} \det(\ y_1(t)\ \cdots\ y_{j-1}(t)\ \ y_j'(t)\ \ y_{j+1}(t)\ \cdots\ y_n(t)\ ) \\ &= \sum_{j=1}^{n} \det(\ y_1(t)\ \cdots\ y_{j-1}(t)\ \ A(t)y_j(t)\ \ y_{j+1}(t)\ \cdots\ y_n(t)\ ). \end{aligned}$$

Setzt man hier $t = \tau$ und beachtet, dass $y_j(\tau) = e_j$ der $j$-te Einheitsvektor ist, so erhält man

$$\begin{aligned}\frac{d}{dt}\det Y(t)\Big|_{t=\tau} &= \sum_{j=1}^{n} \det(\ e_1 \ \cdots \ e_{j-1} \ A(\tau)e_j \ e_{j+1} \ \cdots \ e_n\ ) \\ &= \sum_{j=1}^{n} a_{jj}(\tau) \\ &= \operatorname{tr} A(\tau).\end{aligned}$$

Daher ist

$$\begin{aligned}\frac{d}{dt}\det X(t)\Big|_{t=\tau} &= \frac{d}{dt}\det Y(t)\Big|_{t=\tau} \det X(\tau) \\ &= \operatorname{tr} A(\tau)\det X(\tau).\end{aligned}$$

Dies gilt für jedes $\tau$ und daher genügt $\det X(t)$ der linearen Differentialgleichung erster Ordnung

$$\frac{d}{dt}\det X(t) = \operatorname{tr} A(t)\ \det X(t),$$

woraus die Behauptung folgt. □

**Beispiel:** Gegeben sei die lineare Differentialgleichung $n$-ter Ordnung

$$(*) \qquad x^{(n)} + p_{n-1}(t)x^{(n-1)} + \cdots + p_1(t)x' + p_0(t)x = 0.$$

Als System geschrieben lautet $(*)$:

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= x_3 \\ &\vdots \\ x_{n-1}' &= x_n \\ x_n' &= -p_0(t)x_1 - p_1(t)x_2 - \cdots - p_{n-1}(t)x_n \end{aligned}$$

bzw.

$$(**) \quad x' = A(t)x \qquad \text{mit} \qquad A(t) := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -p_0(t) & -p_1(t) & -p_2(t) & \cdots & -p_{n-1}(t) \end{pmatrix}.$$

Folglich ist $\operatorname{tr} A(t) = -p_{n-1}(t)$. Sind $x^1, \ldots, x^n$ Lösungen von $(*)$, so ist

$$X(t) := \begin{pmatrix} x^1(t) & x^2(t) & \cdots & x^n(t) \\ \dfrac{dx^1}{dt}(t) & \dfrac{dx^2}{dt}(t) & \cdots & \dfrac{dx^n}{dt}(t) \\ \vdots & \vdots & & \vdots \\ \dfrac{d^{n-1}x^1}{dt^{n-1}}(t) & \dfrac{d^{n-1}x^2}{dt^{n-1}}(t) & \cdots & \dfrac{d^{n-1}x^n}{dt^{n-1}}(t) \end{pmatrix}$$

ein Lösungssystem von $(**)$. Wegen Lemma 2.1 ist

$$\det X(t) = \det X(t_0) \exp\Big(-\int_{t_0}^{t} p_{n-1}(s)\,ds\Big).$$

Nun sind $x^1, \cdots, x^n$ genau dann linear abhängig, wenn $c = (c_j) \in \mathbb{R}^n \setminus \{0\}$ existiert mit $\sum_{j=1}^n c_j x^j(t) \equiv 0$. Dies wiederum ist genau dann der Fall, wenn

$$\sum_{j=1}^{n} c_j \frac{d^i x^j}{dt^i}(t) \equiv 0 \qquad (i = 0, \ldots, n-1),$$

was wiederum äquivalent dazu ist, dass $X(t)$ singulär ist. Insbesondere erhält man $n$ linear unabhängige Lösungen $x^1, \ldots, x^n$ von $(*)$, indem man $(*)$ mit den Anfangsbedingungen $x(t_0) = 1$, $x^{(i)}(t_0) = 0$ für $i = 1, \ldots, n-1$ zur Gewinnung von $x^1$, anschließend mit den Anfangsbedingungen $x(t_0) = 0$, $x'(t_0) = 1$, $x^{(i)}(t_0) = 0$ für $i = 2, \ldots, n-1$ zur Bestimmung von $x^2$ löst usw. □

**Beispiel:** Die Differentialgleichung

$$x'' - \frac{1+t}{t}x' + \frac{1}{t}x = 0$$

besitzt die Lösungen $e^t$ und $1 + t$ sowie die allgemeine Lösung

$$x(t) = c_1 e^t + c_2(1+t)$$

auf $(0, \infty)$, da die Wronski-Determinante

$$\det \begin{pmatrix} e^t & 1+t \\ e^t & 1 \end{pmatrix} = -te^t$$

auf $(0, \infty)$ nicht verschwindet. Das selbe Ergebnis erhält man durch

```
dsolve(diff(x(t),t$2)-((1+t)/t)*diff(x(t),t)+(1/t)*x(t)=0,x(t));
```

□

### 2.2.2 Lineare Systeme mit konstanten Koeffizienten

In diesem Unterabschnitt kommt es vor allem darauf an, ein Fundamentalsystem zu $x' = Ax$ zu bestimmen, wobei $A \in \mathbb{R}^{n\times n}$ eine konstante Matrix ist. Ist dies gelungen, so können inhomogene Aufgaben wie im letzten Unterabschnitt behandelt werden.

Zunächst machen wir für eine Lösung von $x' = Ax$ aber den Ansatz $x(t) = ce^{\lambda t}$, wobei der Vektor $c \in \mathbb{R}^n$ vom Nullvektor verschieden sein sollte, damit der Lösungskandidat nichttrivial ist. Wie man sehr leicht nachrechnet ist $x$ genau dann eine Lösung, wenn $Ac = \lambda c$, also $\lambda$ ein Eigenwert von $A$ mit zugehörigem Eigenvektor $c$ ist. Man erkennt hieran eine (kleine) Schwierigkeit: Eine reelle Matrix hat i. allg. nicht nur reelle, sondern auch komplexe Eigenwerte, die dann in konjugiert komplexen Paaren auftreten. Dann sind auch die Eigenvektoren komplex, sie treten in konjugiert komplexen

Paaren auf. Damit ist $x(t) = ce^{\lambda t}$ in diesem Falle komplexwertig, womit man bei reellen Ausgangsdaten nicht immer zufrieden ist. Dies ist aber wirklich nur eine kleine Schwierigkeit. Denn ist $A \in \mathbb{R}^{n\times n}$ eine reelle $n \times n$-Matrix mit dem komplexen Eigenwert $\lambda = \mu + i\nu$ und dem zugehörigen Eigenvektor $c = a + ib$, so ist auch $\bar{\lambda} = \mu - i\nu$ ein Eigenwert von $A$ mit dem zugehörigen Eigenvektor $\bar{c} = a - ib$. Dann ist

$$ce^{\lambda t} = (a + ib)e^{(\mu+i\nu)t} = e^{\mu t}(a\cos\nu t - b\sin\nu t) + ie^{\mu t}(b\cos\nu t + a\sin\nu t).$$

Sowohl der Real- als auch der Imaginärteil hiervon sind Lösungen. Wichtig ist die Bemerkung: Besitzt $A \in \mathbb{R}^{n\times n}$ ein System von $n$ linear unabhängigen Eigenvektoren, so gewinnt man auf diese Weise $n$ linear unabhängige Lösungen von $x' = Ax$ bzw. ein Fundamentalsystem von $x' = Ax$. Insbesondere ist das der Fall, wenn $A$ symmetrisch ist oder die Eigenwerte von $A$ paarweise von einander verschieden sind.

**Beispiel:** Man bestimme die allgemeine Lösung von

$$x''' + 6x'' + 11x' + 6x = 0.$$

Dies ist äquivalent zu

$$x' = Ax \qquad \text{mit} \qquad A := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{pmatrix}.$$

Als charakteristisches Polynom erhält man

$$p_3(\lambda) := \det(A - \lambda I) = -\lambda^3 - 6\lambda^2 - 11\lambda - 6$$

mit den Wurzeln

$$\lambda_1 = -1, \quad \lambda_2 = -2, \quad \lambda_3 = -3$$

und den zugehörigen Eigenvektoren

$$c_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad c_2 = \begin{pmatrix} 1 \\ -2 \\ 4 \end{pmatrix}, \quad c_3 = \begin{pmatrix} 1 \\ -3 \\ 9 \end{pmatrix}.$$

Als allgemeine Lösung der gegebenen Differentialgleichung erster Ordnung hat man also

$$x(t) = ae^{-t} + be^{-2t} + ce^{-3t}.$$

Will man z. B. die Lösung bestimmen, die den Anfangsbedingungen $x(0) = 1$, $x'(0) = x''(0) = 0$ genügt, so hat man das Gleichungssystem

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

zu lösen, woraus man $(a, b, c) = (3, -3, 1)$ erhält. Bei diesen Rechnungen (die hier noch leicht "per hand" durchgeführt werden können) kann wieder Maple helfen, besser geeignet

für Aufgaben der Numerischen linearen Algebra ist aber MATLAB. Zunächst geben wir einige Möglichkeiten von Maple bei der Nullstellenbestimmung von Polynomen bzw. bei Eigenwertaufgaben von Matrizen an. Hierbei ist zu beachten, dasss der Vorteil von Maple gegenüber MATLAB darin besteht, dass symbolisch gerechnet wird, während MATLAB numerisch mit unvermeidbaren Rundungsfehlern rechnet). Die Eigenwerte und Eigenvektoren von $A$ können wir mit

```
with(LinearAlgebra):
A:=Matrix([[0,1,0],[0,0,1],[-6,-11,-6]]):
(lambda,U):=Eigenvectors(A);
```

berechnen, und erhalten den Output

$$\lambda,\, U := \begin{bmatrix} -3 \\ -2 \\ -1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 & 1 \\ -3 & -2 & -1 \\ 9 & 4 & 1 \end{bmatrix},$$

wobei wir uns an die Ausgabe von Maple halten[6]. Will man nur die Eigenwerte bestimmen, so kann man `Eigenvalues` benutzen. Um obiges Resultat wenigstens teilweise zu überprüfen, geben wir

```
A.U[1..-1,1]=lambda[1]*U[1..-1,1];
```

ein und erhalten

$$\begin{bmatrix} -3 \\ 9 \\ 27 \end{bmatrix} = \begin{bmatrix} -3 \\ 9 \\ 27 \end{bmatrix}$$

In MATLAB würden die entsprechenden Befehle folgendermaßen aussehen:

```
A=[0 1 0;0 0 1;-6 -11 -6];
[U,Lambda]=eig(A)
```

liefert

$$U = \begin{pmatrix} -0.5774 & 0.2182 & -0.1048 \\ 0.5774 & -0.4364 & 0.3145 \\ -0.5774 & 0.8729 & -0.9435 \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} -1.0000 & 0 & 0 \\ 0 & -2.0000 & 0 \\ 0 & 0 & -3.0000 \end{pmatrix}.$$

Nach `format long` erhält man das Ergebnis mit mehr Stellen. In Maple sorgt ein Semikolon `;` dafür, dass ein Echo erschallt bzw. das Ergebnis ausgegeben wird, während ein Semikolon in MATLAB gerade ein Echo verhindert. Durch `lambda=diag(Lambda);` kann aus der Diagonalmatrix $\Lambda$ ein Vektor $\lambda$ erzeugt werden (natürlich mit den Diagonaleinträgen als Komponenten). □

[6]Merkwürdigerweise ist die Reihenfolge der Komponenten von $\lambda$ bzw. der Spalten von $A$ sozusagen zeitabhängig. Wenn man zweimal hintereinander dieselben Befehle gibt, können permutierte Ergebnisse erscheinen.

**Beispiel:** Gesucht (siehe W. WALTER (1993, S. 149)) sei die allgemeine Lösung zu

$$x' = Ax \qquad \text{mit} \qquad A := \begin{pmatrix} 1 & -2 & 0 \\ 2 & 0 & -1 \\ 4 & -2 & -1 \end{pmatrix}.$$

Als charakteristisches Polynom zu $A$ erhält man

$$p_3(\lambda) := \det(A - \lambda I) = (1 - \lambda)(\lambda^2 + \lambda + 2).$$

Mit $\alpha := \sqrt{7}/2$ erhält man die Eigenwerte

$$\lambda_1 = -\frac{1}{2} + i\alpha, \quad \lambda_2 = -\frac{1}{2} - i\alpha, \quad \lambda_3 = 1.$$

Diese sind sämtlich voneinander verschieden, so dass man durch die obige Methode ein Fundamentalsystem und damit die allgemeine Lösung der homogenen Aufgabe erhält. Die Eigenvektoren sind (sie sind natürlich jeweils nur bis auf einen Faktor eindeutig bestimmt)

$$c_1 = \begin{pmatrix} \frac{3}{2} + i\alpha \\ 2 \\ 4 \end{pmatrix}, \quad c_2 = \bar{c}_1 = \begin{pmatrix} \frac{3}{2} - i\alpha \\ 2 \\ 4 \end{pmatrix}, \quad c_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

Hieraus erhält man die linear unabhängigen Lösungen

$$\begin{aligned} x^1(t) &= e^{-\frac{1}{2}t}\Big[\begin{pmatrix} \frac{3}{2} \\ 2 \\ 4 \end{pmatrix} \cos \alpha t - \begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix} \sin \alpha t\Big], \\ x^2(t) &= e^{-\frac{1}{2}t}\Big[\begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix} \cos \alpha t + \begin{pmatrix} \frac{3}{2} \\ 2 \\ 4 \end{pmatrix} \sin \alpha t\Big], \\ x^3(t) &= e^t \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \end{aligned}$$

und damit das Fundamentalsystem $X = (\ x^1 \quad x^2 \quad x^3\ )$. Nach (in Maple)

```
A:=Matrix([[1,-2,0],[2,0,-1],[4,-2,-1]]):
(lambda,U):=eigenvectors(A);
```

erhalten wir den Output

$$\lambda,\, U := \begin{bmatrix} -\frac{1}{2} + \frac{1}{2} I\sqrt{7} \\ -\frac{1}{2} - \frac{1}{2} I\sqrt{7} \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{3}{8} + \frac{1}{8} I\sqrt{7} & \frac{3}{8} - \frac{1}{8} I\sqrt{7} & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 1 & 2 \end{bmatrix}$$

Hierbei bedeutet $I$ natürlich die imaginäre Einheit. □

Der Fall einer diagonal ähnlichen (oder diagonalisierbaren) Koeffizientenmatrix $A \in \mathbb{C}^{n\times n}$ (es ist zweckmäßig gleich den komplexen Fall zuzulassen, an den Existenz- und Eindeutigkeitsaussagen ändert sich natürlich nichts) ist also im Prinzip erledigt: Man bestimme die Eigenwerte $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ und zugehörige (linear unabhängige) Eigenvektoren $c_1, \ldots, c_n \in \mathbb{C}^n$. Durch $c_1 e^{\lambda_1 t}, \ldots, c_n e^{\lambda_n t}$ hat man $n$ linear unabhängige Lösungen von $x' = Ax$ bestimmt. Nicht jede Matrix ist diagonalisierbar, im allgemeinen Fall werden wir daher auf die *Jordansche Normalform* zurückgreifen. Zunächst aber eine Bezeichnung:

- Ist $A \in \mathbb{C}^{n\times n}$, so bezeichnen wir das durch $X(0) = I$ normierte Fundamentalsystem $X(t)$ zu $x' = Ax$ mit $e^{At}$.

Im folgenden Lemma wird u. a. diese Bezeichnung gerechtfertigt.

**Lemma 2.2** *Sei $A \in \mathbb{C}^{n\times n}$. Dann gilt:*

1. *Es ist*

$$\frac{d}{dt}e^{At} = Ae^{At} = e^{At}A, \qquad e^{A(t+s)} = e^{At}e^{As}, \qquad (e^{At})^{-1} = e^{(-A)t} = e^{A(-t)}$$

   *für alle $t, s \in \mathbb{R}$.*

2. *Ist $C \in \mathbb{C}^{n\times n}$ nichtsingulär und $J := C^{-1}AC$, so ist $e^{At} = Ce^{Jt}C^{-1}$.*

3. *Mit $\lambda \in \mathbb{C}$ sei $J \in \mathbb{C}^{n\times n}$ definiert durch*

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & & \ddots & 1 \\ 0 & 0 & \cdots & \cdots & \lambda \end{pmatrix}.$$

   *Dann ist*

$$e^{Jt} = e^{\lambda t}\begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n-2}}{(n-2)!} & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{n-3}}{(n-3)!} & \frac{t^{n-2}}{(n-2)!} \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

4. *Die inhomogene Anfangswertaufgabe*

$$x' = Ax + b(t), \qquad x(t_0) = x_0$$

*besitzt die eindeutige Lösung*

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}b(s)\,ds.$$

5. *Sind $A, B \in \mathbb{C}^{n\times n}$ mit $AB = BA$, so ist $e^{(A+B)t} = e^{At}e^{Bt}$ für alle $t$.*

6. *Es ist*

$$e^{At} = \sum_{j=0}^{\infty} \frac{A^j t^j}{j!}.$$

*Genauer ist*

$$\lim_{k\to\infty} \sum_{j=0}^{k} \frac{A^j t^j}{j!} = e^{At} \qquad \text{für jedes } t \in \mathbb{R},$$

*wobei diese Konvergenz auf kompakten Teilmengen von $\mathbb{R}$ gleichmäßig ist.*

**Beweis:** Die erste Gleichung im ersten Teil folgt daraus, dass $e^{At}$ nach Definition ein Fundamentalsystem von $x' = Ax$ ist. Um $Ae^{At} = e^{At}A$ zu erhalten, definieren wir $Y(t) := Ae^{At}$ und $Z(t) := e^{At}A$. Dann ist $Y(0) = Z(0) = A$ und

$$\begin{aligned} \frac{d}{dt}Y(t) &= A\frac{d}{dt}e^{At} = AAe^{At} = AY(t) \\ \frac{d}{dt}Z(t) &= \frac{d}{dt}e^{At}A = Ae^{At}A = AZ(t) \end{aligned}$$

und folglich $Y(t) = Z(t)$ wegen der Eindeutigkeit bei linearen Anfangswertaufgaben.

Die Aussage $e^{A(t+s)} = e^{At}e^{As}$ ist für $t = 0$ (und beliebiges $s$) richtig. Da beide Seiten außerdem der Matrix-Differentialgleichung $Z' = AZ$ genügen, ist die Aussage richtig.

Es ist

$$0 = \frac{d}{dt}[(e^{At})^{-1}e^{At}] = \frac{d}{dt}[(e^{At})^{-1}]e^{At} + A$$

und damit

$$\frac{d}{dt}[(e^{At})^{-1}] = -A(e^{At})^{-1}.$$

Folglich ist $(e^{At})^{-1} = e^{(-A)t}$. Ferner ist $e^{(-A)t} = e^{A(-t)}$, da dies für $t = 0$ richtig ist und beide Seiten der Matrix-Differentialgleichung $Z' = -AZ$ genügen.

Für $t = 0$ stimmt die behauptete Gleichung $e^{At} = Ce^{Jt}C^{-1}$. Ferner ist

$$\frac{d}{dt}[Ce^{Jt}C^{-1}] = CJe^{Jt}C^{-1} = \underbrace{CJC^{-1}}_{=A}Ce^{Jt}C^{-1} = ACe^{Jt}C^{-1}$$

und daraus folgt die Behauptung.

Man definiere

$$x_j(t) := e^{\lambda t} \begin{pmatrix} \dfrac{t^{j-1}}{(j-1)!} \\ \vdots \\ \dfrac{t}{1!} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad j = 1, \ldots, n.$$

Dann ist

$$x_j'(t) = \lambda x_j(t) + x_{j-1}(t) = J x_j(t), \qquad j = 1, \ldots, n.$$

Da außerdem $x_j(0) = e_j$ der $j$-te Einheitsvektor ist, ist die Behauptung bewiesen.

Die Lösung der inhomogenen Anfangswertaufgabe ist durch

$$x(t) = X(t)[X^{-1}(t_0)x_0 + \int_{t_0}^{t} X^{-1}(s)b(s)\,ds]$$

gegeben, wobei $X(t)$ ein beliebiges Fundamentalsystem zu $x' = Ax$ ist. Einsetzen von $X(t) = e^{At}$ und Berücksichtigung der oben nachgewiesenen Rechenregeln liefert die Behauptung.

Zunächst folgt aus der Vertauschbarkeit von $A$ und $B$, dass $e^{At}B = Be^{At}$. Dies ist nämlich für $t = 0$ richtig, ferner ist

$$\frac{d}{dt}[e^{At}B - Be^{At}] = Ae^{At}B - BAeAt = A(e^{At}B - Be^{At}),$$

woraus die Zwischenbehauptung folgt. Die Behauptung selbst sieht man ebenfalls nach vertrautem Muster ein. Sie ist für $t = 0$ richtig und es ist

$$\frac{d}{dt}[e^{At}e^{Bt}] = Ae^{At}e^{Bt} + e^{At}Be^{Bt} = (A + B)e^{At}e^{Bt},$$

woraus die Behauptung folgt.

Man gebe sich $\alpha > 0$ vor und definiere $I_\alpha := [-\alpha, \alpha]$. Ferner sei die Abbildung $F\colon C(I_\alpha; \mathbb{C}^{n\times n}) \longrightarrow C(I_\alpha; \mathbb{C}^{n\times n})$ definiert durch

$$F(X)(t) := I + \int_0^t AX(s)\,ds.$$

Offenbar ist $X(t) = e^{At}$ Fixpunkt von $F$. Auf $C(I_\alpha; \mathbb{C}^{n\times n})$ definiere man die Norm

$$\|X\| := \max_{t\in I_\alpha} e^{-2\,\|A\|\,|t|}\,\|X(t)\|,$$

wodurch $C(I_\alpha; \mathbb{C}^{n\times n})$ zu einem Banach-Raum wird. Hierbei ist rechts $\|\cdot\|$ eine Matrixnorm auf $\mathbb{C}^{n\times n}$. Wie wir schon früher in einer ganz ähnlichen Situation beim Beweis

des Satzes von Picard-Lindelöf ausgerechnet haben, kontrahiert die Abbildung $F$ auf $C(I_\alpha;\mathbb{C}^{n\times n})$ mit einer Lipschitzkonstanten $q := \frac{1}{2}$. Mit

$$X_0(t) := I, \qquad X_{k+1}(t) := F(X_k)(t), \qquad k = 0, 1, \ldots$$

ist

$$X_k(t) = \sum_{j=0}^{k} \frac{A^j t^j}{j!}.$$

Der Fixpunktsatz für kontrahierende Abbildungen liefert die Abschätzung

$$\|X - X_k\| \le \frac{q^k}{1-q} \|X_1 - X_0\|$$

bzw.

$$e^{-2\|A\|\alpha} \left\| e^{At} - \sum_{j=0}^{k} \frac{A^j t^j}{j!} \right\| \le \left(\frac{1}{2}\right)^{k-1} \|A\|\alpha \qquad \text{für alle } t \in I_\alpha.$$

Folglich ist

$$\left\| e^{At} - \sum_{j=0}^{k} \frac{A^j t^j}{j!} \right\| \le \|A\|\alpha e^{2\|A\|\alpha} \left(\frac{1}{2}\right)^{k-1} \qquad \text{für alle } t \in I_\alpha$$

und hieraus folgt die Behauptung. □

Obige Aussagen liefern die Grundlage zur Berechnung des Fundamentalsystems $e^{At}$ auch für den Fall, dass $A$ nicht diagonalisierbar ist. Denn zwar ist nicht jede Matrix ähnlich einer Diagonalmatrix, aber jede Matrix lässt sich durch eine Ähnlichkeitstransformation auf die sogenannte *Jordan'sche Normalform* transformieren, d. h. zu jeder (reellen oder komplexen) $n \times n$-Matrix $A$ existiert eine (im allgemeinen komplexe) nichtsinguläre Matrix $C$ derart, dass $J := C^{-1}AC$ die Jordansche Normalform

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{pmatrix},$$

wobei die "Jordan-Kästen" $J_j$, $j = 1, \ldots, p$, quadratische Matrizen der Form

$$J_j = \begin{pmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & & & 1 \\ 0 & & & & \lambda_j \end{pmatrix} \in \mathbb{C}^{n_j \times n_j}$$

sind und in der Block-Diagonalmatrix $J$ außerhalb der Jordan-Kästen nur Nullen stehen. Dabei ist $\sum_{j=1}^{p} n_j = n$, ferner sind die $\lambda_1, \ldots, \lambda_p$ nicht notwendig voneinander

verschieden. Wegen $e^{At} = Ce^{Jt}C^{-1}$ muss jetzt noch $e^{Jt}$ berechnet werden. Nun ist aber, wie man z. B. aus der Reihendarstellung erkennt,

$$e^{Jt} = \begin{pmatrix} e^{J_1 t} & & \\ & \ddots & \\ & & e^{J_p t} \end{pmatrix},$$

so daß nur noch $e^{J_j t}$, $j = 1, \ldots, p$, zu berechnen ist. Dies ist aber in Teil 3 von Lemma 2.2 geschehen. Hiernach ist

$$e^{J_j t} = e^{\lambda_j t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{n_j-2}}{(n_j-2)!} & \frac{t^{n_j-1}}{(n_j-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{n_j-3}}{(n_j-3)!} & \frac{t^{n_j-2}}{(n_j-2)!} \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Damit haben wir schließlich das Fundamentalsystem $e^{At}$ berechnet, allerdings unter der Voraussetzung, dass uns die Jordan'sche Normalform von $A$ bekannt ist.

**Beispiel:** Die Matrix

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

hat schon Jordan'sche Normalform und es ist daher

$$e^{At} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Komplizierter ist die Situation bei der Matrix

$$A := \begin{pmatrix} 0 & 1 & 0 \\ 4 & 3 & -4 \\ 1 & 2 & -1 \end{pmatrix}.$$

Als Resultat von

```
with(LinearAlgebra}:
A:=Matrix([[0,1,0],[4,3,-4],[1,2,-1]]);
(lambda,U):=Eigenvectors(A);
```

erhalten wir

$$\lambda,\, U := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & \frac{3}{2} & 0 \end{bmatrix}$$

Die Eigenvektoren bilden also keine Basis, die Matrix $A$ ist nicht diagonalisierbar. Es muss also die Jordansche Normalform berechnet werden. Es ist Aufgabe der Linearen Algebra (und nicht einer Vorlesung über gewöhnliche Differentialgleichungen) die entsprechenden Methoden bereit zu stellen. In der folgenden kleinen Sitzung wird die Benutzung des Maple-Befehls `JordanForm` geschildert.

```
> with(LinearAlgebra):
> C:=JordanForm(A,output='Q');
```

$$C := \begin{bmatrix} 5 & 4 & -4 \\ 0 & 4 & 0 \\ 5 & 6 & -5 \end{bmatrix}$$

```
> J:=JordanForm(A);
```

$$J := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

```
> C^(-1).A.C-J;
```

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Es ist dann

$$e^{At} = C \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{pmatrix} C^{-1}.$$

In MATLAB scheint es keine entsprechende Funktion zur Berechnung der Jordanschen Normalform einer Matrix zu geben. Dafür gibt es dort die Funktion `expm`, mit der durch `expm(A)` zu einer gegebenen Matrix $A$ die Matrix $e^A$ berechnet werden kann. □

Wir benötigen nun einige Begriffe und Hilfsmittel aus der linearen Algebra.

- Ist $B \in \mathbb{C}^{n\times n}$, so heißt $\operatorname{Kern}(B) := \{y \in \mathbb{C}^n : By = 0\}$ der *Kern* (oder *Nullraum*) von $B$.
- Sei $A \in \mathbb{C}^{n\times n}$ und $\lambda$ ein Eigenwert von $A$. Dann ist

  $$\{0\} \overset{\subset}{\neq} \operatorname{Kern}(A-\lambda I) \subset \operatorname{Kern}(A-\lambda I)^2 \subset \cdots.$$

  Sei $r(\lambda)$ die kleinste natürliche Zahl $k$ mit $\operatorname{Kern}(A-\lambda I)^{k+1} = \operatorname{Kern}(A-\lambda I)^k$. Dann heißt $M_\lambda(A) := \operatorname{Kern}(A-\lambda I)^{r(\lambda)}$ der *verallgemeinerte Eigenraum* von $A$ zum Eigenwert $\lambda$.
- Es ist $\dim M_\lambda(A)$ gleich der *algebraischen Vielfachheit* von $\lambda$, d. h. der Vielfachheit als Wurzel des charakteristischen Polynoms.
- Man sagt, ein Eigenwert $\lambda$ von $A$ habe *einfache Elementarteiler*, falls $r(\lambda) = 1$ bzw. $M_\lambda(A) = \operatorname{Kern}(A-\lambda I)$. In diesem Fall stimmen algebraische und *geometrische Vielfachheit*, d. h. $\dim \operatorname{Kern}(A-\lambda I)$, überein. I. allg. ist die geometrische Vielfachheit kleiner oder gleich der algebraischen Vielfachheit.

- Sind $\lambda_1, \ldots, \lambda_s$ die paarweise voneinander verschiedenen Eigenwerte von $A$, so sind die zugehörigen verallgemeinerten Eigenräume $M_{\lambda_1}(A), \ldots, M_{\lambda_s}(A)$ *invariant* unter $A$, d. h. es ist $AM_{\lambda_j}(A) \subset M_{\lambda_j}(A)$, und es ist

$$\mathbb{C}^n = M_{\lambda_1}(A) \oplus \cdots \oplus M_{\lambda_s}(A).$$

**Beispiel:** Wir kommen auf obiges Beispiel zurück, es sei also

$$A := \begin{pmatrix} 0 & 1 & 0 \\ 4 & 3 & -4 \\ 1 & 2 & -1 \end{pmatrix}.$$

Die Eigenwerte von $A$ sind 0 und der doppelte Eigenwert 1. Da 0 ein (algebraisch) einfacher Eigenwert ist, ist $r(0) = 1$ und

$$M_0(A) = \text{Kern}\,(A) = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Letzteres kann man z. B. aus `NullSpace(A)` erhalten. Dagegen hat der Eigenwert 1 die algebraische Vielfachheit 2. Als Ergebnis von

```
NullSpace(A-1*IdentityMatrix(3));
```

erhalten wir den Vektor $(1, 1, \frac{3}{2})^T$, der Eigenraum zum Eigenwert 1 ist also eindimensional, die geometrische Vielfachheit des Eigenwertes 1 ist daher 1. Dagegen erhält man als Resultat von

```
NullSpace((A-1*IdentityMatrix(3))^2);
```

die beiden (linear unabhängigen) Vektoren $(0, 4, 1)^T$ und $(1, -5, 0)^T$. Daher ist $r(1) = 2$. Der verallgemeinerte Eigenraum zum Eigenwert 1 ist also

$$M_1(A) = \text{Kern}\,(A - 1 \cdot I)^2 = \text{span}\left\{ \begin{pmatrix} 1 \\ -5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix} \right\}.$$

□

Unter Benutzung dieser Hilfsmittel erhalten wir

**Satz 2.3** *Seien $\lambda_1, \ldots, \lambda_s$ die paarweise voneinander verschiedenen Eigenwerte von $A \in \mathbb{C}^{n\times n}$ und $M_{\lambda_1}(A), \ldots, M_{\lambda_s}(A)$ die zugehörigen verallgemeinerten Eigenräume. Die eindeutige Lösung der Anfangswertaufgabe*

$$x' = Ax, \qquad x(0) = x_0$$

*ist dann gegeben durch*

$$x(t) = \sum_{j=1}^{s} \Big( \sum_{k=0}^{r(\lambda_j)-1} (A - \lambda_j I)^k \frac{t^k}{k!} \Big) x_{0,j} e^{\lambda_j t},$$

*wobei $x_0 = \sum_{j=1}^{s} x_{0,j}$ mit $x_{0,j} \in M_{\lambda_j}(A)$, $j = 1, \ldots, s$.*

**Beweis:** Wegen $\mathbb{C}^n = M_{\lambda_1}(A) \oplus \cdots \oplus M_{\lambda_s}(A)$ existiert eine eindeutige Darstellung $x_0 = \sum_{j=1}^{s} x_{0,j}$ mit $x_{0,j} \in M_{\lambda_j}(A)$, $j = 1, \ldots, s$. Die Lösung von $x' = Ax$, $x(0) = x_0$, ist gegeben durch

$$\begin{aligned} x(t) &= e^{At}x_0 \\ &= \sum_{j=1}^{s} e^{At}x_{0,j} \\ &= \sum_{j=1}^{s} e^{(A-\lambda_j I)t}x_{0,j}e^{\lambda_j t} \\ &= \sum_{j=1}^{s}\Big(\sum_{k=0}^{\infty}(A-\lambda_j I)^k x_{0,j}\frac{t^k}{k!}\Big)e^{\lambda_j t} \\ &= \sum_{j=0}^{s}\Big(\sum_{k=0}^{r(\lambda_j)-1}(A-\lambda_j I)^k\frac{t^k}{k!}\Big)x_{0,j}e^{\lambda_j t}, \end{aligned}$$

da ja $(A - \lambda_j I)^k x_{0,j} = 0$ für alle $k \geq r(\lambda_j)$, womit die Behauptung bewiesen ist. □

**Beispiel:** Sei wieder

$$A := \begin{pmatrix} 0 & 1 & 0 \\ 4 & 3 & -4 \\ 1 & 2 & -1 \end{pmatrix},$$

ferner $x_0 := (0, 1, 2)^T$. Die voneinander verschiedenen Eigenwerte sind $\lambda_1 = 0$ und $\lambda_2 = 1$, die zugehörigen verallgemeinerten Eigenräume sind

$$M_{\lambda_1}(A) = \text{span}\,\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\}, \qquad M_{\lambda_2}(A) = \text{span}\left\{\begin{pmatrix} 1 \\ -5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix}\right\}.$$

Nun müssen wir $x_{0,1} \in M_{\lambda_1}(A)$ und $x_{0,2} \in M_{\lambda_2}(A)$ so bestimmen, dass $x_0 = x_{0,1} + x_{0,2}$. Hierzu lösen wir das lineare Gleichungssystem

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -5 & 4 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

und erhalten mit

```
with(LinearAlgebra):
C:=Matrix([[1,1,0],[0,-5,4],[1,0,1]]):
b:=Vector([0,1,2]):
LinearSolve(C,b);
```

die Lösung

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} -7 \\ 7 \\ 9 \end{pmatrix}.$$

Daher ist

$$x_{0,1} = -7\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -7 \\ 0 \\ -7 \end{pmatrix}, \qquad x_{0,2} = 7\begin{pmatrix} 1 \\ -5 \\ 0 \end{pmatrix} + 9\begin{pmatrix} 0 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \\ 9 \end{pmatrix}.$$

Daher ist die Lösung von $x' = Ax$, $x(0) = x_0$, gegeben durch

$$\begin{aligned} x(t) &= \begin{pmatrix} -7 \\ 0 \\ -7 \end{pmatrix} + \left[ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & 1 & 0 \\ 4 & 2 & -4 \\ 1 & 2 & -2 \end{pmatrix} t \right] \begin{pmatrix} 7 \\ 1 \\ 9 \end{pmatrix} e^t \\ &= \begin{pmatrix} -7 \\ 0 \\ -7 \end{pmatrix} + \left[ \begin{pmatrix} 7 \\ 1 \\ 9 \end{pmatrix} - \begin{pmatrix} 6 \\ 6 \\ 9 \end{pmatrix} t \right] e^t. \end{aligned}$$

□

Zum Schluss dieses Unterabschnitts wollen wir noch auf lineare Differentialgleichungen $n$-ter Ordnung mit konstanten Koeffizienten eingehen.

Gegeben sei der lineare Differentialoperator

$$Lx := \sum_{i=0}^{n} a_i x^{(i)} \qquad \text{mit } a_n := 1,$$

gesucht sei die allgemeine Lösung der homogenen Gleichung $Lx = 0$. Schreibt man diese Gleichung als System, so erhält man die Koeffizientenmatrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix},$$

das zugehörige charakteristische Polynom ist (man entwickele $\det(A - \lambda I)$ nach der letzten Zeile)

$$p_n(\lambda) := (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0).$$

Den folgenden Satz (siehe W. Walter (1993, S. 174)) könnte man aus der allgemeinen Theorie "herausholen", er soll aber direkt und elementar bewiesen werden.

**Satz 2.4** *Gegeben sei die Differentialgleichung*

$$(*) \qquad Lx := \sum_{i=0}^{n} a_i x^{(i)} = 0 \qquad \textit{mit} \quad a_n = 1.$$

*Ist $\lambda$ eine $k$-fache Nullstelle des charakteristischen Polynoms*

$$p_n(\lambda) := (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0),$$

*so entsprechen ihr $k$ Lösungen*

$$(**) \qquad e^{\lambda t},\ te^{\lambda t},\ \ldots,\ t^{k-1}e^{\lambda t}$$

*von $(*)$. Aus den $n$ Nullstellen (jede mit ihrer Vielfachheit gezählt) des charakteristischen Polynoms $p_n$ ergeben sich auf diese Weise $n$ linear unabhängige Lösungen.*

*Sind die $a_i$, $i = 0, \ldots, n-1$, reell, so erhält man reelle linear unabhängige Lösungen, indem man zu einer komplexen Nullstelle $\lambda = \mu + i\nu$ von $k$-ter Ordnung die $k$ Lösungen in $(**)$ in Real- und Imaginärteil aufspaltet:*

$$t^q e^{\mu t}\cos\nu t,\ t^q e^{\mu t}\sin\nu t \qquad (q = 0, \ldots, k-1)$$

*(und die $k$ zu $\bar{\lambda}$ gehörenden Lösungen streicht).*

**Beweis:** Sei $\lambda$ eine $k$-fache Nullstelle des charakteristischen Polynoms $p_n$. Dann ist $t^q e^{\lambda t}$, $q = 0, \ldots, k-1$, eine Lösung von $(*)$, da

$$L(t^q e^{\lambda t}) = L\Big(\frac{d^q}{d\lambda^q}e^{\lambda t}\Big) = \frac{d^q}{d\lambda^q}L(e^{\lambda t}) = (-1)^n\frac{d^q}{d\lambda^q}(e^{\lambda t}p_n(\lambda)) = 0.$$

Nun hat man noch zu zeigen, dass die so erhaltenen Lösungen linear unabhängig sind. Hierzu beweisen wir:

- *Sind $\lambda_1, \ldots, \lambda_m$ paarweise verschieden und $p_1, \ldots, p_m$ Polynome mit*

$$\sum_{i=1}^{m} p_i(t)e^{\lambda_i t} \equiv 0,$$

*so ist $p_i(t) \equiv 0$, $i = 1, \ldots, m$.*

Denn: Man beweist diese Aussage durch vollständige Induktion nach $m$. Für $m = 1$ ist sie offenbar richtig. Angenommen, sie sei für $m$ Summanden schon bewiesen und es sei

$$\sum_{i=1}^{m} p_i(t)e^{\lambda_i t} + p(t)e^{\lambda t} \equiv 0 \qquad \text{mit} \quad \lambda \neq \lambda_i \quad (i = 1, \ldots, m).$$

Eine Multiplikation mit $e^{-\lambda t}$ ergibt

$$\sum_{i=1}^{m} p_i(t)e^{\mu_i t} + p(t) \equiv 0 \qquad \text{mit} \quad \mu_i := \lambda_i - \lambda \neq 0 \quad (i = 1, \ldots, m).$$

Ist $p$ ein Polynom vom Grade $l$, so liefert $(l+1)$-maliges Differenzieren dieser Gleichung, dass

$$\sum_{i=1}^{m} q_i(t)e^{\mu_i t} \equiv 0,$$

wobei die $q_i$ gewisse Polynome und die $\mu_i$ paarweise voneinander verschieden sind. Aus der Induktionsannahme folgt $q_i(t) \equiv 0$, $i = 1, \ldots, m$. Dann sind aber auch die $p_i$ und

folglich auch $p$ das Nullpolynom! Dies erkennt man durch die folgende Überlegung: Ist $r$ ein Polynom, das nicht das Nullpolynom ist, und $\mu \neq 0$, so ist

$$\frac{d}{dt}[r(t)e^{\mu t}] = [r'(t) + \mu r(t)]e^{\mu t}.$$

Da $r \neq 0$ und $\mu \neq 0$, ist auch $q(t) := r'(t) + \mu r(t) \not\equiv 0$. Eine wiederholte Anwendung dieser Behauptung liefert dann die Behauptung. □

**Beispiel:** Gegeben sei die Differentialgleichung (siehe W. Walter (1993, S. 175))

$$x^{(5)} + 4x^{(4)} + 2x^{(3)} - 4x'' + 8x' + 16x = 0.$$

Das zugehörige charakteristische Polynom ist (bis auf den Faktor $(-1)^5 = -1$)

$$p(\lambda) := \lambda^5 + 4\lambda^4 + 2\lambda^3 - 4\lambda^2 + 8\lambda + 16\lambda = (\lambda + 2)^3((\lambda - 1 + i)(\lambda - 1 - i).$$

Ein reelles System linear unabhängiger Lösungen ist daher

$$e^{-2t},\ te^{-2t},\ t^2e^{-2t},\ e^t \sin t,\ e^t \cos t.$$

□

### 2.2.3 Periodische lineare Systeme

Den Fall linearer autonomer Systeme haben wir vollständig behandeln können. Der nächst einfache, und immer noch wichtige, Fall ist der eines linearen periodischen Differentialgleichungssystems.

Als Hilfsmittel für den folgenden Satz benötigen wir

**Lemma 2.5** *Sei $C \in \mathbb{C}^{n\times n}$ nichtsingulär. Dann gibt es eine Matrix $B \in \mathbb{C}^{n\times n}$ mit $C = e^B$.*

**Beweis:** Offenbar können wir o. B. d. A. annehmen, dass $C$ schon die Jordansche Normalform besitzt, also eine Block-Diagonalmatrix der Form $C = \text{diag}\,(C_1, \ldots, C_p)$ ist, wobei die Blöcke $C_j$ durch $C_j = \lambda_j I + R_j$ mit

$$R_j = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

gegeben sind. Da $C$ als nichtsingulär vorausgesetzt ist, ist $\lambda_j \neq 0$. Um das Lemma zu beweisen, genügt es zu zeigen, dass jedes $C_j$ in der Form $C_j = e^{B_j}$ geschrieben werden kann. Wir lassen daher jetzt den Index $j$ fort, nehmen also an, dass $C = \lambda I + R \in \mathbb{C}^{n\times n}$ mit $\lambda \in \mathbb{C} \setminus \{0\}$ und einer $n \times n$-Matrix $R$, deren Einträge alle Null sind bis auf Einsen in der oberen Nebendiagonalen. Man beachte, dass $R^k = 0$ für $k \geq n$. Nun setze man

$$B := (\ln \lambda)I + S$$

mit

$$S := -\sum_{j=1}^{n-1} \frac{(-1)^j R^j}{j\lambda^j} = \frac{R}{\lambda} - \frac{1}{2}\frac{R^2}{\lambda^2} + \frac{1}{3}\frac{R^3}{\lambda^3} - \cdots + (-1)^n \frac{1}{n-1}\frac{R^{n-1}}{\lambda^{n-1}}.$$

Die Motivation hierfür ist die folgende: Beachtet man, dass eine Matrix $B$ mit $C = e^B$ mit gutem Recht auch als $\ln C$ bezeichnet werden kann, so erhält man aus $C = \lambda(I + R/\lambda)$, dass $\ln C = (\ln \lambda)I + \ln(I + R/\lambda)$. Berücksichtigt man noch die bekannte Reihenentwicklung für $\ln(1+\cdot)$, so erscheint die obige Definition der Matrix $B$ sinnvoll zu sein. Nun müssen wir noch zeigen, dass wirklich $e^B = C$ gilt. Hierzu beachten wir, dass $\exp \ln(1 + x) = 1 + x$ und daher

$$\sum_{j=0}^{\infty} \frac{1}{j!}\Big(\sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k}\Big)^j = 1 + x$$

für $|x| < 1$. Weil $R^k = 0$ für $k \geq n$ und Potenzen von $R$ miteinander kommutieren, kann man $R/\lambda$ statt $x$ einsetzen, erhält $e^S = I + R/\lambda$ und damit die Behauptung. □

Das letzte Lemma ist entscheidendes Hilfsmittel für den folgenden Satz:

**Satz 2.6 (Floquet)** *Sei $X(\cdot)$ ein Fundamentalsystem von $x' = A(t)x$ mit stetigem und $T$-periodischem $A(\cdot)$. Dann ist $X(t) = P(t)e^{Bt}$, wobei $B$ eine konstante $n \times n$-Matrix und $P(\cdot)$ stetig und $T$-periodisch.*

**Beweis:** Sei $X(\cdot)$ ein Fundamentalsystem zu $x' = A(t)x$. Dann ist auch durch $Y(t) := X(t + T)$ ein Fundamentalsystem zu $x' = A(t)x$ gegeben, wie man aus

$$Y'(t) = X'(t + T) = A(t + T)X(t + T) = A(t)Y(t).$$

Daher gibt es eine nichtsinguläre Matrix $C$ mit $X(t+T) = X(t)C$. Wegen des vorigen Lemmas gibt es eine Matrix $B$ mit $C = e^{BT}$. Nun definiere man $P(t) := X(t)e^{-Bt}$. Dann ist einerseits natürlich $X(t) = P(t)e^{Bt}$ und andererseits

$$P(t + T) = X(t + T)e^{-B(t+T)} = X(t)e^{BT}e^{-B(t+T)} = P(t),$$

womit alles bewiesen ist. □

**Definition 2.7** Gegeben sei ein lineares System $x' = A(t)x$ mit stetigem, $T$-periodischem $A(\cdot)$. Ist $X(\cdot)$ hierzu ein Fundamentalsystem, so heißt jede nichtsinguläre Matrix $C$ mit $X(t + T) = X(t)C$ eine *Monodromie-Matrix* zu $x' = A(t)x$. Die Eigenwerte $\rho$ einer Monodromie-Matrix heißen *charakteristische Multiplikatoren*, ein $\lambda$ mit $\rho = e^{\lambda T}$ heißt *charakteristischer Exponent* zu $x' = A(t)x$.

**Bemerkung:** Wir wollen uns überlegen, dass alle Monodromie-Matrizen zu einem $T$-periodischem System $x' = A(t)x$ einander ähnlich sind, so dass die charakteristischen Multiplikatoren nicht von der Wahl des Fundamentalsystems abhängen. Denn sei $X$ ein Fundamentalsystem, $C$ eine zugehörige Monodromie-Matrix, also $X(t + T) = X(t)C$. Ist $Y$ ein weiteres Fundamentalsystem, so existiert eine nichtsinguläre Matrix $D$ mit $Y(t) = X(t)D$ und es ist

$$Y(t + T) = X(t + T)D = X(t)CD = X(t)D\,D^{-1}CD = Y(t)D^{-1}CD,$$

die Monodromie-Matrix zu $Y$ ist also $D^{-1}CD$, womit gezeigt ist, dass die Monodromie-Matrizen zu einem linearen, periodischen System einander ähnlich sind. Insbesondere merken wir uns: Ist $X(\cdot)$ ein durch $X(0) = I$ normiertes Fundamentalsystem, so ist $X(T)$ Monodromie-Matrix. □

**Bemerkung:** Gegeben sei die lineare, inhomogene Anfangswertaufgabe

$$x' = A(t)x + b(t), \qquad x(0) = x_0,$$

bei der $A(\cdot)$ und $b(\cdot)$ stetig und $T$-periodisch sind. Sei $X(\cdot)$ ein durch $X(0) = I$ normiertes Fundamentalsystem zu $x' = A(t)x$. Dieses lässt sich wegen des Floquetschen Satzes darstellen als $X(t) = P(t)e^{Bt}$, wobei $P(\cdot)$ stetig und $T$-periodisch. Dann ist

$$\begin{aligned} x(t) &= X(t)\Big[x_0 + \int_0^t X(s)^{-1}b(s)\,ds\Big] \\ &= P(t)e^{Bt}\Big[x_0 + \int_0^t e^{-Bs}P(s)^{-1}b(s)\,ds\Big] \\ &= P(t)\Big[e^{Bt}x_0 + \int_0^t e^{B(t-s)}P(s)^{-1}b(s)\,ds\Big]. \end{aligned}$$

Also ist $x(t) = P(t)y(t)$, wobei $y$ die Lösung von

$$y' = By + P(t)^{-1}b(t), \qquad y(0) = x_0$$

ist. □

### 2.2.4 Zweidimensionale autonome lineare Systeme

Wir betrachten ein lineares, homogenes, autonomes (also zeitunabhängiges) System von zwei Differentialgleichungen erster Ordnung, also die Differentialgleichung $x' = Ax$ mit der Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

die wir als reell voraussetzen. O. B. d. A. ist $\det(A) \neq 0$, da man andernfalls das System auf eine Gleichung erster Ordnung zurückführen kann. Denn ist $A$ singulär, so existiert ein $y \neq 0$ mit $Ay = 0$, dieser konstante Vektor ist eine Lösung von $x' = Ax$. Wir nehmen an, die erste Komponente $y_1$ von $y$ sei von Null verschieden. Für eine weitere Lösung von $x' = Ax$ machen wir den Ansatz

$$x(t) = \phi(t)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \psi(t) \end{pmatrix}$$

mit noch unbekannten skalaren Funktionen $\phi$ und $\psi$. Dann ist

$$\begin{aligned} x'(t) - Ax(t) &= \phi'(t)\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \psi'(t) \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ \psi(t) \end{pmatrix} \\ &= \begin{pmatrix} \phi'(t)y_1 - b\psi(t) \\ \phi'(t)y_2 + \psi'(t) - d\psi(t) \end{pmatrix}. \end{aligned}$$

Daher ist $x$ eine Lösung, wenn

$$\phi'(t) = \frac{b}{y_1}\psi(t), \qquad \psi'(t) + \Big(b\frac{y_2}{y_1} - d\Big)\psi(t) = 0.$$

Nun ist es hiermit leicht möglich, eine weitere (von $y$ linear unabhängige) Lösung zu finden.

Wir setzen

$$D := \det(A) = ad - bc, \qquad S := \operatorname{tr}(A) = a + d.$$

Die beiden Eigenwerte von $A$ sind dann gegeben durch

$$\lambda_{1,2} = \frac{1}{2}\Big(S \pm \sqrt{S^2 - 4D}\Big).$$

Wir machen eine Fallunterscheidung.

(I) Die beiden Eigenwerte $\lambda_1, \lambda_2$ sind reell und voneinander verschieden. Es sei etwa $\lambda_2 < \lambda_1$. Mit $v^1, v^2$ seien auf die (euklidische) Länge 1 normierte Eigenvektoren von $A$ zu den Eigenwerten $\lambda_1, \lambda_2$ bezeichnet. Die allgemeine reelle Lösung zu $x' = Ax$ ist

$$x(t) = c_1 e^{\lambda_1 t} v^1 + c_2 e^{\lambda_2 t} v^2$$

mit beliebigen reellen Konstanten $c_1, c_2$. Sei $L_i := \operatorname{span}\{v_i\}$, $i = 1, 2$.

Hier unterscheiden wir drei Fälle.

(a) Es ist $\lambda_2 < \lambda_1 < 0$ (Stabiler Knoten).

Der Nullpunkt 0 (eine triviale Lösung) ist *asymptotisch stabil* in dem Sinne, dass jede Lösung für $t \to \infty$ gegen den Nullpunkt strebt. In Abbildung 2.1 geben wir zwei Beispiele stabiler Knoten an. Links ist das Richtungsfeld für $A = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}$ eingetragen, rechts für die hierzu ähnliche Matrix $A = \begin{pmatrix} -4 & 2 \\ -3 & 1 \end{pmatrix}$.

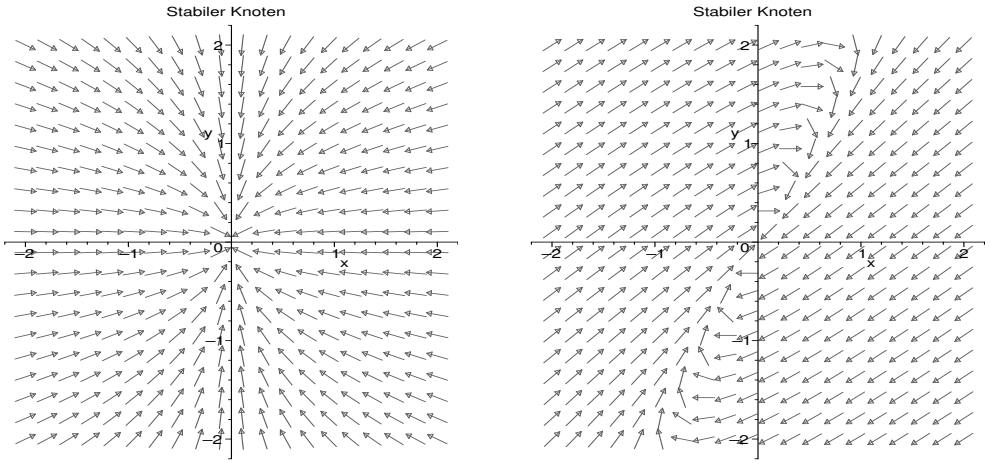


Abbildung 2.1: Zwei Beispiele stabiler Knoten

(b) Es ist $0 < \lambda_2 < \lambda_1$ (Instabiler Knoten).

Der Ursprung ist instabil insofern, dass jede außerhalb des Nullpunkts startende Bahn unbeschränkt ist. Auch hierfür geben wir in Abbildung 2.2 zwei Beispiele an. Links ist

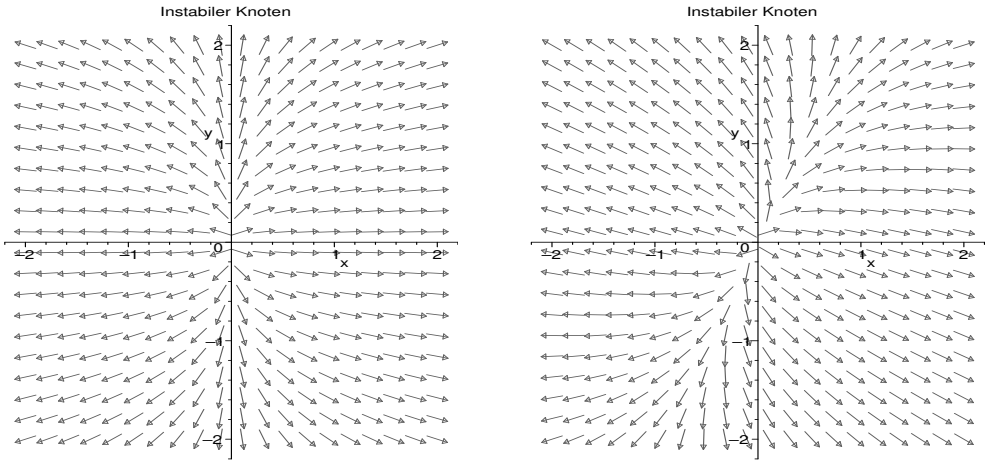


Abbildung 2.2: Zwei Beispiele instabiler Knoten

$A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$, rechts ist $A = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$.

(c) Es ist $\lambda_2 < 0 < \lambda_1$ (Sattelpunkt).

Die Bahnen, die auf $L_2$ starten, streben mit $t \to \infty$ gegen 0, alle anderen Bahnen sind für $t \to \infty$ unbeschränkt. In Abbildung 2.3 geben wir zwei Beispiele hierfür an. Links

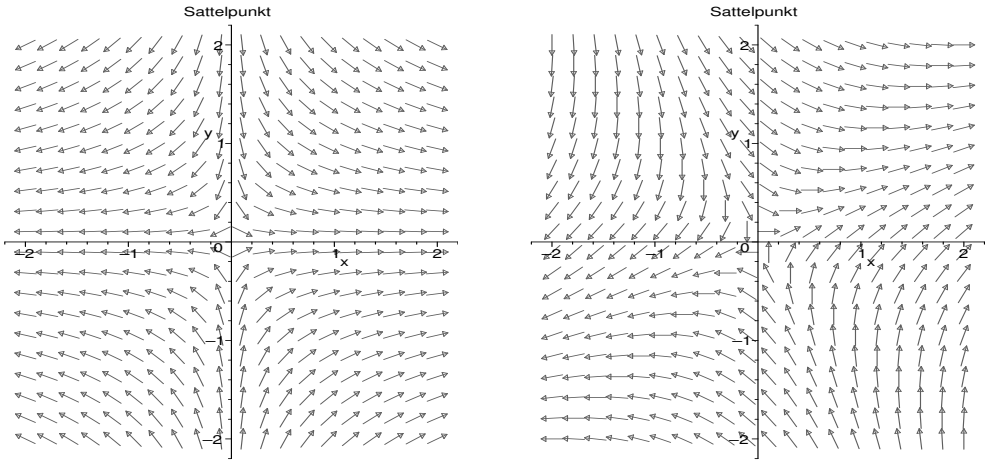


Abbildung 2.3: Zwei Beispiele von Sattelpunkten

ist $A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$, rechts ist $A = \begin{pmatrix} 5 & 1 \\ 1 & -4 \end{pmatrix}$.

(II) Die beiden Eigenwerte $\lambda_1$ und $\lambda_2$ sind komplex.

Da $A$ reell ist, sind $\lambda_1$ und $\lambda_2$ konjugiert komplex, etwa $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$ mit reellen $\alpha, \beta$ und $\beta > 0$. Die Eigenvektoren $v^1, v^2$ können als konjugiert komplex gewählt werden. Die allgemeine reelle Lösung ist

$$x(t) = c_1 e^{(\alpha+i\beta)t} v^1 + \overline{c_1} e^{(\alpha-i\beta)t} \overline{v^1} = 2\Re(c_1 e^{(\alpha+i\beta)t} v^1),$$

wobei $c_1$ eine beliebige komplexe Zahl ist. Sei $v^1 = u + iw$ mit reellen $u, w \in \mathbb{R}^2$ und $c_1 = \gamma e^{i\delta}$ die Polardarstellung von $c_1 \in \mathbb{C}$. Dann ist

$$x(t) = 2\gamma e^{\alpha t} \left[u \cos(\beta t + \delta) - w \sin(\beta t + \delta)\right].$$

Wir unterscheiden jetzt wieder drei Fälle.

(a) Es ist $\lambda_1 = i\beta$, $\lambda_2 = -i\beta$ (Wirbelpunkt).

Die allgemeine Lösung ist

$$x(t) = \gamma \left[u \cos(\beta t + \delta) - w \sin(\beta t + \delta)\right],$$

wobei $a, \delta$ beliebige reelle Konstanten und $u, w$ wie oben sind. Die Bahnen sind geschlossene Kurven und jede Lösung ist $2\pi/\beta$-periodisch.

Als Beispiel betrachten wir $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Hier ist $\lambda_{1,2} = \pm i$, ferner ist

$$v^1 = \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + i \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u + iw.$$

In Abbildung 2.4 links wird das zugehörige Richtungsfeld gezeichnet. Würde man für

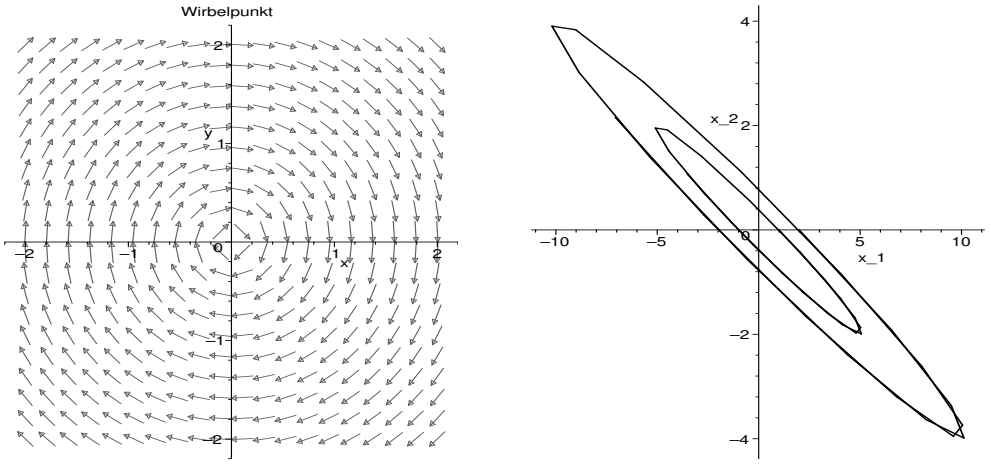


Abbildung 2.4: Wirbelpunkt

$A := \begin{pmatrix} 5 & 13 \\ -2 & -5 \end{pmatrix}$ (auch diese Matrix hat die Eigenwerte $\lambda_{1,2} = \pm i$) das Richtungsfeld wie bisher zeichnen, also etwa durch

```
with(DEtools):
dfieldplot([diff(x_1(t),t)=5*x_1(t)+13*x_2(t),
diff(x_2(t),t)=-2*x_1(t)-5*x_2(t)],
[x_1(t),x_2(t)],t=0..10,x_1=-2..2,x_2=-2..2,arrows=MEDIUM);
```

so würde man gar nichts erkennen, schon gar nicht geschlossene Bahnen. Das liegt daran, dass diese "flache" Ellipsen sind. Daher haben wir in Abbildung 2.4 rechts zwei Phasenbahnen eingetragen. Diese wurden durch

```
phaseportrait([diff(x_1(t),t)=5*x_1(t)+13*x_2(t),
diff(x_2(t),t)=-2*x_1(t)-5*x_2(t)],[x_1(t),x_2(t)],t=0..10,
[[x_1(0)=1,x_2(0)=0],[x_1(0)=2,x_2(0)=0]],arrows=NONE,linecolor=black);
```

erzeugt, wobei natürlich `with(DEtools):` vorangegangen ist.

(b) Es ist $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$ mit $\alpha < 0$ (Stabiler Spiralpunkt).

Der Ursprung ist *asymptotisch stabil*, da jede Lösung mit $t \to \infty$ gegen den Nullpunkt konvergieren, die Bahnen sind Spiralen. Siehe Abbildung 2.5.

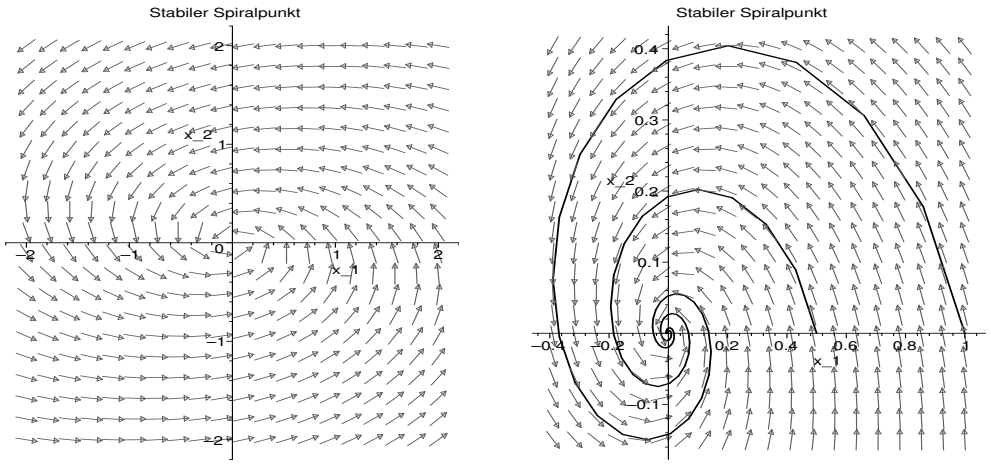


Abbildung 2.5: Stabiler Spiralpunkt

(c) Es ist $\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$ mit $\alpha > 0$ (Instabiler Spiralpunkt).

Der Ursprung ist instabil. Für $A = \begin{pmatrix} 1 & 5 \\ -2 & 1 \end{pmatrix}$ ist z. B. $\lambda_{1,2} = 1 \pm \sqrt{10}i$. In Abbildung 2.6 geben wir links ein Richtungsfeld, rechts zwei Phasenbahnen an.

(III) Die beiden Eigenwerte sind gleich (und reell), $\lambda_1 = \lambda_2 = \lambda$.

Falls zwei linear unabhängige Eigenvektoren $v^1, v^2$ zum Eigenwert $\lambda$ existieren ($\lambda$ also einfache Elementarteiler hat), so ist die allgemeine Lösung $x(t) = (c_1 v^1 + c_2 v^2)e^{\lambda t}$. Die Bahnen liegen auf Geraden, wobei zu unterscheiden ist, ob $\lambda < 0$ (stabiler Fall) oder $\lambda > 0$ (instabiler Fall), siehe Abbildung 2.7. Hier ist links $A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, rechts

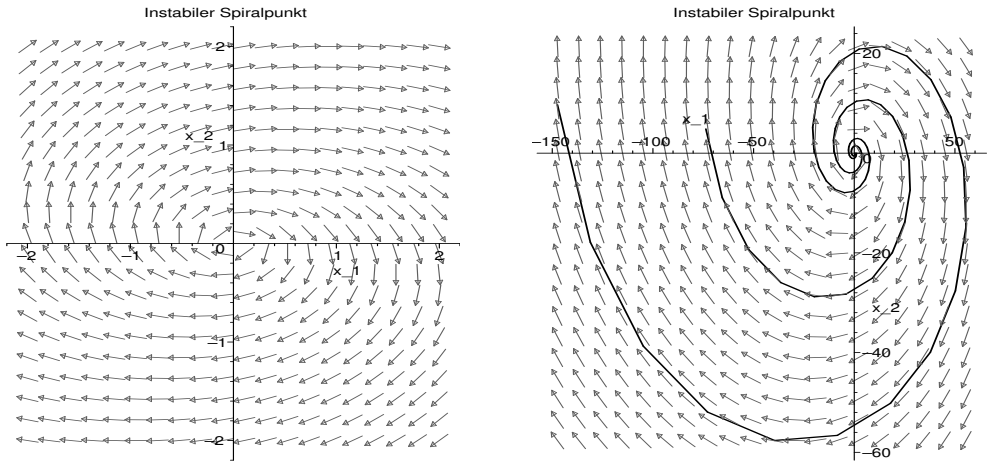


Abbildung 2.6: Instabiler Spiralpunkt

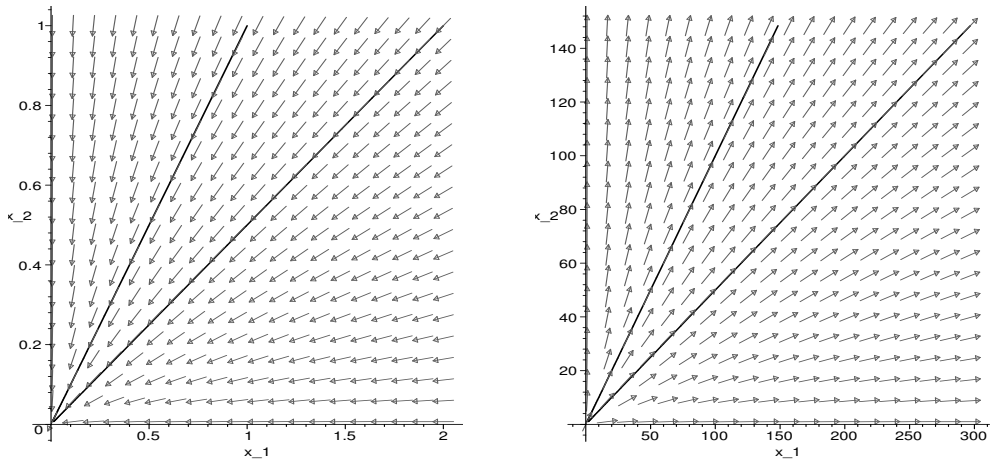


Abbildung 2.7: $\lambda_1 = \lambda_2 < 0$, $A$ diagonalisierbar

ist $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Gibt es dagegen keine zwei linear unabhängigen Eigenvektoren zum Eigenwert $\lambda$, so ist die allgemeine Lösung gegeben durch

$$x(t) = (c_1 + c_2 t)e^{\lambda t}v^1 + c_2 e^{\lambda t}v^2,$$

wobei $v^1$ ein Eigenvektor zum Eigenwert $\lambda$ und $v^2$ hiervon linear unabhängig ist. In Abbildung 2.8 verdeutlichen wir die Bahnen für den stabilen Fall ($\lambda < 0$).

Damit ist der zweidimensionale autonome Fall vollständig durchdiskutiert.

## 2.2.5 Aufgaben

1. Man bestimme mit Hilfe von Maple sämtliche Lösungen von

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} (3t-1) & -(1-t) \\ -(t+2) & (t-2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} te^{t^2} \\ -e^{t^2} \end{pmatrix}.$$

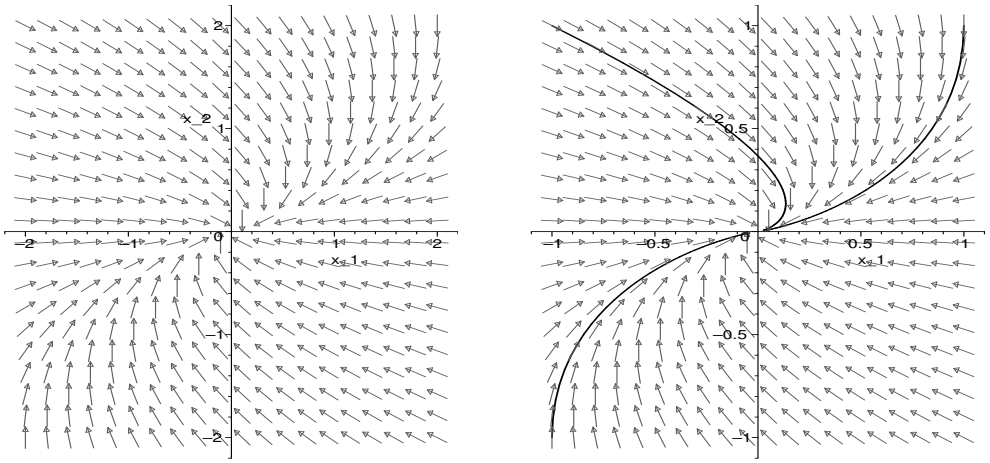


Abbildung 2.8: $\lambda_1 = \lambda_2 < 0$, $A$ nicht diagonalisierbar

2. Sei

$$A := \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix}$$

mit reellem, positiven $\omega$. Man berechne (wie auch immer) $e^{At}$, ferner gebe man eine Darstellung der Lösung von

$$\text{(P)} \qquad x'' + \omega^2 x = g(t), \qquad x(0) = x_0, \quad x'(0) = x_0'$$

an, wobei $x_0, x_0'$ gegebene reelle Zahlen sind und $g(\cdot)$ auf $\mathbb{R}$ stetig ist.

3. Für[7] jede der folgenden Matrizen $A$ berechne man das Fundamentalsystem $e^{At}$ zu $x' = Ax$, wobei Maple benutzt werden darf.

   (a)
   $$A := \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix},$$

   (b)
   $$A := \begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix},$$

   (c)
   $$A := \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

   (d)
   $$A := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

[7] Diese sehr einfache Übungsaufgabe haben wir

D. Betounes (2001) *Differential Equations: Theory and Applications with Maple.* Springer-Verlag, New York-Berlin-Heidelberg

entnommen.

4. Für zwei durch eine Feder gekoppelte Pendel gleicher Masse $m = 1$ und gleicher Länge $l$ lauten die Bewegungsgleichungen

$$\begin{aligned} \ddot{x} &= -\alpha x - k(x-y) \\ \ddot{y} &= -\alpha y - k(y-x), \end{aligned}$$

wobei $g$ die Erdbeschleunigung und $k$ die (positive) Federkonstante bedeuten und $\alpha := g/l$ gesetzt ist. Schreibt man die beiden Differentialgleichungen zweiter Ordnung als ein System von vier Differentialgleichungen erster Ordnung, so erhält man ein homogenes System mit der Koeffizientenmatrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -(\alpha+k) & 0 & k & 0 \\ 0 & 0 & 0 & 1 \\ k & 0 & -(\alpha+k) & 0 \end{pmatrix}.$$

Man bestimme ein zu $x' = Ax$ gehörendes (nicht notwendig normiertes) Fundamentalsystem. Ferner löse man die Anfangswertaufgabe für den Fall, dass zur Zeit $t = 0$ ein Pendel angestoßen wird bzw. die Anfangswerte $x(0) = y(0) = y'(0) = 0$, $x'(0) = 1$ vorgegeben werden.

5. Man bestimme (wie auch immer)[8] ein reelles Fundamentalsystem von Lösungen der Differentialgleichungssysteme $x' = Ax$ mit

$$A := \begin{pmatrix} 3 & 6 \\ -2 & -3 \end{pmatrix}, \qquad A := \begin{pmatrix} 8 & 1 \\ -4 & 4 \end{pmatrix}.$$

6. Sei

$$A := \begin{pmatrix} 0 & 1 & 0 \\ 4 & 3 & -4 \\ 1 & 2 & -1 \end{pmatrix}.$$

Man berechne $e^A$ und vergleiche mit $\sum_{j=0}^{10} A^j/j!$. Man mache hierbei möglichst viele der Rechnungen mit Maple oder MATLAB.

7. Man bestimme die allgemeine Lösung von

$$x'' - 6x' + 25x = e^{2t}.$$

Anschließend bestimme man die Lösung zu den Anfangswerten $x(0) = 1$, $x'(0) = 0$.

8. Sei

$$A(t) := \begin{pmatrix} 0 & 1 \\ -\sin t & \cos t \end{pmatrix}.$$

Man definiere $\phi(t) := \int_0^t e^{-\sin s}\, ds$ und zeige, dass durch

$$X(t) := \begin{pmatrix} e^{\sin t} & e^{\sin t}\phi(t) \\ \cos t e^{\sin t} & 1 + \cos t e^{\sin t}\phi(t) \end{pmatrix}$$

ein Fundamentalsystem zu $x' = A(t)x$ gegeben ist. Anschließend bestimme man die charakteristischen Multiplikatoren, d. h. die Eigenwerte von $C := X(0)^{-1}X(2\pi)$.

[8] Die Aufgabe ist W. Walter (1996, S. 159) entnommen.

## 2.3 Stabilität

### 2.3.1 Definitionen

In diesem Abschnitt behandeln wir eines der für die Anwendungen wichtigsten Probleme, nämlich des Stabilitätsproblem. Eine Anfangsaufgabe $x' = f(t,x)$, $x(t_0) = x_0$, ist i. allg. das mathematische Modell eines realen Prozesses. Ihre Lösung $x(\cdot; t_0, x_0)$ möge für $t \geq t_0$ existieren. Diese Lösung wird man, zunächst vage gesprochen, *stabil* nennen, wenn kleine Änderungen im Anfangszustand für alle $t \geq t_0$, also auch in der fernen Zukunft, nur kleine Änderungen in der Lösung nach sich ziehen. Sie wird *asymptotisch stabil* heißen, wenn die Lösung zu "etwas" verändertem Anfangswert sich im Laufe der Zeit, also für $t \to \infty$, der gegebenen Lösung wieder annähert.

**Beispiel:** Die Anfangswertaufgabe

$$x' = \lambda x, \qquad x(0) = 0$$

hat bei gegebenem $\lambda \in \mathbb{C}$ die Lösung $x(t; 0, 0) \equiv 0$. Die Lösung von

$$x' = \lambda x, \qquad x(0) = x_0$$

ist $x(t; 0, x_0) = e^{\lambda t} x_0$. Daher ist $x(t) \equiv 0$ stabil, wenn $\Re(\lambda) \leq 0$, und asymptotisch stabil, wenn $\Re(\lambda) < 0$, wobei mit $\Re(\lambda)$ der Realteil von $\lambda$ bezeichnet werde. □

Durch eine Variablentransformation kann man die Stabilität einer bestimmten Lösung von $x' = f(t,x)$, $x(t_0) = x_0$, auf die Untersuchung einer trivialen Lösung auf Stabilität zurückführen. Denn ist $x'(t) = f(t, x(t))$ auf $[t_0, \infty)$ und $x(t_0) = x_0$, so definiere man

$$g(t,y) := f(t, x(t) + y) - x'(t) = f(t, x(t) + y) - f(t, x(t))$$

und betrachte die Anfangswertaufgabe

$$y' = g(t,y), \qquad y(t_0) = 0$$

mit der trivialen Lösung $y(t) \equiv 0$. Wir gehen daher im folgenden von einer Gleichung $x' = f(t,x)$ mit $f(t,0) \equiv 0$ aus und nehmen an, $f$ sei so glatt, dass durch jeden Punkt $(t_0, x_0)$ genau eine Lösung geht, die auf $[t_0, \infty)$ existiert und mit $x(\cdot; t_0, x_0)$ bezeichnet wird.

**Definition 3.1** Die Lösung $x = 0$ von $x' = f(t,x)$ mit $f(t,0) \equiv 0$ heißt (*Lyapunov*)-*stabil* auf dem Intervall $I = [\tau, \infty)$, falls es zu jedem $\epsilon > 0$ und jedem $t_0 \in I$ ein $\delta = \delta(\epsilon, t_0) > 0$ gibt mit

$$\|x_0\| \leq \delta \Longrightarrow \|x(t; t_0, x_0)\| \leq \epsilon \quad \text{für alle } t \geq t_0.$$

Die Lösung $x = 0$ heißt *asymptotisch stabil* auf $I = [\tau, \infty)$, wenn sie stabil auf $I$ ist und es zu jedem $t_0 \in I$ ein $b = b(t_0)$ gibt mit

$$\|x_0\| \leq b \Longrightarrow \lim_{t \to \infty} x(t; t_0, x_0) = 0.$$

Sieht man sich noch einmal das obige simple Beispiel an, so erkennt man, dass auch mit der exakten Stabilitätsdefinition die dort gemachten Aussagen richtig sind.

Die beiden oben angegebenen Stabilitätsbegriffe sind die wichtigsten, auch wenn sie nicht für alle Anwendungen ausreichen. Es gibt sehr viele Stabilitätsbegriffe (z. B. findet man bei N. Rouche, J. Mawhin (1980) schon auf S. 10 die Begriffe asymptotically stable, equi-asymptotically stable, uniformly asymptotically stable, globally asymptotically stable, uniformly globally asymptotically stable), auf diese Art von Feinheiten wollen wir hier nicht eingehen.

Nun interessieren natürlich hinreichende, und möglichst auch notwendige Stabilitätbedingungen.

### 2.3.2 Stabilität bei linearen Systemen mit konstanten Koeffizienten

Der entscheidende Satz zur Stabilität der Nulllösung des linearen homogenen Systems $x' = Ax$ mit $A \in \mathbb{R}^{n\times n}$ ist die folgende Aussage. Hierbei gehen wir davon aus, dass das Intervall $I$ , auf dem die Stabilität untersucht wird, durch $I = [0, \infty)$ gegeben ist.

**Satz 3.2** *Die triviale Lösung $x = 0$ ist eine stabile Lösung von $x' = Ax$ genau dann, wenn*

*(a) $\Re(\lambda) \leq 0$ für alle Eigenwerte $\lambda$ von $A$,*

*(b) Ist $\lambda$ ein Eigenwert von $A$ mit $\Re(\lambda) = 0$, so hat $\lambda$ einfache Elementarteiler.*

*Die triviale Lösung $x = 0$ ist genau dann eine asymptoyisch stabile Lösung von $x' = Ax$, wenn $\Re(\lambda) < 0$ für alle Eigenwerte $\lambda$ von $A$.*

**Beweis:** Wir überlegen uns zunächst:

(1) Genau dann ist $x = 0$ eine stabile Lösung von $x' = Ax$, wenn eine Konstante $K > 0$ mit $\|e^{At}\| \leq K$ für alle $t \geq 0$ existiert.

(2) Genau dann ist $x = 0$ eine asymptotisch stabile Lösung von $x' = Ax$, wenn $\lim_{t\to\infty} e^{At} = 0$.

Denn: Zum Beweis von (1) nehmen wir zunächst an, $x = 0$ sei stabil. Nach Definition existiert zu $\epsilon = 1$ und $t_0 = 0$ ein $\delta > 0$ mit

$$\|x_0\| \leq \delta \Longrightarrow \|e^{At}x_0\| \leq 1 \quad \text{für alle } t \geq 0.$$

Hieraus folgt aber, dass

$$\|e^{At}\| = \max_{\|y_0\|=1} \|e^{At}y_0\| \leq \frac{1}{\delta} =: K.$$

Umgekehrt existiere eine Konstante $K > 0$ mit $\|e^{At}\| \leq K$ für alle $t \geq 0$. Zum Nachweis der Stabilität von $x = 0$ gebe man sich $\epsilon > 0$ und $t_0 \geq 0$ vor. Man setze $\delta := \epsilon/K$. Ist dann $\|x_0\| \leq \delta$, so ist

$$\|x(t; t_0, x_0)\| = \|e^{A(t-t_0)}x_0\| \leq K \|x_0\| \leq \epsilon \quad \text{für alle } t \geq t_0,$$

womit die Stabilität von $x = 0$ bewiesen ist.

Zum Nachweis von (2) sei zunächst vorausgesetzt, dass $x = 0$ asymptotisch stabile Lösung von $x' = Ax$ ist. Angenommen $e^{At} \not\to 0$ mit $t \to \infty$. Dann existiert eine Folge $\{t_k\} \subset \mathbb{R}_+$ mit $t_k \to \infty$ und ein $\epsilon > 0$ mit $\|e^{At_k}\| \geq \epsilon$ für alle $k \in \mathbb{N}$. Dann ist

$$\epsilon \leq \|e^{At_k}\| = \max_{\|x_0\|=1} \|e^{At_k} x_0\| = \|e^{At_k} x_{0,k}\|$$

mit $\|x_{0,k}\| = 1$. O. B. d. A. konvergiert die Folge $\{x_{0,k}\}$, etwa gegen ein $x_0$ (und es ist natürlich $\|x_0\| = 1$). Wegen der vorausgesetzten asymptotischen Stabilität gilt $e^{At_k} x_0 \to 0$ und wegen (1) ist $\|e^{At_k}\| \leq K$ mit einer gewissen Konstanten $K > 0$. Dann ist schließlich

$$\|e^{At_k} x_0\| \geq \|e^{At_k} x_{0,k}\| - \|e^{At_k}(x_{0,k} - x_0)\| \geq \epsilon - K \underbrace{\|x_{0,k} - x_0\|}_{\to 0},$$

ein Widerspruch. Dass umgekehrt aus $\lim_{t\to\infty} e^{At} = 0$ die asymptotische Stabilität folgt, ist trivial.

Nun zum Beweis der Aussagen des Satzes. Sei $x = 0$ eine stabile Lösung von $x' = Ax$ und $\lambda$ ein Eigenwert von $A$. Dann ist $x(t) := ce^{\lambda t}$, wobei $c$ ein zu $\lambda$ gehörender Eigenvektor von $A$ ist, eine Lösung von $x' = Ax$. Wäre $\Re(\lambda) > 0$, so wäre diese nicht beschränkt, was ein Widerspruch zu (1) wäre. Sei $\lambda$ ein Eigenwert von $A$ mit $\Re(\lambda) = 0$. Im Widerspruch zur Behauptung nehmen wir an, $\operatorname{Kern}(A - \lambda I)$ sei ein echter Teilraum des verallgemeinerten Eigenraumes $M_\lambda(A)$ von $A$ zum Eigenwert $\lambda$ (bzw. in einer früheren Sprechweise $r(\lambda) \geq 2$). Dann liefert Satz 2.3, dass eine Lösung der Form $te^{\lambda t}c$ mit $c \neq 0$ existiert. Diese wäre nicht beschränkt, was wiederum einen Widerspruch bedeutet. Umgekehrt sei nun $\Re(\lambda) \leq 0$ für alle Eigenwerte $\lambda$ von $A$, diejenigen Eigenwerte $\lambda$ mit $\Re(\lambda) = 0$ mögen einfache Elementarteiler besitzen (d. h. es sei $r(\lambda) = 1$). Aus Satz 2.3 folgt die Beschränktheit jeder Lösung von $x' = Ax$ und hieraus folgt nach (1) die Stabilität.

Sei nun $x = 0$ eine asymptotisch stabile Lösung von $x' = Ax$ und $\lambda$ ein Eigenwert von $A$. Wegen des gerade bewiesenen ersten Teiles des Satzes ist $\Re(\lambda) \leq 0$. Wäre $\Re(\lambda) = 0$, so existierte eine Lösung, die nicht mit $t \to \infty$ gegen 0 konvergiert, ein Widerspruch zu $\lim_{t\to\infty} e^{At} = 0$. Ist umgekehrt $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$, so folgt wiederum aus Satz 2.3, dass jede Lösung mit $t \to \infty$ gegen 0 strebt bzw. $x = 0$ asymptotisch stabil ist. □

**Bemerkung:** Geht man von einer Differentialgleichung $n$-ter Ordnung

$$x^{(n)} + a_{n-1}x^{(n-1)} + \cdots + a_1 x' + a_0 x = 0$$

aus, so ist die Nulllösung nach Satz 3.2 genau dann stabil (bzw. asymptotisch stabil), wenn für alle Nullstellen $\lambda$ von

$$q_n(\lambda) := \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

gilt, dass $\Re(\lambda) \leq 0$ und sogar $\Re(\lambda) < 0$, falls $\lambda$ eine mehrfache Nullstelle ist (bzw. $\Re(\lambda) < 0$).

Ist z. B. $n = 2$, also $q_2(\lambda) = \lambda^2 + a_1\lambda + a_0$, so hat man als Nullstellen

$$\lambda_{1,2} = -\frac{a_1}{2} \pm \sqrt{-a_0 + \frac{a_1^2}{4}}.$$

Als notwendige und hinreichende Bedingung für asymptotische Stabilität erhält man $a_0 > 0$ und $a_1 > 0$ (was positiver Rückstellkraft und positiver Dämpfung entspricht). Für allgemeines $n$ gibt es algebraische Beziehungen, die sogenannten Routh-Hurwitz-Bedingungen, die notwendig und hinreichend dafür sind, dass $\Re(\lambda) < 0$ für jede Nullstelle $\lambda$ des charakteristischen Polynoms $q_n(\cdot)$. Eine verhältnismäßig einfache Notwendige Bedingung wird in Aufgabe 3 angegeben. □

### 2.3.3 Stabilität bei linearen Systemen mit variablen Koeffizienten

In diesem kurzen Unterabschnitt wollen wir nur zwei Ergebnisse über die Stabilität der Nulllösung bei linearen Systemen mit variablen Koeffizienten formulieren und beweisen. In beiden Sätzen ist das lineare System mit variablen Koeffizienten nur eine kleine Störung eines (stabilen bzw. asymptotisch stabilen) linearen Systems mit konstanten Koeffizienten.

**Satz 3.3** *Die Nulllösung $x = 0$ von $x' = (A + B(t))x$ ist stabil, falls die beiden folgenden Bedingungen erfüllt sind:*

*(a) Es ist $A$ stabil, d. h. $\Re(\lambda) \le 0$ für jeden Eigenwert $\lambda$ von $A$, ferner besitzen Eigenwerte $\lambda$ von $A$ mit $\Re(\lambda) = 0$ einfache Elementarteiler.*

*(b) $B(\cdot)$ ist stetig und es ist $\int_0^\infty \|B(t)\|\, dt < \infty$.*

**Beweis:** Sei $x(t) = x(t; t_0, x_0)$ die Lösung von

$$x' = (A + B(t))x, \qquad x(t_0) = x_0,$$

wobei $t_0 \ge 0$ vorausgesetzt wird. Dann ist

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}B(s)x(s)\, ds.$$

Da $A$ stabil ist, existiert eine Konstante $K > 0$ mit $\|e^{At}\| \le K$ für alle $t \ge 0$. Daher ist

$$\|x(t)\| \le K\,\|x_0\| + K \int_{t_0}^{t} \|B(s)\|\,\|x(s)\|\, ds \qquad \text{für alle } t \ge t_0.$$

Wegen der Gronwallschen Ungleichung folgt

$$\begin{aligned} \|x(t)\| &\le K\,\|x_0\| \exp\Big(K \int_{t_0}^{t} \|B(s)\|\, ds\Big) \\ &\le K\,\|x_0\| \exp\Big(K \int_{0}^{\infty} \|B(s)\|\, ds\Big) \end{aligned}$$

für alle $t \geq t_0$, woraus offenbar die behauptete Stabilität folgt. □

**Bemerkung:** Wir wissen: Ist $A$ asymptotisch stabil, d. h. $\Re(\lambda) < 0$ für alle Eigenwerte $\lambda$ von $A$, so gilt $e^{At} \to 0$ für $t \to \infty$. Man kann aber noch etwas mehr aussagen: Ist für jeden Eigenwert $\lambda$ von $A$ sogar $\Re(\lambda) < -\alpha$ mit einer positiven Konstanten $\alpha$, so existiert eine positive Konstante $c$ mit $\|e^{At}\| \leq c\, e^{-\alpha t}$ für alle $t \geq 0$. Denn beim Beweis kann man sich offenbar darauf beschränken, dass $A$ schon Jordan'sche Normalform hat. Dann ist die Behauptung aber leicht einzusehen. Siehe W. Walter (1993, S. 256). □

**Satz 3.4** *Die Nulllösung $x = 0$ von $x' = (A + B(t))x$ ist asymptotisch stabil, falls die folgenden drei Bedingungen erfüllt sind:*

*(a) $A$ ist asymptotisch stabil, d. h. es existiert eine Konstante $\alpha > 0$ mit $\Re(\lambda) < -\alpha$ für alle Eigenwerte $\lambda$ von $A$.*

*(b) Es existieren $\gamma$, $\tau \geq 0$ mit*

$$\int_{t_0}^{t} \|B(s)\|\, ds \leq \gamma\,(t - t_0) + \tau \qquad \text{für alle } t \geq t_0 \geq 0.$$

*(c) Die Konstante $\gamma$ ist hinreichend klein, genauer ist $0 \leq \gamma < \alpha/c$, wobei (siehe obige Bemerkung) $\|e^{At}\| \leq c\, e^{-\alpha t}$ für alle $t \geq 0$.*

**Beweis:** Sei wiederum $x(t) = x(t; t_0, x_0)$ die Lösung von

$$x' = (A + B(t))x, \qquad x(t_0) = x_0,$$

also

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}B(s)x(s)\, ds,$$

woraus wir mit obiger Bemerkung

$$\|x(t)\| \leq c\, e^{-\alpha(t-t_0)}\, \|x_0\| + c \int_{t_0}^{t} e^{-\alpha(t-s)}\, \|B(s)\|\, \|x(s)\|\, ds$$

für alle $t \geq t_0$ erhalten. Setzt, man $\phi(t) := e^{\alpha t}\, \|x(t)\|$, so bedeutet diese Ungleichung, daß

$$\phi(t) \;\leq c\, \phi(t_0) + \int_{t_0}^{t} c\, \|B(s)\|\, \phi(s)\, ds$$

für alle $t \geq t_0$. Eine Anwendung der Gronwallschen Ungleichung ergibt

$$\begin{aligned} \phi(t) &\leq c\, \phi(t_0)\, \exp\Big(c \int_{t_0}^{t} \|B(s)\|\, ds\Big) \\ &\leq c_1\, \phi(t_0) e^{c\gamma(t-t_0)} \qquad \text{mit } c_1 := ce^{c\tau} \end{aligned}$$

für alle $t \geq t_0$. Folglich ist

$$\|x(t)\| \leq c_1\, \|x_0\|\, e^{-(\alpha - c\gamma)(t-t_0)} \qquad \text{für alle } t \geq t_0 \geq 0,$$

woraus wegen $0 \leq \gamma < \alpha/c$ folgt, dass $\lim_{t\to\infty} x(t) = 0$ bzw. die Nulllösung asymptotisch stabil ist. □

### 2.3.4 Periodische lineare Systeme

Wir betrachten ein lineares System $x' = A(t)x$ mit stetigem, $T$-periodischem $A\colon \mathbb{R} \longrightarrow \mathbb{R}^{n\times n}$. Man könnte auf die Idee kommen, dass die Eigenwerte von $A(t)$ dieselbe Rolle spielen wie bei Systemen mit konstanten Koeffizienten. Dies ist nicht der Fall, wie das folgende Beispiel zeigt:

$$A(t) := \begin{pmatrix} -1 + \frac{3}{2}\cos^2 t & 1 - \frac{3}{2}\cos t \sin t \\ -1 - \frac{3}{2}\sin t \cos t & -1 + \frac{3}{2}\sin^2 t \end{pmatrix}$$

($A(\cdot)$ ist $T := \pi$-perodisch) mit den Eigenwerten $\lambda_{1,2}(t) = (-1 \pm i\sqrt{7})/4$. Der Realteil ist jeweils negativ, trotzdem existiert eine für $t \to \infty$ unbeschränkte Lösung von $x' = A(t)x$, nämlich $\begin{pmatrix} -\cos t \\ \sin t \end{pmatrix} e^{t/2}$.

Genau wie bei linearen Systemen mit konstanten Koeffizienten können aber auch bei linearen periodischen Systemen notwendige und hinreichende Stabilitätsbedingungen angegeben werden. Genauer gilt

**Satz 3.5** *Gegeben sei das lineare System $x' = A(t)x$ mit stetigem, $T$-periodischem $A(\cdot)$. Dann gilt:*

*(a) Die Nulllösung ist genau dann stabil, wenn die charakteristischen Multiplikatoren zu $x' = A(t)x$ betragsmäßig kleiner oder gleich 1 sind und diejenigen vom Betrag 1 einfache Elementarteiler haben.*

*(b) Die Nulllösung ist genau dann asymptotisch stabil, wenn die charakteristischen Multiplikatoren zu $x' = A(t)x$ betragsmäßig kleiner als 1 sind.*

**Beweis:** Nach Satz 2.6 ist ein Fundamentalsystem $X(\cdot)$ zu $x' = A(t)x$ in der Form $X(t) = P(t)e^{Bt}$ darstellbar, wobei $P(\cdot)$ stetig und $T$-periodisch. Die charakteristischen Multiplikatoren sind die Eigenwerte von $e^{BT}$.

Sei die Nulllösung stabil. Dann existiert eine Konstante $K > 0$ mit $\|X(t)\| \le K$ für alle $t \ge 0$ (dieser Schluss wurde beim Beweis von Satz 3.2 für autonome lineare Systeme gemacht, gilt aber natürlich auch allgemein). Dann ist

$$\|e^{Bt}\| \le \|P(t)^{-1}\|\,\|P(t)e^{Bt}\| \le K \max_{t\in[0,T]} \|P(t)^{-1}\| =: \tilde{K}$$

für alle $t \ge 0$. Das wiederum bedeutet, dass die Nulllösung des autonomen linearen Systems $x' = Bx$ stabil ist. Wegen Satz 3.2 bedeutet dies: Es ist $\Re(\lambda) \le 0$ für jeden Eigenwert $\lambda$ von $B$ und die Eigenwerte $\lambda$ mit $\Re(\lambda) = 0$ haben einfache Elementarteiler. Die charakteristischen Multiplikatoren $\mu$ sind als Eigenwerte von $e^{BT}$ gegeben durch $\mu = e^{\lambda T}$, wobei $\lambda$ Eigenwert von $B$. Hieraus folgt dann, dass die charakteristischen Multiplikatoren zu $x' = A(t)x$ betragsmäßig kleiner oder gleich 1 sind und diejenigen vom Betrag 1 einfache Elementarteiler haben. Ist umgekehrt dies der Fall, so ist $x = 0$ eine stabile Lösung von $x' = Bx$, daher existiert eine Konstante $\tilde{K} > 0$ mit $\|e^{Bt}\| \le \tilde{K}$ für alle $t \ge 0$. Ist dann $t_0 \ge 0$, $x_0 \in \mathbb{R}^n$ und $x(\cdot; t_0, x_0)$ die Lösung von $x' = A(t)x$, $x(t_0) = x_0$, so ist

$$\|x(t; t_0, x_0)\| = \|X(t)X(t_0)^{-1}x_0\| = \|P(t)e^{B(t-t_0)}P(t_0)^{-1}x_0\| \le c_0 c_1 \tilde{K}\,\|x_0\|$$

für alle $t \geq t_0$, wobei

$$c_0 := \max_{t\in[0,T]} \|P(t)\|, \qquad c_1 := \max_{t\in[0,T]} \|P(t)^{-1}\|.$$

Hieraus liest man sofort die Stabilität der Nulllösung ab.

Der zweite Teil des Satzes kann praktisch genau so bewiesen werden. □

### 2.3.5 Stabilität bei nichtlinearen Systemen

Der folgende Satz (siehe auch W. Walter (1993, S. 258)) ist eines der klassischen Ergebnisse von Lyapunov. Sein Beweis ist denen der beiden Sätze 3.3 und 3.4 in Unterabschnitt 2.3.3 sehr ähnlich.

**Satz 3.6** *Die Nulllösung $x = 0$ von $x' = Ax + g(t, x)$ ist asymptotisch stabil, wenn die beiden folgenden Bedingungen erfüllt sind:*

*(a) Es ist $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$.*

*(b) Es ist $\lim_{\|x\|\to 0} \|g(t, x)\|/\|x\| = 0$ gleichmäßig auf $[0, \infty)$ und damit insbesondere $g(t, 0) = 0$.*

**Beweis:** Sei $\alpha > 0$ so gewählt, dass $\Re(\lambda) < -\alpha$ für jeden Eigenwert $\lambda$ von $A$. Dann existiert eine Konstante $c > 0$ mit $\|e^{At}\| \leq c\, e^{-\alpha t}$ für alle $t \geq 0$. Wegen (b) existiert ein $\delta > 0$ mit

$$\|g(t, x)\| \leq \frac{\alpha}{2c} \|x\| \qquad \text{für alle } (t, x) \text{ mit } t \geq 0 \text{ und } \|x\| \leq \delta.$$

Für die Lösung $x(t) = x(t; t_0, x_0)$ von

$$x' = Ax + g(t, x), \qquad x(t_0) = x_0$$

gilt

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)} g(s, x(s))\, ds.$$

Sei nun $\epsilon \in (0, \delta/c)$ beliebig und $\|x_0\| \leq \epsilon$ (da notwendig $c \geq 1$ ist dann $\|x_0\| < \delta$). Für alle $t \geq t_0$ mit $\|x(s)\| \leq \delta$ für $s \in [t_0, t]$ ist dann

$$\|x(t)\| \leq c\, e^{-\alpha(t-t_0)} \|x_0\| + \int_{t_0}^{t} c\, e^{-\alpha(t-s)} \frac{\alpha}{2c} \|x(s)\|\, ds.$$

Setzt man wieder $\phi(t) := e^{\alpha t} \|x(t)\|$, so lautet diese Ungleichung

$$\phi(t) \leq c\, \phi(t_0) + \int_{t_0}^{t} \frac{\alpha}{2} \phi(s)\, ds,$$

also nach Anwendung der Gronwallschen Ungleichung

$$\phi(t) \leq c\, \phi(t_0)\, e^{\alpha(t-t_0)/2}$$

bzw.

$$\|x(t)\| \leq c\,\|x_0\|\,e^{-\alpha(t-t_0)/2}.$$

Hieraus erkennen wir: Für alle $t \geq t_0$ mit $\|x(s)\| \leq \delta$ für alle $s \in [t_0, t]$ und $\|x_0\| \leq \epsilon < \delta/c$ ist

$$\|x(t)\| \leq c\,\|x_0\|\,e^{-\alpha(t-t_0)/2} \leq c\,\|x_0\| \leq \epsilon c < \delta.$$

Daher gilt: Ist $\|x_0\| \leq \epsilon < \delta/c$, so ist $\|x(t)\| \leq c\,\|x_0\|\,e^{-\alpha(t-t_0)/2}$ für alle $t \geq t_0$, woraus wiederum die asymptotische Stabilität der Nulllösung folgt. □

**Bemerkung:** Als Ergänzung zu Satz 3.6 sei die folgende Aussage ohne Beweis formuliert[9].

- Über $g$ seien die Voraussetzungen von Satz 3.6 erfüllt. Ferner sei $A \in \mathbb{R}^{n\times n}$ eine konstante Matrix und $\Re(\lambda) > 0$ für mindestens einen Eigenwert $\lambda$ von $A$. Dann ist die Nulllösung des Differentialgleichungssystems $x' = Ax + g(t, x)$ nicht stabil.

Diese Aussage nennen wir den Instabilitätssatz. □

Zum Schluss dieses Unterabschnitts wollen wir eine kleine Einführung in die Lyapunovsche direkte Methode geben, siehe etwa auch W. Walter (1993, S. 265 ff.). Wir betrachten nur autonome Systeme der Form $x' = f(x)$ mit $f(0) = 0$ und wollen die Stabilität der Nulllösung untersuchen. Es soll uns hier keineswegs auf größtmögliche Allgemeinheit ankommen, hierfür sei auf die Spezialliteratur verwiesen.

**Definition 3.7** Sei $\mathcal{U} \subset \mathbb{R}^n$ eine Umgebung von 0. Eine Abbildung $V{:}\mathcal{U} \longrightarrow \mathbb{R}$ heißt *positiv (negativ) definit auf* $\mathcal{U}$, falls $V(x) > 0$ ($V(x) < 0$) für alle $x \in \mathcal{U} \setminus \{0\}$ und $V(0) = 0$. Die Abbildung $V$ heißt *positiv (negativ) semidefinit auf* $\mathcal{U}$, falls $V(x) \geq 0$ ($V(x) \leq 0$) für alle $x \in \mathcal{U} \setminus \{0\}$ und $V(0) = 0$.

Ist $x$ eine Lösung von $x' = f(x)$, so ist bei hinreichend glattem $V$ offenbar

$$\frac{d}{dt}V(x(t)) = \sum_{j=1}^{n} \frac{\partial V}{\partial x_j}(x(t))x_j'(t) = \sum_{j=1}^{n} \frac{\partial V}{\partial x_j}(x(t))f_j(x(t)) = \nabla V(x(t))^T f(x(t)),$$

wobei

$$\nabla V(x) = \left(\frac{\partial V}{\partial x_1}(x), \ldots, \frac{\partial V}{\partial x_n}(x)\right)^T$$

den Gradienten von $V$ in $x$ bedeutet. Für auf $\mathcal{U}$ stetig partiell differenzierbares $V$ definieren wir daher die Abbildung $\dot{V}{:}\mathcal{U} \longrightarrow \mathbb{R}$ durch

$$\dot{V}(x) := \nabla V(x)^T f(x).$$

Man beachte, dass $\dot{V}$ bezüglich des Differentialgleichungssystems $x' = f(x)$ definiert ist. Wir geben einen ersten Stabilitätssatz[10] an.

[9] Einen Beweis findet man z. B. bei W. Walter (1993, S. 259).

[10] Beim Beweis orientieren wir uns an

J. K. Hale (1969, S. 293 ff.) *Ordinary Differential Equations.* Wiley-Interscience, New York-London-Sydney-Toronto.

**Satz 3.8** *Man betrachte das autonome Differentialgleichungssystem $x' = f(x)$, wobei $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ stetig und so glatt ist, dass durch jeden Punkt $(t_0, x_0)$ eine eindeutige Lösung geht, die stetig von den Anfangsdaten abhängt. Ferner sei $f(0) = 0$. Dann gilt:*

*(a) Gibt es eine stetig partiell differenzierbare, auf einer Umgebung $\mathcal{U}$ von 0 positiv definite Funktion $V: \mathcal{U} \longrightarrow \mathbb{R}$ derart, dass $\dot{V}$ auf $\mathcal{U}$ negativ semidefinit ist, so ist die Nulllösung $x = 0$ stabil.*

*(b) Gilt über (a) hinaus, dass $\dot{V}$ auf $\mathcal{U}$ negativ definit ist, so ist die Nulllösung $x = 0$ asymptotisch stabil.*

**Beweis:** Sei $B_r := \{x \in \|R^n : \|x\| \leq r\}$ die abgeschlossene Kugel um 0 mit dem Radius $r$. Da $\mathcal{U}$ eine Umgebung von 0 ist, gibt es ein $r > 0$ mit $B_r \subset \mathcal{U}$. Für jedes $\epsilon \in (0, r]$ ist $k(\epsilon) := \min_{\|x\|=\epsilon} V(x) > 0$. Wegen $V(0) = 0$ und der Stetigkeit von $V$ im Nullpunkt gibt es ein $\delta = \delta(\epsilon) \in (0, \epsilon)$ mit $V(x) < k(\epsilon)$ für alle $x$ mit $\|x\| \leq \delta$.

Nun sei $t_0 \geq 0$, $\|x_0\| \leq \delta$ und $x(t) = x(t; t_0, x_0)$ die Lösung von $x' = f(x)$ durch $(t_0, x_0)$. Dann ist $\|x(t)\| \leq \epsilon$ für alle $t \geq t_0$ bzw. die Nulllösung stabil. Zur Begründung definieren wir

$$\hat{t} := \sup\{\tilde{t} \geq t_0 : \|x(t)\| \leq \epsilon \text{ für alle } t \in [t_0, \tilde{t}]\}.$$

Wegen $\|x(t_0)\| \leq \delta < \epsilon$ ist $\hat{t} > t_0$. Angenommen, es ist $\hat{t} < \infty$ (andernfalls wären wir fertig) und damit $\|x(t)\| \leq \epsilon$ für alle $t \in [t_0, \hat{t}]$. Für diese $t$ ist $\dot{V}(x(t)) \leq 0$ bzw.

$$\frac{d}{dt} V(x(t)) \leq 0,$$

daher $V(x(\hat{t})) \leq V(x(t_0)) = V(x_0) < k(\epsilon)$ und folglich $\|x(\hat{t})\| < \epsilon$ (wegen $V(x(\hat{t})) < k(\epsilon)$ und der Definition von $k(\epsilon)$ ist zunächst $\|x(\hat{t})\| \neq \epsilon$, andererseits ist $\|x(\hat{t})\| \leq \epsilon$) und dies ist ein Widerspruch zur Definition von $\hat{t}$. Also gilt die Implikation

$$t_0 \geq 0,\ \|x_0\| \leq \delta \Longrightarrow \|x(t; t_0, x_0)\| \leq \epsilon \quad \text{für alle } t \geq t_0.$$

Damit ist die Stabilität der Nulllösung bewiesen.

Zum Beweis von (b) bleibt zu zeigen: Zu jedem $t_0 \geq 0$ gibt es ein $b_0 = b(t_0)$ mit

$$\|x_0\| \leq b_0 \Longrightarrow \lim_{t \to \infty} x(t; t_0, x_0) = 0.$$

Wir hatten beim Beweis von (a) gezeigt, dass es zu vorgegebenem $\epsilon \in (0, r]$ ($r > 0$ war so gewählt, dass $B_r \subset \mathcal{U}$) ein $\delta = \delta(\epsilon) \in (0, \epsilon)$ mit

$$t_0 \geq 0,\ \|x_0\| \leq \delta \Longrightarrow \|x(t; t_0, x_0)\| \leq \epsilon \quad \text{für alle } t \geq t_0$$

gibt. (Wir konnten $\delta(\epsilon)$ unabhängig von $t_0$ bestimmen, man spricht dann auch von *gleichmäßiger Stabilität*.) Insbesondere existiert also zu beliebig vorgegebenem $H > 0$ ein $b_0 > 0$ mit

$$\|x_0\| \leq b_0 \Longrightarrow \|x(t; t_0, x_0)\| \leq H \quad \text{für alle } t \geq t_0.$$

Um die asymptotische Stabilität nachzuweisen, zeigen wir:

- Ist $\epsilon \in (0, r]$ und $t_0 \geq 0$, $\|x_0\| \leq b_0$, so existiert ein $T \geq t_0$ mit $\|x(t; t_0, x_0)\| \leq \delta(\epsilon) < \epsilon$ für alle $t \geq T$. Mit anderen Worten: Die Nulllösung ist asymptotisch stabil.

Hierzu zeigen wir zunächst:

- Ist $\epsilon \in (0, r]$ und $t_0 \geq 0$, $\|x_0\| \leq b_0$, so existiert ein $T \geq t_0$ mit $\|x(T; t_0, x_0)\| \leq \delta(\epsilon)$.

Ist der Nachweis hierfür gelungen, so folgt die Behauptung. Denn auf Grund der Stabilität ist $\|x(t; t_0, x_0)\| = \|x(t; T, x(T; t_0, x_0))\| \leq \epsilon$ für alle $t \geq T$.

Angenommen, es gibt ein $x_0$ mit $\|x_0\| \leq b_0$ und $\|x(t; t_0, x_0)\| > \delta(\epsilon)$ für alle $t \geq t_0$. Da $\dot{V}$ negativ definit ist, gibt es ein $\gamma > 0$ mit $\dot{V}(x) < -\gamma$ für alle $x$ mit $\delta(\epsilon) \leq \|x\| \leq H$. Hieraus wiederum folgt

$$\frac{d}{dt}V(x(t)) < -\gamma \quad \text{bzw.} \quad V(x(t)) < V(x_0) - \gamma(t - t_0) \quad \text{für alle } t \geq t_0.$$

Da andererseits $V$ positiv definit ist, existieren positive Konstanten $c_1, c_2$ mit $0 < c_1 \leq V(x) \leq c_2$ für alle $x$ mit $\delta(\epsilon) \leq \|x\| \leq H$. Damit folgt $V(x(t)) < c_2 - \gamma(t - t_0) \leq c_1$, falls $t \geq T := t_0 + (c_2 - c_1)/\gamma$, woraus wiederum $\|x(t)\| < \delta(\epsilon)$ für alle $t \geq T$ folgt. Dies ist ein Widerspruch zur Annahme. Wie wir uns oben schon überlegt haben, ist damit der Beweis abgeschlossen. □

Als eines der ersten Stabilitätsresultate haben wir bewiesen, dass die Nulllösung von $x' = Ax$ mit einer konstanten $n \times n$-Matrix genau dann asymptotisch stabil ist, wenn $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$. Nun wollen wir dies mit Hilfe der obigen Ergebnisse beweisen. Hierzu setze man $V$ als quadratische Form $V(x) := x^T Bx$ mit symmetrischem $B$ an. Dann ist

$$\dot{V}(x) = 2(Bx)^T Ax = 2x^T BAx = x^T(A^T B + BA)x.$$

Die Suche nach einem positiv definiten $V$, für das $\dot{V}$ negativ definit ist, führt also auf die Frage:

- Gibt es zu beliebigem $A \in \mathbb{R}^{n\times n}$ eine symmetrische, positiv definite Matrix $B \in \mathbb{R}^{n\times n}$ derart, dass $A^T B + BA$ negativ definit ist?

Diese Frage kann bejaht werden, wie das folgende Lemma zeigt.

**Lemma 3.9** *Sei $A$ eine reelle $n \times n$-Matrix mit der Eigenschaft, dass $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$. Dann besitzt die Matrixgleichung $A^T B + BA = -C$ für jede positiv definite Matrix $C$ eine positiv definite Lösung.*

**Beweis:** Da $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$ (und dann auch von $A^T$, denn die Eigenwerte von $A$ und $A^T$ stimmen überein), existieren positive Konstanten $c$ und $\alpha$ mit

$$\|e^{At}\|,\ \|e^{A^T t}\| \leq c\, e^{-\alpha t} \quad \text{für alle } t \geq t_0.$$

Bei vorgegebener positiv definiter Matrix $C$ definiere man $B := \int_0^\infty e^{A^T t} C e^{At}\, dt$. Die Matrix $B$ ist offenbar wohldefiniert, symmetrisch und positiv definit. Man beachte hierzu, dass $(e^{At})^T = e^{A^T t}$. Ferner ist

$$\begin{aligned} A^T B + BA &= \int_0^\infty [A^T e^{A^T t} C e^{At} + e^{A^T t} C e^{At} A]\, dt \\ &= \int_0^\infty \frac{d}{dt}[e^{A^T t} C e^{At}]\, dt \\ &= -C. \end{aligned}$$

Damit ist das Lemma bewiesen. □

**Bemerkung:** Auch bei nichtlinearen, autonomen Systemen kann mit Hilfe einer Lyapunov-Funktion unter geeigneten Voraussetzungen die Stabilität bewiesen werden. Zu untersuchen sei z. B. die Stabilität der Nulllösung von $x' = Ax + g(x)$, wobei wieder vorausgesetzt wird, dass $\Re(\lambda) < 0$ für alle Eigenwerte $\lambda$ von $A$. Ferner sei $g \in C^1(\mathbb{R}^n; \mathbb{R}^n)$, $g(0) = 0$ und $g'(0) = 0$ (dies impliziert, daß $\lim_{\|x\|\to 0} \|g(x)\|/\|x\| = 0$). Durch die Anwendung von Lemma 3.9 erhält man, dass die Matrixgleichung $A^T B + BA = -I$ eine positiv definite Lösung $B$ besitzt. Hiermit definiere man die (auf dem ganzen $\mathbb{R}^n$) positiv definite Funktion $V$ durch $V(x) := x^T B x$. Dann wird

$$\dot{V}(x) = 2(Bx)^T(Ax + g(x)) = x^T(A^T B + BA) + 2x^T B g(x) = -x^T x + 2x^T B g(x).$$

Wegen $\lim_{\|x\|\to 0} \|g(x)\|/\|x\| = 0$ ist $\dot{V}$ in einer Umgebung von 0 negativ definit. Denn es gibt eine Umgebung $\mathcal{U}$ des Ursprungs derart, dass $\|g(x)\| \le \frac{1}{4}\|x\|$ für alle $x \in \mathcal{U}$ (hier sei jetzt, da gleich die Cauchy-Schwarzsche Ungleichung angewandt wird, $\|\cdot\|$ die euklidische Vektornorm). Daher ist $\dot{V}(x) \le -\frac{1}{2}\|x\|^2$ für alle $x \in \mathcal{U}$, so dass wieder die asymptotische Stabilität der Nulllösung folgt. □

**Beispiel:** Wir betrachten die sogenannte van der Polsche Differentialgleichung

$$x'' + \mu(x^2 - 1)x' + x = 0$$

und wollen durch Anwendung der gerade eben gemachten Bemerkung nachweisen, dass die Nulllösung für $\mu < 0$ (das ist für die Anwendungen allerdings gerade der uninteressante Fall) asymptotisch stabil ist. Denn die gegebene Gleichung zweiter Ordnung ist äquivalent zu dem System

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & \mu \end{pmatrix}}_{=:A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ -\mu x_1^2 x_2 \end{pmatrix}}_{=:g(x)}.$$

Die Eigenwerte von $A$ sind $\lambda_{1,2} = \frac{1}{2}(\mu \pm \sqrt{\mu^2 - 4})$, sie haben für $\mu < 0$ negativen Realteil. Weiter ist $g(0) = 0$ und $g'(0) = 0$, obige Bemerkung zeigt dann die Behauptung.

Wir wollen dieses Beispiel illustrieren. Durch

```
phaseportrait([diff(x_1(t),t)=x_2(t),
diff(x_2(t),t)=-x_1(t)-(x_1(t)^2-1)*x_2(t)],[x_1(t),x_2(t)],
t=0..10,[[x_1(0)=0.5,x_2(0)=0],[x_1(0)=1,x_2(0)=0]],
arrows=NONE,linecolor=black,stepsize=0.1,thickness=0);
```

haben wir die Lösungen von

$$x'' - (x^2 - 1)x' + x = 0, \qquad x(0) = 0.5 \text{ bzw. } x(0) = 1,\ x'(0) = 0$$

in der $(x, x')$-Phasenebene geplottet, siehe Abbildung 2.9 links. Man erkennt, wie sich

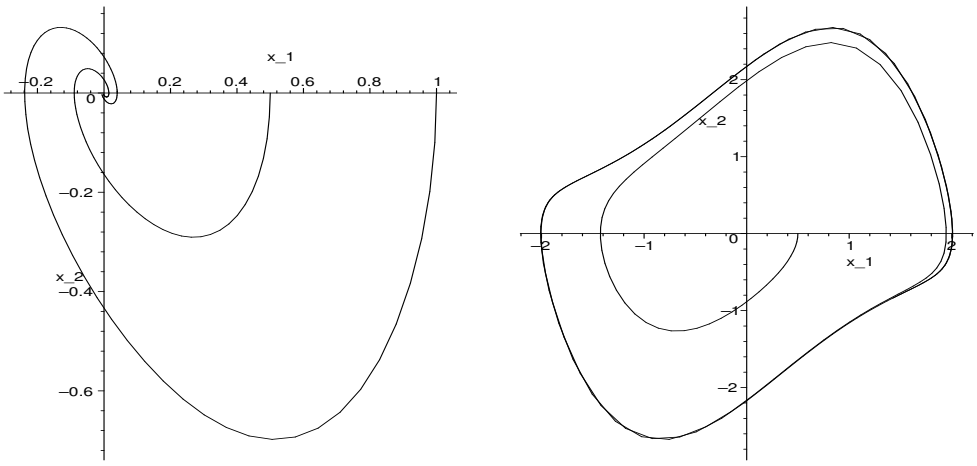


Abbildung 2.9: Asymptotische Stabilität bei $x'' - (x^2 - 1)x' + x = 0$

die Bahnen dem Ursprung annähern. Ganz anders ist das Verhalten der van der Polschen Gleichung für positives $\mu$. In Abbildung 2.9 rechts zeichnen wir die Phasenbahn zu

$$x'' + (x^2 - 1)x' + x = 0, \qquad x(0) = 0.5,\ x'(0) = 0$$

in der $(x, x')$-Phasenebene auf. Offenbar nähert sich diese einem "Grenzzyklus" an. □

**Beispiel:** Durch Anwendung von Satz 3.8 wollen wir zeigen, dass die Nulllösung des Differentialgleichungssystems

$$\begin{aligned} x_1' &= -2x_2 + x_2x_3, \\ x_2' &= x_1(1 - x_3), \\ x_3' &= x_1x_2 \end{aligned}$$

stabil, aber nicht asymptotisch stabil ist. Mit noch nicht festgelegten positiven Zahlen $a_1, a_2, a_3$ definieren wir $V\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ durch

$$v(x_1, x_2, x_3) := \frac{1}{2}(a_1x_1^2 + a_2x_2^2 + a_3x_3^2).$$

Dann ist

$$\begin{aligned} \dot{V}(x_1, x_2, x_3) &= \begin{pmatrix} a_1x_1 \\ a_2x_2 \\ a_3x_3 \end{pmatrix} \begin{pmatrix} -2x_2 + x_2x_3 \\ x_1(1 - x_3) \\ x_1x_2 \end{pmatrix} \\ &= (a_1 - a_2 + a_3)x_1x_2x_3 + (a_2 - 2a_1)x_1x_2. \end{aligned}$$

Wählen wir also $a_2 = 2a_1$ und $a_1 = a_3$, so ist $V > 0$ auf $\mathbb{R}^3 \setminus \{0\}$ und $\dot{V} = 0$ auf $\mathbb{R}^3$, also die Nulllösung stabil. Für eine Lösung $x$ ist $(d/dt)V(x(t)) = 0$. Daher ist $V(x(t)) = V(x(0))$ für alle $t$, so dass jede Trajektorie auf einer Ellipse

$$E_r := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + 2x_2^2 + x_3^2 = r^2\}$$

liegt. Die Nulllösung ist also nicht asymptotisch stabil. □

### 2.3.6 Aufgaben

1. Sei
$$A := \begin{pmatrix} 3 & 6 \\ -2 & -3 \end{pmatrix} \qquad \text{bzw.} \qquad A := \begin{pmatrix} -2 & -1 \\ 4 & -1 \end{pmatrix}.$$
Man untersuche die Stabilität der Nulllösung des Differentialgleichungssystems $x' = Ax$ und veranschauliche beide Fälle durch Phasenportraits.

2. Sei $A := \text{tridiag}(1, -2, 1) \in \mathbb{R}^{n\times n}$ die Tridiagonalmatrix, bei der alle Hauptdiagonaleinträge gleich $-2$ und alle Nebendiagonaleinträge gleich 1 sind. Man zeige, dass die Nulllösung von $x' = Ax$ asymptotisch stabil ist.

3. Man zeige: Hat das Polynom
$$p(z) := z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$
mit den reellen Koeffizienten $a_0, \ldots, a_{n-1}$ nur Nullstellen mit negativem Realteil, so sind die Koeffizienten $a_0, \ldots, a_{n-1}$ von $p$ notwendigerweise positiv.

4. Man betrachte die sogenannte Hillsche Differentialgleichung
$$x'' + b(t)x = 0,$$
wobei $b(\cdot)$ stetig und $T$-periodisch. Man zeige, dass die Nulllösung zum äquivalenten System
$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b(t) & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$
*nicht* asymptotisch stabil ist, da das Produkt der beiden charakteristischen Multiplikatoren 1 ist.

5. Wie in Aufgabe 4 betrachte man die Hillsche Differentialgleichung
$$x'' + b(t)x = 0,$$
wobei $b(\cdot)$ stetig und $T$-periodisch. Ferner sei $b(\cdot)$ eine gerade Funktion. Sei
$$X(t) = \begin{pmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{pmatrix}$$
das durch $X(0) = I$ normierte Fundamentalsystem zu dem äquivalenten System
$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b(t) & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$
Man zeige:

(a) Es ist $x_1(\cdot)$ gerade und $x_2(\cdot)$ ungerade.

(b) Es ist $X(T)^{-1} = X(-T)$.

(c) Es ist $x_1(T) = x_2'(T)$.

(d) Die beiden chrakteristischen Multiplikatoren sind Nullstellen der quadratischen Gleichung $\mu^2 - 2x_1(T)\mu + 1 = 0$. Das Stabilitätsverhalten der Nulllösung der Hillschen Gleichung wird also im wesentlichen nur durch $x_1(T)$ bestimmt. Genauer gilt: Ist $|x_1(T)| < 1$, so ist die Nulllösung stabil, ist $|x_1(T)| > 1$ so ist sie instabil.

(e) Gegeben sei die spezielle Mathieusche Differentialgleichung

$$x'' + (\delta + \gamma \cos 2t)x = 0,$$

hier ist also $T = \pi$. Man bestimme numerisch das Stabilitätsverhalten der Nulllösung für $(\delta, \gamma) = (1, 2)$, $(\frac{1}{4}, 1)$ und illustriere dies durch Bahnen in der Phasenebene.

6. Man zeige in Lemma 3.9 die Umkehrung. Genauer beweise man: Gibt es zu $A \in \mathbb{R}^{n\times n}$ eine symmetrische, positiv definite Matrix $B \in \mathbb{R}^{n\times n}$ derart, dass $A^T B + BA$ negativ definit ist, so ist $\Re(\lambda) < 0$ für alle Eigenwerte $\lambda$ von $A$.

7. Man betrachte die Differentialgleichung zweiter Ordnung

$$x'' + h(x) = 0,$$

wobei $h \in C(\mathbb{R})$ und $xh(x) > 0$ für alle $x \neq 0$ (woraus $h(0) = 0$ folgt). Man zeige, dass die Nulllösung zu dieser Differentialgleichung bzw. dem äquivalenten Differentialgleichungssystem

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= -h(x_1) \end{aligned}$$

stabil ist.

8. Man zeige: Die Nulllösungen der Differentialgleichungssysteme

$$\begin{aligned} x_1' &= -x_2\sqrt{x_1^2 + x_2^2} \\ x_2' &= x_1\sqrt{x_1^2 + x_2^2}, \end{aligned} \qquad \text{bzw.} \qquad \begin{aligned} x_1' &= (x_1^2 + x_2^2 - 1)x_1 - x_2 \\ x_2' &= x_1 + (x_1^2 + x_2^2 - 1)x_2 \end{aligned}$$

sind stabil bzw. asymptotisch stabil. Weiter zeige man, dass die Nulllösung von

$$\begin{aligned} x_1' &= (1 - x_1^2 - x_2^2)x_1 - x_2 \\ x_2' &= x_1 + (1 - x_1^2 - x_2^2)x_2 \end{aligned}$$

nicht stabil ist.

9. Man zeige, dass $(\frac{1}{2}, \frac{1}{2})$ eine asymptotisch stabile Lösung des Differentialgleichungssystems

$$\begin{aligned} x_1' &= x_1(1 - x_1 - x_2) \\ x_2' &= x_2(\tfrac{3}{4} - x_2 - \tfrac{1}{2}x_1) \end{aligned}$$

ist.

10. Gegeben sei das Differentialgleichungssystem (Lorenz-Attraktor)

$$\begin{aligned} x_1' &= -\sigma x_1 + \sigma x_2 \\ x_2' &= r x_1 - x_2 - x_1 x_3 \\ x_3' &= -b x_3 + x_1 x_2, \end{aligned}$$

wobei $b, r, \sigma$ positive Konstanten sind. Man zeige, dass die triviale Lösung asymptotisch stabil ist, wenn $r \in (0, 1)$. Mit Hilfe des im Anschluss von Satz 3.6 (ohne Beweis) angegebenen Instabilitätssatzes begründe man, dass die Nullösung für $r > 1$ nicht stabil ist.

11. Gegeben sei das Differentialgleichungssystem

$$\begin{aligned} x' &= ax - bxy - ex^2, \\ y' &= -cy + dxy - fy^2 \end{aligned}$$

mit positiven Konstanten $a, b, c, d, e, f$. Dieses hat den Gleichgewichtspunkt bzw. die konstante Lösung

$$(\hat{x}, \hat{y}) = \frac{1}{ef + bd}(af + bc, ad - ce).$$

Man zeige, dass dies für $(a, b, c, d, e, f) := (2, 0.01, 1, 0.01, 0.001, 0.001)$ (siehe Abbildung 1.6) eine asymptotisch stabile Lösung ist.

# Kapitel 3

# Die numerische Behandlung gewöhnlicher Anfangswertaufgaben

In diesem Kapitel[1] gehen wir auf die numerische Behandlung von Anfangswertaufgaben bei einem System gewöhnlicher Differentialgleichungen erster Ordnung ein, also die Aufgabe, die Lösung $x$ von

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

zu bestimmen. Bekanntlich läßt sich die Anfangswertaufgabe bei einer Differentialgleichung $n$-ter Ordnung

$$x^{(n)} = f(t, x, x', \ldots, x^{(n-1)}), \qquad x^{(i)}(t_0) = x_{i,0} \quad (i = 0, \ldots, n-1)$$

in der Form (P) als System schreiben, nämlich als

$$\begin{pmatrix} x_1' \\ \vdots \\ x_{n-1}' \\ x_n' \end{pmatrix} = \begin{pmatrix} x_2 \\ \vdots \\ x_n \\ f(t, x_1, \ldots, x_n) \end{pmatrix}, \qquad \begin{pmatrix} x_1(t_0) \\ \vdots \\ x_n(t_0) \end{pmatrix} = \begin{pmatrix} x_{0,0} \\ \vdots \\ x_{n-1,0} \end{pmatrix}.$$

Damit ist ein Verfahren, das für (P) entwickelt wurde, auch für eine Anfangswertaufgabe für eine Differentialgleichung $n$-ter Ordnung anwendbar. Allerdings werden

[1] Die numerische Behandlung gewöhnlicher Anfangswertaufgaben wird in einigen Büchern über numerische Mathematik ausführlich geschildert. Zu nennen sind hier z. B. die Bücher von Stoer-Bulirsch, Deuflhard-Bornemann. An Spezialliteratur geben wir an:

Hairer, E., S. P. Nørsett, G. Wanner (1993) *Solving Ordinary Differential Equations I. Nonstiff Problems. Second Revised Edition.* Springer-Verlag, Berlin-Heidelberg-New York.

Hairer, E., G. Wanner (1991) *Solving ordinary Differential Equations II. Stiff and Differential-Algebraic Problems.* Springer-Verlag, Berlin-Heidelberg-New York.

U. M. Ascher, L. R. Petzold (1998) *Computer Methods for Differential Equations and Differential-Algebraic Equations.* SIAM, Philadelphia.

K. Strehmel, R. Weiner (1992) *Linear-implizite Runge-Kutta=Methoden und ihre Anwendung.* B. G. Teubner, Stuttgart-Leipzig.

K. Strehmel, R. Weiner (1995) *Numerik gewöhnlicher Differentialgleichungen.* B. G. Teubner, Stuttgart.

dann auch Näherungen nicht nur für die gesuchte Lösung, sondern auch ihrer Ableitungen bis zur Ordnung $n-1$ berechnet. Im wesentlichen werden wir uns daher mit (P) beschäftigen und nur exemplarisch auf die Adaption der Verfahren etwa auf eine Differentialgleichung zweiter Ordnung eingehen.

## 3.1 Einschrittverfahren

Ziel dieses Abschnittes ist es, einige der sogenannten *Einschrittverfahren* zur Lösung der Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

aufzustellen und zu motivieren, um anschließend eine Konvergenztheorie hierfür zu entwickeln.

Gegeben sei die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0.$$

Wir werden im folgenden annehmen, dass (P) eine eindeutige Lösung auf dem Intervall $[t_0, T]$ besitzt, die dort approximiert werden soll. Dies ist z. B. der Fall, wenn $f$ bezüglich des zweiten Arguments global lipschitzstetig ist. Wir setzen daher generell voraus:

(V) Es ist $f\colon [t_0,T] \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ stetig und es existiert eine Konstante $L > 0$ mit

$$\|f(t,x) - f(t,y)\| \le L\,\|x-y\| \qquad \text{für alle } (t,x),\ (t,y) \in [t_0,T] \times \mathbb{R}^n.$$

Dies impliziert die Existenz genau einer Lösung von (P) auf dem Intervall $[t_0, T]$. Wir brauchen uns also keine Gedanken darüber zu machen, dass wir vielleicht versuchen etwas zu berechnen, was es gar nicht gibt.

### 3.1.1 Beispiele von Einschrittverfahren

Sei $x$ die Lösung von (P), also $x(t)$ der Wert der Lösung zur Zeit $t$, und $h > 0$ eine gewisse Schrittweite. Bei einem Einschrittverfahren ist eine *Näherung* $u(t)$ für $x(t)$ bekannt und es wird eine Vorschrift angegeben, wie $u(t+h)$, eine Näherung für $x(t+h)$, zu berechnen ist. Ausgangspunkt ist stets die Identität

$$x(t+h) = x(t) + \int_t^{t+h} f(s, x(s))\,ds.$$

Indem man rechts das Integral durch einen durch eine *Quadraturformel* ermittelten Wert ersetzt, erhält man geeignete Verfahren.

(a) Das Eulersche Polygonzugverfahren.

Wendet man die sehr simple Quadraturformel

$$\int_a^b f(s)\,ds \approx (b-a)f(a)$$

an, so erhält man $x(t+h) \approx x(t) + hf(t, x(t))$. Ersetzt man hier auf der rechten Seite die unbekannte exakte Lösung $x(t)$ zur Zeit $t$ durch $u(t)$, so ist die Vorschrift des Euler'schen Polygonzugverfahrens durch

$$u(t+h) := u(t) + hf(t, u(t))$$

gegeben.

(b) Verfahren von Heun (auch Verfahren von Euler-Cauchy genannt).

Es ist

$$x(t+h) \approx x(t) + \tfrac{1}{2}\,h[f(t, x(t)) + f(t+h, x(t+h))],$$

was einer Anwendung der Trapezregel

$$\int_a^b f(s)\,ds \approx \tfrac{1}{2}\,[f(a) + f(b)]$$

entspricht. Hieraus könnte man natürlich das *implizite* Einschrittverfahren

$$u(t+h) := u(t) + \tfrac{1}{2}\,h[f(t, u(t)) + f(t+h, u(t+h))]$$

gewinnen, das sogenannte verbesserte Euler-Verfahren oder Euler-Heun-Verfahren. Zur Bestimmung von $u(t+h)$ muss hier also ein nichtlineares Gleichungssystem (bzw. eine Fixpunktaufgabe) gelöst werden. Unter der Voraussetzung (V) hat diese Fixpunktaufgabe wegen des Kontraktionssatzes für $\frac{1}{2}\,hL < 1$, also für eine hinreichend kleine Schrittweite $h$, eine eindeutige Lösung. Man kann aber auch aus

$$x(t+h) \approx x(t) + hx'(t) = x(t) + hf(t, x(t))$$

die Verfahrensvorschrift

$$u(t+h) := u(t) + \tfrac{1}{2}\,h\,[f(t, u(t)) + f(t+h, u(t) + hf(t, u(t)))]$$

erhalten, das Verfahren von Heun.

Die bisher angegebenen Verfahren spielen i. allg. in der Praxis keine Rolle. Wir haben sie trotzdem angegeben und hergeleitet, weil man an ihnen einige theoretische Ergebnisse besonders einfach erläutern kann. Die bisher benutzten Quadraturformeln zur näherungsweisen Berechnung von $I(f) := \int_a^b f(s)\,ds$ haben die Form

$$Q(f) := (b-a)f(a), \quad Q(f) := \tfrac{1}{2}\,(b-a)\,[f(a) + f(b)].$$

Bezeichnet man mit $\Pi_k$ die Menge der Polynome vom Grad $\leq k$, so stellt man sofort fest, dass die erste Quadraturformel auf $\Pi_0$ (der Menge der Konstanten) exakt ist, während die zweite auf $\Pi_1$ exakt ist.

(c) Verfahren von Runge-Kutta (vierter Ordnung).

Ausgangspunkt ist

$$x(t+h) \approx x(t) + \tfrac{1}{6}\, h\, [f(t,x(t)) + 4f(t+\tfrac{1}{2}\, h, x(t+\tfrac{1}{2}\, h)) + f(t+h, x(t+h))].$$

Dies entspricht der Anwendung der Simpson-Regel (siehe Aufgabe 1)

$$\int_a^b f(s)\, ds \approx \frac{b-a}{6}\, [f(a) + 4\, f(\tfrac{1}{2}\, (a+b)) + f(b)].$$

Es ist

$$x(t+\tfrac{1}{2}\, h) \approx x(t) + \tfrac{1}{2}\, h x'(t) = x(t) + \tfrac{1}{2}\, h k_1(t)$$

mit

$$k_1(t) := f(t, x(t)).$$

Entsprechend ist

$$\begin{aligned} x(t+\tfrac{1}{2}\, h) &\approx x(t) + \tfrac{1}{2}\, h\, x'(t+\tfrac{1}{2}\, h) \\ &= x(t) + \tfrac{1}{2}\, h f(t+\tfrac{1}{2}\, h, x(t+\tfrac{1}{2}\, h)) \\ &\approx x(t) + \tfrac{1}{2}\, h f(t+\tfrac{1}{2}\, h, x(t) + \tfrac{1}{2}\, h k_1(t)) \\ &= x(t) + \tfrac{1}{2}\, h k_2(t) \end{aligned}$$

mit

$$k_2(t) := f(t+\tfrac{1}{2}\, h, x(t) + \tfrac{1}{2}\, h k_1(t)).$$

Ferner ist

$$\begin{aligned} x(t+h) &\approx x(t) + h x'(t+\tfrac{1}{2}\, h) \\ &= x(t) + h f(t+\tfrac{1}{2}\, h, x(t+\tfrac{1}{2}\, h)) \\ &\approx x(t) + h f(t+\tfrac{1}{2}\, h, x(t) + \tfrac{1}{2}\, h k_2(t)) \\ &= x(t) + h k_3(t) \end{aligned}$$

mit

$$k_3(t) := f(t+\tfrac{1}{2}\, h, x(t) + \tfrac{1}{2}\, h k_2(t)).$$

Schließlich ist

$$f(t+h, x(t+h)) \approx f(t+h, x(t) + h k_3(t)) =: k_4(t).$$

Damit wird

$$\begin{aligned} x(t+h) &\approx x(t) + \tfrac{1}{6}\, h[f(t,x(t)) + 2f(t+\tfrac{1}{2}\, h, x(t+\tfrac{1}{2}\, h)) \\ &\qquad + 2f(t+\tfrac{1}{2}\, h, x(t+\tfrac{1}{2}\, h)) + f(t+h, x(t+h))] \\ &\approx x(t) + \tfrac{1}{6}\, h\, [k_1(t) + 2k_2(t) + 2k_3(t) + k_4(t)] \end{aligned}$$

Das (klassische) Runge-Kutta Verfahren ist daher durch die folgende Vorschrift gegeben:

$$u(t+h) := u(t) + \frac{h}{6}\, [k_1(t) + 2k_2(t) + 2k_3(t) + k_4(t)]$$

mit

$$\begin{aligned} k_1(t) &:= f(t, u(t)), \\ k_2(t) &:= f(t + \tfrac{1}{2}h, u(t) + \tfrac{1}{2}hk_1(t)), \\ k_3(t) &:= f(t + \tfrac{1}{2}h, u(t) + \tfrac{1}{2}hk_2(t)), \\ k_4(t) &:= f(t + h, u(t) + hk_3(t)). \end{aligned}$$

**Beispiel:** Wir schreiben MATLAB-Programme zum Euler-, Heun- und zum Runge-Kutta-Verfahren. Das Intervall $[t_0, t_{\max}]$, auf dem die Lösung der gegebenen Anfangswertaufgabe zu berechnen ist, sei äquidistant unterteilt. Zunächst das Programm zum Euler-Verfahren.

```
function [tvals,xvals]=FixedEuler(fname,x_0,t_0,t_max,m);
%
%Pre:  fname is a string that names a function of the form
%      f(t,x). More exactly
%              f_1(t,x_1,...,x_n)
%      f(t,x)= ................
%              f_n(t,x_1,...,x_n)
%      [t_0,t_max] is the intervall where the solution of
%      the IVP x'=f(t,x), x(t_0)=x_0 is computed using
%      mesh-width h=(t_max-t_0)/(m-1) and the classical Euler
%      method
%Post: tvals(k)=t_0+(k-1)h, where h=(t_max-t_0)/(m-1) and xvals(:,k) is an
%      approximation of the true solution x(tvals(k)) for k=1:m
%
  t_c=t_0; x_c=x_0; tvals=t_c; xvals=x_c; f_c=feval(fname,t_c,x_c);
  h=(t_max-t_0)/(m-1);
  for k=1:m-1
     x_c=x_c+h*f_c; t_c=t_c+h; f_c=feval(fname,t_c,x_c);
     xvals=[xvals x_c];  tvals=[tvals t_c];
  end;
```

Wir wollen nun die Anfangswertaufgabe $x' = x - t^2 + 1$, $x(0) = 0.5$ auf $[0, 2]$ näherungsweise mit dem Euler-Verfahren lösen, wobei wir die verhältnismäßig grobe Schrittweite $h = 0.2$ benutzen. Die Lösung ist $x(t) = (1 + t)^2 - \frac{1}{2}e^t$. Hierzu definieren wir zunächst eine Funktion `f1`, welche die rechte Seite berechnet, durch

```
function out=f1(t,x);
  out=x-t^2+1;
```

Nach dem Aufruf `[t,x]=FixedEuler('f1',0.5,0,2,11);` erhalten wir die Werte in der folgenden Tabelle (wir benutzen `format short`). Zum Vergleich geben wir auch noch die entsprechenden, durch das Heun-Verfahren und das Runge-Kutta-Verfahren

gewonnenen Werte an.

| $t$ | Euler | Heun | Runge-Kutta | $x(t)$ |
|---|---|---|---|---|
| 0.0000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 |
| 0.2000 | 0.8000 | 0.8260 | 0.8293 | 0.8293 |
| 0.4000 | 1.1520 | 1.2069 | 1.2141 | 1.2141 |
| 0.6000 | 1.5504 | 1.6372 | 1.6489 | 1.6489 |
| 0.8000 | 1.9885 | 2.1102 | 2.1272 | 2.1272 |
| 1.0000 | 2.4582 | 2.6177 | 2.6408 | 2.6409 |
| 1.2000 | 2.9498 | 3.1496 | 3.1799 | 3.1799 |
| 1.4000 | 3.4518 | 3.6937 | 3.7323 | 3.7324 |
| 1.6000 | 3.9501 | 4.2351 | 4.2834 | 4.2835 |
| 1.8000 | 4.4282 | 4.7556 | 4.8151 | 4.8152 |
| 2.0000 | 4.8658 | 5.2331 | 5.3054 | 5.3055 |

Hierbei unterscheidet sich das Programm für das Heun-Verfahren nur unwesentlich von dem für das Euler-Verfahren, siehe auch das gleich folgende Programm zum klassischen Runge-Kutta-Verfahren mit konstanter Schrittweite[2].

```
function [tvals,xvals]=FixedRK(fname,x_0,t_0,t_max,m);
%
%Produces an approximate solution to the initial value problem
%x'=f(t,x), x(t_0)=x_0 using the classical Runge-Kutta-method
%with fixed meshsize h=(t_max-t_0)/(m-1).
%Pre:   fname=string that names the function f.
%       x_0=initial condition vector.
%       t_0=initial time
%       t_max=final time.
%       m=number of steps to be taken.
%Post:  tvals(k)=t_0+(k-1)h, k=1:m, where h=(t_max-t_0)/(m-1)
%       xvals(:,k)=approximate solution at t=tvals(k), k=1..m
%
  t_c=t_0; x_c=x_0; tvals=t_c; xvals=x_c; f_c=feval(fname,t_c,x_c);
  h=(t_max-t_0)/(m-1);
  for k=1:m-1
     k_1=f_c;
     k_2=feval(fname,t_c+(h/2),x_c+(h/2)*k_1);
     k_3=feval(fname,t_c+(h/2),x_c+(h/2)*k_2);
     k_4=feval(fname,t_c+h,x_c+h*k_3);
     x_c=x_c+(h/6)*(k_1+2*k_2+2*k_3+k_4);
     t_c=t_c+h;
     f_c=feval(fname,t_c,x_c);
     xvals=[xvals x_c];
     tvals=[tvals t_c];
  end;
```

[2]Siehe auch

C. F. van Loan (1997) *Introduction to Scientific Computing. A Matrix-Vector Approach using MATLAB*. Prentice Hall, Upper Saddle River.

Hier ist fast kein Unterschied zwischen exakter Lösung und Runge-Kutta-Lösung feststellbar (zumindestens bei der Formatierung `format short`). □

Nun wollen wir durch ein Beispiel feststellen, ob die obigen Funktionen ohne Änderung auch eine Anfangswertaufgabe für ein System von Differentialgleichungen lösen können.

**Beispiel:** Zu lösen sei die Anfangswertaufgabe (Lotka-Volterra-Modell)

$$\begin{aligned} x' &= 2x - 0.01xy, & x(0) &= 300, \\ y' &= -y + 0.01xy, & y(0) &= 150. \end{aligned}$$

Wir definieren zunächst eine Funktion, die die rechte Seite des Differentialgleichungssystems liefert (dies kann ein eigenes Function-File sein oder der Funktion `FixedRK` angehängt sein), etwa

```
function out=Lotka(t,x);
   out=[2*x(1)-0.01*x(1)*x(2);-x(2)+0.01*x(1)*x(2)];
```

Anschließend gibt man z. B. ein:

```
x_0=[300;150];
[t,x]=FixedRK('Lotka',x_0,0,10,200);
plot(x(1,:),x(2,:))
title('Phasenbahn im Lotka-Volterra-Modell')
xlabel('Beute')
ylabel('Raeuber')
```

Wir erhalten das in Abbildung 3.1 angegebene Bild für die Phasenbahn. Die Übereinstimmung mit Abbildung 1.3 ist offensichtlich. □

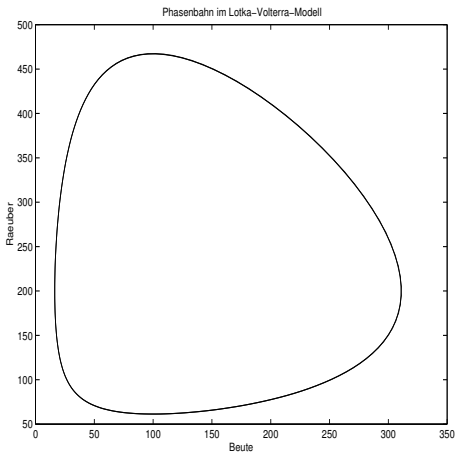


Abbildung 3.1: Phasenbahn beim Lotka-Volterra-Modell

Mit einem $s \in \mathbb{N}$ spricht man allgemein von einem (expliziten) *s-stufigen Runge-Kutta Verfahren*, wenn es folgendermaßen gebildet wird: Es ist

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

wobei

$$\begin{aligned} k_1(t,u) &= f(t,u), \\ k_2(t,u) &= f(t+c_2h, u+ha_{21}k_1(t,u)), \\ &\vdots \\ k_s(t,u) &= f(t+c_sh, u+h\textstyle\sum_{i=0}^{s-1} a_{si}k_i(t,u)). \end{aligned}$$

Dabei sind die $a_{ki}$, $b_i$ und $c_i$ geeignet gewählte reelle Zahlen, die das Verfahren vollständig festlegen. Zur Beschreibung könnte man sie in einem Schema der folgenden Art anordnen:

$$\begin{array}{c|cccc|c} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} & \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array}$$

Man überzeugt sich leicht, dass die bisher erhaltenen Verfahren in dieses Schema hineinpassen. Z. B. ist das klassische Runge-Kutta-Verfahren durch das Schema

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

gegeben.

### 3.1.2 Konsistenz von Einschrittverfahren

Nun kommen wir zur Definition von Einschrittverfahren und überlegen uns dazu, was die im letzten Unterabschnitt angegebenen Verfahren gemeinsam haben. Gemeinsam ist, dass aus der Kenntnis einer Näherung $u(t)$ für die Lösung $x(t)$ zur Zeit $t$ bei vorgegebener Schrittweite $h > 0$ und rechter Seite $f$ eine Näherung $u(t+h)$ für $x(t+h)$ aus der Vorschrift

$$u(t+h) := u(t) + h\Phi(h,f)(t,u(t))$$

berechnet wird. Hierbei heißt $\Phi$ die *Verfahrensfunktion* des zugehörigen *Einschrittverfahrens*. In den bisher angegebenen Beispielen ist die Verfahrensfunktion gegeben durch:

(a) Euler: $\Phi(h,f)(t,u) := f(t,u)$.

(b) Heun: $\Phi(h,f)(t,u) := \frac{1}{2}\left[f(t,u) + f(t+h, u+hf(t,u))\right]$.

(c) Runge-Kutta: $\Phi(h,f)(t,u) := \frac{1}{6}\,[k_1 + 2k_2 + 2k_3 + k_4]$ mit $k_1 := f(t,u)$, $k_2 := f(t+\frac{1}{2}\,h, f(u+\frac{1}{2}\,hk_1))$, $k_3 := f(t+\frac{1}{2}\,h, u+\frac{1}{2}\,hk_2)$, $k_4 := f(t+h, u+hk_3)$.

Im folgenden sei $\|\cdot\|$ eine beliebig gegebene Norm auf dem $\mathbb{R}^n$. Die folgenden Funktionenräume werden in der Konvergenztheorie eine Rolle spielen:

(a) Sei $\mathrm{Lip}[t_0,T]$ die Menge aller stetigen Abbildungen $f\colon [t_0,T]\times\mathbb{R}^n \longrightarrow \mathbb{R}^n$, zu denen eine Konstante $L>0$ mit

$$\|f(t,x)-f(t,y)\| \le L\,\|x-y\| \qquad \text{für alle } (t,x),\,(t,y)\in[t_0,T]\times\mathbb{R}^n$$

existiert.

(b) Sei $F_N[t_0,T]$ die Menge aller gleichmäßig stetigen Abbildungen $f\colon [t_0,T]\times\mathbb{R}^n \longrightarrow \mathbb{R}^n$, deren partielle Ableitungen bis zur Ordnung $N$ auf $[t_0,T]\times\mathbb{R}^n$ existieren und dort gleichmäßig stetig und beschränkt sind.

Dann ist offenbar $F_1[t_0,T] \subset \mathrm{Lip}\,[t_0,T]$. Natürlich ist die Voraussetzung, dass die rechte Seite $f$ einer Anfangswertaufgabe $x' = f(t,x)$, $x(t_0)=x_0$, zu einer Funktionenklasse $F_N[t_0,T]$ gehört, ziemlich einschränkend. "Lokale Versionen" (lokal bezieht sich hier auf die Variable $x$) der obigen Funktionenräume wären angebrachter. Da die Argumentation aber nicht wesentlich anders würde, bleiben wir bei obigen Funktionsklassen.

**Definition 1.1** Für $(t,u)\in[t_0,T]\times\mathbb{R}^n$ sei $z=z(s)$ die Lösung der Anfangswertaufgabe $z' = f(s,z)$, $z(t)=u$. Dann heißt

$$\Delta(h,f)(t,u) := \frac{z(t+h)-z(t)}{h} - \Phi(h,f)(t,u)$$

der *lokale Diskretisierungsfehler*. Das durch die Verfahrensfunktion $\Phi$ definierte Einschrittverfahren heißt *konsistent* (zur gegebenen Anfangswertaufgabe), wenn

$$\lim_{h\to 0+} \Delta(h,f)(t,u) = 0$$

gleichmäßig auf kompakten Teilmengen von $[t_0,T]\times\mathbb{R}^n$ für alle $f\in F_1[t_0,T]$. Das Einschrittverfahren hat die *Konsistenzordnung* $p$, falls $\Delta(h,f)(t,u) = O(h^p)$ gleichmäßig auf kompakten Teilmengen von $[t_0,T]\times\mathbb{R}^n$ für alle $f\in F_p[t_0,T]$.

Es folgt eine einfach nachprüfbare (notwendige und hinreichende) Bedingung für Konsistenz.

**Satz 1.2** *Das Einschrittverfahren mit der Verfahrensfunktion $\Phi$ ist genau dann konsistent zu der gegebenen Anfangswertaufgabe, wenn*

$$\lim_{h\to 0+} \Phi(h,f)(t,u) = f(t,u)$$

*gleichmäßig auf kompakten Teilmengen von $[t_0,T]\times\mathbb{R}^n$.*

**Beweis:** Sei $z$ die Lösung der Anfangswertaufgabe $z' = f(s,z)$, $z(t) = u$. Es ist

$$\begin{aligned}
\|\Delta(h,f)(t,u) + \Phi(h,f)(t,u) - f(t,u)\| &= \left\| \frac{z(t+h)-z(t)}{h} - z'(t) \right\| \\
&= \frac{1}{h} \left\| \int_0^h [z'(t+\tau) - z'(t)]\, d\tau \right\| \\
&= \frac{1}{h} \left\| \int_0^h [f(t+\tau, z(t+\tau)) - f(t,z(t))]\, d\tau \right\| \\
&\leq \max_{\tau\in[0,h]} \|f(t+\tau,z(t+\tau)) - f(t,z(t))\| \\
&\to 0.
\end{aligned}$$

Hieraus liest man die Behauptung ab. □

Die bisher angegebenen Einschrittverfahren sind wegen des eben angegebenen Satzes offenbar sämtlich konsistent.

**Beispiele:** Wir wollen für die bisher angegebenen Einschrittverfahren die Konsistenzordnung berechnen. Bei vorgegebenem $(t,u) \in [t_0,T] \times \mathbb{R}^n$ sei $z$ jeweils die Lösung der Anfangswertaufgabe $z' = f(s,z)$, $z(t) = u$. Beim Eulerschen Polygonzugverfahren ist der lokale Diskretisierungsfehler für $f \in F_1[t_0,T]$ durch

$$\begin{aligned}
\Delta(h,f)(t,u)) &= \frac{z(t+h)-z(t)}{h} - f(t,u) \\
&= z'(t) + O(h) - f(t,u) \\
&= O(h)
\end{aligned}$$

gegeben. Das Euler'sche Polygonzugverfahren ist also ein Verfahren erster Ordnung.

Für $f \in F_2[t_0,T]$ ist beim Verfahren von Heun

$$\begin{aligned}
\Delta(h,f)(t,u) &= \frac{z(t+h)-z(t)}{h} - \frac{1}{2}[f(t,u) + f(t+h, u+hf(t,u))] \\
&= z'(t) + \frac{h}{2} z''(t) + O(h^2) - \frac{1}{2}[f(t,u) + f(t+h,u+hf(t,u))] \\
&= f(t,u) + \frac{h}{2}\frac{d}{ds} f(s,z(s))\Big|_{s=t} + O(h^2) \\
&\qquad - \frac{1}{2}[f(t,u) + f(t,u) + f_t(t,u)h + f_x(t,u)hf(t,u)] + O(h^2) \\
&= f(t,u) + \frac{h}{2}[f_t(t,u) + f_x(t,u) \underbrace{z'(t)}_{=f(t,u)}] + O(h^2) \\
&\qquad - \frac{1}{2}[f(t,u) + f(t,u) + f_t(t,u)h + f_x(t,u)hf(t,u)] + O(h^2) \\
&= O(h^2).
\end{aligned}$$

Das Verfahren von Heun ist also ein Verfahren zweiter Ordnung.

Man kann (mit ziemlich mühsamer Rechnung) zeigen, das das klassische Runge-Kutta-Verfahren die Ordnung vier hat. Ein Verfahren der Ordnung drei ist in Aufgabe 3 angegeben.

Natürlich stellt sich die Frage, ob die "ziemlich mühsame Rechnung" von Maple übernommen werden kann. Dies ist in der Tat der Fall. Wir beginnen mit dem Eulerschen-Polygonzugverfahren, bei dem wir folgendermaßen vorgehen können:

```
> g:=t->f(t,z(t));
```

$$g := t \to \mathrm{f}(t,\, \mathrm{z}(t))$$

```
> Delta:=h->(z(t+h)-z(t))/h-f(t,z(t));
```

$$\Delta := h \to \frac{\mathrm{z}(t+h) - \mathrm{z}(t)}{h} - \mathrm{f}(t,\, \mathrm{z}(t))$$

```
> s:=series(Delta(h),h,2);
```

$$s := (\mathrm{D}(z)(t) - \mathrm{f}(t,\, \mathrm{z}(t))) + \mathrm{O}(h)$$

```
> s1:=subs(D(z)(t)=g(t),s);
```

$$s1 := \mathrm{O}(h)$$

Durch diese Rechnung erhalten wir, dass

$$\Delta(h, f)(t, u)) = O(h),$$

das Euler-Verfahren also ein Verfahren erster Ordnung ist. Entsprechend gehen wir beim Heun-Verfahren vor (wir verzichten diesmal auf das Echo):

```
> g:=t->f(t,z(t)):
> k_1:=h->f(t,z(t)):
> k_2:=h->f(t+h,z(t)+h*k_1(h)):
> Delta:=h->(z(t+h)-z(t))/h-(1/2)*(k_1(h)+k_2(h)):
> s:=series(Delta(h),h,3):
> s1:=subs((D@@2)(z)(t)=D(g)(t),s):
> s2:=subs(D(z)(t)=g(t),s1);
```

$$s2 := \mathrm{O}(h^2)$$

Das Heun-Verfahren ist also ein Verfahren der Ordnung 2. Dass das klassische Runge-Kutta-Verfahren die Ordnung 4 hat, erhält man durch:

```
> restart;
> g:=t->f(t,z(t)):
> k_1:=h->f(t,z(t)):
> k_2:=h->f(t+(h/2),z(t)+(h/2)*k_1(h)):
> k_3:=h->f(t+(h/2),z(t)+(h/2)*k_2(h)):
> k_4:=h->f(t+h,z(t)+h*k_3(h)):
> Delta:=h->(z(t+h)-z(t))/h-(1/6)*(k_1(h)+2*k_2(h)+2*k_3(h)+k_4(h)):
> s:=series(Delta(h),h,5):
> s1:=subs((D@@4)(z)(t)=(D@@3)(g)(t),s):
> s2:=subs((D@@3)(z)(t)=(D@@2)(g)(t),s1):
> s3:=subs((D@@2)(z)(t)=D(g)(t),s2):
> s4:=subs(D(z)(t)=g(t),s3):
> simplify(%);
```

$$\mathrm{O}(h^4)$$

In Aufgabe 3 kann "zu Fuß" und mit Hilfe von Maple nachgewiesen werden, dass ein gewisses Einschrittverfahren die Konsistenzordnung 3 besizt. □

### 3.1.3 Konvergenz von Einschrittverfahren

Nun wollen wir uns mit der *Konvergenz* von Einschrittverfahren zur Lösung der Anfangswertaufgabe

$$x' = f(t,x), \qquad x(t_0) = x_0$$

beschäftigen, wobei wir uns zunächst darüber klar werden müssen, was unter Konvergenz zu verstehen ist. Hierzu geben wir uns ein beliebiges $t \in [t_0, T]$ vor. Man erreicht $t$ von $t_0$ ausgehend in $m$ äquidistanen Schritten der Länge $h_m := (t-t_0)/m$, erhält also eine Näherung $u_m$ für die Lösung $x(t)$ zur Zeit $t$ durch

$$u_0 := x_0, \quad u_{i+1} := u_i + h_m \Phi(h_m, f)(t_0 + ih_m, u_i) \quad (i = 0, \ldots, m-1).$$

Es werden daher die Schrittweiten

$$H_t := \left\{ \frac{t-t_0}{m} : m \in \mathbb{N} \right\}$$

eine besondere Rolle spielen. Mit einem durch die Verfahrensfunktion $\Phi$ definierten Einschrittverfahren kann mit einer Maschenweite $h \in H_t$ die Näherung $u(t;h)$ für die Lösung $x(t)$ berechnet werden. Der Fehler

$$e(t;h) := u(t;h) - x(t)$$

heißt *globaler Diskretisierungsfehler*.

Nun ist naheliegend, was unter Konvergenz zu verstehen ist.

**Definition 1.3** Sei $x(\cdot)$ die Lösung der Anfangswertaufgabe $x' = f(t,x)$, $x(t_0) = x_0$. Ein durch eine Verfahrensfunktion $\Phi$ definiertes Einschrittverfahren heißt *konvergent*, falls

$$\lim_{m\to\infty} u(t;h_m) = x(t) \qquad \text{für alle } t \in [t_0, T] \text{ und alle } f \in F_1[t_0, T].$$

Hierbei ist $h_m := (t-t_0)/m$, während $u(t;h_m)$ aus

- $u_0 := x_0$,
- Für $i = 0, \ldots, m-1$:

$$u_{i+1} := u_i + h_m \Phi(h_m, f)(t_i, u_i), \qquad t_{i+1} := t_i + h_m,$$

- $u(t;h_m) := u_m$

berechnet wird.

Nun kommen wir zum Konvergenzsatz für Einschrittverfahren. In seiner einfachsten Version sagt er aus, dass ein mit der gegebenen Anfangswertaufgabe konsistentes Einschrittverfahren konvergent ist[3]. Diese Version geben wir zunächst an.

[3]Es gilt hier auch die Umkehrung, worauf wir aber nicht eingehen wollen.

**Satz 1.4** *Mit $f \in \mathrm{Lip}[t_0, T]$ und $x_0 \in \mathbb{R}^n$ sei die Anfangswertaufgabe*

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

*gegeben, $x(\cdot)$ sei die eindeutige Lösung. Die Verfahrensfunktion $\Phi(h;t,u)$ (wir werden die Abhängigkeit von $f$ unterdrücken) des betrachteten Einschrittverfahrens sei stetig auf $G := [0,h_0] \times [t_0,T] \times \mathbb{R}^n$, wobei $h_0 > 0$, und lipschitzstetig, es existiere also eine Konstante $M > 0$ mit*

$$\|\Phi(h;t,u) - \Phi(h;t,v)\| \le M \|u - v\| \quad \text{für alle } (h,t,u),\ (h,t,v) \in G.$$

*Ist das Einschrittverfahren konsistent, so ist es auch konvergent.*

**Beweis:** Wir nehmen zunächst an, das Verfahren sei konsistent. Sei $t \in (t_0, T]$ fest gewählt und $h := (t - t_0)/m$ mit einem so großen $m \in \mathbb{N}$, dass $h \le h_0$. Sei $t_i := t_0 + ih$, $i = 0, \ldots, m$. Wie schon mehrfach beschrieben, erhält man $u(t;h)$ aus

- $u_0 := x_0$,
- Für $i = 0, \ldots, m-1$:
  $u_{i+1} := u_i + h\Phi(h;t_i,u_i)$
- $u(t;h) := u_m$

Zur Abkürzung setze man $x_i := x(t_i)$ und $e_i := u_i - x_i$. Wegen

$$\Delta(h;t_i,x_i) = \frac{x_{i+1} - x_i}{h} - \Phi(h;t_i,x_i)$$

ist

$$x_{i+1} = x_i + h[\Phi(h;t_i,x_i) + \Delta(h;t_i,x_i)]$$

und folglich

$$e_{i+1} = u_{i+1} - x_{i+1} = e_i + h[\Phi(h;t_i,u_i) - \Phi(h;t_i,x_i)] - h\Delta(h;t_i,x_i).$$

Mit

$$\sigma(h) := \max_{t \in [t_0,T]} \|\Delta(h;t,x(t))\|$$

ist dann

$$\|e_{i+1}\| \le (1 + hM)\|e_i\| + h\sigma(h), \qquad i = 0, \ldots.$$

Wegen $e_0 = 0$ erhält man durch Zurückspulen

$$\begin{aligned}
\|u(t;h) - x(t)\| &= \|e_m\| \\
&\le h\sigma(h) \sum_{i=0}^{m-1} (1 + hM)^i \\
&= h\sigma(h) \frac{(1+hM)^m - 1}{hM} \\
&\le h\sigma(h) \frac{e^{mhM} - 1}{hM} \\
&= \sigma(h) \frac{e^{M(t-t_0)} - 1}{M}.
\end{aligned}$$

Wegen der vorausgesetzten Konsistenz ist $\lim_{h\to 0+} \sigma(h) = 0$, aus der gerade eben bewiesenen Ungleichungskette folgt die Konvergenz des Einschrittverfahrens. □

**Korollar 1.5** *Mit $f \in F_p[t_0, T]$ (mit $p \in \mathbb{N}$) und $x_0 \in \mathbb{R}^n$ sei die Anfangswertaufgabe*

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

*gegeben, $x(\cdot)$ sei die eindeutige Lösung. Die Verfahrensfunktion $\Phi(h;t,u)$ (wir werden die Abhängigkeit von $f$ unterdrücken) des betrachteten Einschrittverfahrens sei stetig auf $G := [0,h_0] \times [t_0,T] \times \mathbb{R}^n$, wobei $h_0 > 0$. Es gebe positive Konstanten $M$ und $N$ derart, dass*

*(a) Es ist*

$$\|\Phi(h;t,u) - \Phi(h;t,v)\| \le M\,\|u-v\| \quad \text{für alle } (h,t,u),\ (h,t,v) \in G.$$

*(b) Für den lokalen Diskretisierungsfehler (auch hier lassen wir die Abhängigkeit von $f$ fort) gilt die Abschätzung*

$$\|\Delta(h;t,x(t))\| \le Nh^p \quad \text{für alle } t \in [t_0,T],\ h \in [0,h_0].$$

*Dann läßt sich der globale Diskretisierungsfehler $e(t;h) := u(t;h) - x(t)$ abschätzen durch*

$$\|e(t;h)\| \le Nh^p\,\frac{e^{M(t-t_0)}-1}{M}$$

*für alle $t \in [t_0,T]$ und $h = (t-t_0)/m$, wobei $m \in \mathbb{N}$ so groß sei, dass $h \le h_0$.*

**Beweis:** Die Aussage folgt offenbar aus dem Beweis des letzten Satzes, da $\sigma(h) \le Nh^p$ vorausgesetzt wird. □

Theoretisch gibt der letzte Satz die Möglichkeit, nicht nur die qualitative Aussage zu machen, dass der globale Diskretisierungsfehler die gleiche Ordnung wie der lokale Diskretisierungsfehler hat, sondern sogar eine quantitative Fehlerabschätzung anzugeben, zumindestens dann, wenn man geeignete Konstanten $M$ und $N$ kennt, was aber sehr selten der Fall ist.

### 3.1.4 Einschrittverfahren und Extrapolation

Statt vom *Lösen* einer Differentialgleichung spricht man häufig auch von ihrer *Integration*, was naürlich ist, denn die Lösung $x$ von $x' = f(t)$, $x(t_0) = x_0$ erhält man durch das Integral $x(t) = x_0 + \int_{t_0}^t f(s)\,ds$. Durch jedes der vorgestellten Verfahren kann man also auch bestimmte Integrale berechnen. Aus der numerischen Integration wissen wir (eventuell), dass man die Ordnung der (zusammengesetzten) Trapezregel und der Simpson-Regel sukzessive durch *Extrapolation* erhöhen kann, was z. B. auf das Romberg-Verfahren führt. Dass ein solches Vorgehen auch bei der numerischen Behandlung von Anfangswertaufgaben möglich ist, folgt aus dem nächsten Satz. Dieser stammt

von W. Gragg (1963), sein sehr hübscher Beweis von E. Hairer, Ch. Lubich (1984)[4]. In ihm wird die Existenz einer asymptotischen Entwicklung des lokalen Diskretisierungsfehlers vorausgesetzt und auf eine des globalen Diskretisierungsfehlers geschlossen.

**Satz 1.6** *Sei $f \in F_{N+2}[t_0, T]$ und $u(t;h)$ die von dem Einschrittverfahren mit der (hinreichend glatten) Verfahrensfunktion $\Phi$ gelieferte Näherungslösung für $x(t)$, wobei $x(\cdot)$ die Lösung der Anfangswertaufgabe*

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

*ist. Es sei $\Phi(0;t,u) = f(t,u)$ (Konsistenzbedingung), ferner gelte mit einem $p \geq 1$ die folgende Entwicklung des lokalen Diskretisierungsfehlers:*

$$(*)\ \ \Delta(h,t,x(t)) = \frac{x(t+h)-x(t)}{h} - \Phi(h;t,x(t)) = d_p(t)h^p + \cdots + d_N(t)h^N + O(h^{N+1}),$$

*insbesondere habe das Verfahren also die Ordnung $p$. Dann besitzt $u(t;h)$ eine asymptotische Entwicklung der Form*

$$u(t;h) = x(t) + e_p(t)h^p + e_{p+1}(t)h^{p+1} + \cdots + e_N(t)h^N + E_{N+1}(t;h)h^{N+1}$$

*für alle $t \in [t_0, T]$, $h = h_m = (t-t_0)/m$, $m \in \mathbb{N}$. Dabei sind die Funktionen $e_i$ von $h$ unabhängig und das Restglied $E_{N+1}(t;h)$ ist bei festem $t$ für alle $h = h_m$, $m \in \mathbb{N}$, beschränkt.*

**Beweis:** Zunächst wird nur ausgenutzt, dass wegen $(*)$ insbesondere

$$\Delta(h;t,x(t)) = \frac{x(t+h)-x(t)}{h} - \Phi(h;t,x(t)) = d_p(t)h^p + O(h^{p+1}).$$

Wir zeigen, dass eine stetig differenzierbare Funktion $e_p$ existiert mit

$$u(t;h) - x(t) = e_p(t)h^p + O(h^{p+1}).$$

Hierzu definieren wir

$$\hat{u}(t;h) := u(t;h) - e_p(t)h^p$$

mit noch unbestimmter Funktion $e_p$. Dann ist

$$\begin{aligned}
\hat{u}(t+h;h) &= u(t+h;h) - e_p(t+h)h^p \\
&= u(t;h) + h\Phi(h;t,u(t;h)) - e_p(t+h)h^p \\
&= \hat{u}(t;h) + e_p(t)h^p + h\Phi(h;t,\hat{u}(t;h) + e_p(t)h^p) - e_p(t+h)h^p \\
&= \hat{u}(t;h) + h\,\hat{\Phi}(h;t,\hat{u}(t;h))
\end{aligned}$$

mit

$$\hat{\Phi}(h;t,u) := \Phi(h;t,u + e_p(t)h^p) - [e_p(t+h) - e_p(t)]h^{p-1}.$$

[4] Hairer, E., C. Lubich (1984) "Asymptotic expansions of the global error of fixed-stepsize methods." Numer. Math. 45, 345–360.

Also ist $\hat{u}(t;h)$ Resultat eines Schrittes eines Einschrittverfahrens mit der Verfahrensfunktion $\hat{\Phi}$. Nun entwickeln wir den lokalen Diskretisierungsfehler des zu $\hat{\Phi}$ gehörenden Einschrittverfahrens:

$$\begin{aligned}
\hat{\Delta}(h;t,x(t)) &= \frac{x(t+h)-x(t)}{h} - \hat{\Phi}(h;t,x(t)) \\
&= \Delta(h;t,x(t)) + [\Phi(h;t,x(t)) - \hat{\Phi}(h;t,x(t))] \\
&= d_p(t)h^p + O(h^{p+1}) + [\Phi(h;t,x(t)) - \hat{\Phi}(h;t,x(t))] \\
&= [\Phi(h;t,x(t)) - \Phi(h;t,x(t)+e_p(t)h^p)] \\
&\qquad + [d_p(t) + e_p'(t)]h^p + O(h^{p+1}) \\
&= [d_p(t) - f_x(t,x(t))e_p(t) + e_p'(t)]h^p + O(h^{p+1}).
\end{aligned}$$

Hierbei haben wir ausgenutzt, daß

$$\begin{aligned}
\Phi(h;t,x(t)+e_p(t)h^p) &= \Phi(h;t,x(t)) + \Phi_x(h;t,x(t))e_p(t)h^p + O(h^{p+1}) \\
&= \Phi(h;t,x(t)) + \Phi_x(0;t,x(t))e_p(t)h^p + O(h^{p+1}) \\
&= \Phi(h;t,x(t)) + f_x(t,x(t))e_p(t)h^p + O(h^{p+1}).
\end{aligned}$$

Daher bestimme man $e_p$ als Lösung der linearen Anfangswertaufgabe

$$e_p' = f_x(t,x(t))e_p - d_p(t), \qquad e_p(t_0) = 0.$$

Dann ist durch die Verfahrensfunktion $\hat{\Phi}$ ein Einschrittverfahren der Ordnung $p+1$ gegeben, so dass die zugehörige Konvergenzordnung auch mindestens $p+1$ ist. Also ist

$$u(t;h) - x(t) = e_p(t)h^p + \hat{u}(t;h) - x(t) = e_p(t)h^p + O(h^{p+1}).$$

Ist $N = p$, so ist der Beweis abgeschlossen, andernfalls kann die obige Konstruktion fortgesetzt werden, wobei $\Phi$ durch $\hat{\Phi}$ und $p$ durch $p+1$ zu ersetzen ist. □

Eine asymptotische Entwicklung des globalen Diskretisierungsfehlers ist aus mindestens zwei Gründen wichtig. Zum einen kann man in diesem Falle den globalen Diskretisierungsfehler abschätzen bzw. genauer schätzen. Es gelte also etwa

$$u(t;h) - x(t) = e_p(t)h^p + O(h^{p+1}).$$

Hat man mit der Schrittweite $h$ den Näherungswert $u(t;h)$ für $x(t)$ bestimmt, so berechne man anschließend mit einer anderen Schrittweite, z. B. mit der Schrittweite $\frac{1}{2}h$, für dasselbe $t$ den Näherungswert $u(t;\frac{1}{2}h)$. Aus

$$\begin{aligned}
u(t;h) - x(t) &= e_p(t)h^p + O(h^{p+1}), \\
u(t;\tfrac{1}{2}h) - x(t) &= e_p(t)\Big(\frac{h}{2}\Big)^p + O(h^{p+1})
\end{aligned}$$

folgt durch Subtraktion

$$u(t;h) - u(t;\tfrac{1}{2}h) = e_p(t)\Big(\frac{h}{2}\Big)^p(2^p-1) + O(h^{p+1})$$

und anschließend

$$e_p(t)\Big(\frac{h}{2}\Big)^p = \frac{u(t;h) - u(t;\frac{1}{2}h)}{2^p - 1} + O(h^{p+1}).$$

Dann ist schließlich durch

$$u(t;\tfrac{1}{2}h) - x(t) = \frac{u(t;h) - u(t;\frac{1}{2}h)}{2^p - 1} + O(h^{p+1})$$

eine Schätzung des Fehlers $e(t;\frac{1}{2}h)$ gelungen. Beim Runge-Kutta Verfahren ist $p = 4$, hier lautet die entsprechende Schätzung

$$u(t;\tfrac{1}{2}h) - x(t) \approx \frac{u(t;h) - u(t;\frac{1}{2}h)}{15}.$$

**Beispiel:** Wir wollen die Qualität des gerade eben vorgestellten Fehlerschätzers an einem Beispiel testen. Hierzu lösen wir die Anfangswertaufgabe $x' = x - t^2 + 1$, $x(0) = 0.5$, deren Lösung $x(t) = (1+t)^2 - \frac{1}{2}e^t$ ist, mit dem Runge-Kutta-Verfahren mit $h := 0.2$ und $\frac{1}{2}h = 0.1$ auf dem Intervall $[0, 2]$. Wir erhalten die folgenden Werte für $t = 2$:

| $u(2;h)$ | $u(2;\frac{1}{2}h)$ | $x(2)$ |
|---|---|---|
| 5.30536300069265 | 5.30546496022735 | 5.30547195053467 |

und damit

$$\begin{aligned} \frac{u(2;h) - u(t;\frac{1}{2}h)}{15} &= -6.797302313129213e-06, \\ u(2;\tfrac{1}{2}h) - x(2) &= -6.990307323206935e-06. \end{aligned}$$

Das ist also schon eine verblüffend gute Schätzung. □

Die Idee der Extrapolation wollen wir nur ganz kurz beschreiben. Angenommen, mit dem Runge-Kutta-Verfahren seien $u(t;h)$ und $u(t;\frac{1}{2}h)$ berechnet worden. Dann ist

$$\begin{aligned} u(t;h) &= x(t) + e_4(t)h^4 + e_5(t)h^5 + \cdots \\ u(t;\tfrac{1}{2}h) &= x(t) + \tfrac{1}{16}e_4(t)h^4 + \tfrac{1}{32}e_5(t)h^5 + \cdots, \end{aligned}$$

woraus man den (hoffentlich) verbesserten Wert

$$u^*(t;\tfrac{1}{2}h) := \frac{16u(t;\frac{1}{2}h) - u(t;h)}{15} = x(t) - \tfrac{1}{30}e_5(t)h^5 + \cdots$$

erhält.

**Beispiel:** Wir setzen das eben angegebene Beispiel fort und geben auch noch den extrapolierten Wert $u^*(2;\frac{1}{2}h)$ an:

| $u(2;h)$ | $u(2;\frac{1}{2}h)$ | $u^*(2;\frac{1}{2}h)$ | $x(2)$ |
|---|---|---|---|
| 5.30536300069265 | 5.30546496022735 | 5.30547175752966 | 5.30547195053467 |

Der Erfolg ist offensichtlich. □

### 3.1.5 Schrittweitensteuerung

Nun wollen wir einiges zu dem wichtigen Problem der Schrittweitensteuerung sagen, natürlich nach wie vor bei Einschrittverfahren. Einer der Vorteile von Einschrittverfahren gegenüber den im nächsten Abschnitt zu behandelnden Mehrschrittverfahren besteht gerade darin, dass eine Schrittweitensteuerung im Prinzip sehr einfach durchführbar ist. Das Problem besteht darin, dass man zwei Ziele verfolgen will, die sich gegenseitig ausschließen:

- Zum einen will man die Schrittweite möglichst groß wählen, um den Arbeitsaufwand zur Integration einer Anfangswertaufgabe über ein vorgegebenes Zeitintervall möglichst klein zu halten,
- andererseits will man die Lösung möglichst genau berechnen, wozu i. allg. eine kleine Schrittweite nötig ist,
- außerdem will man die Schrittweite den jeweiligen Verhältnissen anpassen, also dort feiner diskretisieren, wo es nötig ist, sonst aber mit einer gröberen Schrittweite arbeiten.

Wir beschreiben nun eine mögliche Form der Schrittweitensteuerung (und folgen hier J. Stoer, R. Bulirsch (1990, S. 127 ff.)).

Gegeben seien $(t_0, x_0)$ und $\epsilon > 0$, gesucht ist eine möglichst große Schrittweite $h > 0$ mit $\|e(t_0+h;h\| \leq \epsilon$ oder wenigstens $\|e(t_0+h;h)\| \approx \epsilon$. Hierbei bedeutet $e(t;h)$ wieder den globalen Diskretisierungsfehler in $t$ bei Verwendung eines Einschrittverfahrens mit der Schrittweite $h$. Hierbei sollte $\epsilon$ nicht zu klein gewählt werden, etwa

$$\epsilon \approx \operatorname{eps} K \qquad \text{mit} \quad K := \max_{t\in[t_0,t_0+h]} \|x(t)\|,$$

wobei eps die Maschinengenauigkeit bedeutet. Die Maschinengenauigkeit $u$ (kleinste positive Zahl mit $1+u \neq 1$) wird häufig durch das folgende Programm approximiert:

```
u=1;
while 1+u~=1
   u=u/2;
end;
u=2*u;
```

Als Resultat erhalten wir (mit `format long g`) `u=2.22044604925031e-16`, was genau mit dem Resultat der MATLAB-Funktion `eps` übereinstimmt. Wir nehmen an, dass ein Verfahren der Ordnung $p$ benutzt wird, so dass $e(t;h) = e_p(t)h^p + O(h^{p+1})$. Wegen $e_p(t_0) = 0$ ist $e_p(t) \approx (t-t_0)e_p'(t_0)$. Es ist

$$e(t_0+h;h) \approx e_p(t_0+h)h^p \approx e_p'(t_0)h^{p+1},$$

wenn also $\|e(t_0+h;h)\| \approx \epsilon$ sein soll, hat man $h$ so zu wählen, dass $\|e_p'(t_0)\|\, h^{p+1} \approx \epsilon$. Um hieraus $h$ zu berechnen, müßte man $\|e_p'(t_0)\|$ kennen, was aber i. allg. nicht der Fall ist. Daher versuchen wir, $\|e_p'(t_0)\|$ zu schätzen.

Man wähle eine Schrittweite $H > 0$ und berechne $u(t_0 + H; H)$ und $u(t_0 + H; \frac{1}{2}H)$. Dann ist (siehe oben)

$$e(t_0 + H; \tfrac{1}{2}H) \approx \frac{u(t_0 + H; H) - u(t_0 + H; \frac{1}{2}H)}{2^p - 1}.$$

Andererseits ist

$$e(t_0 + H; \tfrac{1}{2}H) \approx e_p(t_0 + H)\Big(\frac{H}{2}\Big)^p \approx e_p'(t_0)H\Big(\frac{H}{2}\Big)^p$$

und daher

$$\|e_p'(t_0)\| \approx \frac{1}{H^{p+1}} \frac{2^p}{2^p - 1} \|u(t_0 + H; H) - u(t_0 + H; \tfrac{1}{2}H)\|.$$

Aus $\|e_p'(t_0)\| h^{p+1} \approx \epsilon$ erhält man also $h$ aus

$$\frac{H}{h} \approx \sqrt[p+1]{\frac{2^p}{2^p - 1} \frac{\|u(t_0 + H; H) - u(t_0 + H; \frac{1}{2}H)\|}{\epsilon}}.$$

Ist $H/h \gg 2$, so ist $\|e(t_0 + H; \frac{1}{2}H)\| \gg 2\epsilon$, da

$$\frac{H}{h} \approx \sqrt[p+1]{\frac{2^p}{\epsilon} \|e(t_0 + H; \tfrac{1}{2}H)\|}.$$

Dann ersetzt man $H$ durch den (wesentlich) kleineren Wert $2h$ und beginnt die Rechnung noch einmal. Dies sieht dann also folgendermaßen aus:

- Gegeben $(t_0, x_0)$ (Anfangszeit, Anfangszustand), $H > 0$ (Anfangsschrittweite), $\epsilon > 0$ (gewünschte Genauigkeit), $T > t_0$ (Endzeit, die Lösung soll auf $[t_0, T]$ berechnet werden).

(1) Berechne $u(t_0 + H; H)$ und $u(t_0 + H; \frac{1}{2}H)$. Anschließend berechne man

$$h := H \sqrt[p+1]{\frac{2^p - 1}{2^p} \frac{\epsilon}{\|u(t_0 + H; H) - u(t_0 + H; \frac{1}{2}H)\|}}.$$

(2) Falls $H/h \gg 2$, dann setze $H := 2h$ und gehe zu (1).

(3) Setze $(t_0, x_0) := (t_0 + H, u(t_0 + H; \frac{1}{2}H))$, $H := 2h$.

(4) Falls $t_0 \geq T$, dann: STOP, die Anfangswertaufgabe ist auf $[t_0, T]$ numerisch gelöst. Andernfalls gehe man nach (1).

### 3.1.6 Aufgaben

1. Man zeige: Ist $p \in \Pi_3$ (kubisches Polynom), so ist

$$\int_a^b p(t)\,dt = \frac{b-a}{6}\,[p(a) + 4\,p(\tfrac{1}{2}\,(a+b)) + p(b)].$$

2. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := a_1 f(t,u) + a_2 f(t + b_1 h, u + b_2 h f(t,u))$$

und zeige, dass dieses die Ordnung 2 besitzt, falls

$$a_1 + a_2 = 1, \qquad a_2 b_1 = \frac{1}{2}, \qquad a_2 b_2 = \frac{1}{2}.$$

Spezialfälle erhält man übrigens für $a_1 = 0$, $a_2 = 1$, $b_1 = b_2 = \frac{1}{2}$ (modifiziertes Euler-Verfahren) und für $a_1 = a_2 = \frac{1}{2}$, $b_1 = b_2 = 1$ (Heun-Verfahren).

3. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{4}k_1 + \frac{3}{4}k_3,$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t + \tfrac{1}{3}h, u + \tfrac{1}{3}hk_1), \qquad k_3 := f(t + \tfrac{2}{3}h, u + \tfrac{2}{3}hk_2).$$

Man zeige, dass dies ein Verfahren der Ordnung 3 ist. Hierbei darf man sich auf den Fall einer Differentialgleichung erster Ordnung, also $n = 1$, beschränken. Anschließend löse man diese Aufgabe mit Maple.

4. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{6}(k_1 + 4k_2 + k_3),$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t + \tfrac{1}{2}h, u + \tfrac{1}{2}hk_1), \qquad k_3 := f(t + h, u - hk_1 + 2hk_2).$$

Man zeige, dass dies ein Verfahren der Ordnung 3 ist. Hierbei darf man sich auf den Fall einer Differentialgleichung erster Ordnung, also $n = 1$, beschränken und die Aufgabe mit Maple lösen.

5. Man erläutere, weshalb das Eulersche Polygonzugverfahren, das Verfahren von Heun und auch das klassische Runge-Kutta-Verfahren nicht gut zur numerischen Lösung der einfachen Anfangswertaufgabe $x' = \lambda x$, $x(0) = x_0$, mit kleinem *negativen* $\lambda$ geeignet sind. Man überlege sich, dass dies für das *implizite* Euler-Verfahren

$$u(t+h) = u(t) + hf(t+h, u(t+h))$$

wesentlich besser aussieht. Für $\lambda = -1000$, $x_0 = 1$ und eine Schrittweite $h = 0.01$ mache man sich klar, mit welchen Werten man beim Runge-Kutta-Verfahren bzw. dem impliziten Euler-Verfahren zu rechnen hat.

6. Man bestimme die exakte Lösung der Anfangswertaufgabe $x' = -x^2$, $x(0) = 1$, und vergleiche diese Werte mit dem durch das Runge-Kutta-Verfahren mit den Schrittweite $h = 0.1$ und $h = 0.05$ auf dem Intervall $[0, 1]$ erhaltenen Werte.

7. Man bestimme[5] die exakte Lösung der Anfangswertaufgabe $x' = (2/t)x$, $x(1) = 1$. Anschließend bestimme man einen analytischen Ausdruck für die durch das Eulersche Polygonzugverfahren erhaltene Näherung und gebe den globalen und den lokalen Diskretisierungsfehler an.

8. Man schreibe ein MATLAB-Programm für das klassische Runge-Kutta-Verfahren mit automatischer Schrittweitensteuerung. Anschließend teste man das Programm an der Anfangswertaufgabe (siehe Stoer-Bulirsch)

$$x' = -200tx^2, \qquad x(-3) = \frac{1}{901},$$

welche auf $[-3, 0]$ zu lösen sei.

## 3.2 Mehrschrittverfahren

Bei den bisher betrachteten Einschrittverfahren zur Lösung der Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t, x), \qquad x(t_0) = x_0$$

mit der Lösung $x(\cdot)$ benötigt man zur Berechnung von $u(t + h)$, einer Näherung für $x(t + h)$, lediglich $(t, u(t))$ (und die Schrittweite $h$). Diese Methoden sind einfach und bequem zu programmieren, eine Steuerung der Schrittweite macht keine große Mühe. Andererseits ist es plausibel, dass man Methoden höherer Genauigkeit erhalten kann, wenn man bei der Berechnung von $u(t + h)$ neben $u(t)$ etwa auch $u(t - h)$, $u(t - 2h)$ berücksichtigt. Verfahren dieser Art nennt man *Mehrschrittverfahren*. Sie haben den Nachteil, dass eine gewisse Anzahl von Startwerten durch ein anderes Verfahren zu berechnen sind (z. B. durch ein Runge-Kutta Verfahren mit kleiner Schrittweite). Dieser Nachteil wird i. allg. dadurch aufgewogen, dass die Mehrschrittverfahren bei gleichem Aufwand (dieser wird durch die Anzahl der Funktionswertberechnungen gemessen) eine größere Genauigkeit als entsprechende Einschrittverfahren besitzen. Oder anders gesagt: Um die gleiche Genauigkeit zu erreichen, kommt man bei Mehrschrittverfahren mit einer größeren Schrittweite aus.

Zunächst werden wir einige Beispiele von Mehrschrittverfahren angeben. Danach werden wir genau definieren, was wir unter einem Mehrschrittverfahren verstehen und entsprechend dem Vorgehen bei Einschrittverfahren Konsistenz und Konvergenz definieren. Während bei Einschrittverfahren aus der Konsistenz die Konvergenz folgt, muss bei Mehrschrittverfahren noch eine *Stabilitätsbedingung* erfüllt sein.

[5] Diese Aufgabe ist dem Lehrbuch

R. Kress (1998) *Numerical Analysis*. Springer-Verlag, New York-Berlin-Heidelberg.

entnommen.

### 3.2.1 Beispiele von Mehrschrittverfahren

Gegeben sei die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0,$$

wobei wir über $f \in C([t_0,T]\times\mathbb{R}^n;\mathbb{R}^n)$ wie im letzten Abschnitt voraussetzen, dass $f$ bezüglich der zweiten Variablen global lipschitzstetig ist. Die Lösung $x$ von (P) existiert also eindeutig auf $[t_0,T]$ und sei dort näherungsweise zu berechnen.

Bei einer festen Schrittweite $h > 0$ seien $t_k := t_0 + kh$ äquidistante Stützstellen, $u_k$ seien Näherungen für $x(t_k)$. In einem Mehrschrittverfahren, genauer einem $r$-Schrittverfahren mit $r \in \mathbb{N}$, $r \geq 2$, wird zu $r$ gegebenen Näherungswerten $u_j, \ldots, u_{j+r-1}$ ein Näherungswert $u_{j+r}$ für $x(t_{j+r})$ berechnet. Für den *Start* benötigt man natürlich $r$ Startwerte $u_0 := x_0, u_1, \ldots, u_{r-1}$, die man sich auf andere Weise, etwa ein Runge-Kutta Verfahren mit kleiner Schrittweite oder durch Taylor-Entwicklung verschaffen muss. Letzteres wollen wir uns durch ein Beispiel klar machen.

**Beispiel:** Gegeben sei eine Anfangswertaufgabe beim Lotka-Volterra Modell, also die Aufgabe

$$\begin{array}{rclcrcl} x' &=& ax - bxy, & & x(0) &=& x_0, \\ y' &=& -cy + dxy, & & y(0) &=& y_0. \end{array}$$

Durch Taylor-Entwicklung kann man für kleine $t$ gute Approximationen für $x(t)$ und $y(t)$ erhalten. Denn es ist

$$\begin{array}{rclcl} x'(0) &=& ax_0 - bx_0y_0 &=:& x_0', \\ y'(0) &=& -cy_0 + dx_0y_0 &=:& y_0'. \end{array}$$

Durch Differenzieren des Differentialgleichungssystems erhält man weiter

$$\begin{array}{rclcl} x''(0) &=& ax_0' - b(x_0'y_0 + x_0y_0') &=:& x_0'', \\ y''(0) &=& -cy_0' + d(x_0'y_0 + x_0y_0') &=:& y_0''. \end{array}$$

Damit erhält man

$$\begin{array}{rcl} x(t) &=& x_0 + x_0't + \frac{1}{2}x_0''t^2 + O(t^3), \\ y(t) &=& y_0 + y_0't + \frac{1}{2}y_0''t^2 + O(t^3). \end{array}$$

Dies wollen wir nun noch durch ein Zahlenbeispiel illustrieren. Hierzu sei, wie auch früher schon, $(a,b,c,d) := (2, 0.01, 1, 0.01)$ und $(x_0,y_0) := (300,150)$. Dann ist

$$\begin{array}{rclcr} x_0' &=& ax_0 - bx_0y_0 &=& 150, \\ y_0' &=& -cy_0 + dx_0y_0 &=& 300, \\ x_0'' &=& ax_0' - b(x_0'y_0 + x_0y_0') &=& -825, \\ y_0'' &=& -cy_0' + d(x_0'y_0 + x_0y_0') &=& 825. \end{array}$$

Wir setzen

$$\begin{array}{rclcl} x_0(t) &:=& x_0 + x_0't + \frac{1}{2}x_0''t^2 &=& 300 + 150t - 412.5t^2, \\ y_0(t) &:=& y_0 + y_0't + \frac{1}{2}y_0''t^2 &=& 150 + 300t + 412.5t^2. \end{array}$$

Anschließend vergleichen wir die hierdurch gewonnenen Werte auf dem "kleinen" Intervall $[0, 0.1]$ mit denen, die durch das Runge-Kutta-Verfahren mit der festen Schrittweite $h = 0.001$ gefunden wurden (da wir die exakte Lösung nicht kennen, können wir allerdings nicht beurteilen, was die besseren Werte sind).

| $t$ | $x_0(t)$ | RK | $y_0(t)$ | RK |
|---|---|---|---|---|
| 0.00 | 300.00000000000 | 300.000000000000 | 150.000000000000 | 150.000000000000 |
| 0.02 | 302.83500000000 | 302.829938554061 | 156.165000000000 | 156.166702289822 |
| 0.04 | 305.34000000000 | 305.299462142311 | 162.660000000000 | 162.673193742525 |
| 0.06 | 307.51500000000 | 307.378193603684 | 169.485000000000 | 169.527943609429 |
| 0.08 | 309.36000000000 | 309.036159655878 | 176.640000000000 | 176.737656494651 |
| 0.10 | 310.87500000000 | 310.244238744248 | 184.125000000000 | 184.306888529072 |

Die Werte stimmen einigermaßen überein, wobei die durch das Runge-Kutta-Verfahren gewonnenen Werte etwas "vertrauenswürdiger" zu sein scheinen, da das Intervall $[0, 0.1]$ für die quadratische Taylor-Entwicklung etwas zu groß scheint. Das bestätigt sich, wenn man das Runge-Kutta-Verfahren auch noch mit der Schrittweite $h = 0.0001$ anwendet und praktisch die selben Ergebnisse erhält wie mit der Schrittweite $h = 0.001$. □

Eine Integration von $x'(t) = f(t, x(t))$ über das Intervall $[t_{p-j}, t_{p+k}]$ ergibt

$$x(t_{p+k}) = x(t_{p-j}) + \int_{t_{p-j}}^{t_{p+k}} f(t, x(t))\, dt.$$

Nun ersetze man den Integranden $f(t, x(t))$ durch dasjenige Polynom $P_q \in \Pi_q$ vom Grade $\leq q$, das $f(t, x(t))$ an den Stellen $t_i$, $i = p, p-1, \ldots, p-q$, interpoliert, also den Bedingungen

$$(*) \qquad P_q(t_i) = f(t_i, x(t_i)), \qquad i = p, p-1, \ldots, p-q,$$

genügt (da $f$ eine $n$-Vektorfunktion ist, handelt es sich bei $P_q$ eigentlich um ein "Vektorpolynom", also ein Polynom, dessen Koeffizienten Vektoren sind). Es ist klar, dass $P_q \in \Pi_q$ durch die Interpolationsbedingungen $(*)$ eindeutig bestimmt ist[6]. Das Polynom $P_q$ lässt sich darstellen (die sogenannte Lagrange-Darstellung) in der Form

$$P_q(t) = \sum_{i=0}^{q} f(t_{p-i}, x(t_{p-i}))L_i(t) \qquad \text{mit} \qquad L_i(t) := \prod_{\substack{l=0 \\ l \neq i}}^{q} \frac{t - t_{p-l}}{t_{p-i} - t_{p-l}}.$$

Dies liegt einfach daran, dass $L_i \in \Pi_q$, $i = 0, \ldots, q$, und $L_i(t_{p-j}) = \delta_{ij}$. Für eine Lösung $x$ der Differentialgleichung $x' = f(t, x)$ und *äquidistante* Stützstellen $t_k := t_0 + kh$,

[6]Die Interpolationsbedingungen $(*)$ können äquivalent geschrieben werden als ein lineares Gleichungssystem (etwa für die Koeffizienten einer Monombasis von $\Pi_q$) mit ebenso vielen Gleichungen wie Unbekannten. Daher genügt es nachzuweisen, dass das homogene Interpolationsproblem ($f = 0$) nur die triviale Lösung hat. Das ist aber klar, da ein Polynom aus $\Pi_q$ mit $q+1$ paarweise verschiedenen Nullstellen notwendigerweise das Nullpolynom ist.

$k = 0, 1, \ldots$, bei vorgegebener Maschenweite $h > 0$ ist daher

$$\begin{aligned} x(t_{p+k}) &= x(t_{p-j}) + \int_{t_{p-j}}^{t_{p+k}} f(t, x(t))\, dt \\ &\approx x(t_{p-j}) + \int_{t_{p-j}}^{t_{p+k}} P_q(t)\, dt \\ &= x(t_{p-j}) + \sum_{i=0}^{q} f(t_{p-i}, x(t_{p-i})) \int_{t_{p-j}}^{t_{p+k}} L_i(t)\, dt \\ &= x(t_{p-j}) + h \sum_{i=0}^{q} \beta_{qi} f(t_{p-i}, x(t_{p-i})) \end{aligned}$$

mit (jetzt nützen wir zum ersten Mal aus, dass die Stützstellen äquidistant sind)

$$\beta_{qi} := \frac{1}{h} \int_{t_{p-j}}^{t_{p+k}} L_i(t)\, dt = \frac{1}{h} \int_{t_p - jh}^{t_p + kh} \prod_{\substack{l=0 \\ l \neq i}}^{q} \frac{t + lh - t_p}{(l-i)h}\, dt = \int_{-j}^{k} \prod_{\substack{l=0 \\ l \neq i}}^{q} \frac{s+l}{-i+l}\, ds,$$

wobei wir die letzte Gleichung mit der Variablentransformation $s = (t - t_p)/h$ erhalten haben. Damit erhält man das von $k, j, q$ abhängende Mehrschrittverfahren

$$u_{p+k} = u_{p-j} + h \sum_{i=0}^{q} \beta_{qi} f(t_{p-i}, u_{p-i}).$$

Je nach Wahl von $k, j, q$ erhält man die folgenden Verfahren.

- Adams-Bashforth (explizit).

Hier ist $k = 1$, $j = 0$, $q = 0, 1, \ldots$ und daher

$$\beta_{qi} = \int_0^1 \prod_{\substack{l=0 \\ l \neq i}}^{q} \frac{s+l}{-i+l}\, ds, \qquad i = 0, \ldots, q.$$

Man erhält die folgende Tabelle.

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\beta_{0i}$ | 1 | | | | |
| $2\beta_{1i}$ | 3 | $-1$ | | | |
| $12\beta_{2i}$ | 23 | $-16$ | 5 | | |
| $24\beta_{3i}$ | 55 | $-59$ | 37 | $-9$ | |
| $720\beta_{4i}$ | 1901 | $-2774$ | 2616 | $-1274$ | 251 |

und hiermit das $(q+1)$-Schrittverfahren

$$u_{p+1} = u_p + h \sum_{i=0}^{q} \beta_{qi} f(t_{p-i}, u_{p-i}).$$

Im einfachsten Fall ($q = 0$) hat man das Euler'sche Polygonzugverfahren, der nächst einfache Fall ($q = 1$) ist

$$u_{p+1} = u_p + \frac{h}{2}[3f(t_p, u_p) - f(t_{p-1}, u_{p-1})].$$

- Adams-Moulton (implizit).

Hier ist $k = 0$, $j = 1$, $q = 0, 1, \ldots$ und daher

$$\beta_{qi} = \int_{-1}^{0} \prod_{\substack{l=0 \\ l \neq i}}^{q} \frac{s+l}{-i+l}\, ds.$$

Man erhält die folgende Tabelle.

| $i$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\beta_{0i}$ | 1 | | | | |
| $2\beta_{1i}$ | 1 | 1 | | | |
| $12\beta_{2i}$ | 5 | 8 | $-1$ | | |
| $24\beta_{3i}$ | 9 | 19 | $-5$ | 1 | |
| $720\beta_{4i}$ | 251 | 646 | $-264$ | 106 | $-19$ |

und damit das Verfahren

$$u_p = u_{p-1} + h \sum_{i=0}^{q} \beta_{qi} f(t_{p-i}, u_{p-i}).$$

Ersetzt man hier $p$ durch $p + 1$, so lautet das Verfahren

$$u_{p+1} = u_p + h \sum_{i=0}^{q} \beta_{qi} f(t_{p+1-i}, u_{p+1-i}).$$

Man erkennt, dass das Adams-Moulton Verfahren *implizit* ist, d. h. zur Berechnung von $u_{p+1}$ hat man ein nichtlineares Gleichungssystem zu lösen. Wenn man z. B. die globale Lipschitzstetigkeit von $f$ bezüglich der zweiten Komponente voraussetzen, also die Existenz einer Konstanten $L > 0$ mit

$$\|f(t, x) - f(t, y)\| \leq L \|x - y\| \qquad \text{für alle } (t, x),\ (t, y) \in [t_0, T] \times \mathbb{R}^n,$$

wie wir es im letzten Abschnitt getan haben, so folgt aus dem Kontraktionssatz für hinreichend kleines $h$ die Existenz einer eindeutigen Lösung dieses nichtlinearen Gleichungssystems. Im einfachsten Fall erhält man das implizite Eulersche Verfahren, der nächst einfache Fall ($q = 1$) ist die Trapezregel

$$u_{p+1} = u_p + \frac{h}{2}[f(t_{p+1}, u_{p+1}) + f(t_p, u_p)].$$

Gewöhnlich kombiniert man ein explizites Adams-Bashforth Verfahren und ein implizites Adams-Moulton Verfahren zu einem sogenannten *Prädiktor-Korrektor Verfahren*.

Durch das explizite Verfahren wird ein Startwert $\tilde{u}_{p+1}$ für das implizite Verfahren gewonnen, dieser wird durch (gewöhnlich nur eine) Iteration mit dem impliziten Verfahren verbessert.

**Beispiel:** Kombiniert man Adams-Bashforth mit $q = 0$ (bzw. das Euler'sche Polygonzugverfahren) mit Adams-Moulton mit $q = 1$, so erhält man

$$\begin{aligned} \tilde{u}_{p+1} &= u_p + hf(t_p, u_p), \\ u_{p+1} &= u_p + \frac{h}{2}[f(t_{p+1}, \tilde{u}_{p+1}) + f(t_p, u_p)]. \end{aligned}$$

Das entstehende Verfahren ist offenbar genau das Verfahren von Heun. □

**Beispiel:** Ein häufig benutztes Prädiktor-Korrektor Verfahren ist das folgende:

$$\begin{aligned} \tilde{u}_{p+1} &= u_p + \frac{h}{12}[23f(t_p, u_p) - 16f(t_{p-1}, u_{p-1}) + 5f(t_{p-2}, u_{p-2})] \\ &\quad \text{(Adams-Bashforth, } q = 2) \\ u_{p+1} &= u_p + \frac{h}{24}[9f(t_{p+1}, \tilde{u}_{p+1}) + 19f(t_p, u_p) - 5f(t_{p-1}, u_{p-1}) + f(t_{p-2}, u_{p-2})] \\ &\quad \text{(Adams-Moulton, } q = 3). \end{aligned}$$

Bei jedem Schritt benötigt man hier zwei Funktionsauswertungen, in seiner Komplexität ist es also mit dem Heun-Verfahren gleichzusetzen. Zum Start benötigt man die Werte $u_0, u_1, u_2$. □

I. allg. verwendet man Mehrschrittverfahren in der Form von Prädiktor-Korrektor Verfahren. Einer der Vorteile gegenüber Einschrittverfahren besteht dann darin, dass selbst bei hoher Konsistenzordnung (Definition folgt unten) nur zwei Funktionsauswertungen pro Zeitschritt zu machen sind. Zu den Nachteilen zählt, dass eine Startrechnung erforderlich und eine gute Schrittweitensteuerung schwierig ist, da nach Konstruktion die Zeitschritte äquidistant sind.

### 3.2.2 Konsistenz, Konvergenz und Stabilität: Definitionen

Zunächst wollen wir genauer ein Mehrschritt- bzw. ein $r$-Schrittverfahren definieren. Gegeben sei die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t, x), \qquad x(t_0) = x_0$$

mit der Lösung $x(\cdot)$. Sei $h > 0$ eine feste Schrittweite und $t_j := t_0 + jh$. Mit $u_k$ wird eine Näherung für $x_k := x(t_k)$ bezeichnet. Bei einem $r$-Schrittverfahren sind

$$\begin{aligned} u_0 &= x_0 + \epsilon_0 \\ u_1 &= x_1 + \epsilon_1 \\ &\vdots \\ u_{r-1} &= x_{r-1} + \epsilon_{r-1} \end{aligned}$$

gegeben, wobei i. allg. $\epsilon_0 = 0$ (bis auf Rundungsfehler) und $u_1, \ldots, u_{r-1}$ durch eine Startrechnung gewonnen sind. Man berechne $u_r, u_{r+1}, \ldots$ aus

$$\text{(MV)} \quad u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j = hF(h,f)(u_{j+r}, u_{j+r-1}, \ldots, u_j, t_j), \quad j = 0, 1, \ldots.$$

Man spricht von einem *linearen $r$-Schrittverfahren*, wenn, wie in den obigen Beispielen, die Verfahrensfunktion $F$ linear von $f$ abhängt, wenn also

$$F(h,f)(u_{j+r}, u_{j+r-1}, \ldots, u_j, t_j) = b_r f(t_{j+r}, u_{j+r}) + \cdots + b_0 f(t_j, u_j).$$

Nun übertragen wir die Definitionen (Konsistenz, lokaler Diskretisierungsfehler, Konvergenz) des letzten Abschnitts auf allgemeine (nicht notwendig lineare) Mehrschrittverfahren.

**Definition 2.1** Bei vorgegebenem $(t,u) \in [t_0, T] \times \mathbb{R}^n$ sei $z(\cdot)$ die Lösung von $z' = f(s,z)$, $z(t) = u$. Dann heißt

$$\Delta(h,f)(t,u) := \frac{1}{h}\Big[z(t+rh) + \sum_{i=0}^{r-1} a_i z(t+ih)\Big] - F(h,f)(z(t+rh), \ldots, z(t), t)$$

der *lokale Diskretisierungsfehler* des Mehrschrittverfahrens (MV) an der Stelle $(t,u)$. Das Mehrschrittverfahren (MV) heißt *konsistent* (zur gegebenen Anfangswertaufgabe), wenn

$$\lim_{h \to 0+} \Delta(h,f)(t,u) = 0$$

gleichmäßig auf kompakten Teilmengen von $[t_0, T] \times \mathbb{R}^n$ für alle $f \in F_1[t_0, T]$. Das Einschrittverfahren hat die *Konsistenzordnung* $p$, falls $\Delta(h,f)(t,u) = O(h^p)$ gleichmäßig auf kompakten Teilmengen von $[t_0, T] \times \mathbb{R}^n$ für alle $f \in F_p[t_0, T]$.

Diese Definition verallgemeinert die entsprechende Definition 1.1 von Ein- auf Mehrschrittverfahren. Genau wie bei Einschrittverfahren gibt der lokale Diskretisierungsfehler an, wie gut die exakte Lösung der Differentialgleichung der Differenzengleichung genügt.

Wir wollen nur einige wenige Beispiele betrachten.

- Adams-Bashforth.

Für $q = 0$ hat man das Euler'sche Polygonzugverfahrern, welches die Ordnung 1 hat. Für $q = 1$ lautet das Verfahren

$$u_{j+2} = u_{j+1} + \frac{h}{2}[3f(t_{j+1}, u_{j+1}) - f(t_j, u_j)].$$

Für den lokalen Diskretisierungsfehler erhält man

$$\begin{aligned} \Delta(h,f)(t,u) &= \frac{1}{h}[z(t+2h) - z(t+h)] - \frac{1}{2}[3f(t+h, z(t+h)) - f(t, z(t))] \\ &= \frac{1}{h}[z(t+2h) - z(t+h)] - \frac{1}{2}[3z'(t+h) - z'(t)] \end{aligned}$$

$$\begin{aligned} &= \frac{1}{h}[z(t) + 2hz'(t) + 2h^2z''(t) + \tfrac{4}{3}h^3z'''(t) \\ &\qquad - z(t) - hz'(t) - \tfrac{1}{2}h^2z''(t) - \tfrac{1}{6}h^3z'''(t) + O(h^4)] \\ &\qquad - \frac{1}{2}[2z'(t) + 3hz''(t) + \tfrac{3}{2}h^2z'''(t) + O(h^3)] \\ &= \tfrac{5}{12}h^2z'''(t) + O(h^3). \end{aligned}$$

Hieraus liest man die Konsistenzordnung 2 ab. Für $q = 2$ hat man das Verfahren

$$u_{j+3} = u_{j+2} + \frac{h}{12}[23f(t_{j+2}, u_{j+2}) - 16f(t_{j+1}, u_{j+1}) + 5f(t_j, u_j)],$$

nach geringer Mühe erhält man für den lokalen Diskretisierungsfehler

$$\begin{aligned} \Delta(h, f)(t, u) &= \frac{1}{h}[z(t + 3h) - z(t + 2h)] \\ &\qquad - \frac{1}{12}[23z'(t + 2h) - 16z'(t + h) + 5z'(t)] \\ &\vdots \\ &= \tfrac{3}{8}h^3z^{(4)}(t) + O(h^4), \end{aligned}$$

das Verfahren hat also die Ordnung 3. Allgemein kann man zeigen, dass das $r$-Schritt Adams-Bashforth Verfahren die Ordnung $r$ hat.

**Beispiel:** Wir wollen mit Hilfe von Maple nachweisen, dass obiges Adams-Bashforth-Verfahren mit $q = 2$ die Konsistenzordnung 3 besitzt. Das ist hier wesentlich einfacher als bei etwa dem klassischen Runge-Kutta-Verfahren: Nach

```
Delta:=h->(z(t+3*h)-z(t+2*h))/h-(23*D(z)(t+2*h)-16*D(z)(t+h)+D(z)(t))/12:
series(Delta(h),h,5);
```

erhalten wir den Output

$$\frac{3}{8}(D^{(4)})(z)(t)h^3 + O(h^4).$$

Hier wird also noch einmal das oben angegebene Ergebnis bestätigt. □

- Adams-Moulton.

Für $q = 0$ lautet das Verfahren

$$u_{j+1} = u_j + hf(t_{j+1}, u_{j+1}).$$

Wegen

$$\begin{aligned} \Delta(h, f)(t, u) &= \frac{1}{h}[z(t + h) - z(t)] - z'(t + h) \\ &= -\tfrac{1}{2}hz''(t) + O(h^2) \end{aligned}$$

ist die Ordnung 1. Für $q = 1$ hat man das Verfahren

$$u_{j+1} = u_j + \frac{h}{2}[f(t_{j+1}, u_{j+1}) + f(t_j, u_j)]$$

mit dem lokalen Diskretisierungsfehler

$$\Delta(h, f)(t, u) = -\tfrac{1}{12}h^2 z'''(t) + O(h^3).$$

Für $q = 2$ lautet das Verfahren

$$u_{j+2} = u_{j+1} + \frac{h}{12}[5f(t_{j+2}, u_{j+2}) + 8f(t_{j+1}, u_{j+1}) - f(t_j, u_j)],$$

es hat den lokalen Diskretisierungsfehler

$$\Delta(h, f)(t, u) = -\tfrac{1}{24}h^3 z^{(4)}(t) + (h^4)$$

und damit die Ordnung 3. Für $q = 3$ hat man das Verfahren

$$u_{j+3} = u_{j+2} + \frac{h}{24}[9f(t_{j+3}, u_{j+3}) + 19f(t_{j+2}, u_{j+2}) - 5f(t_{j+1}, u_{j+1}) + f(t_j, u_j)],$$

den lokalen Diskretisierungsfehler

$$\Delta(h, f)(t, u) = -\tfrac{19}{720}h^4 z^{(5)}(t) + O(h^5)$$

und damit die Ordnung 4. Allgemein kann man zeigen, dass das $r$-Schritt Adams-Moulton Verfahren die Ordnung $r + 1$ besitzt (bis auf $q = 0$, hier ergibt sich ein implizites Einschrittverfahren der Ordnung 1).

In einem Beispiel hatten wir ein Prädiktor-Verfahren der Ordnung 3 (Adams-Bashforth mit $q = 2$) mit einem Korrektor-Verfahren der Ordnung 4 (Adams-Moulton mit $q = 3$) kombiniert:

$$\begin{aligned}
\tilde{u}_{j+3} &= u_{j+2} + \frac{h}{12}[23f(t_{j+2}, u_{j+2}) - 16f(t_{j+1}, u_{j+1}) + 5f(t_j, u_j)], \\
u_{j+3} &= u_{j+2} + \frac{h}{24}[9f(t_{j+3}, \tilde{u}_{j+3}) + 19f(t_{j+2}, u_{j+2}) - 5f(t_{j+1}, u_{j+1}) + f(t_j, u_j)].
\end{aligned}$$

Wir wollen uns überlegen, dass dieses Verfahren die Ordnung 4 besitzt. Wir bezeichnen mit $\Delta_{PK}$ den lokalen Diskretisierungsfehler dieses Prädiktor-Korrektor Verfahrens, mit $\Delta_P$ den des Prädiktor-Verfahrens und mit $\Delta_K$ den des Korrektor-Verfahrens. Es ist

$$\begin{aligned}
u_{j+3} &= u_{j+2} + \frac{h}{24}[9f(t_{j+3}, u_{j+3}) + 19f(t_{j+2}, u_{j+2}) - 5f(t_{j+1}, u_{j+1}) + f(t_j, u_j)] \\
&\quad + \frac{9h}{24}[f(t_{j+3}, \tilde{u}_{j+3}) - f(t_{j+3}, u_{j+3})]
\end{aligned}$$

und daher

$$\begin{aligned}
\Delta_{PK}(h, f)(t, u) &= \Delta_K(f, h)(t, u) + \frac{9}{24}[f(t + 3h, z(t + 3h)) \\
&\quad - f(t + 3h, z(t + 2h) + \tfrac{1}{12}h[23z'(t + 2h) - 16z'(t + h) + 5z'(t)])]
\end{aligned}$$

und daher

$$\|\Delta_{PK}(h,f)(t,u)\| \le \|\Delta_K(h,f)(t,u)\| + \frac{9}{24}Lh\,\|\Delta_P(h,f)(t,u)\|,$$

wobei $L$ die (globale) Lipschitzkonstante von $f$ bezüglich des zweiten Arguments ist. Hieraus liest man nicht nur die Behauptung, sondern auch das allgemeine Prinzip ab: Die Kombination eines Adams-Bashforth-Verfahrens der Ordnung $p$ mit einem Adams-Moulton-Verfahren der Ordnung $p+1$ zu einem Prädiktor-Korrektor-Verfahren liefert ein Verfahren der Ordnung $p+1$.

Wir kommen nun zur Definition der Konvergenz eines Mehrschrittverfahrens.

**Definition 2.2** Gegeben sei die Anfangswertaufgabe

$$\text{(P)} \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

mit der Lösung $x(\cdot)$ auf $[t_0,T]$. Das Mehrschrittverfahren

$$\text{(MV)} \quad u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j = hF(h,f)(u_{j+r},\ldots,u_j,t_j), \qquad j = 0,1,\ldots$$

heißt *konvergent*, falls für alle $f \in F_1[t_0,T]$ gilt: Ist $t \in [t_0,T]$, $h_m := (t-t_0)/m$, $m = r, r+1, \ldots$, $t_i := t_0 + ih$,

$$\epsilon(h_m) = \begin{pmatrix} \epsilon_0(h_m) \\ \vdots \\ \epsilon_{r-1}(h_m) \end{pmatrix} \to 0 \quad \text{mit} \quad m \to \infty,$$

und wird $u_m = u(t;\epsilon,h_m)$ gewonnen aus

$$\begin{aligned} u_0 &= x(t_0) + \epsilon_0(h_m) \\ &\vdots \\ u_{r-1} &= x(t_{r-1}) + \epsilon_{r-1}(h_m) \end{aligned}$$

und

$$u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j = hF(h,f)(u_{j+r},\ldots,u_j,t_j), \qquad j = 0,\ldots,m-r,$$

so gilt

$$\lim_{m\to\infty} u(t;\epsilon,h_m) = x(t).$$

**Bemerkung:** Ein konsistentes Einschrittverfahren ist auch konvergent und die Ordnung des lokalen Diskretisierungsfehlers (bzw. die Konsistenzordnung) ist gleich der Ordnung des globalen Diskretisierungsfehlers (bzw. der Konvergenzordnung). Eine entsprechende Aussage ist bei Mehrschrittverfahren nicht richtig! Als Beispiel betrachte man (siehe J. Stoer, R. Bulirsch (1990, S. 141 ff.)) das 2-Schrittverfahren

$$u_{j+2} + 4u_{j+1} - 5u_j = h[4f(t_{j+1},u_{j+1}) + 2f(t_j,u_j)].$$

Bei den folgenden Untersuchungen benutzen wir Maple, wobei die Rechnungen natürlich auch "per hand" gemacht werden könnten. Der lokale Diskretisierungsfehler ist

$$\begin{aligned}\Delta(h,f)(t,u) &= \frac{1}{h}[z(t+2h)+4z(t+h)-5z(t)]-[4z'(t+h)+2z'(t)]\\ &= \frac{1}{6}h^3 z^{(4)}(t)+O(h^4).\end{aligned}$$

Dies erhalten wir durch

```
Delta:=h->(z(t+2*h)+4*z(t+h)-5*z(t))/h-(4*D(z)(t+h)+2*D(z)(t)):
s:=series(Delta(h),h,5);
```

Es handelt sich hier also um ein 2-Schrittverfahren der Ordnung 3. Dieses Verfahren werde auf die Anfangswertaufgabe

$$x' = -x, \qquad x(0) = 1$$

mit der exakten Lösung $x(t) = e^{-t}$ angewandt, wobei wir als Startwerte $u_0, u_1$ die exakten Werte $u_0 = 1$, $u_1 = e^{-h}$ nehmen. Die $u_j$ werden also berechnet aus

$$u_0 := 1, \quad u_1 := e^{-h}$$

sowie

$$(*) \qquad u_{j+2} + 4(1+h)u_{j+1} + (-5+2h)u_j = 0, \qquad j = 0, 1, \ldots.$$

Für die Lösung der Differenzengleichung $(*)$ mache man den Ansatz $u_j = \lambda^j$. Dies ist genau dann eine Lösung von $(*)$, wenn

$$\lambda^j[\lambda^2 + 4(1+h)\lambda + (-5+2h)] = 0$$

und diese Gleichung besitzt die beiden nichttrivialen Lösungen

$$\begin{aligned}\lambda_1(h) &:= -2-2h+\sqrt{9+6h+4h^2},\\ \lambda_2(h) &:= -2-2h-\sqrt{9+6h+4h^2}.\end{aligned}$$

Zur Not (wenn man nicht weiß, wie man quadratische Gleichungen löst) hilft hier

```
solve(lambda^2+4*(1+h)*lambda+(-5+2*h),lambda);
```

Da die Differenzengleichung linear ist, erhält man $u_j$ als Linearkombination von $\lambda_1(h)^j$ und $\lambda_2(h)^j$:

$$u_j = \alpha(h)\lambda_1(h)^j + \beta(h)\lambda_2(h)^j, \qquad j = 0, 1, \ldots.$$

Hierbei sind $\alpha, \beta$ durch die Anfangsbedingungen $u_0 = 1$, $u_1 = e^{-h}$ festgelegt sind. Dies ergibt

$$\alpha(h) := \frac{\lambda_2(h) - e^{-h}}{\lambda_2(h) - \lambda_1(h)}, \qquad \beta(h) := \frac{e^{-h} - \lambda_1(h)}{\lambda_2(h) - \lambda_1(h)}.$$

Insgesamt ist also

$$u_j = \frac{\lambda_2(h) - e^{-h}}{\lambda_2(h) - \lambda_1(h)}\lambda_1(h)^j + \frac{e^{-h} - \lambda_1(h)}{\lambda_2(h) - \lambda_1(h)}\lambda_2(h)^j.$$

Mit dem Maple-Befehl `series` erhält man sehr leicht (wir geben mehr Terme als nötig an), dass

$$\lambda_1(h) = 1 - h + \frac{1}{2}h^2 - \frac{1}{6}h^3 + \frac{1}{72}h^4 + O(h^5), \qquad \lambda_2(h) = -5 - 3h - \frac{1}{2}h^2 + O(h^3)$$

und

$$\alpha(h) = 1 + \frac{1}{216}h^4 + O(h^5), \qquad \beta(h) = -\frac{1}{216}h^4 + O(h^5).$$

Sei nun $t \neq 0$ fest, $h_m = t/m$, $m = 1, 2, \ldots$. Dann ist

$$\begin{aligned} u_m &= u(t; h_m) \\ &= \frac{\lambda_2(t/m) - e^{-t/m}}{\lambda_2(t/m) - \lambda_1(t/m)}\lambda_1(t/m)^m + \frac{e^{-t/m} - \lambda_1(t/m)}{\lambda_2(t/m) - \lambda_1(t/m)}\lambda_2(t/m)^m \\ &= [1 + O((t/m)^4)][1 - t/m + O((t/m)^2)]^m \\ &\qquad - \tfrac{1}{216}(t/m)^4[1 + O(t/m)][-5 - 3t/m + O((t/m)^2)]^m. \end{aligned}$$

Der erste Term strebt für $m \to \infty$ gegen $e^{-t}$, der zweite verhält sich für $m \to \infty$ wie

$$-\frac{t^4}{216}\,\frac{(-5)^m}{m^4}e^{3t/5}$$

und dieser "strebt oszillierend" gegen $\pm\infty$, es liegt also Divergenz vor. Der Grund hierfür ist, dass $|\lambda_2(0)| = |-5| > 1$. Es kommt also offenbar auf die Nullstellen von $\mu^2 + 4\mu - 5$ bzw., im allgemeinen Fall, von $\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0$ an. □

Die folgende Definition wird sich als entscheidend für Konvergenzuntersuchungen herausstellen.

**Definition 2.3** Das Mehrschrittverfahren

$$\text{(MV)} \quad u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j = hF(h, f)(u_{j+r}, \ldots, u_j, t_j), \qquad j = 0, 1, \ldots$$

genügt der *Stabilitätsbedingung*, falls für das Polynom

$$\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0$$

gilt:

1. Alle Nullstellen von $\psi$ sind dem Betrage nach kleiner oder gleich 1.
2. Ist $\lambda \in \mathbb{C}$ eine Nullstelle von $\psi$ mit $|\lambda| = 1$, so ist $\lambda$ eine einfache Nullstelle bzw. $\psi'(\lambda) \neq 0$.

Unschwer stellt man fest, dass die Mehrschrittverfahren von Adams-Bashforth und Adams-Moulton die Stabilitätsbedingung erfüllen, während das Verfahren aus dem letzten Beispiel dies nicht tut.

### 3.2.3 Der Äquivalenzsatz

Ziel dieses Unterabschnittes ist es, den folgenden Satz zu beweisen.

**Satz 2.4** *Das Mehrschrittverfahren*

$$\text{(MV)} \qquad u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j = hF(h,f)(u_{j+r}, \ldots, u_j, t_j)$$

*sei konsistent. $F$ genüge den folgenden Bedingungen:*

*(a) Ist $f \in F_1[t_0, T]$, so existieren Konstanten $h_0 > 0$ und $M$ derart, daß*

$$\|F(h,f)(u_r, \ldots, u_0, t) - F(h,f)(v_r, \ldots, v_0, t)\| \le M \sum_{i=0}^{r} \|u_i - v_i\|$$

$$\text{für alle } t \in [t_0, T],\ h \in [0, h_0],\ u_i, v_i \in \mathbb{R}^n,\ i = 0, \ldots, r.$$

*(b) $F(h,0)(u_r, \ldots, u_0, t) \equiv 0$.*

*Dann ist das Mehrschrittverfahren* (MV) *(für jedes $f \in F_1[t_0, T]$) genau dann konvergent, wenn es der Stabilitätsbedingung genügt.*

**Bemerkung:** Ist

$$F(h,f)(u_{j+r}, \ldots, u_j, t_j) = b_r f(t_{j+r}, u_{j+r}) + \cdots + b_0 f(t_j, u_j),$$

das Mehrschrittverfahren (MV) also linear, so sind die Bedingungen (a), (b) im obigen Satz offenbar erfüllt. Das gleiche gilt auch für Prädiktor-Korrektor-Verfahren vom Adams-Bashforth-Moulton Typ. □

Zum Beweis des obigen, sogenannten *Äquivalenzsatzes* benötigen wir einen Hilfssatz über Matrizen, den wir zunächst formulieren und beweisen. Für eine Matrix $A \in \mathbb{C}^{r\times r}$ bezeichnen wir hierbei mit $\rho(A)$ den *Spektralradius* der Matrix $A$, d. h. es sei

$$\rho(A) := \max\{|\lambda| : \lambda \text{ ist Eigenwert von } A\}.$$

**Lemma 2.5** *Sei $A \in \mathbb{C}^{r\times r}$ eine Matrix mit den folgenden beiden Eigenschaften:*

*(a) Es ist $\rho(A) = 1$, d. h. alle Eigenwerte von $A$ sind betragsmäßig kleiner oder gleich 1, mindestens einer hat den Betrag 1.*

*(b) Ist $\lambda$ ein Eigenwert von $A$ mit $|\lambda| = 1$, so ist $\lambda$ ein einfacher (die Vielfachheit von $\lambda$ als Nullstelle des charakteristischen Polynoms ist 1) Eigenwert von $A$.*

*Dann existiert eine Vektornorm $\|\cdot\|$ auf $\mathbb{C}^r$ derart, dass $\|A\| = \rho(A) = 1$ für die zugehörige Matrixnorm gilt.*

**Beweis:** Die Matrix $A$ läßt sich durch eine Ähnlichkeitstransformation auf Jordansche Normalform bringen. Also existiert eine nichtsinguläre Matrix $P \in \mathbb{C}^{r\times r}$ mit $P^{-1}AP = J$, wobei

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & J_p \end{pmatrix}$$

eine Blockdiagonalmatrix mit

$$J_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix} \in \mathbb{R}^{r_i\times r_i}, \qquad \sum_{i=1}^{p} r_i = r,$$

ist. Nach Voraussetzung sind die Eigenwerte von $A$ auf dem Rande des Einheitskreises einfach, die zugehörigen Jordan-Blöcke sind also $1 \times 1$-"Blöcke". Wir machen eine Fallunterscheidung.

- Alle Eigenwerte von $A$ liegen auf dem Rand des Einheitskreises.

In diesem Falle sind alle Eigenwerte von $A$ einfach und daher $A$ diagonalisierbar. Daher ist $p = r$ und $J = \text{diag}\,(\lambda_1, \ldots, \lambda_r)$. Man definiere die transformierte Maximumnorm $\|x\| := \|P^{-1}x\|_\infty$. Die zugehörige Maximumnorm ist dann

$$\|A\| = \|P^{-1}AP\|_\infty = \|J\|_\infty = \max_{i=1,\ldots,r} |\lambda_i| = 1,$$

in diesem Fall ist der Satz also richtig.

- Mindestens ein Eigenwert von $A$ liegt im Innern des Einheitskreises.

Die Eigenwerte $\lambda_1, \ldots, \lambda_r$ von $A$ seien so angeordnet, daß

$$1 = |\lambda_1| = \cdots = |\lambda_s| > |\lambda_{s+1}| \geq \cdots \geq |\lambda_r|.$$

Also ist in der Jordan'schen Normalform $r_1 = \cdots = r_s = 1$. Man setze $\epsilon := 1 - |\lambda_{s+1}|$ und definiere anschließend $D := \text{diag}\,(1, \epsilon, \ldots, \epsilon^{r-1})$, $\hat{J} := D^{-1}JD$. Dann ist (siehe auch J.W. Numerische Mathematik 1, S. 24)

$$\hat{J} = \begin{pmatrix} \hat{J}_1 & 0 & \cdots & 0 \\ 0 & \hat{J}_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{J}_p \end{pmatrix} \quad \text{mit} \quad \hat{J}_i = \begin{pmatrix} \lambda_i & \epsilon & & & \\ & \lambda_i & \epsilon & & \\ & & \ddots & \ddots & \\ & & & \ddots & \epsilon \\ & & & & \lambda_i \end{pmatrix} \in \mathbb{R}^{r_i\times r_i}.$$

Nun definiere man die transformierte Maximumnorm $\|\cdot\|$ durch $\|x\| := \|(PD)^{-1}x\|_\infty$. Als zugehörige Matrixnorm berechnet man

$$\|A\| = \|(PD)^{-1}A(PD)\|_\infty = \|D^{-1}JD\|_\infty = \|\hat{J}\|_\infty = \max(1, \max_{i=s+1,\ldots,r} |\lambda_i| + \epsilon) = 1,$$

womit das Lemma vollständig bewiesen ist. □

**Beweis von Satz 2.4:** Der Beweis zerfällt natürlich in zwei Teile. Zunächst zeigen wir, dass aus der Konvergenz die Gültigkeit der Stabilitätsbedingung folgt.

Das Mehrschrittverfahren (MV) sei für jedes $f \in F_1[t_0, T]$ konvergent. Speziell liegt Konvergenz bei Integration der Anfangswertaufgabe $x' = 0$, $x(t_0) = 0$ mit der exakten Lösung $x = 0$ vor. Sei $t \in (t_0, T]$ fest vorgegeben und $h_m := (t - t_0)/m$, $m \in \mathbb{N}$. Seien $v_0, \ldots, v_{r-1} \in \mathbb{R}^n$ beliebig und

$$\epsilon(h_m) := h_m \begin{pmatrix} v_0 \\ \vdots \\ v_{r-1} \end{pmatrix} =: \begin{pmatrix} \epsilon_0 \\ \vdots \\ \epsilon_{r-1} \end{pmatrix},$$

so dass $\lim_{m\to\infty} \epsilon(h_m) = 0$. Man berechne $u_m = u(t; \epsilon, h_m)$ aus

$$\begin{aligned} u_0 &= \epsilon_0 \\ &\vdots \\ u_{r-1} &= \epsilon_{r-1} \\ u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0 u_j &= h_m \underbrace{F(h_m, 0)(u_{j+r}, \ldots, u_j, t_j)}_{=0 \text{ wegen (b)}} \\ &= 0. \end{aligned}$$

Dann ist $u_m = h_m v_m$, wobei $v_m$ rekursiv aus

$$v_{j+r} + a_{r-1}v_{j+r-1} + \cdots + a_0 v_j = 0, \qquad j = 0, \ldots, m - r,$$

zu berechnen ist. Da das Mehrschrittverfahren (MV) nach Voraussetzung konvergent ist, gilt

$$\lim_{m\to\infty} u_m = 0 \qquad \text{bzw.} \qquad \lim_{m\to\infty} \frac{v_m}{m} = 0.$$

Dies gilt zunächst für beliebige $v_0, \ldots, v_{r-1} \in \mathbb{R}^n$, aber dann auch (man trenne in Real- und Imaginärteil!) für beliebige $v_0, \ldots, v_{r-1} \in \mathbb{C}^r$. Wir wollen zeigen, dass hieraus die Stabilitätsbedingung folgt.

(a) Sei $\lambda$ eine Nullstelle von $\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0$. Man setze $v_i := \lambda^i e$, $i = 0, \ldots, r-1$, wobei $e := (1, \ldots, 1)^T$. Dann ist $v_j = \lambda^j e$, $j = 0, \ldots, m$ und daher insbesondere $v_m = \lambda^m e$. Aus

$$0 = \lim_{m\to\infty} \frac{v_m}{m} = \lim_{m\to\infty} \frac{\lambda^m}{m} e$$

folgt $|\lambda| \leq 1$.

(b) Angenommen, ein $\lambda$ mit $|\lambda| = 1$ sei eine mehrfache Nullstelle von $\psi$, es sei also $0 = \psi(\lambda) = \psi'(\lambda)$ bzw.

$$(*) \qquad \psi(\lambda) = \lambda^r + a_{r-1}\lambda^{r-1} + \cdots + a_1\lambda + a_0 = 0$$

und

$$(**) \qquad \psi'(\lambda) = r\lambda^{r-1} + (r-1)a_{r-1}\lambda^{r-2} + \cdots + a_1 = 0.$$

Man setze $v_i := i\lambda^{i-1}e$, $i = 0, \ldots, r-1$. Dann ist $v_j = j\lambda^{j-1}e$ für $j = 0, \ldots, m$, wie man leicht feststellt. Denn multipliziert man $(*)$ mit $j\lambda^{j-1}$, $(**)$ mit $\lambda^j$ und addiert die Ergebnisse, so erhält man

$$(j+r)\lambda^{j+r-1} + (j+r-1)a_{r-1}\lambda^{j+r-2} + \cdots + (j+1)a_1\lambda^j + ja_0\lambda^{j-1} = 0,$$

woraus man die Behauptung abliest. Insbesondere ist $v_m = m\lambda^{m-1}e$. Wegen

$$0 = \lim_{m\to\infty} \frac{v_m}{m} = \lim_{m\to\infty} \lambda^{m-1}e$$

erhält man einen Widerspruch zu $|\lambda| = 1$.

Insgesamt ist nachgewiesen, dass aus der Konvergenz des Mehrschrittverfahrens (MV) die Gültigkeit der Stabilitätsbedingung folgt.

Nun sei die Stabilitätsbedingung erfüllt. Es soll die Konvergenz des konsistenten Mehrschrittverfahrens (MV) nachgewiesen werden. Hierzu wird im Prinzip ähnlich vorgegangen wie beim Beweis für die Konvergenz konsistenter Einschrittverfahren.

Sei $f \in F_1[t_0, T]$ und hiermit die Anfangswertaufgabe

$$(P) \qquad x' = f(t,x), \qquad x(t_0) = x_0$$

mit der exakten Lösung $x(\cdot)$ auf $[t_0, T]$ gegeben. Sei $t \in (t_0, T]$ fest vorgegeben, $h_m := (t - t_0)/m$ und $t_i := t_0 + ih_m$, $i = 1, \ldots, m$. Die Näherungen $u_i$ für $x_i := x(t_i)$ seien gewonnen aus

$$\begin{aligned} u_i &= x_i + \epsilon_i, \qquad i = 0, \ldots, r-1, \\ u_{j+r} + a_{r-1}u_{j+r-1} + \cdots + a_0u_j &= hF(h,f)(u_{j+r}, \ldots, u_j, t_j), \qquad j = 0, \ldots, m-r. \end{aligned}$$

Für den Fehler $e_i := u_i - x_i$ an der Stelle $t_i$ haben wir dann $e_i = \epsilon_i$, $i = 0, \ldots, r-1$. Nach Definition des lokalen Diskretisierungsfehlers ist

$$\Delta(h_m, f)(t_j, x_j) = \frac{1}{h_m}[x_{j+r} + a_{r-1}x_{j+r-1} + \cdots + a_0x_j] - F(h_m, f)(x_{j+r}, \ldots, x_j, t_j).$$

Folglich ist

$$\begin{aligned} e_{j+r} + a_{r-1}e_{j+r-1} + \cdots + a_0e_j &= h_m\,[F(h_m,f)(u_{j+r}, \ldots, u_j, t_j) \\ &\qquad - F(h_m,f)(x_{j+r}, \ldots, x_j, t_j)] \\ &\qquad - h_m\Delta(h_m,f)(t_j,x_j) \\ &=: c_{j+r}. \end{aligned}$$

Wegen der Konsistenz von (MV) existiert eine reellwertige Funktion $\sigma$ mit

$$\|\Delta(h,f)(t_j,x_j)\| \le \sigma(h) \qquad \text{mit} \quad \lim_{h\to 0+}\sigma(h)=0.$$

Da ferner die Verfahrensfunktion $F$ nach Voraussetzung (a) lipschitzstetig ist, ist

$$\|c_{j+r}\| \le h_m M \sum_{i=0}^{r} \|e_{j+i}\| + h_m\sigma(h_m).$$

Die Gleichungen

$$\begin{aligned} e_i &= \epsilon_i, \qquad i=0,\dots,r-1,\\ e_{j+r} &= -a_0e_j - \cdots - a_{r-1}e_{j+r-1} + c_{j+r}, \qquad j=0,\dots,m-r \end{aligned}$$

werden nun komponentenweise betrachtet. Die $k$-te Komponente, etwa von $e_j$, wird mit $e_{j,k}$ bezeichnet, $k=1,\dots,n$, für andere Vektoren gelte entsprechendes. Für $k=1,\dots,n$ hat man also die Gleichungen

$$\begin{aligned} e_{i,k} &= \epsilon_{i,k}, \qquad i=0,\dots,r-1,\\ e_{j+r,k} &= -a_0e_{j,k} - \cdots - a_{r-1}e_{j+r-1,k} + c_{j+r,k}, \qquad j=0,\dots,m-r. \end{aligned}$$

Dann ist

$$\underbrace{\begin{pmatrix} e_{0,k}\\ e_{1,k}\\ \vdots\\ e_{r-1,k}\end{pmatrix}}_{=E_{0,k}} = \begin{pmatrix} \epsilon_{0,k}\\ \epsilon_{1,k}\\ \vdots\\ \epsilon_{r-1,k}\end{pmatrix}$$

und

$$\underbrace{\begin{pmatrix} e_{j+1,k}\\ e_{j+2,k}\\ \vdots\\ e_{j+r,k}\end{pmatrix}}_{=E_{j+1,k}} = \underbrace{\begin{pmatrix} 0 & 1 & \cdots & 0\\ 0 & 0 & \ddots & 0\\ \vdots & \vdots & \ddots & 1\\ -a_0 & -a_1 & \cdots & -a_{r-1}\end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} e_{j,k}\\ e_{j+1,k}\\ \vdots\\ e_{j+r-1,k}\end{pmatrix}}_{=E_{j,k}} + c_{j+r,k}\underbrace{\begin{pmatrix} 0\\ 0\\ \vdots\\ 1\end{pmatrix}}_{=b}$$

für $k=1,\dots,n$. Hierbei ist $A$ die sogenannte *Begleitmatrix* zum Polynom

$$\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_1\mu + a_0.$$

Die Nullstellen von $\psi$ sind genau die Eigenwerte von $A$. Wegen der Stabilitätsbedingung sind alle Eigenwerte betragsmäßig kleiner gleich 1 bzw. $\rho(A)\le 1$. Es ist sogar $\rho(A)=1$, da 1 eine Nullstelle von $\psi$ ist, wie man folgender Argumentation entnimmt: Bei vorgegebenem $u\in\mathbb{R}^n$ ist $z(s)\equiv u$ die Lösung von $z'=0$, $z(t)=u$. Daher ist

$$\Delta(h,0)(t,u) = \frac{1}{h}u\Big[1+\sum_{i=0}^{r-1}a_i\Big] - \underbrace{F(h,0)(u,\dots,u,t)}_{=0}.$$

Nach Voraussetzung ist (MV) konsistent und folglich

$$\lim_{h\to 0+} \frac{1}{h} u\Big[1 + \sum_{i=0}^{r-1} a_i\Big] = 0 \qquad \text{für jedes } u \in \mathbb{R}^n.$$

Hieraus folgt

$$\psi(1) = 1 + \sum_{i=0}^{r-1} a_i = 0.$$

O. B. d. A. ist die bei der vorausgesetzten Lipschitzstetigkeit der Verfahrensfunktion auftretende Vektornorm die 1-Norm bzw. die Betragssummennorm $\|\cdot\|_1$. Nach Lemma 2.5 existiert eine Vektornorm $\|\cdot\|$ (diese Bezeichnung haben wir gerade eben ermöglicht) auf $\mathbb{C}^r$ derart, dass für die zugeordnete Matrixnorm gilt $\|A\| = \rho(A) = 1$. Aus

$$E_{j+1,k} = AE_{j,k} + c_{j+r,k} b$$

folgt

$$\|E_{j+1,k}\| \le \underbrace{\|A\|}_{=1} \|E_{j,k}\| + |c_{j+r,k}|\,\|b\|, \qquad k = 1, \ldots, n,$$

und daher durch Summation dieser $n$ Ungleichungen

$$\sum_{k=1}^{n} \|E_{j+1,k}\| \le \sum_{k=1}^{n} \|E_{j,k}\| + \|c_{j+r}\|_1\,\|b\|.$$

Wegen

$$c_{j+r} = h_m[F(h_m, f)(u_{j+r}, \ldots, u_j, t_j) - F(h_m, f)(x_{j+r}, \ldots, x_j, t_j)] - h_m \tau(h_m, f)(x_j, t_j)$$

erhalten wir mit der Lipschitzstetigkeit der Verfahrensfunktion, daß

$$\begin{aligned} \|c_{j+r}\|_1 &\le h_m M \sum_{i=0}^{r} \|e_{j+i}\|_1 + h_m \sigma(h_m) \\ &= h_m M \sum_{k=1}^{n} \sum_{i=0}^{r} |e_{j+i,k}| + h_m \sigma(h_m) \\ &\le h_m M \sum_{k=1}^{n} (\|E_{j,k}\|_1 + \|E_{j+1,k}\|_1) + h_m \sigma(h_m). \end{aligned}$$

Zur letzten Ungleichung bemerken wir, dass alleine durch $\|E_{j,k}\|_1$ der Summand $|e_{j+r,k}|$ nicht berücksichtigt wird.

Alle Normen auf $\mathbb{C}^r$ sind äquivalent, wie wir z. B. aus der Numerischen Mathematik wissen. Daher existiert eine Konstante $c > 0$ derart, dass

$$\frac{1}{c}\,\|v\| \le \|v\|_1 \le c\,\|v\| \qquad \text{für alle } v \in \mathbb{C}^r.$$

Damit wird

$$\sum_{k=1}^{n} \|E_{j+1,k}\| \le \sum_{k=1}^{n} \|E_{j,k}\| + \Big[h_m Mc\Big(\sum_{k=1}^{n} \|E_{j,k}\| + \sum_{k=1}^{n} \|E_{j+1,k}\|\Big) + h_m\sigma(h_m)\Big]\,\|b\|$$

bzw.

$$(1 - h_m Mc\,\|b\|) \sum_{k=1}^{n} \|E_{j+1,k}\| \le (1 + h_m Mc\,\|b\|) \sum_{k=1}^{n} \|E_{j,k}\| + h_m\sigma(h_m)\,\|b\|.$$

Nun wähle man $m$ so groß, dass

$$h_m \le \frac{1}{2Mc\,\|b\|}.$$

Wegen

$$\frac{1+a}{1-a} \le 1 + 4a \qquad \text{für } 0 \le a \le \tfrac{1}{2}$$

ist dann

$$\underbrace{\sum_{k=1}^{n} \|E_{j+1,k}\|}_{|\xi_{j+1}|} \le (1 + \underbrace{4h_m Mc\,\|b\|}_{=:\,\delta}) \underbrace{\sum_{k=1}^{n} \|E_{j,k}\|}_{|\xi_j|} + \underbrace{2h_m\sigma(h_m)\,\|b\|}_{=:\,B}, \qquad j = 0, 1 \ldots.$$

Nach Voraussetzung geht der Anfangsfehler, der in der Startrechnung gemacht wird, gegen Null. Es existiert also eine Funktion $\rho(\cdot)$, natürlich nicht zu verwechseln mit dem Spektralradius, derart dass

$$\sum_{k=1}^{n} \|E_{0,k}\| \le \rho(h_m) \qquad \text{und} \qquad \lim_{m\to\infty} \rho(h_m) = 0.$$

Nun überlegen wir uns die Gültigkeit der folgenden Aussage:

- *Genügen die Zahlen $\xi_0, \xi_1, \ldots$ mit Konstanten $\delta > 0$, $B > 0$ einer Abschätzung der Form*

  $$|\xi_{i+1}| \le (1+\delta)|\xi_i| + B, \qquad i = 0, 1, \ldots,$$

  *so gilt*

  $$|\xi_m| \le e^{m\delta}|\xi_0| + B\,\frac{e^{m\delta} - 1}{\delta}, \qquad m = 0, 1, \ldots.$$

Denn: Die nun folgende erste Ungleichung erhält man leicht durch vollständige Induktion, danach folgt schnell die Behauptung:

$$\begin{aligned} |\xi_m| &\le (1+\delta)^m|\xi_0| + B \sum_{j=0}^{m-1} (1+\delta)^j \\ &= (1+\delta)^m|\xi_0| + B\,\frac{(1+\delta)^m - 1}{\delta} \\ &\le e^{m\delta}|\xi_0| + B\,\frac{e^{m\delta} - 1}{\delta} \\ &\qquad \text{(wegen } 1 + \delta \le e^{\delta}\text{)}. \end{aligned}$$

Eine Anwendung dieser Aussage liefert

$$\sum_{k=1}^{n} \|E_{m,k}\| \leq e^{4Mc\|b\|(t-t_0)}\rho(h_m) + \frac{e^{4Mc\|b\|(t-t_0)}-1}{2Mc}\,\sigma(h_m)$$

für alle hinreichend großen $m$. Wegen $\lim_{m\to\infty}\rho(h_m) = \lim_{m\to\infty}\sigma(h_m) = 0$ ist die Konvergenz des Mehrschrittverfahrens (MV) bewiesen.

**Bemerkung:** Aus dem Beweis des letzten Satzes liest man ab: Ist $p$ die Konsistenzordnung eines Mehrschrittverfahrens, das der Stabilitätsbedingung genügt, und werden die Startwerte mit einem (Einschritt-) Verfahren der Ordnung $p$ berechnet, so hat der globale Diskretisierungsfehler des Mehrschrittverfahrens die Ordnung $p$. □

**Bemerkung:** Das Polynom

$$\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0$$

genüge der Stabilitätsbedingung, d. h. alle Wurzeln seien betragsmäßig kleiner oder gleich 1 und diejenigen auf dem Rande des Einheitskreises seien einfach. Weiter können wir annehmen, dass 1 eine Wurzel von $\psi$ ist (andernfalls könnte ein entsprechendes Verfahren nicht konsistent sein). Dann gilt:

- Wie auch immer $u_0, \ldots, u_{r-1}$ vorgegeben sind, die Folge $\{u_{j+r}\}_{j=0,1,\ldots}$, die aus

$$u_{j+r} + \sum_{i=0}^{r-1} a_i u_{j+i} = 0, \qquad j = 0, 1, \ldots,$$

  gewonnen ist, ist beschränkt.

Denn: Es ist

$$\begin{pmatrix} u_{j+1} \\ u_{j+2} \\ \vdots \\ u_{j+r} \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ -a_0 & -a_1 & \cdots & -a_{r-1} \end{pmatrix} \begin{pmatrix} u_j \\ u_{j+1} \\ \vdots \\ u_{j+r-1} \end{pmatrix}$$

bzw.

$$U_{j+1} = AU_j, \qquad j = 0, 1, \ldots,$$

wobei

$$U_j := \begin{pmatrix} u_j \\ u_{j+1} \\ \vdots \\ u_{j+r-1} \end{pmatrix}, \qquad A := \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ -a_0 & -a_1 & \cdots & -a_{r-1} \end{pmatrix}.$$

Nun haben wir in Lemma 2.5 die Existenz einer Vektornorm $\|\cdot\|$ nachgewiesen mit der Eigenschaft, dass $\|A\| = 1$ für die zugeordnete Matrixnorm. Dann folgt aber $\|U_j\| \leq \|U_0\|$, $j = 0, 1, \ldots$ und dies impliziert die Beschränktheit der Folge $\{u_{j+r}\}_{j=0,1,\ldots}$. Es ist nicht schwierig, auch die Umkehrung der obigen Aussage nachzuweisen:

- Wie auch immer $u_0, \dots, u_{r-1}$ vorgegeben, sei die Folge $\{u_{j+r}\}_{j=0,1,\dots}$, die aus

$$u_{j+r} + \sum_{i=0}^{r-1} a_i u_{j+i} = 0, \qquad j = 0, 1, \dots,$$

gewonnen ist, beschränkt. Dann genügt

$$\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0$$

der Stabilitätsbedingung.

Denn: Es genügt praktisch, den Beweis des ersten Teiles von Satz 2.4 zu wiederholen. Ist $\lambda$ ein Nullstelle von $\psi$, so setze $u_i := \lambda^i$, $i = 0, \dots, r-1$. Dann ist $u_{j+r} = \lambda^{j+r}$, aus der Beschränktheit von $\{u_{j+r}\}_{j=0,1,\dots}$ folgt $|\lambda| \le 1$. Ist $\lambda$ eine doppelte Nullstelle von $\psi$, so setze man $u_i := t\lambda^{i-1}$, $i = 0, \dots, r-1$. Dann ist $u_{j+r} = (j+r)\lambda^{j+r-1}$ (siehe Beweis von Satz 2.4), aus der Beschränktheit von $\{u_{j+r}\}_{j=0,1,\dots}$ folgt $|\lambda| < 1$. Insgesamt genügt $\psi$ der Stabilitätsbedingung. □

### 3.2.4 Lineare Mehrschrittverfahren

Wie schon früher erwähnt, ist bei einem linearen Mehrschrittverfahren die Verfahrensfunktion gegeben durch

$$F(h, f)(u_{j+r}, \dots, u_j, t_j) = b_r f(t_{j+r}, u_{j+r}) + \cdots + b_0 f(t_j, u_j).$$

Das Mehrschrittverfahren ist also durch Angabe von $a_0, \dots, a_{r-1}$ sowie von $b_0, \dots, b_r$ festgelegt. Das Verfahren ist explizit (Prädiktor-Verfahren), wenn $b_r = 0$, andernfalls implizit (Korrektor-Verfahren). Wie schon früher bemerkt, sind bei einem linearen Mehrschrittverfahren die Voraussetzungen (a) und (b) in Satz 2.4 erfüllt. Die Untersuchung der Konsistenz ist besonders einfach. Zweckmäßig ist es, die beiden Polynome

$$\begin{aligned} \psi(\mu) &:= a_0 + a_1\mu + \cdots + a_{r-1}\mu^{r-1} + \mu^r, \\ \chi(\mu) &:= b_0 + b_1\mu + \cdots + b_{r-1}\mu^{r-1} + b_r\mu^r \end{aligned}$$

einzuführen. Bei vorgegebenem $(t, u) \in [t_0, T] \times \mathbb{R}^n$ sei $z(\cdot)$ wieder die Lösung von $z' = f(s, z)$, $z(t) = u$. Ist $f$ glatt, so ist es auch $z$. Der lokale Diskretisierungsfehler ist gegeben durch

$$\begin{aligned} \Delta(h, f)(t, u) &= \frac{1}{h}\Big[z(t + rh) + \sum_{i=0}^{r-1} a_i z(t + ih)\Big] - \sum_{i=0}^{r} b_i f(t + ih, z(t + ih)) \\ &= \frac{1}{h}\Big[z(t + rh) + \sum_{i=0}^{r-1} a_i z(t + ih)\Big] - \sum_{i=0}^{r} b_i z'(t + ih) \\ &= \frac{1}{h} z(t) \underbrace{\Big[1 + \sum_{i=0}^{r-1} a_i\Big]}_{=: C_0} + z'(t) \underbrace{\Big[r + \sum_{i=1}^{r-1} i a_i - \sum_{i=0}^{r} b_i\Big]}_{=: C_1} + O(h) \end{aligned}$$

mit den beiden von $h$ und $z$ unabhängigen Konstanten

$$C_0 := 1 + \sum_{i=0}^{r-1} a_i, \qquad C_1 := r + \sum_{i=1}^{r-1} i a_i - \sum_{i=0}^{r} b_i.$$

Offenbar ist $C_0 = \psi(1)$, $C_1 = \psi'(1) - \chi(1)$. Das gegebene lineare Mehrschrittverfahren ist genau dann konsistent, wenn $C_0 = C_1 = 0$. Insbesondere hat ein konsistentes lineares Mehrschrittverfahren mindestens die Ordnung 1.

Wir benutzen die eben eingeführten Bezeichnungen und beweisen als Ergänzung zum Äquivalenzsatz 2.4 das folgende Ergebnis.

**Satz 2.6** *Ist ein lineares Mehrschrittverfahren konvergent, so ist es auch konsistent.*

**Beweis:** Wir haben zu zeigen, dass die Konstanten

$$C_0 := 1 + \sum_{i=0}^{r-1} a_i, \qquad C_1 := r + \sum_{i=1}^{r-1} i a_i - \sum_{i=0}^{r} b_i$$

verschwinden. Zum Nachweis dafür, dass $C_0 = 0$ betrachte man die skalare Anfangswertaufgabe

$$x' = 0, \qquad x(0) = 1$$

mit der exakten Lösung $x(t) \equiv 1$. Als Startwerte nehme man die exakten Werte $u_i := 1$, $i = 0, \ldots, r-1$. Die Werte $u_{j+r}$, $j = 0, 1, \ldots$, sind zu berechnen aus

$$(*) \qquad u_{j+r} + a_{r-1} u_{j+r-1} + \cdots + a_0 u_j = 0.$$

Setzt man $h_m := t/m$ mit vorgegebenem $t > 0$, so ist $u(t; h_m) = u_m$. Wegen der vorausgesetzten Konvergenz des linearen Mehrschrittverfahrens gilt $\lim_{m\to\infty} u_m = x(t) = 1$. Für $j \to \infty$ folgt daher aus $(*)$, dass

$$C_0 = 1 + a_{r-1} + \cdots + a_0 = 0.$$

Um auch $C_1 = 0$ zu beweisen, betrachte man die skalare Anfangswertaufgabe

$$x' = 1, \qquad x(0) = 0$$

mit der exakten Lösung $x(t) \equiv t$. Wir wissen bereits, dass $C_0 = 0$ bzw. 1 eine Nullstelle von $\psi$ ist. Wegen des Äquivalenzsatzes und der vorausgesetzten Konvergenz ist die Stabilitätsbedingung erfüllt (man beachte, dass man zu diesem Teil des Äquivalenzsatzes *nicht* die Konsistenz des betrachteten Mehrschrittverfahrens benötigt). Hiernach ist 1 eine einfache Nullstelle von $\psi$, so dass $\psi'(1) \neq 0$ und die Konstante

$$K := \frac{\chi(1)}{\psi'(1)}$$

wohldefiniert ist. Mit $h_m := t/m$ nehme man nun $u_j := j h_m K$, $j = 0, \ldots, r-1$, als Startwerte für das gegebene lineare Mehrschrittverfahren. Wegen $t_j = j h_m$ und $x(t) = t$ ist

$$\epsilon_j(h_m) := u_j - x(t_j) = j h_m (K - 1), \qquad j = 0, \ldots, r-1.$$

Offensichtlich ist $\lim_{m\to\infty}\epsilon_j(h_m)=0$, $j=0,\ldots,r-1$. Zu den angegebenen Startwerten liefert das Verfahren Werte $u_{j+r}$, $j=0,1,\ldots$, aus

$$u_{j+r}+a_{r-1}u_{j+r-1}+\cdots+a_0u_j=h_m(b_0+b_1+\cdots+b_r)=h_m\chi(1).$$

Hieraus erhält man $u_j=jh_mK$ für alle $j$, wie man durch Einsetzen unter Benutzung von $C_0=0$ leicht bestätigt. Es ist $u(t;h_m)=u_m$, wegen der vorausgesetzten Konvergenz ist

$$\lim_{m\to\infty}u_m=\lim_{m\to\infty}\underbrace{mh_mK}_{tK}=t,$$

daher $K=1$ und $C_1=0$. Damit ist der Satz bewiesen. □

Mit der Aussage des folgenden Satzes (siehe auch J. Stoer, R. Bulirsch (1990, S. 155)) kann die Konsistenzordnung eines linearen Mehrschrittverfahrens bestimmt werden. Wieder bezeichnen $\psi,\chi$ die oben eingeführten, das lineare Mehrschrittverfahren bestimmenden Polynome aus $\Pi_r$.

**Satz 2.7** *Ein lineares Mehrschrittverfahren besitzt genau dann die Konsistenzordnung $p$, wenn die Funktion*

$$\phi(\mu):=\frac{\psi(\mu)}{\ln\mu}-\chi(\mu)$$

*die Zahl $\mu=1$ als $p$-fache Nullstelle besitzt.*

**Beweis:** Das gegebene lineare Mehrschrittverfahren habe die Ordnung $p$. Der zugehörige lokale Diskretisierungsfehler ist gegeben durch

$$(*)\qquad \Delta(h,f)(t,u)=\frac{1}{h}\Big[z(t+rh)+\sum_{i=0}^{r-1}a_iz(t+ih)\Big]-\sum_{i=0}^{r}b_iz'(t+ih),$$

wobei, wie in diesem Zusammenhang üblich, $z(\cdot)$ die Lösung von $z'=f(s,z)$, $z(t)=u$ ist. Nach Voraussetzung ist $\Delta(h,f)(t,u)=O(h^p)$ für alle $f\in F_p[t_0,T]$. Setzt man in $(*)$ speziell $z(s):=e^s$ (nimm $f(s,z):=z$, $u:=e^t$), so ist

$$\Delta(h,f)(t,u)=e^t\Big[\frac{\psi(e^h)}{h}-\chi(e^h)\Big].$$

Da die Konsistenzordnung $p$ ist, ist $h=0$ eine $p$-fache Nullstelle von $\phi(e^h)$ bzw. $\mu=1$ eine $p$-fache Nullstelle von $\phi$. Umgekehrt sei $\mu=1$ eine $p$-fache Nullstelle von $\phi$ bzw. $h=0$ eine $p$-fache Nullstelle von $\psi(e^h)/h-\chi(e^h)$. Für $f\in F_p[t_0,T]$ ist (mit den üblichen Bezeichnungen) nach einfacher Rechnung $\Delta(h,f)(t,u)=O(h^p)$. Denn: Es ist

$$\begin{aligned}z(t+ih) &= \sum_{j=0}^{p}\frac{(ih)^j}{j!}z^{(j)}(t)+O(h^{p+1}),\\ z'(t+ih) &= \sum_{j=1}^{p}\frac{(ih)^{j-1}}{(j-1)!}z^{(j)}(t)+O(h^p).\end{aligned}$$

Der lokale Diskretisierungsfehler ist also gegeben durch (es sei $a_r := 1$)

$$\begin{aligned}\Delta(h,f)(t,u) &= \frac{1}{h}\sum_{i=0}^{r} a_i z(t+ih) - \sum_{i=0}^{r} b_i z'(t+ih)\\ &= \sum_{i=0}^{r}\Big[a_i \sum_{j=0}^{p} \frac{i^j h^{j-1}}{j!} z^{(j)}(t) - b_i \sum_{j=1}^{p} \frac{i^{j-1}h^{j-1}}{(j-1)!} z^{(j)}(t)\Big] + O(h^p)\\ &= \frac{1}{h} z(t) \sum_{i=0}^{r} a_i + \sum_{j=1}^{p} h^{j-1} z^{(j)}(t) \sum_{i=0}^{r}\Big(\frac{i^j}{j!}a_i - \frac{i^{j-1}}{(j-1)!} b_i\Big) + O(h^p).\end{aligned}$$

Daher ist die Konsistenzordnung des linearen Mehrschrittverfahrens genau dann $p$, wenn

$$\sum_{i=0}^{r} a_i = 0, \qquad \sum_{i=0}^{r}\Big(\frac{i^j}{j!}a_i - \frac{i^{j-1}}{(j-1)!} b_i\Big) = 0 \quad (j = 1,\dots,p),$$

diese Beziehungen sind also nachzuweisen. Andererseits ist $h = 0$ eine $p$-fache Nullstelle von $\phi(e^h)$. Daher ist

$$\begin{aligned}O(h^p) &= \phi(e^h)\\ &= \frac{\psi(e^h)}{h} - \chi(e^h)\\ &= \frac{1}{h}\sum_{i=0}^{r} a_i e^{hi} - \sum_{i=0}^{r} b_i e^{hi}\\ &= \frac{1}{h}\sum_{i=0}^{r} a_i + \sum_{j=1}^{p} h^{j-1} \sum_{i=0}^{r}\Big(\frac{i^j}{j!}a_i - \frac{i^{j-1}}{(j-1)!} b_i\Big) + O(h^p)\end{aligned}$$

Hieraus folgt dann die Behauptung. □

**Bemerkung:** Die Konstruktion linearer Mehrschrittverfahren möglichst hoher Ordnung könnte nun folgendermaßen verlaufen. Gegeben seien $a_0,\dots,a_{r-1}$ und hiermit das Polynom

$$\psi(\mu) := \mu^r + a_{r-1}\mu^{r-1} + \cdots + a_0.$$

Es sei $\psi(1) = 0$, was ja eine notwendige Bedingung für Konsistenz ist. Die dann in einer Umgebung von $\mu = 1$ holomorphe Funktion $\psi(\mu)/\ln\mu$ denke man sich um $\mu = 1$ in eine Potenzreihe entwickelt:

$$\frac{\psi(\mu)}{\ln\mu} = c_0 + c_1(\mu-1) + \cdots + c_{r-1}(\mu-1)^{r-1} + c_r(\mu-1)^r + \cdots.$$

Wählt man

$$\begin{aligned}\chi(\mu) &:= c_0 + c_1(\mu-1) + \cdots + c_{r-1}(\mu-1)^{r-1} + c_r(\mu-1)^r\\ &= b_0 + b_1\mu + \cdots + b_{r-1}\mu^{r-1} + b_r\mu^r,\end{aligned}$$

so erhält man ein Korrektor-Verfahren (dieses ist implizit, was man an $b_r \neq 0$ erkennt) von mindestens der Ordnung $r+1$. Wählt man dagegen

$$\begin{aligned}\chi(\mu) &:= c_0 + c_1(\mu-1) + \cdots + c_{r-1}(\mu-1)^{r-1} \\ &= b_0 + b_1\mu + \cdots + b_{r-1}\mu^{r-1},\end{aligned}$$

so erhält man ein (explizites) Prädiktor-Verfahren von mindestens der Ordnung $r$. □

**Beispiel:** Sei $r = 2$, es soll ein stabiles und konsistentes Verfahren möglichst hoher Ordnung konstruiert werden. Es sind $a_0, a_1$ so vorzugeben, dass $1 + a_0 + a_1 = 0$, was auf den Ansatz

$$\psi(\mu) = \mu^2 - (1+a)\mu + a = (\mu-1)(\mu-a)$$

führt. Damit das resultierende Mehrschrittverfahren der Stabilitätsbedingung genügt, muß $a \in [-1, 1)$ sein. Eine Taylor-Entwicklung von $\psi(\mu)/\ln\mu$ um $\mu = 1$ liefert, wenn man

```
series((mu-1)*(mu-a)/ln(mu),mu=1,5);
```

eingibt

$$\frac{\psi(\mu)}{\ln\mu} = 1 - a + \frac{3-a}{2}(\mu-1) + \frac{a+5}{12}(\mu-1)^2 - \frac{1+a}{24}(\mu-1)^3 + O(\mu-1)^4.$$

Setzt man

$$\chi(\mu) := 1 - a + \frac{3-a}{2}(\mu-1) + \frac{a+5}{12}(\mu-1)^2,$$

so hat das resultierende lineare Mehrschrittverfahren für $a \neq -1$ die Ordnung 3 und für $a = -1$ die Ordnung 4. Z. B. erhält man für $a = 0$ die Polynome

$$\psi(\mu) = \mu^2 - \mu, \qquad \chi(\mu) = \frac{1}{12}(5\mu^2 + 8\mu - 1)$$

bzw. das lineare (implizite) Mehrschrittverfahren

$$u_{j+2} - u_{j+1} = \frac{h}{12}\left[5f(t_{j+2}, u_{j+2}) + 8f(t_{j+1}, u_{j+1}) - f(t_j, u_j)\right], \qquad j = 0, 1, \ldots.$$

Dies ist gerade das Adams-Moulton-Verfahren für $q = 2$. □

Die Frage nach der Existenz stabiler, linearer $r$-Schrittverfahren möglichst hoher Konsistenzordnung ist naheliegend. Hier gilt ein (nicht einfach beweisbares) Resultat von Dahlquist:

- *Sei $p$ die Ordnung eines linearen $r$-Schrittverfahrens, welches der Stabilitätsbedingung genügt. Dann gilt*

$$p \leq \begin{cases} r+2 & \text{für } r \text{ gerade,} \\ r+1 & \text{für } r \text{ ungerade,} \\ r & \text{für ein explizites Verfahren.} \end{cases}$$

Einen Beweis kann man z. B. bei E. Haier et al. (1993, S. 384) finden. Man spricht in diesem Zusammenhang von der sogenannten Dahlquist-Grenze.

### 3.2.5 Aufgaben

1. Die Folge $\{f_k\}$ (die sogenannte Fibonacci-Folge) sei gegeben durch

$$f_0 := 1, \quad f_1 := 1, \quad f_{k+2} := f_{k+1} + f_k.$$

   Man zeige, dass $\lim_{k\to\infty}(f_{k+1} - \tau f_k) = 0$, wobei $\tau := (1 + \sqrt{5})/2$.

2. Man[7] löse folgende Differenzengleichungen:

   (a) $u_{j+2} - 2u_{j+1} - 3u_j = 0$, $u_0 = 0$, $u_1 = 1$,

   (b) $u_{j+1} - u_j = 2^j$, $u_0 = 0$,

   (c) $u_{j+2} - 2u_{j+1} - 3u_j = 1$, $u_0 = 0$, $u_1 = 0$,

   (d) $u_{j+1} - u_j = j$, $u_0 = 0$.

   Hierbei kann auch der Maple-Befehl `rsolve` benutzt werden.

3. Man[8] löse die Differenzengleichung

$$u_{j+4} - 6u_{j+3} + 14u_{j+2} - 16u_{j+1} + 8u_j = j$$

   mit den Anfangsbedingungen

$$u_0 = 1, \quad u_1 = 2, \quad u_2 = 3, \quad u_3 = 4.$$

4. Man[9] bestimme die allgemeine Lösung der Differenzengleichung

$$u_{j+2} - 2au_{j+1} + au_j = 1$$

   mit $a \in (0, 1)$ im Komplexen und im Reellen. Man zeige, dass $\lim_{j\to\infty} u_j = 1/(1 - a)$

5. Man[10] bestimme $\alpha$, $\beta$ und $\gamma$ so, dass das lineare Mehrschrittverfahren

$$u_{j+3} - u_{j+1} + \alpha(u_{j+2} - u_j) = h\{\beta[f(t_{j+2}, u_{j+2}) - f(t_j, u_j)] + \gamma f(t_{j+1}, u_{j+1})\}$$

   die Konsistenzordnung 3 hat. Ist das so gewonnene Verfahren stabil?

6. Es werde das durch

$$u_{j+2} + a_1u_{j+1} + a_0u_j = h[b_0f(t_j, u_j) + b_1f(t_{j+1}, u_{j+1})]$$

   gegebene explizite Zweischrittverfahren betrachtet.

---

[7] Diese Aufgabe haben wir aus

K. Strehmel, R. Weiner (1995) *Numerik gewöhnlicher Differentialgleichungen*. B. G. Teubner, Stuttgart.

[8] Diese Aufgabe haben wir aus

A. Quarteroni, R. Sacco, F. Saleri (2000) *Numerical Mathematics*. Springer, New York-Berlin-Heidelberg.

[9] Diese Aufgabe haben wir aus

R. Kress (1998) *Numerical Analysis*. Springer, New York-Berlin-Heidelberg.

[10] Diese und die nächsten beiden Aufgaben haben wir J. Stoer, R. Bulirsch (1990, S. 246) entnommen.

(a) Man bestimme $a_0$, $b_0$ und $b_1$ in Abhängigkeit von $a_1$ so, dass man ein Verfahren mindestens zweiter Konsistenzordnung hat.

(b) Für welche $a_1$-Werte ist das so gewonnene Verfahren stabil?

(c) Welche speziellen Verfahren erhält man für $a_1 = 0$ und $a_1 = -1$?

(d) Lässt sich $a_1$ so wählen, dass man ein stabiles Verfahren der Konsistenzordnung 3 erhält?

7. Man prüfe, ob das lineare Mehrschrittverfahren

$$u_{j+4} - u_j = \frac{h}{3}[8f(t_{j+3}, u_{j+3}) - 4f(t_{j+2}, u_{j+2}) + 8f(t_{j+1}, u_{j+1})]$$

konvergent ist.

8. Man[11] bestimme das $\alpha$-Intervall, für das das explizite lineare 3-Schrittverfahren

$$u_{j+3} + \alpha(u_{j+2} - u_{j+1}) - u_j = \frac{h}{2}(3+\alpha)[f(t_{j+2}, u_{j+2}) + f(t_{j+1}, u_{j+1})], \quad \alpha \in \mathbb{R},$$

der Stabilitätsbedingung genügt. Ferner zeige man, dass ein $\alpha$ existiert, für das das Verfahren die Konsistenzordnung 4 hat, dass aber für ein stabiles Verfahren die Konsistenzordnung höchstens 2 sein kann.

9. Eine Anfangswertaufgabe

$$\text{(P)} \qquad x'' = f(t,x), \qquad x(t_0) = x_0,\ x'(t_0) = x_0'$$

für eine Differentialgleichung zweiter Ordnung kann man natürlich dadurch numerisch lösen, dass man die Aufgabe als ein System von zwei Differentialgleichungen erster Ordnung schreibt und dieses mit einem Ein- oder Mehrschrittverfahren löst. Die folgenden zu beweisenden Aussagen sollen Hinweise dafür geben, wie man Mehrschrittverfahren zur Lösung von (P) konstruieren kann, die diesen Umweg nicht gehen.

(a) Eine Lösung $x(\cdot)$ von (P) genügt der Identität

$$x(t+h) - 2x(t) + x(t-h) = h^2 \int_0^1 (1-s)[f(t+sh, x(t+sh)) + f(t-sh, x(t-sh))]\, ds.$$

(b) Welche Mehrschrittverfahren zur Lösung von (P) suggeriert Teil (a) dieser Aufgabe? Man gebe ein explizites und ein implizites Verfahren an.

10. Man schreibe eine MATLAB-Funktion zur Lösung der Anfangswertaufgabe $x' = f(t,x)$, $x(t_0) = x_0$, welche ein Prädiktor-Korrektor-Verfahren (mit Adams-Bashforth-Formeln als Prädiktor und Adams-Moulton-Formeln als Korrektor) benutzt. Diese Funktion[12] könnte so deklariert werden:

[11] Diese Aufgabe haben wir aus

K. Strehmel, R. Weiner (1995) *Numerik gewöhnlicher Differentialgleichungen*. B. G. Teubner, Stuttgart.

[12] Wir halten uns eng an

C. F. van Loan (1997) *Introduction to Scientific Computing. A Matrix-Vector Approach using MATLAB*. Prentice Hall, Upper Saddle River.

```
function [tvals,xvals]=FixedPC(fname,t_0,x_0,h,p,m);
```

Hierbei seien die Eingabe Daten:

```
fname   string that names the function f.
  t_0   initial time.
  x_0   initial condition vector.
    h   stepsize.
    p   order of method.  (1<=p<=4).
    m   numberof steps to be taken.
```

Ausgabedaten seien

```
tvals   tvals(j)=t_0+(j-1)h, j=1:m+1.
xvals   approximate solution at t=tvals(j), j=1:m+1.
```

Hierzu sollten die folgenden Funktionen bereitgestellt werden (wir geben Input- und Ouputparameter an, ihre Bedeutung sollte sich fast von alleine erschließen):

```
function [tvals,xvals,fvals]=StartAB(fname,t_0,x_0,h,p);
function [t_new,x_new,f_new]=PCstep(fname,t_c,x_c,fvals,h,p);
```

Für die Funktion `StartAB` kann es zweckmäßig sein, noch eine Funktion

```
function [t_new,x_new,f_new]=RKstep(fname,t_c,x_c,f_c,h,p);
```

bereitzustellen. Als Test löse man das folgende Zweikörperproblem

$$\ddot{x} = -\frac{x}{(x^2+y^2)^{3/2}}, \qquad x(0) = 0.4, \qquad \dot{x}(0) = 0,$$
$$\ddot{y} = -\frac{y}{(x^2+y^2)^{3/2}}, \qquad y(0) = 0, \qquad \dot{y}(0) = 2$$

über dem Zeitintervall $[0, 2\pi]$ und plotte die Bahn $\{(x(t), y(t)) : t \in [0, 2\pi]\}$

## 3.3 MATLAB-Funktionen für nicht-steife Differentialgleichungen

Wir wissen noch nicht, was steife Differentialgleichungen, damit auch nicht, was nicht-steife Differentialgleichungen sind. Sagen wir hier zunächst einfach, dass nicht-steife Differentialgleichungen "harmlose" Differentialgleichungen sind, bei denen keine "pathologischen" Eigenschaften festzustellen sind. Wir werden hierauf im nächsten Abschnitt genauer eingehen. MATLAB stellt für Anfangswertaufgaben bei solchen Differentialgleichungssysteme vor allem zwei Funktionen zur Verfügung, die wir in diesem Abschnitt kurz vorstellen wollen, so dass sie leicht selbst benutzt werden können.

### 3.3.1 ODE45

Wir geben zunächst einen Teil der Information wieder, die man nach Eingabe von `help ode45` erhält:

```
ODE45  Solve non-stiff differential equations, medium order method.
    [T,Y] = ODE45('F',TSPAN,Y0) with TSPAN = [T0 TFINAL] integrates the
    system of differential equations y' = F(t,y) from time T0 to TFINAL with
    initial conditions Y0.  'F' is a string containing the name of an ODE
    file.  Function F(T,Y) must return a column vector.  Each row in
    solution array Y corresponds to a time returned in column vector T.  To
    obtain solutions at specific times T0, T1, ..., TFINAL (all increasing
    or all decreasing), use TSPAN = [T0 T1 ... TFINAL].

    [T,Y] = ODE45('F',TSPAN,Y0,OPTIONS) solves as above with default
    integration parameters replaced by values in OPTIONS, an argument
    created with the ODESET function.  See ODESET for details.  Commonly
    used options are scalar relative error tolerance 'RelTol' (1e-3 by
    default) and vector of absolute error tolerances 'AbsTol' (all
    components 1e-6 by default).

    [T,Y] = ODE45('F',TSPAN,Y0,OPTIONS,P1,P2,...) passes the additional
    parameters P1,P2,... to the ODE file as F(T,Y,FLAG,P1,P2,...) (see
    ODEFILE).  Use OPTIONS = [] as a place holder if no options are set.
```

Wir beginnen mit einem Beispiel.

**Beispiel:** Es soll eine Anfangswertaufgabe für die sogenannte van der Polsche Differentialgleichung gelöst werden, genauer sei mit vorgegebenem $\mu > 0$ die Aufgabe

$$x'' - \mu(1-x^2)x' + x = 0, \qquad x(0) = x_0,\ x'(0) = x_0'.$$

Zunächst wird diese Aufgabe in ein System umgeschrieben:

$$\begin{aligned} x_1' &= x_2, & x_1(0) &= x_0, \\ x_2' &= \mu(1-x_1^2)x_2 - x_1, & x_2(0) &= x_0'. \end{aligned}$$

Anschließend schreiben wir ein Function-File `VDpol.m` mit dem folgenden Inhalt:

```
function xprime=VDpol(t,x);
%VDpol(t,x) returns the state derivative of the van der Pol equation:
%
%               x''-mu(1-x^2)x'+x=0
%
mu=2;%choose 0<mu
xprime=[x(2);mu*(1-x(1)^2)*x(2)-x(1)];
```

Gibt man MATLAB anschließend die Befehle

```
tspan=[0 30];
x_0=[1;0];
ode45('VDpol',tspan,x_0);
```

so wird (wir haben keine Output-Parameter) die Lösung (d. h. Lösung und Ableitung) der Anfangswertaufgabe

$$x'' - 2(1 - x^2)x' + x = 0, \qquad x(0) = 1,\ x'(0) = 0$$

über dem Zeitinervall $[0, 30]$ geplottet, siehe Abbildung 3.2 links. Um Zugriff auf die

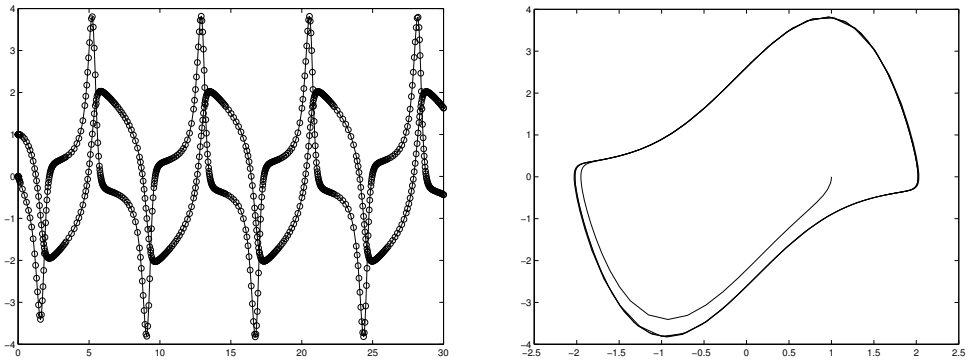

Abbildung 3.2: Lösung der van der Polschen Gleichung

Lösung selber zu haben, gibt man Output-Parameter an:

```
[t,x]=ode45('VDpol',tspan,x_0);
```

Hierbei ist `t` ein Spaltenvektor, der die Zeitpunkte enthält, zu denen die Lösung berechnet wurde. Weiter ist `x` eine Matrix mit zwei Spalten und `length(t)` Zeilen. Die erste Spalte enthält Näherungen für die gesuchte Lösung in den durch `t` gegebenen Zeitpunkten, die zweite Spalte entsprechend die Ableitung. Die Phasenbahn kann man sich durch

```
plot(x(:,1),x(:,2));
```

plotten, siehe Abbildung 3.2 rechts. Man kann natürlich auch die Spalten von $x$ gegen die Zeitachse plotten, was etwa durch

```
plot(t,x(:,1),'-',t,x(:,2),'.');
```

geschehen kann. Das Resultat findet man in Abbildung 3.3. □

Man kann sich die Funktion `ode45` ansehen, da sie selbst in MATLAB geschrieben ist und durch `edit ode45` betrachtet (und natürlich auch ausgedruckt) werden kann. Grundlage ist ein Aufsatz von J. R. Dormand, P. J. Prince (1980)[13]. Ab Zeile 264 stehen die relevanten Informationen über das benutzte Einschrittverfahren. Mit Hilfe von 6 Funktionsberechnungen wird ein Paar von Runge-Kutta-Formeln der Ordnung

[13] J. R. Dormand, P. J. Prince (1980) "A family of embedded Runge-Kutta formulae". J. Comp. Appl. Math. 6, 19–26.

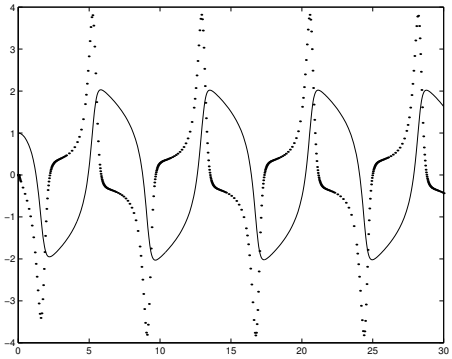


Abbildung 3.3: Lösung der van der Polschen Gleichung

vier und fünf aufgestellt. Diese Formeln geben wir jetzt an. Im Prinzip werden die folgenden Rechnungen durchgeführt:

$$\begin{aligned}
k_1 &:= f(t,u)\\
k_2 &:= f(t+\tfrac{1}{5}h, u+h\tfrac{1}{5}k_1)\\
k_3 &:= f(t+\tfrac{3}{10}h, u+h(\tfrac{3}{40}k_1+\tfrac{9}{40}k_2))\\
k_4 &:= f(t+\tfrac{4}{5}h, u+h(\tfrac{44}{45}k_1-\tfrac{56}{15}k_2+\tfrac{32}{9}k_3))\\
k_5 &:= f(t+\tfrac{8}{9}h, u+h(\tfrac{19372}{6561}k_1-\tfrac{25360}{2187}k_2+\tfrac{64448}{6561}k_3-\tfrac{212}{729}k_4))\\
k_6 &:= f(t+h, u+h(\tfrac{9017}{3168}k_1-\tfrac{355}{33}k_2+\tfrac{46732}{5247}k_3+\tfrac{49}{176}k_4-\tfrac{5103}{18656}k_5))\\
u_+ &= u+h(\tfrac{35}{384}k_1+\tfrac{500}{1113}k_3+\tfrac{125}{192}k_4-\tfrac{2187}{6784}k_5+\tfrac{11}{84}k_6)
\end{aligned}$$

Anschließend berechnet man

$$\begin{aligned}
t_+ &:= t+h\\
k_7 &:= f(t_+,u_+)\\
d_+ &:= \tfrac{71}{57600}k_1-\tfrac{71}{16695}k_3+\tfrac{71}{1920}k_4-\tfrac{17253}{339200}k_5+\tfrac{22}{525}k_6-\tfrac{1}{40}k_7.
\end{aligned}$$

Setzt man $\tilde{u}_+ := u_+ - hd$, so ist

$$\tilde{u}_+ = u+h(\tfrac{5179}{57600}k_1+\tfrac{7571}{16695}k_3+\tfrac{393}{640}k_4-\tfrac{92097}{339200}k_5+\tfrac{187}{2100}k_6+\tfrac{1}{40}k_7).$$

Die Formeln für $(u_+, \tilde{u}_+)$ bilden das Dormand-Prince-Paar der Ordnung 4 bzw. 5, offenbar sind pro Iterationsschritt 6 Funktionsauswertungen nötig[14]. Etwas kompakter hätten wir obige Formeln natürlich auch in einem Runge-Kutta-Schema anordnen können. Je nach der Größe von $d$ (bzw. $\|d\|$) wird $u_+$ als neue Näherung akzeptiert. Auf Einzelheiten, auch zur Schrittweitensteuerung, können wir nicht mehr eingehen. Der Vorteil eine "eingebetteten" Paares von Runge-Kutta-Verfahren besteht gerade darin, dass verhältnisnäßig einfach eine Schrittweitensteuerung möglich ist, siehe z. B. E.Hairer et al. (1993, S. 167).

Will man die Lösung der Anfangswertaufgabe

$$x' = f(t,x), \qquad x(t_0) = x_0$$

[14]Die Bestimmung der Konsistenzordnung (wie beim klassischen Runge-Kutta) stellt an Maple allerdings schon ganz erhebliche Anforderungen. Nach einer Stunde Rechenzeit hatten wir noch kein Ergebnis erreicht.

zu bestimmten Zeiten $t_0, \ldots, t_5$ bestimmen, so kann man dies durch

```
tspan=[t_0 t_1 t_2 t_3 t_4 t_5];
[t,x]=ode45('VDpol',tspan,x_0);
```

erreichen. Jetzt haben `t` und `x` natürlich nur 6 Komponenten bzw. Zeilen. Man kann ferner Parameter an die rechte Seite der Differentialgleichung übergeben. Z. B. könnte man das File `VDpol.m` folgendermaßen ändern:

```
function varargout = VDpol(t,x,flag,mu);
%VDpol(t,x) returns the state derivative of the van der Pol equation:
%
%               x''-mu(1-x^2)x'+x=0
%
%VDpol(t,x) or VDpol(t,x,[],mu) returns the derivatives vector for the
%   van der Pol equation.  By default, mu is 1, and the problem is not
%   stiff.  Optionally, pass in the mu parameter as an additional parameter
%   to an ODE Suite solver.  The problem becomes more stiff as MU is
%   increased.
if nargin < 4 | isempty(mu)
  mu = 1;
end
switch flag
case ''                                  % Return dx/dt = f(t,x).
  varargout{1} = f(t,x,mu);
otherwise
  error(['Unknown flag ''' flag '''.']);
end

function dxdt = f(t,x,mu)
dxdt = [x(2); (mu*(1-x(1)^2)*x(2) - x(1))];
```

und anschließend den Aufruf

```
[t,x]=ode45('VDpol',[0 30],[1;0],[],10);
```

Hiermit wird die van der Polsche Differentialgleichung mit $\mu = 10$ numerisch gelöst. Diverse weitere Optionen sind möglich, hierzu verweisen wir auf die MATLAB-Hilfen. Im MATLAB Function Reference kann man lesen: In general, `ode45` is the best function to apply as a "first try" for most problems.

### 3.3.2 ODE23

Für die MATLAB-Funktion `ode23` gilt praktisch alles, was im vorigen Unterabschnitt über `ode45` gesagt wurde. Die Syntax ist vollständig identisch, daher können wir uns ganz kurz fassen. Fast der einzige Unterschied besteht darin, das ein Paar von Runge-

Kutta-Formeln der Ordnung 2 bzw. 3 benutzt wird. Genauer wird berechnet

$$\begin{aligned} k_1 &:= f(t,u) \\ k_2 &:= f(t+\tfrac{1}{2}h, u+h\tfrac{1}{2}k_1) \\ k_3 &:= f(t+\tfrac{3}{4}h, u+\tfrac{3}{4}k_2) \\ u_+ &:= u+h(\tfrac{2}{9}k_1+\tfrac{1}{3}k_2+\tfrac{4}{9}k_3) \\ t_+ &:= t+h \\ k_4 &:= f(t_+,u_+) \\ d &:= -\tfrac{5}{72}k_1+\tfrac{1}{12}k_2+\tfrac{1}{9}k_3-\tfrac{1}{8}k_4 \end{aligned}$$

Dann ist

$$\tilde{u}_+ := u_+ - hd = u + h(\tfrac{7}{24}k_1+\tfrac{1}{4}k_2+\tfrac{1}{3}k_3+\tfrac{1}{8}k_4).$$

Mit $(u_+, \tilde{u}_+)$ hat man ein Paar von Runge-Kutta-Formeln, wobei pro Iterationsschritt 3 Funktionsauswertungen benötigt werden. Diesmal ist es nicht schwierig, mit Maple die jeweilige Konsistenzordnung zu bestimmen. Wir erhalten, dass

$$\Phi(h,f)(t,u) := \frac{2}{9}k_1+\frac{1}{3}k_2+\frac{4}{9}k_3$$

Verfahrensfunktion eines Einschrittverfahrens der Konsistenzordnung 3 ist, während

$$\tilde{\Phi}(h,f)(t,u) := \frac{7}{24}k_1+\frac{1}{4}k_2+\frac{1}{3}k_3+\frac{1}{8}k_4$$

zu einem Einschrittverfahren der Ordnung 2 führt. Hierbei sind $k_1, k_2, k_3, k_4$ wie oben angegeben definiert.

### 3.3.3 Aufgaben

1. Bei E. Hairer et al. (1993) und W. Walter (1993) findet man folgendes System von zwei Differentialgleichungen erster Ordnung (siehe auch Aufgabe 4 in Abschnitt 1.2):

$$\begin{aligned} \ddot{x} &= x+2\dot{y}-\mu'\frac{x+\mu}{[(x+\mu)^2+y^2]^{3/2}}-\mu\frac{x-\mu'}{[(x-\mu')^2+y^2]^{3/2}}, \\ \ddot{y} &= y-2\dot{x}-\mu'\frac{y}{[(x+\mu)^2+y^2]^{3/2}}-\mu\frac{y}{[(x-\mu')^2+y^2]^{3/2}}. \end{aligned}$$

   Hierbei ist $\mu$ eine gegebene Konstante und $\mu' := 1-\mu$. Für $\mu := 0.01212277471$ und die Anfangsbedingungen

$$x(0)=1.2, \quad \dot{x}(0)=0, \quad y(0)=0, \quad \dot{y}(0)=-1.04936$$

   sowie

$$x(0)=0.994, \quad \dot{x}(0)=0, \quad y(0)=0, \quad \dot{y}(0)=-2.0015851063791$$

   berechne man mit Hilfe der MATLAB-Funktion `ode45` jeweils eine Lösung und plotte die Phasenbahn $\{(x(t),y(t)) : t \in I\}$, wobei das Intervall $I$ einmal $[0,7]$ und einmal $[0,17.1]$ ist.

2. Man löse die folgende Anfangswertaufgabe (Euler-Gleichungen für die Bewegung eines Festkörpers ohne äußere Kräfte, siehe z. B. L. F. Shampine, M. K. Gordon (1975, S. 243)[15])

$$\begin{array}{rclrcl} x_1' &=& x_2x_3 & x_1(0) &=& 0, \\ x_2' &=& -x_1x_3 & x_2(0) &=& 1, \\ x_3' &=& -0.51x_1x_2 & x_3(0) &=& 1 \end{array}$$

auf dem Zeitintervall $[0, 12]$ mit Hilfe der MATLAB-Funktion `ode45`. Ferner plotte man die Lösungskomponenten auf diesem Intervall.

## 3.4 Steife Differentialgleichungen

### 3.4.1 Beispiele, Motivation

Wir wollen noch gar nicht eine Definition steifer Differentialgleichungen versuchen, sondern zunächst nur Beispiele von Anfangswertaufgaben angeben, mit denen die bisher geschilderten Verfahren (also explizite Einschrittverfahren vom Runge-Kutta-Typ und lineare Mehrschrittverfahren) Schwierigkeiten haben.

**Beispiel:** Wir betrachten das System von zwei Differentialgleichungen erster Ordnung:

$$x' = Ax \qquad \text{mit} \qquad A := \begin{pmatrix} \frac{1}{2}(\lambda_1+\lambda_2) & \frac{1}{2}(\lambda_1-\lambda_2) \\ \frac{1}{2}(\lambda_1-\lambda_2) & \frac{1}{2}(\lambda_1+\lambda_2) \end{pmatrix}$$

und *negativen* Konstanten $\lambda_1, \lambda_2$. Die symmetrische Matrix $A$ besitzt die Eigenwerte $\lambda_1$, $\lambda_2$ mit zugehörigen Eigenvektoren $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Die allgemeine Lösung lautet daher

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} C_1e^{\lambda_1 t} + C_2e^{\lambda_2 t} \\ C_1e^{\lambda_1 t} - C_2e^{\lambda_2 t} \end{pmatrix}$$

mit Integrationskonstanten $C_1, C_2$, woraus man noch einmal abliest, dass das vorgegebene System asymptotisch stabil ist. Benutzt man zur Integration der Anfangswertaufgabe

$$x' = Ax, \qquad \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} C_1 + C_2 \\ C_1 - C_2 \end{pmatrix} =: x_0$$

das (explizite) Euler'sche Polygonzugverfahren

$$u_0 := x_0, \qquad u_{j+1} := u_j + hAu_j, \qquad j = 0, 1, \ldots,$$

so erhält man

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

[15] L. F. Shampine, M. K. Gordon (1975) *Computer Solution of Ordinary Differential Equations. The Initial Value Problem.* W. H. Freeman and Company, San Francisco.

Aus der Konvergenztheorie für Einschrittverfahren wissen wir, dass der globale Diskretisierungsfehler mit $h \to 0+$ gegen Null konvergiert. Das asymptotische Verhalten wird aber nur dann richtig wiedergegeben, wenn die Schrittweite $h > 0$ so klein ist, dass

$$|1 + h\lambda_1| < 1 \qquad \text{und} \qquad |1 + h\lambda_2| < 1.$$

Nun nehmen wir an, es sei $\lambda_2 < \lambda_1$ und $|\lambda_2|$ sei wesentlich größer als $|\lambda_1|$. Der Einfluss von $e^{\lambda_2 t}$ in der Lösung ist dann, zumal für große $t > 0$, vernachlässigbar klein gegenüber $e^{\lambda_1 t}$. Trotzdem bestimmt $\lambda_2$ durch die Forderung $|1 + h\lambda_2| < 1$ bzw. $h < 2/|\lambda_2|$ wesentlich die Schrittweite. □

Wir geben nun ein spezielles Beispiel für eine lineare Anfangswertaufgabe an und wenden bisher entwickelte MATLAB-Funktionen auf dieses Beispiel an.

**Beispiel:** Gegeben sei die Anfangswertaufgabe

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} -51 & 49 \\ 49 & -51 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Die exakte Lösung (wir haben hier einen Spezialfall des ersten Beispiels) ist

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} 2e^{-2t} - e^{-100t} \\ 2e^{-2t} + e^{-100t} \end{pmatrix}.$$

Wir wollen die Lösung numerisch über das Intervall $[0, 10]$ berechnen. Wenden wir `FixedRK` mit $m = 101$ bzw. der Maschenweite $h = 0.1$ an, so "explodieren" die Werte, während sich für $m = 1001$ bzw. $h = 0.01$ einigermaßen vernünftige Werte ergeben (für die erste Komponente in $t = 10$ z. B. `4.122307356653245e-09` statt `4.122307244877116e-09`, was in Ordnung ist). Macht man den entsprechenden Vergleich mit `FixedEuler`, so erhält man im Prinzip dasselbe Ergebnis. Wenden wir `ode45` an, so werden 1229 Zeitschritte gemacht und für $t = 10$ erhält man für die erste Komponente die Näherung `-7.709504006196912e-07`, was nicht ganz richtig ist. Wir erkennen hier also in der Praxis, was wir im ersten Beispiel zumindestens für das explizite Euler-Verfahren auch theoretisch eingesehen haben: Um das asymptotische Verhalten der Lösung auch nur einigermaßen richtig zu reproduzieren, sind wir gezwungen, sehr kleine Zeitschritte zu benutzen. □

**Beispiel:** Man betrachte das Differentialgleichungssystem $x' = Ax$, wobei $A \in \mathbb{R}^{n \times n}$ durch

$$A := \frac{1}{(n+1)^2} \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix}$$

gegeben ist. Anfangswertaufgaben für dieses Differentialgleichungssystem treten auf, wenn man eine gewisse Anfangsrandwertaufgabe für die Wärmeleitungsgleichung in

einer Raumdimension bezüglich der Ortsvariablen diskretiert. Das aber nur am Rande, hierauf wollen wir nicht näher eingehen. Die Matrix $A \in \mathbb{R}^{n\times n}$ besitzt die Eigenwerte

$$\lambda_i := -4(n+1)^2 \sin^2\Big(\frac{i\pi}{2(n+1)}\Big), \qquad i = 1, \ldots, n.$$

Alle Eigenwerte sind negativ, daher ist die Nulllösung asymptotisch stabil. Für $n \to \infty$ gilt $\lambda_1 \approx -\pi^2$ und $\lambda_n \approx -4(n+1) \to -\infty$. Wir können jetzt für das explizite Euler-Verfahren (mit etwas mehr Mühe auch für andere Einschrittverfahren) das feststellen, was wir im ersten Beispiel gesehen haben: Da ein Eigenwert von $A$ für großes $n$ sehr klein ist, sind wir gezwungen, sehr kleine Schrittweiten zu benutzen. Genauer: Zu lösen sei die Anfangswertaufgabe $x' = Ax$, $x(0) = x_0$, für $t > 0$. Mit $m \in \mathbb{N}$ wählen wir die Schrittweite $h_m := t/m$. Wenden wir das Euler-Verfahren an, so erhalten wir

$$u(t; h_m) = (I + h_m A)^m u_0$$

als zugehörigen Näherungswert, wobei wir natürlich $u_0 := x_0$ setzen. Da $A$ symmetrisch ist, existiert eine orthogonale Matrix $V \in \mathbb{R}^{n\times n}$ derart, dass $A = V^T \Lambda V$, wobei $\Lambda := \text{diag}\,(\lambda_1, \ldots, \lambda_n)$. Dann ist

$$u(t; h_m) = V^T (I + h_m \Lambda)^m V u_0,$$

während die Lösung selber gegeben ist durch $x(t) = V^T e^{\Lambda t} V$. Daher ist

$$\begin{aligned} u(t; h_m) - x(t) &= V^T[(I + (t/m)\Lambda)^m - e^{\Lambda t}]Vx_0 \\ &= V^T \text{diag}\,((1 + (t/m)\lambda_i)^m - e^{\lambda_i t})Vx_0. \end{aligned}$$

Zwar wissen wir wegen der Konvergenz konsistenter Einschrittverfahren (und lesen es aus der letzten Gleichung noch einmal ab), dass $\lim_{m\to\infty} u(t; h_m) = x(t)$. Damit das asymptotische Verhalten aber richtig wiedergespiegelt wird, muss $m$ sehr gross sein. Um dies zu verdeutlichen, tragen wir in der folgenden Tabelle $m$ und $(1 - 1000/m)^m$ ein (etwa $t = 10$ und $\lambda = -100$, z. B. ist $\lambda_{30} \approx -123.7$, wenn $n = 30$):

| $m$ | $(1 - 1000/m)^m$ |
|---|---|
| 100 | $2.656139888758748e + 95$ |
| 250 | $1.906837481167966e + 119$ |
| 500 | 1 |

Dagegen ist $e^{-1000} = 0$ (jedenfalls für MATLAB). □

Das folgende Beispiel findet man bei K. Strehmel, R. Weiner (1995, S. 208). In ihm wird gezeigt, dass das explizite Euler-Verfahren und auch das klassische Runge-Kutta-Verfahren völlig versagen können, während das implizite Euler-Verfahren befriedigende Werte liefert.

**Beispiel:** Wir betrachten das einfache Anfangswertproblem

$$x' = \lambda(x - e^{-t}) - e^{-t}, \qquad x(0) = 1,$$

mit der von $\lambda$ unabhängigen exakten Lösung $x(t) = e^{-t}$. Wir wenden das explizite Euler-Verfahren und das klassische Runge-Kutta-Verfahren für $\lambda = -10$ und $\lambda = -1000$ an, wobei wir die Schrittweite $h = 2^{-k}$, $k = 4, 6, 8, 10$, benutzen und den Fehler zur Endzeit $t_e := 1$ eintragen. Als Vergleich geben wir noch die entsprechenden Ergebnisse für das implizite Euler-Verfahren an. Hierzu haben wir eine MATLAB-Funktion geschrieben, welche zu vorgegebenen $k \in \mathbb{N}$ und $\lambda$ die durch das implizite Euler-Verfahren zur Schrittweite $h = 2^{-k}$ für $t_e = 1$ gewonnenene Näherungslösung zur obigen Anfangswertaufgabe berechnet. Dies kann etwa folgendermaßen geschehen:

```
function out=ImplEuler(k,lambda);

u_c=1;n=2^k;h=1/n;t_c=0;
for i=0:n-1
   t_c=t_c+h;u_c=(u_c-h*(1+lambda)*exp(-t_c))/(1-h*lambda);
end;
out=u_c-exp(-1);
```

Wir erhalten folgende Werte für den Fehler $u(1; 2^{-k}) - e^{-1}$.

| | $\lambda = -10$ | | | $\lambda = -1000$ | | |
|---|---|---|---|---|---|---|
| $k$ | expl. Euler | Runge-Kutta | impl. Euler | expl. Euler | Runge-Kutta | impl. Euler |
| 4 | $1.2e-003$ | $8.7e-006$ | $1.3e-003$ | $1.3e+024$ | $4.3e+087$ | $1.1e-005$ |
| 6 | $3.2e-004$ | $2.7e-008$ | $3.2e-004$ | $2.9e+069$ | $2.1e+205$ | $2.9e-006$ |
| 8 | $8.0e-005$ | $9.8e-011$ | $8.0e-005$ | $8.0e+112$ | $2.6e+161$ | $7.2e-007$ |
| 10 | $2.0e-005$ | $3.8e-013$ | $2.0e-005$ | $1.8e-007$ | $5.5e-009$ | $1.8e-007$ |

Für $\lambda = -1000$ müssen wir beim expliziten Euler-Verfahren und dem (expliziten) Runge-Kutta-Verfahren sehr kleine Schrittweiten (etwa $2^{-10}$) wählen, um vernünftige Werte zu erhalten, während das implizite Euler-Verfahren stets gute Werte liefert. Das gilt auch für die MATLAB-Funktion `ode45`. Denn nach dem Aufruf `[t,x]=ode45('f',[0 1],1);` stellen wir durch `length(t)` fest, dass 1205 Zeitschritte benutzt wurden (zum Vergleich: $2^{10} = 1024$). Der Fehler ist sogar nur in der Größenordnung $10^{-4}$. □

Wir betrachten nun zunächst das explizite Euler-Verfahren mit der Schrittweite $h$, angewandt auf die Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$. Bezeichnen wir mit $e(t) := u(t; h) - x(t)$ den globalen Diskretisierungsfehler zur Zeit $t$ und schreiben wir $u(t)$ statt $u(t; h)$, so ist

$$\begin{aligned} e(t+h) &= u(t+h) - x(t+h) \\ &= u(t) + hf(t, u(t)) - x(t+h) \\ &= u(t) - x(t) - [x(t+h) - x(t) - hf(t, x(t)] + h[f(t, u(t)) - f(t, x(t))] \\ &= e(t) - h\Delta(h, f)(t, x(t)) + h[f(t, x(t) + e(t)) - f(t, x(t))] \\ &= [I + hf_x(t, x(t))]e(t) + O(h^2), \end{aligned}$$

wobei $\Delta(h, f)$ der lokale Diskretisierungsfehler ist (das Euler-Verfahren hat die Konsistenzordnung 1, es ist also $\Delta(h, f) = O(h)$). Der zweite Summand $O(h^2)$ ist bestimmt durch die Genauigkeit des Verfahrens, während der erste Term angibt, wie der globale Fehler von Schritt zu Schritt angefacht wird. Zum Vergleich betrachten wir die

Situation beim impliziten Euler-Verfahren. Diesmal ist

$$\begin{aligned} e(t+h) &= u(t+h) - x(t+h) \\ &= u(t) + hf(t+h, u(t+h)) - x(t+h) \\ &= u(t) - x(t) - [x(t+h) - x(t) - hf(t+h, x(t+h))] \\ &\qquad + h[f(t+h, u(t+h)) - f(t+h, x(t+h))] \\ &= e(t) + hf_x(t+h, x(t+h))e(t+h) + O(h^2). \end{aligned}$$

Diesmal ist also (etwas lax argumentiert)

$$e(t+h) = [I - hf_x(t+h, x(t+h))]^{-1} e(t) + O(h^2).$$

Hier erkennt man also genau den prinzipiellen Unterschied zwischen explizitem und implizitem Euler-Verfahren.

Es gibt in der Literatur keine exakte Definition für die *Steifheit* einer Differentialgleichung. I. Allg. versteht man darunter eine Differentialgleichung $x' = f(t, x)$, für welche $(T - t_0)\|f_x(t, x(t))\| \gg 1$, wobei $[t_0, T]$ das Integrationsintervall ist, bzw. mit einer Lipschitzkonstanten $L$ (für $f$ bezüglich der zweiten Variablen) gilt, dass $(T - t_0)L \gg 1$. Aus der Abschätzung für den globalen Diskretisierungsfehler eines Einschrittverfahrens der Konsistenzordnung $p$ in Abschnitt 3.1, Korollar 1.5

$$\|e(t)\| \leq Nh^p \, \frac{e^{M(t-t_0)} - 1}{M},$$

wobei $M$ Lipschitzkonstante der Verfahrensfunktion ist, erkennen wir, dass wir mit Schwierigkeiten zu rechnen haben, wenn $(T - t_0)M \gg 1$.

### 3.4.2 Stabilitätsgebiet expliziter Runge-Kutta-Verfahren

Ein $s$-stufiges explizites Runge-Kutta-Verfahren zur Lösung der Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, mit der Lösung $x(\cdot)$ berechnet bekanntlich aus einer Näherung $u(t)$ für $x(t)$ und einer Schrittweite $h > 0$ eine Näherung $u(t + h)$ für $x(t + h)$ durch die folgende Vorschrift: Es ist

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

wobei

$$\begin{aligned} k_1(t,u) &= f(t,u), \\ k_2(t,u) &= f(t + c_2 h, u + h a_{21} k_1(t,u)), \\ &\vdots \\ k_s(t,u) &= f(t + c_s h, u + h \textstyle\sum_{i=0}^{s-1} a_{si} k_i(t,u))). \end{aligned}$$

Dabei sind die $a_{ki}$, $b_i$ und $c_i$ geeignet gewählte reelle Zahlen, die das Verfahren vollständig festlegen.

Wendet man das obige explizite $s$-stufige Runge-Kutta-Verfahren auf die sogenannte *Test-Differentialgleichung* $x' = \lambda x$ mit dem Skalar $\lambda$ an, so erhält man, dass

$$u(t+h) = R(h\lambda)u(t),$$

wobei $R$ ein Polynom vom Grad $\leq s$ ist. Das Polynom $R$ heißt zugehörige Stabilitätsfunktion, die Menge

$$S := \{z \in \mathbb{C} : |R(z)| \leq 1\}$$

das zum Verfahren gehörende *Stabilitätsgebiet*. Ist in der Testgleichung $\lambda < 0$ bzw. $\Re(\lambda) < 0$ für komplexes $\lambda$, so ist die Nulllösung asymptotisch stabil. Damit auch die Näherungslösung mit wachsender Zeit abklingt, sollte also $h\lambda$ im Innern des Stabilitätsgebietes liegen.

**Beispiele:** Das explizite Euler-Verfahren, angewandt auf $x' = \lambda x$, ist

$$u(t+h) = u(t) + h\lambda u(t) = (1 + h\lambda)u(t),$$

die zugehörige Stabilitätsfunktion ist also $R(z) := 1 + z$, das Stabilitätsgebiet ist der Kreis um $-1$ mit dem Radius 1.

Das Verfahren von Heun, wieder angewandt auf $x' = \lambda x$, ergibt

$$u(t+h) = u(t) + \frac{h}{2}[\lambda u(t) + \lambda(u(t) + h\lambda u(t))] = R(h\lambda)u(t)$$

mit

$$R(z) := 1 + z + \frac{1}{2}z^2.$$

Das Stabilitätsgebiet bzw. seinen Rand geben wir in Abbildung 3.4 links an. Beim klassischen Runge Kutta-Verfahren berechnet man

$$k_1 := \lambda u(t), \quad k_2 := \lambda(u(t) + \tfrac{1}{2}hk_1), \quad k_3 := \lambda(u(t) + \tfrac{1}{2}hk_2), \quad k_4 := \lambda(u(t) + hk_3)$$

und anschließend

$$u(t+h) := u(t) + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4).$$

Einsetzen ergibt $u(t+h) = R(h\lambda)u(t)$ mit

$$R(z) := 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4.$$

Das Stabilitätsgebiet bzw. seinen Rand geben wir in Abbildung 3.4 rechts an. □

**Satz 4.1** *Die Stabilitätsfunktion eines $s$-stufigen expliziten Runge-Kutta-Verfahrens der Konsistenzordnung $p$ ist ein Polynom vom Grad $\leq s$, welches sich in der Form*

$$R(z) = \sum_{i=0}^{p} \frac{z^i}{i!} + O(z^{p+1})$$

*darstellen lässt.*

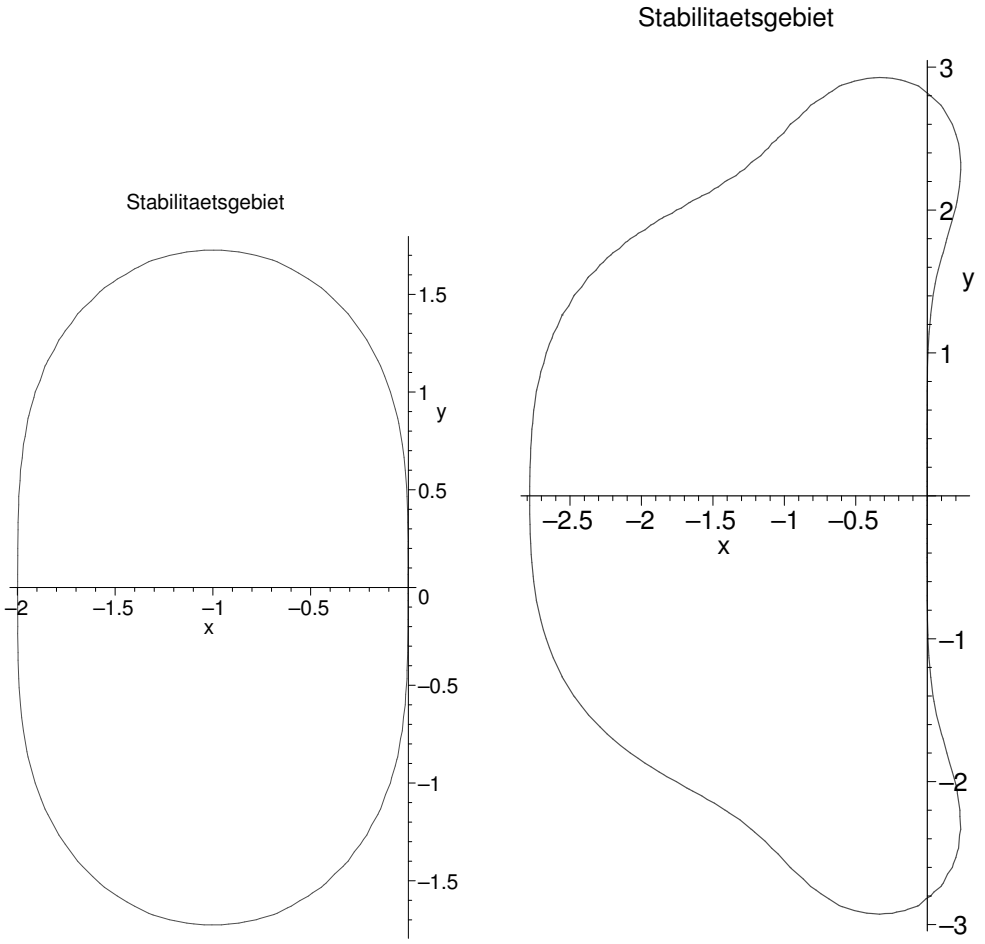


Abbildung 3.4: Stabilitätsgebiet des Heun- bzw. Runge-Kutta-Verfahrens

**Beweis:** Die exakte Lösung von $x' = \lambda x$, $x(0) = 1$, zur Zeit $h$ ist $e^{\lambda h}$, während $u(h) = R(h\lambda)$ der entsprechende durch das Verfahren gewonnene Näherungswert ist. Wegen $u(h) - e^{\lambda h} = O(h^{p+1})$ (dies erkennt man aus Korollar 1.5 in Abschnitt 3.1) folgt die Behauptung. □

Ein explizites $s$-stufiges Runge-Kutta-Verfahren der Konsistenzordnung $p = s$ besitzt also die Stabilitätsfunktion

$$R(z) = 1 + z + \cdots + \frac{z^s}{s!}.$$

Das Stabilitätsgebiet expliziter Runge-Kutta-Verfahren ist beschränkt. Insbesondere kann nicht die gesamte linke Halbebene zum Stabilitätsgebiet gehören.

**Beispiel:** Ein 6-stufiges explizites Runge-Kutta-Verfahren der Konsistenzordnung 5 ist das in `ode45` benutzte Verfahren von Dormand-Prince. Unter Benutzung von Maple erhalten wir, dass

$$R(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24} + \frac{z^5}{120} + \frac{z^6}{600}$$

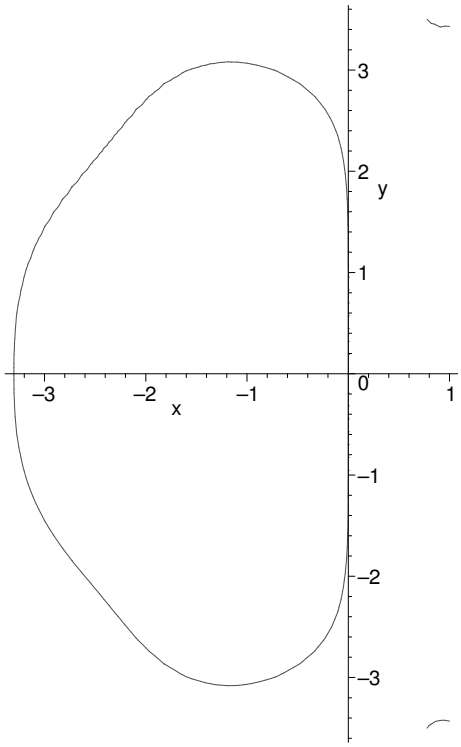


Abbildung 3.5: Stabilitätsgebiet des Dormand-Prince-Verfahrens

die zugehörige Stabilitätsfunktion ist. In Abbildung 3.5 wird das Stabilitätsgebiet bzw. sein Rand geplottet. Für negatives $\lambda$ sollte die Schrittweite also etwa so klein sein, dass $-3.3 < h\lambda$. Für $\lambda = -1000$ führt dies auf die Forderung, dass $h < 0.0033$ sein sollte. □

### 3.4.3 Stabilitätsgebiet linearer Mehrschrittverfahren

Ein allgemeines lineares $r$-Schrittverfahren hat die Form

$$a_r u(t+rh) + \cdots + a_0 u(t) = h[b_r f(t+rh, u(t+rh)) + \cdots + b_0 f(t, u(t))],$$

wobei $a_r \neq 0$ und daher o. B. d. A. $a_r = 1$. Wenden wir dieses Verfahren auf die Test-Differentialgleichung $x' = \lambda x$ an, so erhalten wir die lineare Differenzengleichung

$$\sum_{i=0}^{r} (a_i - h\lambda b_i) u(t+ih) = 0.$$

Deren Lösungen sind bestimmt durch das zugehörige charakteristische Polynom

$$\rho(\mu, z) := \sum_{i=0}^{r} (a_i - z b_i)\mu^i = \psi(\mu) - z\chi(\mu),$$

wobei $\psi, \chi$ die das obige lineare Mehrschrittverfahren bestimmenden Polynome

$$\psi(\mu) := \sum_{i=0}^{r} a_i \mu^i, \qquad \chi(\mu) := \sum_{i=0}^{r} b_i \mu^i$$

sind. Daher liegt die folgende Definition nahe:

**Definition 4.2** Gegeben sei ein lineares $r$-Schrittverfahren, das durch die Polynome

$$\psi(\mu) := \sum_{i=0}^{r} a_i \mu^i, \qquad \chi(\mu) := \sum_{i=0}^{r} b_i \mu^i$$

bestimmt ist. Dann heißt für $z \in \mathbb{C}$ das Polynom

$$\rho(\mu, z) := \psi(\mu) - z\chi(\mu)$$

das zugehörige *Stabilitätspolynom*. Die Menge

$$S := \left\{ z \in \mathbb{C} : \begin{array}{l} \rho(\mu, z) = 0 \Longrightarrow |\mu| \leq 1, \\ \rho(\mu, z) = 0,\ |\mu| = 1 \Longrightarrow \mu \text{ einfache Nullstelle} \end{array} \right\}$$

heißt das zum Mehrschrittverfahren gehörende *Stabilitätsgebiet*. Ein lineares Mehrschrittverfahren heißt *A-stabil*, wenn die gesamte linke Halbebene zum Stabilitätsgebiet gehört. Äquivalent hierzu ist: Ist $\Re(\lambda) \leq 0$, so ist die durch das Mehrschrittverfahren, angewandt auf die Testgleichung $x' = \lambda x$, gewonnene Näherungslösung beschränkt (siehe auch die Bemerkung am Schluss von Unterabschnitt 3.2.3..

**Beispiel:** Das Stabilitätsgebiet der Adams-Bashforth- oder Adams-Moulton-Verfahren zu bestimmen, ist ziemlich kompliziert. Man stellt fest, dass diese mit wachsender Konsistenzordnung kleiner werden und weit davon entfernt sind, A-stabil zu sein (bis auf die ersten beiden Adams-Moulton-Verfahren). Sie sind bei E. Hairer, G. Wanner (1991, S. 260) abgebildet. Wir wollen daher nur ein verhältnismäßig einfaches Beispiel betrachten, das aber immerhin die beiden ersten Adams-Moulton-Verfahren enthält.

Mit $\theta \in \mathbb{R}$ betrachten wir das Verfahren

$$u(t+h) = u(t) + h[(1-\theta)f(t, u(t)) + \theta f(t+h, u(t+h))].$$

Als Sonderfälle enthält dieses Verfahren das explizite Euler-Verfahren ($\theta = 0$), das implizite Euler-Verfahren ($\theta = 1$) und die Trapezregel ($\theta = \frac{1}{2}$). Die Konsistenzordnung ist 2 für $\theta = \frac{1}{2}$, ansonsten 1. Das Stabilitätspolynom ist

$$\rho(\mu, z) = (1 - \theta z)\mu - 1 - (1-\theta)z,$$

das Stabilitätsgebiet sei mit $S$ bezeichnet. Die Wurzel des Stabilitätspolynoms ist

$$\mu(z) = \frac{1 + (1-\theta)z}{1 - \theta z}.$$

Daher ist $z = x + iy \in S$ genau dann, wenn

$$|\mu(z)|^2 = \frac{(1 + (1-\theta)x)^2 + (1-\theta)^2 y^2}{(1-\theta x)^2 + \theta^2 y^2} \leq 1$$

bzw. nach einfacher Rechnung

$$2x + (1 - 2\theta)(x^2 + y^2) \leq 0.$$

Für $\theta = \frac{1}{2}$ ist das Stabilitätsgebiet also genau die linke komplexe Halbebene. Für $\theta > \frac{1}{2}$ gehört $z$ genau dann zum Stabilitätsgebiet, wenn

$$\frac{1}{(2\theta-1)^2} \leq \Big(x - \frac{1}{2\theta-1}\Big)^2 + y^2,$$

d. h. $S$ ist das Äußere des (offenen) Kreises um $(1/(2\theta-1), 0)$ mit dem Radius $1/(2\theta-1)$. Die linke komplexe Halbebene ist im Stabilitätsgebiet enthalten. Daher ist das Verfahren für $\theta \geq \frac{1}{2}$ A-stabil. Für $\theta < \frac{1}{2}$ gehört $z$ genau dann zum Stabilitätsgebiet, wenn

$$\Big(x + \frac{1}{1-2\theta}\Big)^2 + y^2 \leq \frac{1}{(1-2\theta)^2},$$

d. h. $S$ ist der (abgeschlossene) Kreis um $(-1/(1-2\theta), 0)$ mit dem Radius $1/(1-2\theta)$. In diesem Falle ist das Verfahren also nicht A-stabil. □

Im vorigen Unterabschnitt haben wir gesehen, dass explizite Runge-Kutta-Verfahren nicht A-stabil sind, da sie ein beschränktes Stabilitätsgebiet besitzen. Wir wollen uns überlegen, dass auch explizite lineare $r$-Schrittverfahren nicht A-stabil sein können. Es gilt nämlich:

**Satz 4.3** *1. Ist ein zu $(\psi, \chi)$ gehörendes lineares Mehrschrittverfahren A-stabil, so gilt*

$$(*) \qquad \Re\Big(\frac{\psi(\mu)}{\chi(\mu)}\Big) > 0 \qquad \text{für alle } \mu \text{ mit } |\mu| > 1.$$

*2. Sind $\psi$ und $\chi$ irreduzibel, besitzen $\psi$ und $\chi$ also keine gemeinsame Nullstelle, so gilt auch die Umkehrung der obigen Aussage, d. h. $(*)$ impliziert die A-Stabilität des Verfahrens.*

*3. Ein explizites lineares $r$-Schrittverfahren ist nicht A-stabil.*

**Beweis:** Das zu $(\psi, \chi)$ gehörende lineare Mehrschrittverfahren sei A-stabil. Angenommen, es existiert ein $\mu_0$ mit

$$|\mu_0| > 1, \qquad \Re(\psi(\mu_0)/\chi(\mu_0)) \leq 0.$$

Mit $z_0 := \psi(\mu_0)/\chi(\mu_0)$ hat man eine komplexe Zahl gefunden, die nichtpositiven Realteil besitzt, aber nicht zum Stabilitätsgebiet gehört, was ein Widerspruch zur A-Stabilität ist.

Nun seien $\psi$ und $\chi$ irreduzibel, ferner gelte $(*)$. Wir wollen zeigen, dass dies die A-Stabilität des zu $(\psi, \chi)$ gehörenden Verfahrens impliziert. Denn sei $z_0 \in \mathbb{C}$ mit $\Re(z_0) \leq 0$ beliebig gegeben und $\mu_0$ eine Wurzel von $\rho(\cdot, z_0)$. Wir haben zu zeigen, dass $|\mu_0| \leq 1$ ist, und aus $|\mu_0| = 1$ folgt, dass $\mu_0$ eine einfache Wurzel von $\rho(\cdot, z_0)$ ist. Denn dann ist gezeigt, dass $z_0$ zum Stabilitätsgebiet des Verfahrens gehört. Es ist $\chi(\mu_0) \neq 0$ (wegen der Irreduzibilität von $\psi$ und $\chi$), daher ist $z_0 = \psi(\mu_0)/\chi(\mu_0)$. Wegen $(*)$ und $\Re(z_0) \leq 0$

folgt $|\mu_0| \le 1$. Wir nehmen nun an, es sei $|\mu_0| = 1$. Dies ist nur möglich, wenn $\Re(z_0) = 0$. Denn aus $(*)$ folgt durch ein Stetigkeitsargument, dass

$$\Re\Big(\frac{\psi(\mu)}{\chi(\mu)}\Big) \ge 0 \qquad \text{für alle } \mu \text{ mit } |\mu| \ge 1.$$

Es bleibt daher zu zeigen: Ist $\Re(z_0) = 0$ und $\mu_0$ eine Wurzel von $\rho(\cdot, z_0)$ mit $|\mu_0| = 1$, so ist $\mu_0$ einfach. In einer Umgebung von $\mu_0$ ist

$$\frac{\psi(\mu)}{\chi(\mu)} - z_0 = c_1(\mu - \mu_0) + c_2(\mu - \mu_0)^2 + \cdots.$$

Es ist $c_1 \neq 0$ wegen $(*)$ (???) und dies impliziert wiederum die Einfachheit von $\mu_0$.

Bei einem expliziten $r$-Schrittverfahren hat das Stabilitätspolynom die Form

$$\rho(\mu, z) = \psi(\mu) - z\chi(\mu),$$

wobei $\psi \in \Pi_r$ und $\chi \in \Pi_q \setminus \Pi_{q-1}$ mit $q < r$. Wegen

$$\frac{\psi(\mu)}{\chi(\mu)} = \frac{a_r\mu^r + \cdots + a_0}{b_q\mu^q + \cdots + b_0} = \frac{a_r}{b_q}\mu^{r-q}\Big[1 + O\Big(\frac{1}{|\mu|}\Big)\Big]$$

kann $\psi(\mu)/\chi(\mu)$ nicht für alle $\mu$ mit $|\mu| > 1$ positiven Realteil besitzen. Denn sei $a_r/b_q = \rho_0 e^{i\phi}$, man setze $\mu_\rho := \rho e^{i\psi}$ mit $\psi := (\pi - \phi)/(r - q)$ und $\rho > 1$ hinreichend groß. Wegen

$$\frac{\psi(\mu_\rho)}{\chi(\mu_\rho)} = -\rho_0\rho^{r-q}[1 + O(1/\rho)]$$

ist $\Re(\psi(\mu_\rho)/\chi(\mu_\rho)) \le 0$ für alle hinreichend großen $\rho > 1$. Explizite lineare Mehrschrittverfahren sind also nicht A-stabil. □

Ein berühmter Satz von G. Dahlquist (1963) (man spricht auch von der zweiten Dahlquist Schranke) sagt aus, dass es kein A-stabiles lineares Mehrschrittverfahren mit einer höheren Konsistenzordnung als 2 gibt und dass das oben als A-stabil erkannte Trapezverfahren unter allen Verfahren der Ordnung zwei in einem gewissen Sinne optimal ist. Der lokale Diskretisierungsfehler ist

$$\begin{aligned}\frac{z(t+h) - z(t)}{h} - \frac{1}{2}[z'(t) + z'(t+h)] &= z'(t) + \frac{h}{2}z''(t) + \frac{h^2}{6}z'''(t) \\ &\quad - \Big[z'(t) + \frac{h}{2}z''(t) + \frac{h^2}{4}z'''(t)\Big] + O(h^3) \\ &= -\frac{1}{12}h^2x'''(t) + O(h^3).\end{aligned}$$

"In einem gewissen Sinne optimal" bedeutet, dass die Konstante $\frac{1}{12}$ kleinst möglich ist. Einen verhältnismäßig elementaren Beweis dieses Satzes findet man bei R. D. Grigorieff (1977, S. 218) oder auch E. Hairer, G. Wanner (1991, S. 265). Dies ist natürlich insgesamt ein enttäuschendes Resultat.

### 3.4.4 BDF-Methoden

Die zweite Dahlquist-Schranke schränkt die Ordnung A-stabiler linearer Mehschrittverfahren auf 2 ein, was für eine effiziente Lösung steifer systeme nicht ausreichend ist. Daher wird häufig der Stabilitätsbegriff abgeschwächt.

**Definition 4.4** Ein lineares Mehrschrittverfahren mit dem Stabilitätsgebiet $S$ heißt A($\alpha$)-stabil mit einem $\alpha \in [0, \pi/2]$, wenn

$$\{z \in \mathbb{C} : |\arg(z) - \pi| \leq \alpha\} \subset S.$$

Hierbei ist $z = re^{i \arg(z)}$ mit $\arg(z) \in [0, 2\pi)$.

**Bemerkung:** Ein A($\pi/2$)-stabiles Verfahren ist natürlich A-stabil. Je größer bei einem A($\alpha$)-stabilen Verfahren der Winkel $\alpha$ ist, desto besser. □

Bei Einschrittverfahren ist die L-Stabilität eine wichtige Eigenschaft. Hierbei heißt ein (implizites) Einschrittverfahren L-stabil, wenn es A-stabil ist und zusätzlich

$$\lim_{z \to \infty} R(z) = 0$$

gilt, wobei $R$ die zugehörige Stabilitätsfunktion ist, siehe Aufgabe 6. Z. B. ist das implizite Euler-Verfahren

$$u(t+h) = u(t) + hf(t+h, u(t+h))$$

L-stabil, nicht aber die Trapezregel

$$u(t+h) = u(t) + \frac{h}{2}[f(t, u(t)) + f(t+h, u(t+h))].$$

In Aufgabe 2 kann man erkennen, dass bei einer sehr steifen Differentialgleichung das implizite Euler-Verfahren bessere Ergebnisse als die Trapezregel liefert, obwohl sie eine kleinere Konsistenzordnung besitzt. Für ein lineares $r$-Schrittverfahren, das durch die Polynome $\psi$ und $\chi$ bestimmt ist, entspricht diese Eigenschaft der, dass alle Wurzeln $\mu(z)$ des Stabilitätspolynoms $\rho(\cdot, z) := \psi(\cdot) - z\chi(\cdot)$ mit $|z| \to \infty$ gegen Null gehen. Dies impliziert (???, siehe K. Strehmel, R. Weiner (1995, S. 333)), dass $\chi(\mu) = b_r \mu^r$ mit $b_r \neq 0$. Die Normierung $b_r = 1$ führt dann zu einem Verfahren der Form

$$a_r u(t+rh) + \cdots + a_0 u(t) = hf(t+rh, u(t+rh)).$$

Für die Bestimmung der Koeffizienten $a_0, \ldots, a_r$ kann man ähnlich wie bei den Adams-Verfahren vorgehen. Wir nehmen an, dass Näherungen $u_j, \ldots, u_{j+r-1}$ für die Lösung $x$ an den Stellen $t_j, \ldots, t_{j+r-1}$ bekannt sind. Nun bestimmen wir das Polynom $P_r \in \Pi_r$ und den noch unbekannten Wert $u_{j+r}$ (im skalaren Fall sind dies $r+2$ Unbekannte) durch die Forderungen, dass

$$P_r(t_{j+i}) = u_{j+i} \quad (i = 0, \ldots, r), \qquad P_r'(t_{j+r}) = f(t_{j+r}, u_{j+r}).$$

Wir wollen dies für $r = 1$ illustrieren.

**Beispiel:** Sei $r = 1$. Mit dem Ansatz $P_1(s) = \alpha_1 s + \alpha_0$ haben wir die Gleichungen

$$\alpha_1 t_j + \alpha_0 = u_j, \qquad \alpha_1 t_{j+1} + \alpha_0 = u_{j+1}$$

sowie

$$\alpha_1 = f(t_{j+1}, u_{j+1}).$$

Subtrahiert man die erste Gleichung von der zweiten und berücksichtigt die dritte Gleichung, so erhält man das Verfahren

$$u_{j+1} - u_j = hf(t_{j+1}, u_{j+1}),$$

das implizite Euler-Verfahren. □

Natürlich kann man hier systematisch vorgehen. Man kann nämlich (siehe z. B. K. Strehmel, R. Weiner (1995, S. 334)) zeigen, dass bei vorgegebenem $r \in \mathbb{N}$ ein Verfahren der Form

$$(*) \qquad \sum_{l=1}^{r} \frac{1}{l} \nabla^l u_{j+r} = hf(t_{j+r}, u_{j+r})$$

entsteht, wobei die sogenannten *rückwärtsgenommenen Differenzen* (backward differentiation formulas, BDF) induktiv durch

$$\nabla^0 v_j := v_j, \qquad \nabla^i v_j := \nabla^{i-1} v_j - \nabla^{i-1} v_{j-1}$$

definiert sind.

**Beispiel:** Sei $r = 2$. Das resultierende Verfahren ist

$$\begin{aligned} hf(t_{j+2}, u_{j+2}) &= \nabla^1 u_{j+2} + \frac{1}{2}\nabla^2 u_{j+2} \\ &= (u_{j+2} - u_{j+1}) + \frac{1}{2}(\nabla^1 u_{j+2} - \nabla^1 u_{j+1}) \\ &= (u_{j+2} - u_{j+1}) + \frac{1}{2}[(u_{j+2} - u_{j+1}) - (u_{j+1} - u_j)] \\ &= \frac{1}{2}(3u_{j+2} - 4u_{j+1} + u_j). \end{aligned}$$

Da das Polynom $\psi(\mu) := \frac{1}{2}(3\mu^2 - 4\mu + 1)$ die Nullstellen 1 und 1/3 besitzt, ist die Stabilitätsbedingung erfüllt (oder das Mehrschrittverfahren *nullstabil*). Der lokale Diskretisierungsfehler ist

$$\frac{1}{2h}[3z(t+2h) - 4z(t+h) + z(t)] - z'(t+2h) = -\frac{1}{3}z'''(t)h^2 + O(h^3),$$

das Verfahren hat also die Konsistenzordnung 2. Das Stabilitätspolynom ist

$$\rho(\mu, z) = (\tfrac{3}{2} - z)\mu^2 - 2\mu + \tfrac{1}{2}.$$

Bei vorgegebenem $z \in \mathbb{C}$ sind

$$\mu_{1,2}(z) = \frac{2 \pm \sqrt{1+2z}}{3-2z}$$

die beiden Wurzeln des Stabilitätspolynoms. Man kann zeigen (Beweis?), dass das Verfahren A-stabil ist, d. h. der linke komplexe Halbraum im Stabilitätsgebiet enthalten ist. Weiter ist offenbar $\lim_{|z|\to\infty} |\mu_{1,2}(z)| = 0$. □

Für $r = 1, \ldots, 6$ erfüllt das Verfahren $(*)$ die Stabilitätbedingung, was für $r > 6$ nicht mehr der Fall ist. Für $r = 1, 2$ ist das Verfahren A-stabil, für $r \geq 3$ ist das Verfahren nur noch A($\alpha$)-stabil, wobei der Winkel $\alpha$ mit wachsenden $r$ kleiner, das Verfahren also "immer weniger stabil" wird. Die Konstenzordnung des Verfahrens ist $r$.

### 3.4.5 Implizite Runge-Kutta-Verfahren

Ein Einschrittverfahren nennt man ein $s$-stufiges implizites Runge-Kutta-Verfahren, wenn es die Form

$$u(t+h) = u(t) + h\Phi(h, f)(t, u(t))$$

hat und die Verfahrensfunktion durch

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

gegeben ist, wobei

$$\begin{aligned} k_1(t,u) &= f(t + c_1 h, u + h \textstyle\sum_{j=1}^{s} a_{1j} k_j(t,u)), \\ &\vdots \\ k_s(t,u) &= f(t + c_s h, u + h \textstyle\sum_{j=1}^{n} a_{sj} k_j(t,u)). \end{aligned}$$

Dies stellt ein nichtlineares Gleichungssystem von $s$ Gleichungen zur Bestimmung der $s$ Unbekannten $k_1(t,u), \ldots, k_s(t,u)$ dar, welches bei vorausgesetzter Lipschitzstetigkeit von $f$ (bezüglich des zweiten Argumentes) für hinreichend kleines[16] $h$ wegen des Fixpunktsatzes für kontrahierende Abbildungen eine eindeutige Lösung besitzt. Festgelegt ist ein solches Verfahren durch die Angabe des Vektors $c = (c_i) \in \mathbb{R}^s$, die Matrix $A = (a_{ij}) \in \mathbb{R}^{s\times s}$ und den Vektor $b = b_i \in \mathbb{R}^s$. Diese können wir uns in der Form

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} \qquad \text{bzw.} \qquad \begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}$$

[16] Wegen dieser Beschränkung der Schrittweite sollte aber gerade bei steifen Differentialgleichungen nicht die einfache Fixpunktiteration sondern ein intelligenteres Verfahren benutzt werden. Siehe z. B. K. Strehmel, R. Weiner (1995, S. 265 ff.).

aufgeschrieben denken. Wir wollen nun die zugehörige Stabilitätsfunktion berechnen, indem wir das implizite Runge-Kutta-Verfahren auf die Testgleichung $x' = \lambda x$ anwenden. Das Gleichungssystem zur Berechnung von $k(t,u)$ ist

$$k(t,u) = \lambda(ue + hAk(t,u)),$$

wobei $e$ der Vektor (des $\mathbb{R}^s$) ist, dessen Komponenten alle gleich 1 sind. Mit $z := h\lambda$ führt dies auf

$$k(t,u) = \lambda(I - zA)^{-1}eu.$$

Wegen

$$u(t+h) = u + hb^T k(t,u) = [1 + zb^T(I - zA)^{-1}e]u$$

ist die zugehörige Stabilitätsfunktion also

$$R(z) := 1 + zb^T(I - zA)^{-1}e.$$

Mit $S := \{z \in \mathbb{C} : |R(z)| \leq 1\}$ wird das zum Verfahren gehörende *Stabilitätsgebiet* bezeichnet. Ferner heißt dieses *A-stabil*, wenn die linke Halbebene im Stabilitätsgebiet enthalten ist.

Wir wollen zunächst zwei Beispiele angeben.

**Beispiel:** Wir beschränken uns auf eine skalare Differentialgleichung $x' = f(t,x)$. Ein einstufiges implizites Runge-Kutta-Verfahren hat eine Verfahrensfunktion

$$\Phi(h,f)(t,u) = bk(t,u),$$

wobei $k(t,u)$ implizit durch

$$k(t,u) = f(t + ch, u + hak(t,u))$$

mit gewissen Konstanten $a, b, c$ festgelegt ist. Wir wollen das (wie wir sehen werden eindeutige) einstufige implizite Runge-Kutta-Verfahren zweiter Ordnung bestimmen. Sei $z$ die Lösung von $z' = f(s,z)$, $z(t) = u$. Der lokale Diskretisierungsfehler ist

$$\begin{aligned}
\Delta(h,f)(t,u) &= \frac{z(t+h) - z(t)}{h} - \Phi(h,f)(t,u) \\
&= f(t,u) + [f_t(t,u) + f_x(t,u)f(t,u)]\frac{h}{2} + O(h^2) - bk(t,u) \\
&= f(t,u) + [f_t(t,u) + f_x(t,u)f(t,u)]\frac{h}{2} + O(h^2) \\
&\qquad - b[f(t,u) + f_t(t,u)ch + f_x(t,u)hak(t,u)] \\
&= f(t,u) + [f_t(t,u) + f_x(t,u)f(t,u)]\frac{h}{2} + O(h^2) \\
&\qquad - b[f(t,u) + f_t(t,u)ch + f_x(t,u)f(t,u)ha] \\
&\qquad (\text{da } k(t,u) = f(t,u) + O(h)) \\
&= (1-b)f(t,u) + \frac{h}{2}[(1-2bc)f_t(t,u) + (1-2ba)f_x(t,u)f(t,u)] \\
&\qquad + O(h^2).
\end{aligned}$$

Um ein Verfahren der Ordnung 2 zu erhalten ist also notwendig $b = 1$, $a = c = \frac{1}{2}$. In der obigen Schreibweise ist das Verfahren durch

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$$

gegeben. Bei diesem Verfahren ist also $k(t,u)$ zu bestimmen aus

$$k(t,u) = f(t + \tfrac{1}{2}h, u + \tfrac{1}{2}hk(t,u))$$

anschließend ist $u(t+h) = u + hk(t,u)$ die neue Näherung zur Zeit $t+h$. Naheliegenderweise nennt man das Verfahren die implizite Mittelpunktsregel. Wir wollen noch das zugehörige Stabilitätsgebiet berechnen. Die Stabilitätsfunktion ist

$$R(z) = 1 + \frac{z}{1 - \frac{1}{2}z}.$$

Daher ist das Stabilitätsgebiet gegeben durch

$$S = \Big\{ z \in \mathbb{C} : \Big| \frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z} \Big| \leq 1 \Big\}.$$

Man rechnet leicht nach, dass das Stabilitätsgebiet genau die linke (abgeschlossene) Halbebene in $\mathbb{C}$ ist, so dass das Verfahren A-stabil ist. □

**Beispiel:** Man kann zeigen, dass es genau ein zweistufiges implizites Runge-Kutta-Verfahren der Konsistenzordnung 4 gibt. Das Gleichungssystem zur Bestimmung von $k_1, k_2$ (wir lassen hier und im folgenden das Argument $(t,u)$ fort) lautet

$$\begin{aligned} k_1 &= f(t + c_1 h, u + h a_{11} k_1 + h a_{12} k_2), \\ k_2 &= f(t + c_2 h, u + h a_{21} k_1 + h a_{22} k_2), \end{aligned}$$

die Verfahrensfunktion ist

$$\Phi(h, f) = b_1 k_1 + b_2 k_2.$$

Bei H. Werner, H. Arndt (1986, S. 140 ff.) wird genauer durchgeführt, dass die 8 Unbekannten $b_1, b_2$ sowie $c_1, c_2$ und $a_{11}, a_{12}, a_{21}, a_{22}$ 10 Gleichungen zu genügen haben, damit das resultierende Verfahren die Konsistenzordnung 4 hat. Dieses überbestimmte nichtlineare Gleichungssystem ist erstaunlicherweise (im wesentlichen: die beiden Lösungen führen zum selben Verfahren) eindeutig lösbar, die Lösung kann man mit Maple erhalten. Hierauf wollen wir nicht näher eingehen, sondern im Anschluss nur noch die Lösung angeben. Interessanter wäre die Frage, wie man auch das Aufstellen des Gleichungssystems einem mathematischen Anwendersystem überlassen könnte. Die Lösung ist in unserer Notation:

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|cc} \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Die zugehörige Stabilitätsfunktion (bei der Berechnung kann wieder Maple helfen) ist

$$R(z) = 1 + zb^T(I - zA)^{-1}e = \frac{12 + 6z + z^2}{12 - 6z + z^2}.$$

Wir wollen nachweisen, dass das angegebene implizite Runge-Kutta-Verfahren A-stabil ist. Mit Hilfe der Maple-Befehle

```
z:=x+I*y;
evalc(abs((12+6*z+z^2)/(12-6*z+z^2))^2);
factor(%);
```

erhalten wir, dass

$$|R(z)|^2 = \frac{144 + 144x + 60x^2 + 12x^3 + x^4 + y^2(12 + 12x + 2x^2 + y^2)}{144 - 144x + 60x^2 - 12x^3 + x^4 + y^2(12 - 12x + 2x^2 + y^2)},$$

wobei $x := \Re(z)$ und $y := \Im(z)$. Daher ist $|R(z)| \leq 1$ genau dann, wenn

$$x(12 + x^2 + y^2) \leq 0$$

bzw. $\Re(z) \leq 0$. Das Stabilitätsgebiet ist also genau die linke komplexe Halbebene, das Verfahren ist A-stabil. □

In Verallgemeinerung der letzten beiden Beispiele kann man zeigen, dass es genau ein $s$-stufiges implizites Runge-Kutta-Verfahren der Ordnung $2s$ gibt. Dieses ist A-stabil. Man spricht von einem *$s$-stufigen Gauß-Verfahren*, siehe E. Hairer, G. Wanner (1991, S. 75). Im obigen Beispiel ist das Stabilitätsgebiet des 2-stufigen Gauß-Verfahrens genau die linke komplexe Halbebene. Dagegen nennt man ein $s$-stufiges implizites Runge-Kutta-Verfahren der Konsistenzordnung $2s - 1$ ein *Radau-Verfahren*. Ein Beispiel ist in Aufgabe 5 angegeben. Es kann bei der Bearbeitung dieser Aufgabe gezeigt werden, dass das zugehörige Stabilitätsgebiet wesentlich größer als die linke komplexe Halbebene ist. Weiter heißt ein $s$-stufiges implizites Runge-Kutta-Verfahren der Konsistenzordnung $2s - 2$ ein *Lobatto-Verfahren*.

Einige wenige Bemerkungen zu impliziten Runge-Kutta-Verfahren sollen diesen Unterabschnitt abschließen. Für wesentlich ausführlichere Darstellungen konsultiere man insbesondere E. Hairer, G. Wanner (1991) und K. Strehmel, R. Weiner (1995).

Zunächst geben wir eine andere Darstellung der Stabilitätsfunktion eines impliziten Runge-Kutta-Verfahrens an.

**Lemma 4.5** *Gegeben sei ein $s$-stufiges Runge-Kutta-Verfahren zu den Daten*

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$$

*Die zugehörige Stabilitätsfunktion*

$$R(z) = 1 + zb^T(I - zA)^{-1}e,$$

*wobei $e \in \mathbb{R}^s$ nur Einsen als Komponenten hat, lässt sich dann auch in der Form*

$$R(z) = \frac{\det(I - zA + zeb^T)}{\det(I - zA)}$$

*schreiben.*

**Beweis:** Es ist (siehe obige Herleitung der Stabilitätsfunktion)

$$R(z)u = u + hb^T k = u + zb^T(k/\lambda),$$

wobei

$$k/\lambda = (I - zA)^{-1}eu.$$

Daher ist

$$\begin{pmatrix} I - zA & 0 \\ -zb^T & 1 \end{pmatrix} = \begin{pmatrix} k/\lambda \\ R(z)u \end{pmatrix} = \begin{pmatrix} e \\ 1 \end{pmatrix} u$$

bzw.

$$\begin{pmatrix} I - zA & 0 \\ -zb^T & 1 \end{pmatrix} \begin{pmatrix} k/(\lambda u) \\ R(z) \end{pmatrix} = \begin{pmatrix} e \\ 1 \end{pmatrix}.$$

Nun erinnern wir an die *Cramersche Regel*:

- *Sei $A = (\ a_1 \ \cdots \ a_n\ ) \in \mathbb{R}^{n\times n}$ nichtsingulär. Dann ist die $k$-te Komponente $x_k$ der Lösung von $Ax = b$ gegeben durch*

$$x_k = \frac{\det(\ a_1 \ \cdots \ a_{k-1} \ \ b \ \ a_{k+1} \ \cdots \ a_n\ )}{\det(A)}.$$

Wenden wir dieses Ergebnis auf unseren Fall an, so erhalten wir, dass

$$R(z) = \frac{\det \begin{pmatrix} I - zA & e \\ -zb^T & 1 \end{pmatrix}}{\det(I - zA)}.$$

Die Behauptung folgt, wenn man in der Zählerdeterminante die letzte Zeile von den anderen Zeilen subtrahiert. □

Wegen des eben bewiesenen Lemmas ist die Stabilitätsfunktion eines impliziten $s$-stufigen Runge-Kutta-Verfahrens eine rationale Funktion, wobei Zähler und Nenner höchstens den Grad $s$ haben.

Die Stabilitätsfunktion vieler impliziter Runge-Kutta-Verfahren steht in einem engen Zusammenhang mit der sogenannten *Padé-Approximation* der Exponentialfunktion. Dies wollen wir zum Schluss dieses Unterabschnitts andeuten.

**Definition 4.6** Sei $g$ eine in der Umgebung von $z = 0$ analytische Funktion, ferner seien $j, k$ nichtnegative ganze Zahken. Dann heißt die rationale Funktion

$$R_{jk}(z) = \frac{P_{jk}(z)}{Q_{jk}(z)} = \frac{\sum_{l=0}^{j} a_l z^l}{\sum_{l=0}^{k} b_l z^l}$$

mit $b_0 = 1$ (also Zählergrad $\leq j$, Nennergrad[17] $\leq k$) *Padé-Approximation* an $g$ vom Index $(j,k)$, wenn

$$R_{jk}^{(l)}(0) = g^{(l)}(0), \qquad l = 0, \ldots, j+k.$$

**Bemerkung:** Eine rationale Funktion der angegebenen Art ist offenbar genau dann eine Padé-Approximation an $g$ vom Index $(j,k)$, wenn

$$R_{jk}(z) = g(z) + O(z^{j+k+1}) \quad \text{für } z \to 0.$$

Man sagt, diese Padé-Approximation besitze die Approximationsordnung $r = j+k$ an $g$. □

Es ist leicht zu zeigen, dass die Padé-Approximation eindeutig ist, wenn sie existiert.

**Lemma 4.7** *Eine Padé-Approximation vom Index $(j,k)$ an eine in der Umgebung von $z = 0$ analytische Funktion $g$ ist, wenn sie existiert, eindeutig bestimmt.*

**Beweis:** Seien $R_{jk} = P_{jk}/Q_{jk}$ und $R_{jk}^* = P_{jk}^*/Q_{jk}^*$ zwei Padé-approximationen zum selben Index. Dann gilt

$$w(z) := P_{jk}(z)Q_{jk}^*(z) - Q_{jk}(z)P_{jk}^*(z) = O(z^{j+k+1}) \quad \text{für } z \to 0.$$

Da $w$ ein Polynom vom Grad $\leq j+k$ ist, muss $w = 0$ sein, was die Eindeutigkeit einer Padé-Approximation impliziert. □

**Beispiel:** Es soll die $(2,1)$-Padé-Approximation an $g(z) := e^z$ bestimmt werden. Wir machen den Ansatz

$$R_{21}(z) = \frac{a_0 + a_1 z + a_2 z^2}{1 + b_1 z}.$$

Zur Bestimmung der vier unbekannten Parameter hat man die 4 Gleichungen

$$R(0) = 1, \quad R'(0) = 1, \quad R''(0) = 1, \quad R'''(0) = 1.$$

Die Bedingung $R(0) = 1$ liefert zunächst $a_0 = 1$. Die restlichen drei Gleichungen lauten dann

$$\begin{array}{rcl} R'(0) &=& a_1 - b_1 \qquad\qquad = 1, \\ R''(0) &=& 2(a_2 - a_1 b_1 + b_1^2) \quad = 1, \\ R'''(0) &=& 6b_1(a_1 b_1 - a_2 - b_1^2) \quad = 1. \end{array}$$

Hierzu haben wir natürlich auch Maple benutzt. Die folgende Sequenz gibt die gesuchten Ableitungen:

```
R:=z->(1+a_1*z+a_2*z^2)/(1+b_1*z);
R_1:=z->D(R)(z);R_1(0);
R_2:=z->D(R_1)(z);R_2(0);
R_3:=z->D(R_2)(z);R_3(0);
```

[17] Man muss hier beim Lesen in der Literatur aufpassen: Bei uns (genau wie bei Hairer-Wanner) gibt der erste Index $j$ den Zählergrad, der zweite Index $k$ den Nennergrad an. Bei Strehmel-Weiner ist es genau umgekehrt.

Hieraus erhält man (z. B. mit Maple) sofort $a_1 = \frac{2}{3}$, $b_1 = -\frac{1}{3}$ und $a_2 = \frac{1}{6}$. Es ist also

$$R_{21}(z) = \frac{1 + \frac{2}{3}z + \frac{1}{6}z^2}{1 - \frac{1}{3}z}$$

die gesuchte Padé-Approximation. Dies erhalten wir auch sofort nach

```
with(numapprox):
pade(exp(z),z,[2,1]);
```

Während dies noch relativ einfach zu erhalten ist, wäre die $(4,6)$-Padé-Approximation von $g(z) := \cos z \exp(z^2)$ schon schwerer zu berechnen. Dagegen liefert

```
pade(cos(z)*exp(z^2),z,[4,6]);
```

sofort das Ergebnis

$$R_{46}(z) = \frac{1 - \frac{18649}{39381}z^2 - \frac{9908657}{66160080}z^4}{1 - \frac{76679}{78762}z^2 + \frac{19539853}{66160080}z^4 - \frac{1695041}{26464032}z^6}.$$

Das möchte man wahrscheinlich nicht zu Fuß ausrechnen! □

Im folgenden Satz werden wir eine geschlossene Darstellung der Padé-Approximation an die Exponentialfunktion angeben. Auf einen Beweis verzichten wir (siehe z. B. E. Hairer, G. Wanner (1991, S. 50)).

**Satz 4.8** *Die $(j,k)$-Padé-Approximation an $e^z$ existiert und ist gegeben durch*

$$R_{jk}(z) = \frac{P_{jk}(z)}{Q_{jk}(z)},$$

*wobei*

$$\begin{aligned} P_{jk}(z) &= 1 + \frac{j}{k+j}z + \frac{j(j-1)}{(k+j)(k+j-1)}\frac{z^2}{2!} + \cdots + \frac{j(j-1)\cdots 1}{(k+j)\cdots(k+1)}\frac{z^j}{j!}, \\ Q_{jk}(z) &= 1 - \frac{k}{j+k}z + \frac{k(k-1)}{(j+k)(j+k-1)}\frac{z^2}{2!} - \cdots + (-1)^j\frac{k(k-1)\cdots 1}{(j+k)\cdots(j+1)}\frac{z^k}{k!}. \end{aligned}$$

**Beispiel:** In der folgenden Tabelle geben wir einige Padé-Approximationen an $e^z$ an.

| | | |
|---|---|---|
| $\frac{1}{1}$ | $\frac{1+z}{1}$ | $\frac{1+z+\frac{1}{2}z^2}{1}$ |
| $\frac{1}{1-z}$ | $\frac{1+\frac{1}{2}z}{1-\frac{1}{2}z}$ | $\frac{1+\frac{2}{3}z+\frac{1}{6}z^2}{1-\frac{1}{3}z}$ |
| $\frac{1}{1-z+\frac{1}{2}z}$ | $\frac{1+\frac{1}{3}z}{1-\frac{2}{3}z+\frac{1}{6}z^2}$ | $\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$ |

Man erkennt einige Einträge als Stabilitätsfunktionen schon behandelter impliziter Runge-Kutta-Verfahren wieder. □

Zum Schluss geben wir, wieder ohne Beweis, den folgenden Satz an.

**Satz 4.9** *Die Stabilitätsfunktion des s-stufigen Gauß-Verfahrens ist die Padé-Approximation vom Index* $(s, s)$*, des s-stufigen Radau-Verfahrens die Padé-Approximation vom Index* $(s-1, s)$*, des s-stufigen Lobatto-Verfahrens die Padé-Approximation vom Index* $(s-1, s-1)$*.*

### 3.4.6 MATLAB-Funktionen

In MATLAB werden einige Funktionen zur Lösung von Anfangswertaufgaben bei Differentialgleichungen bereitgestellt. Über die entsprechenden Funktionen wird gesagt:

- ode45 is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. It is a one-step solver - in computing y(tn), it needs only the solution at the immediately preceding time point, y(tn-1). In general, ode45 is the best function to apply as a first try for most problems.

- ode23 is an implementation of an explicit Runge-Kutta (2,3) pair of Bogacki and Shampine. It may be more efficient than ode45 at crude tolerances and in the presence of moderate stiffness. Like ode45, ode23 is a one-step solver.

- ode113 is a variable order Adams-Bashforth-Moulton PECE solver. It may be more efficient than ode45 at stringent tolerances and when the ODE file function is particularly expensive to evaluate. ode113 is a multistep solver - it normally needs the solutions at several preceding time points to compute the current solution.

The above algorithms are intended to solve nonstiff systems. If they appear to be unduly slow, try using one of the stiff solvers below.

- ode15s is a variable order solver based on the numerical differentiation formulas (NDFs). Optionally, it uses the backward differentiation formulas (BDFs, also known as Gear's method) that are usually less efficient. Like ode113, ode15s is a multistep solver. Try ode15s when ode45 fails, or is very inefficient, and you suspect that the problem is stiff, or when solving a differential-algebraic problem.

- ode23s is based on a modified Rosenbrock formula of order 2. Because it is a one-step solver, it may be more efficient than ode15s at crude tolerances. It can solve some kinds of stiff problems for which ode15s is not effective.

- ode23t is an implementation of the trapezoidal rule using a freeïnterpolant. Use this solver if the problem is only moderately stiff and you need a solution without numerical damping. ode23t can solve DAEs.

- ode23tb is an implementation of TR-BDF2, an implicit Runge-Kutta formula with a first stage that is a trapezoidal rule step and a second stage that is a backward differentiation formula of order two. By construction, the same iteration matrix is used in evaluating both stages. Like ode23s, this solver may be more efficient than ode15s at crude tolerances.

Einige nähere Informationen findet man bei L. F. Shampine, M. W. Reichelt (1997)[18]. Im Prinzip arbeiten die Löser steifer Differentialgleichungen wie die entsprechenden für nichtsteife Systeme, wobei nur einige weitere Punkte zu beachten sind. Wir wollen dies durch Beispiele andeuten.

**Beispiel:** Die van der Polsche Differentialgleichung

$$x'' - \mu(1 - x^2)x' + x = 0$$

mit großem $\mu > 0$, etwa $\mu = 1000$, ist eines der bekanntesten Beispiele einer steifen Differentialgleichung. Sie spielt für elektrische Schwingkreise eine wichtige Rolle. Wir geben die Anfangsbedingungen $x(0) = 2$, $x'(0) = 0$ vor, haben also das System

$$\begin{array}{rclcrcl} x_1' &=& x_2 & & x_1(0) &=& 2, \\ x_2' &=& 1000(1 - x_1^2)x_2 - x_1 & & x_2(0) &=& 0 \end{array}$$

zu lösen. In ein File `vdp1000.m` (das von MATLAB aus erreicht werden kann) schreiben wir z. B. den folgenden Inhalt:

```
function dx=vdp1000(t,x);
dx=[x(2);1000*(1-x(1)^2)*x(2)-x(1)];
```

Will man die gegebene van der Polsche Differentialgleichung auf dem Zeitintervall $[0, 3000]$ numerisch lösen und die Default-Parameter für relative und absolute Toleranzen ($10^{-3}$ bzw. $10^{-6}$) benutzen, so macht man z. B. den Aufruf

```
[t,x]=ode15s('vdp1000',[0 3000],[2; 0]);
```

Durch `plot(t,x(:,1));` bzw. `plot (x(:,1),x(:,2))` kann man sich die Lösung bzw. die Phasenbahn plotten, siehe Abbildung 3.6 links bzw. rechts. Es wurden 592 Zeit-

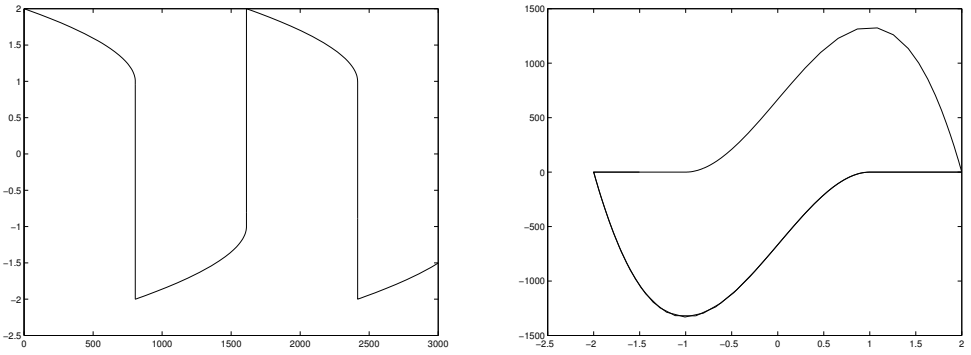


Abbildung 3.6: Lösung bzw. Phasenbahn erhalten mit `ode15s`

schritte gemacht, während `ode45` in angemessener Zeit nicht fertig wird. □

Verfahren für steife Differentialgleichungen sind implizit, in jedem Zeitschritt muss ein nichtlineares Gleichungssystem gelöst werden. Dies geschieht mit Hilfe des Newton-Verfahrens. Der Benutzer hat die Möglichkeit, die Jacobi-Matrix $f_x(t, x)$ bereitzustellen

[18] L. F. Shampine, M. W. Reichelt (1995) "The MATLAB ODE Suite". SIAM J. Sci. Comput. 18, 1–22.

(andernfalls werden die benötigten partiellen Ableitungen durch entsprechende Differenzenquotienten ersetzt, siehe die Funktion `numjac`). Dieser Wunsch kann mit Hilfe von

```
options=odeset('Jacobian','on')
```

der entsprechenden MATLAB-Funktion mitgeteilt werden, ferner muss die Jacobi-Matrix analytisch bereitgestellt werden. Wir geben am besten wieder ein Beispiel an.

**Beispiel:** Wir betrachten noch einmal das letzte Beispiel, wollen diesmal nur die Jacobi-Matrix dem Löser mitteilen. Hierzu schreiben wir in ein File `vdpt.m` z. B. den folgenden Inhalt:

```
function varargout=vdpt(t,x,flag);
switch flag
case ''
   varargout{1}=f(t,x);
case 'jacobian'
   varargout{1}=jacobian(t,x);
end;

function dx=f(t,x);
dx=[x(2);1000*(1-x(1)^2)*x(2)-x(1)];

function dfdx=jacobian(t,x);
dfdx=[0 1;-2000*x(1)*x(2)-1 1000*(1-x(1)^2)];
```

Anschließend wird `options=odeset('Jacobian','on')` gesetzt und der Aufruf

```
[t,x]=ode15s('vdpt',[0 3000],[2;0],options);
```

gemacht. Nun interessiert natürlich (die Zahl der Zeitschritte bleibt dieselbe), ob das Verfahren dadurch schneller wird. Hierzu schreiben wir ein Script-File `Test.m` mit dem Inhalt

```
tic;
for k=1:100
   [t,x]=ode15s('vdp1000',[0, 3000],[2; 0]);
end;
toc
```

durch welches hundert mal der obige Aufruf gemacht und eine Stopuhr aktiviert und angehalten wird. Hier erhalten wir die Auskunft:

```
elapsed_time = 85.5993
```

Bei der entsprechenden Version mit Bereitstellung der Jacobi-Matrix erhält man die Auskunft

```
elapsed_time = 72.6135
```

Die Zeitersparnis kann wesentlich größer sein. □

**Beispiel:** Der zeitliche Verlauf chemischer Reaktionen lässt sich gut mittels gewöhnlicher Differentialgleichungen modellieren. Da hier einige Reaktionen schnell, andere dagegen sehr langsam erfolgen, führt dies häufig auf steife Differentialgleichungssysteme, wobei die rechte Seite i. Allg. polynomial ist. Dies kann z. B. auf die folgende Aufgabe (siehe E. Hairer, G. Wanner (1991, S. 3)) führen:

$$\begin{array}{rclrcl} x_1' &=& -0.04x_1 + 10^4 x_2 x_3 & x_1(0) &=& 1 \\ x_2' &=& 0.04x_1 - 10^4 x_2 x_3 - 3\cdot 10^7 x_2^2 & x_2(0) &=& 0 \\ x_3' &=& 3\cdot 10^7 x_2^2 & x_3(0) &=& 0. \end{array}$$

Hierzu schreiben wir ein Function-File `chemrea.m` mit dem Inhalt:

```
function varargout=chemrea(t,x,flag);
switch flag
case ''
    varargout{1}=f(t,x);
case 'jacobian'
    varargout{1}=jacobian(t,x);
end;

function dx=f(t,x);
dx=[-0.04*x(1)+1e4*x(2)*x(3);0.04*x(1)-1e4*x(2)*x(3)-3e7*x(2)^2;3e7*x(2)^2];

function dfdx=jacobian(t,x);
dfdx=[-0.04,1e4*x(3),1e4*x(2);0.04,-(1e4*x(3)+6e7*x(2)),-1e4*x(2);0,6e7*x(2),0];
```

der Aufruf erfolgt (wenn wir das Zeitintervall $[0, 5]$ zu grunde legen) durch

```
[t,x]=ode15s('chemrea',[0,5],[1;0;0],options);
```

Die drei Komponenten geben wir in Abbildung 3.7 an. Diesen Plot haben wir mit MATLABs `subplot` Befehl hergestellt. Genauer durch

```
subplot(2,2,1);
plot(t,x(:,1)), title('x_1')
subplot(2,2,2);
plot(t,x(:,2)), title('x_2')
subplot(2,2,3);
plot(t,x(:,3)), title('x_3')
```

Es wurden 36 Zeitschritte durchgeführt, während es bei der Anwendung von `ode45` sogar 13 889 Zeitschritte sind! Für die zweite Komponente beobachtet man ferner sehr starke Schwankungen. Das sieht man in Abbildung 3.7 rechts unten. □

In den letzten beiden Beispielen wurden den Lösern analytische Ausdrücke für die Jacobi-Matrix mitgeteilt. Oft ist es allerdings mühsam, die Jacobi-Matrix zu berechnen (selbst wenn es im Prinzip möglich ist), außerdem macht man dabei leicht Fehler. Dagegen ist häufig eine gewisse "Dünnbesetzheitsstruktur" (sparsity pattern) in der Jacobi-Matrix offensichtlich. Diese sollte dem Löser mitgeteilt werden, damit das Lösen der linearen Gleichungssysteme im Newton-Verfahren vereinfacht wird.

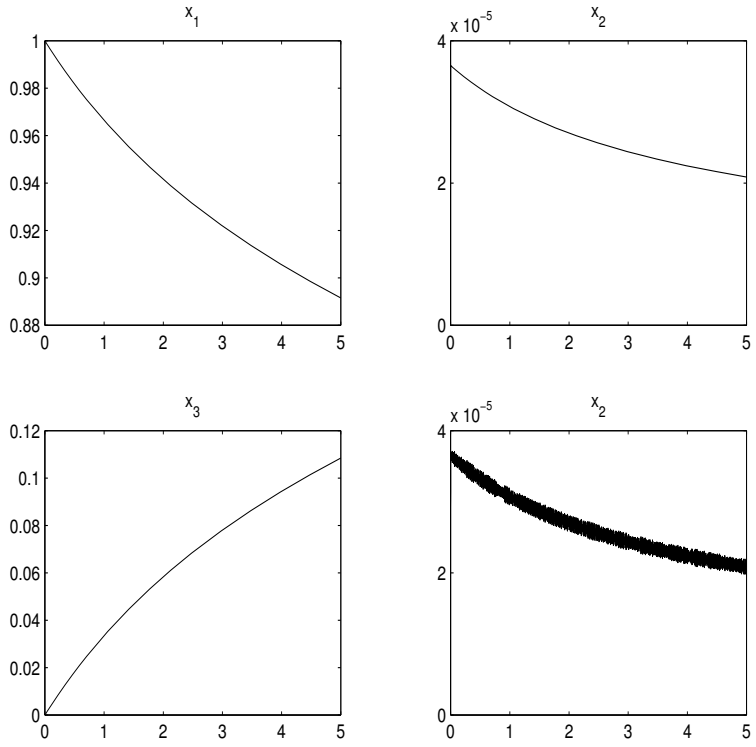


Abbildung 3.7: Die drei Lösungs-Komponenten

**Beispiel:** Wir reproduzieren ein Beispiel bei E. HAIRER, G. WANNER (1991, S. 5 ff.). Ursprünglich ist eine Anfangs-Randwertaufgabe für eine partielle Differentialgleichung mit einer Ortsvariablen $x \in [0,1]$ und einer Zeitvariablen $t > 0$ gegeben. Genauer bestimme man eine Lösung $(u, v)$ des partiellen Differentialgleichungssystems

$$\begin{aligned} \frac{\partial u}{\partial t} &= 1 + u^2 v - 4u + \alpha \frac{\partial^2 u}{\partial t^2} \\ \frac{\partial v}{\partial t} &= 3u - u^2 v + \alpha \frac{\partial^2}{\partial x^2} \end{aligned}$$

(mit $\alpha := \frac{1}{50}$) für $(x,t) \in [0,1] \times \mathbb{R}_{>0}$, welche den Randbedingungen

$$u(0,t) = u(1,t) = 1, \qquad v(0,t) = v(1,t) = 3$$

für $t \geq 0$ sowie den Anfangsbedingungen

$$u(x,0) = 1 + \sin(2\pi x), \qquad v(x,0) = 3$$

für $x \in [0,1]$ genügt. Man diskretisiere bezüglich der Ortsvariablen, mit $N \in \mathbb{N}$ setze man also $\Delta x := 1/(N+1)$ und $x_i := i\Delta x$, $i = 0, \ldots, N$. Mit $u_i(t)$ bezeichne man eine Näherung für $u(x_i, t)$. Ersetzt man die zweiten partiellen Ableitungen nach der Ortsvariablen $x$ durch einen zentralen Differenzenquotienten, so erhalten wir ein System von $2N$ gewöhnlichen Anfangswertaufgaben für $u_1, v_1, \ldots, u_N, v_N$. Unter Berücksichtigung von

$$u_0(t) = u_{N+1}(t) = 1, \qquad v_0(t) = v_{N+1}(t) = 3$$

lautet dieses

$$\begin{aligned}
u_1' &= 1 + u_1^2 v_1 - 4u_1 + \alpha(N+1)^2(1 - 2u_1 + u_2) & u_1(0) &= 1 + \sin(2\pi x_1),\\
v_1' &= 3u_1 - u_1^2 v_1 + \alpha(N+1)^2(3 - 2v_1 + v_2) & v_1(0) &= 3,\\
&\vdots & &\vdots\\
u_i' &= 1 + u_i^2 v_i - 4u_i + \alpha(N+1)^2(u_{i-1} - 2u_i + u_{i+1}) & u_i(0) &= 1 + \sin(2\pi x_i),\\
v_i' &= 3u_i - u_i^2 v_i + \alpha(N+1)^2(v_{i-1} - 2v_i + v_{i+1}) & v_i(0) &= 3,\\
&\vdots & &\vdots\\
u_N' &= 1 + u_N^2 v_N - 4u_N + \alpha(N+1)^2(u_{N-1} - 2u_N + 1) & u_N(0) &= 1 + \sin(2\pi x_N),\\
v_N' &= 3u_N - u_N^2 v_N + \alpha(N+1)^2(v_{N-1} - 2v_N + 3) & v_N(0) &= 3.
\end{aligned}$$

Ordnet man die Gleichungen in dieser Weise an, so hat die Jacobi-Matrix der rechten Seite Bandstruktur. Genauer sind in jeder Zeile der Jacobi-Matrix höchstens vier von Null verschiedene Einträge und zwar sind höchstens die Hauptdiagonale und zwei untere und zwei obere Nebendiagonalen besetzt. Wir geben eine etwas vereinfachte Version von `brussode.m` (ein File, das man nach dem Aufruf von MATLAB mit Hilfe von `edit brussode` ansehen und ausdrucken an) wieder. Bei der dünnbesetzten Jacobi-Matrix und der Festlegung der Dünnbesetztheitsstruktur benutzen wir die MATLAB-Funktion `spdiags`, über die man u. a. nachlesen kann:

```
A=SPDIAGS(B,d,m,n) creates an m-by-n sparse matrix from the
columns of B and places them along the diagonals specified by d.
```

Und nun die entsprechende Funktion.

```
function varargout=bruss(t,x,flag,N);
switch flag
case ''
   varargout{1}=f(t,x,N);
case 'init'
   [varargout{1:2}]=init(N);
case 'jpattern'
   varargout{1}=jpattern(t,x,N);
case 'jacobian'
   varargout{1}=jacobian(t,x,N);
end;
%-----------------------------------------------------------
function dx=f(t,x,N);
c=0.02*(N+1)^2;          %=alpha*(N+1)^2
dx=zeros(2*N,1);         %Platz fuer dx
i=1;                     %Zwei Komponenten am linken Rand
dx(i)=1+x(i+1).*x(i).^2-4*x(i)+c*(1-2*x(i)+x(i+2));
dx(i+1)=3*x(i)-x(i+1).*x(i).^2+c*(3-2*x(i+1)+x(i+3));
i=3:2:2*N-3;             %Zwei innere Komponenten
dx(i)=1+x(i+1).*x(i).^2-4*x(i)+c*(x(i-2)-2*x(i)+x(i+2));
dx(i+1)=3*x(i)-x(i+1).*x(i).^2+c*(x(i-1)-2*x(i+1)+x(i+3));
i=2*N-1;                 %Zwei Komponenten am rechten Rand
```

```
dx(i)=1+x(i+1).*x(i).^2-4*x(i)+c*(x(i-2)-2*x(i)+1);
dx(i+1)=3*x(i)-x(i+1).*x(i).^2+c*(x(i-1)-2*x(i+1)+3);
%-------------------------------------------------------
function [tspan,x_0]=init(N);
tspan=[0;10];
x_0=[1+sin((2*pi/(N+1))*(1:N));3+zeros(1,N)];
x_0=x_0(:);
%-------------------------------------------------------
function S=jpattern(t,x,N);
B=ones(2*N,5);
B(2:2:2*N,2)=zeros(N,1);
B(1:2:2*N-1,4)=zeros(N,1);
S=spdiags(B,-2:2,2*N,2*N);
%-------------------------------------------------------
function dfdx=jacobian(t,x,N);
c=0.02*(N+1)^2;          %=alpha*(N+1)^2
B=zeros(2*N,5);          %Enthaelt Haupdiagonale
                         %und vier Nebendiagonalen
B(1:2*(N-1),1)=c;B(3:2*N,5)=c;
i=1:2:2*N-1;
B(i,2)=3-2*x(i).*x(i+1);
B(i,3)=2*x(i).*x(i+1)-4-2*c;
B(i+1,3)=-x(i).^2-2*c;
B(i+1,4)=x(i).^2;
dfdx=spdiags(B,-2:2,2*N,2*N);
%-------------------------------------------------------
```

Die Option, dass die Jacobi-Matrix in analytischer Weise übergeben wird, wird durch `options=odeset('jacobian','on');` aktiviert. Ein Aufruf erfolgt z. B. durch

```
[t,x]=ode15s('bruss',[],[],options,100);
```

wenn mit $N = 100$ und dem im obigen File spezifizierten Zeitintervall und Anfangszustand gerechnet werden soll. Wir wollen einen Test machen und dadurch feststellen, welchen Zeitgewinn wir erhalten, wenn wir die Jacobi-Matrix analytisch bereitstellen oder nur den "sparsity pattern". Macht man hundertmal den obigen Aufruf, so brauchten wir bei der analytischen Bereitstellung oder der Angabe der Dünnbesetzheitsstruktur jeweils etwa 36 Sekunden, andernfalls etwa 133 Sekunden. Diesmal stoppten wir durch

```
s=cputime;
for k=1:100
   [t,x]=ode15s('bruss',[],[],options,100);
end;
cputime-s
```

In Abbildung 3.8 veranschaulichen wir die Lösungskomponenten auf $[0,1] \times [0,10]$. Den links stehenden Plot haben wir z. B. erzeugt durch

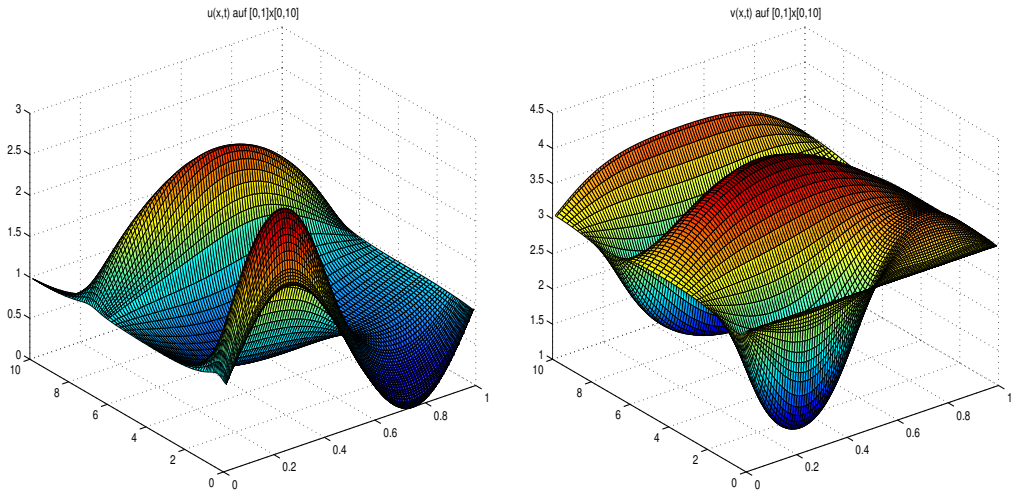


Abbildung 3.8: Die Lösungskomponenten $u$ und $v$ auf $[0,1] \times [0,10]$

```
u=x(:,1:2:199);
surf(y,t,u);
title('u(x,t) auf [0,1]x[0,10]');
```

□

In MATLAB kann man den Befehl `odedemo` angeben und hierdurch eine Vielzahl von Beispielen mit verschiedenen Lösern bearbeiten.

### 3.4.7 Aufgaben

1. Gegeben sei das Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{6}(k_1 + 4k_2 + k_3),$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t+\tfrac{1}{2}h, u+\tfrac{1}{2}hk_1), \qquad k_3 := f(t+h, u-hk_1+2hk_2).$$

Man berechne die zugehörige Stabilitätsfunktion und plotte mit Hilfe von Maple das Stabilitätsgebiet.

2. Man bestimme das Stabilitätsgebiet zum 2-Schrittverfahren

$$u(t+2h) - u(t) = 2hf(t+h, u(t+h)).$$

3. Auf die Anfangswertaufgabe

$$x' = -2000(x - \cos t), \qquad x(0) = 1$$

wende man die (implizite) Trapezregel und das implizite Euler-Verfahren mit Schrittweite $h = 1.5/40$ an. Man plotte die erhaltenen Ergebnisse über dem Intervall $[0, 1.5]$ und vergleiche diese mit der exakten Lösung.

4. Man zeige, dass das Stabilitätsgebiet des 2-Schrittverfahrens

$$u(t+2h) - u(t) = \frac{1}{2}h[f(t+h, u(t+h)) + 3f(t, u(t))]$$

im Kreis um $(-\frac{2}{3}, 0)$ mit dem Radius $\frac{2}{3}$ enthalten ist. Ferner zeige man, dass das reelle Intervall $[-\frac{4}{3}, 0]$ im Stabilitätsgebiet enthalten ist[19].

5. Gegeben sei das 2-stufige implizite Runge-Kutta-Verfahren mit

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

Man zeige, dass dieses die Konsistenzordnung 3 besitzt, berechne die Stabilitätsfunktion und beweise, dass das Verfahren A-stabil ist. So weit wie möglich sollte zur Bearbeitung dieser Aufgabe Maple eingesetzt werden.

6. Ein implizites Runge-Kutta-Verfahren heißt *L-stabil*, wenn es A-stabil ist und zusätzlich $\lim_{z\to\infty} R(z) = 0$, wobei $R$ die zum Verfahren gehörende Stabilitätsfunktion ist.

Gegeben sei ein A-stabiles implizites $s$-stufiges Runge-Kutta-Verfahren mit den Daten

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

Hierbei sei $A$ nichtsingulär und

$$a_{sj} = b_j, \qquad j = 1, \ldots, s,$$

oder

$$a_{i1} = b_1, \qquad i = 1, \ldots, s.$$

Man zeige, dass das zugehörige Runge-Kutta-Verfahren L-stabil ist. Weiter zeige man, dass das implizite Euler-Verfahren L-stabil ist, nicht aber die Trapezregel.

7. Bei einem $s$-stufigen impliziten Runge-Kutta-Verfahren für ein System von $n$ Differentialgleichungen erster Ordnung muss in jedem Zeitschritt zur Berechnung von $k_1, \ldots, k_s$ ein nichtlineares Gleichungssystem mit $ns$ Gleichungen und ebenso vielen Unbekannten gelöst werden. Ist die Verfahrensmatrix $A \in \mathbb{R}^{s\times s}$ eine untere Dreiecksmatrix, so spricht man von einem *diagonal-impliziten Runge-Kutta-Verfahren*. Jetzt können $k_1, \ldots, k_s$ sukzessive nach einander berechnet werden, so dass jetzt $s$ nichtlineare Gleichungssysteme mit $n$ Gleichungen und Unbekannten zu lösen sind. Man spricht von einem *einfach-diagonal-impliziten Runge-Kutta-Verfahren*, wenn die Diagonalelemente zusätzlich alle gleich sind, wenn also $a_{ii} = \gamma$, $i = 1, \ldots, s$.

[19] In einer Aufgabe bei K. Strehmel, R. Weiner (1995, S. 348) wird behauptet, dass das Stabilitätsgebiet des obigen 2-Schrittverfahrens ein Kreis mit dem Mittelpunkt $(-\frac{2}{3}, 0)$ und dem Radius $\frac{2}{3}$ ist. Wer kann das beweisen?

Man betrachte das einfach-diagonal-implizite Runge-Kutta-Verfahren zu den Daten

$$\begin{array}{c|cc} \gamma & \gamma & \\ c_2 & c_2-\gamma & \gamma \\ \hline & b_1 & b_2 \end{array}$$

mit $b_2 \neq 0$. Man gebe Bedingungen dafür, dass dieses Verfahren die Konsistenzordnung 2 oder sogar 3 besitzt. Schließlich gebe man ein 2-stufiges einfach-diagonal-implizites Runge-Kutta-Verfahren der Konsistenzordnung 3 an, welches $A$ stabil ist.

8. Man löse die Anfangswertaufgabe

$$x' = -100(x - \sin t), \qquad x(0) = 1,$$

auf dem Zeitintervall $[0, 10]$ mit den MATLAB-Funktionen `ode45` und `ode15s`. Man vergleiche die Anzahl der benötigten Zeitschritte.

9. Die folgende Definition spielt in einem vertieften Studium steifer Differentialgleichungen eine wichtige Rolle (siehe z. B. K. Strehmel, R. Weiner (1995, S. 199) und E. Hairer et al. (1993, S. 61)).

Sei $\|\cdot\|$ eine beliebige Vektornorm im $\mathbb{R}^n$. Ist dann $\|\cdot\|$ die zugeordnete Matixnorm, so heißt

$$\mu(A) := \lim_{\delta \to 0+} \frac{\|I + \delta A\| - 1}{\delta}$$

die zugeordnete *logarithmische Norm*. Man zeige:

(a) Die logarithmische Norm existiert, da die Abbildung $\delta \mapsto (\|I + \delta A\| - 1)/\delta$ auf $(0, 1]$ nach unten beschränkt und monoton nicht fallend ist.

(b) Die der euklidischen Norm $\|\cdot\|$ zugeordnete Matrixnorm ist bekanntlich die sogenannte Spektralnorm $\|\cdot\|_2$, wobei $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$. Man berechne die zugehörige logarithmische Norm.

(c) Die der Maximumnorm $\|\cdot\|_\infty$ zugeordnete Matrixnorm ist bekanntlich die maximale Betragssummennorm $\|\cdot\|_\infty$, wobei $\|A\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^n |a_{ij}|$. Man berechne die zugehörige logarithmische Norm.

(d) Man beweise die folgenden Eigenschaften der logarithmischen Norm:

i. $\mu(\alpha A) = \alpha \mu(A)$ für $\alpha \geq 0$.
ii. $|\mu(A)| \leq \|A\|$.
iii. $\mu(A + B) \leq \mu(A) + \mu(B)$.
iv. $|\mu(A) - \mu(B)| \leq \|A - B\|$.

# Kapitel 4

# Die Theorie gewöhnlicher Rand- und Eigenwertaufgaben

Bisher haben wir uns mit Anfangswertaufgaben gewöhnlicher Differentialgleichungen beschäftigt. Hier ist eine Lösung einer Differentialgleichung bzw. eines Differentialgleichungssystems zu finden, welche in einem Anfangszeitpunkt einer Anfangsbedingung genügt. In diesem Kapitel werden wir uns im wesentlichen auf Zweipunkt-Randwertaufgaben beschränken, bei welchen Zusatzbedingungen am Rande des betrachteten Lösungsintervalls gestellt werden. Mit Rücksicht auf den physikalischen Hintergrund werden wir eine Umbezeichnung vornehmen. Die unabhängige Variable, die i. Allg. den Charakter einer Ortsvariablen hat, werden wir jetzt mit $x$ statt mit $t$ bezeichnen, während die gesuchte Lösung jetzt $u$ statt $x$ genannt wird.

Es wird sich herausstellen, dass die Lösungstheorie für gewöhnliche Randwertaufgaben wesentlich komplizierter als bei Anfangswertaufgaben ist. Fast ausschließlich werden wir uns auf eine lineare Differentialgleichung zweiter Ordnung mit sogenannten Sturmschen Randwertaufgaben beschränken. Es ist naheliegend und verlockend, auch einen funktionalanalytischen Zugang zu Eigenwertaufgaben zu schildern. Aus Zeitgründen verzichten wir hierauf und verweisen z. B. auf W. Walter (1993, S. 237 ff.).

## 4.1 Sturmsche Randwertaufgaben

### 4.1.1 Beispiele, Definitionen

Wir beginnen mit zwei Beispielen.

**Beispiel:** Wir betrachten einen Stab der Länge $L$, dessen linkes Ende die konstante Temperatur $u_0$ und dessen rechtes Ende die konstante Temperatur $u_L$ besitze. Die Außentemperatur sei $u_A$, sie sei ebenfalls konstant. Die Temperaturverteilung $u(\cdot)$ längs des Stabes ist dann (siehe H. Heuser (1989, S. 371)) Lösung der Zweipunkt-Randwertaufgabe

$$u'' = \alpha^2(u - u_A), \qquad u(0) = u_0, \quad u(L) = u_L.$$

Hierbei ist $\alpha > 0$ eine gewisse Materialkonstante. Auch diese Gleichung kann Maple übrigens lösen:

```
> dsolve({(D@@2)(u)(x)=alpha^2*(u(x)-u_A),u(0)=u_0,u(L)=u_L},u(x));
```

$$\mathrm{u}(x) = u\_A + \frac{(u\_A\, e^{(-\alpha\, L)} - u\_A + u\_L - u\_0\, e^{(-\alpha\, L)})\, e^{(\alpha\, x)}}{e^{(\alpha\, L)} - e^{(-\alpha\, L)}}$$
$$- \frac{(-u\_A + e^{(\alpha\, L)}\, u\_A - e^{(\alpha\, L)}\, u\_0 + u\_L)\, e^{(-\alpha\, x)}}{e^{(\alpha\, L)} - e^{(-\alpha\, L)}}$$

Es ist leicht, auf diese Lösung in systematischer Weise zu kommen und zu erkennen, dass es die einzige ist. Denn die allgemeine Lösung der inhomogenen Differentialgleichung

$$u'' - \alpha^2 u = -\alpha^2 u_A$$

ist offenbar

$$u(x) = u_A + c_1 e^{\alpha x} + c_2 e^{-\alpha x}.$$

Die beiden Randbedingungen führen auf das lineare Gleichungssystem

$$\begin{pmatrix} 1 & 1 \\ e^{\alpha L} & e^{-\alpha L} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} u_0 - a_A \\ u_L - u_A \end{pmatrix},$$

welches eindeutig lösbar ist und die obige Lösung ergibt. Wir haben hier ein Beispiel einer linearen Randwertaufgabe kennengelernt. □

Auch das nächste Beispiel findet man bei H. Heuser (1989, S. 346).

**Beispiel:** Eine Säule mit Länge $l$ und Elastizitäts modul $E$ sei fest im Boden verankert und trage auf ihrem oberen Ende eine Last $P$. Ihr Flächenträgheitsmoment $I$ nehme mit wachsendem $x$ nach dem Gesetz

$$I(x) = I_0 e^{-(k/l)x} \qquad (k > 0 \text{ konstant})$$

ab. Es kann gezeigt werden, dass die Bestimmung der Auslenkung der Säule auf eine Randwertaufgabe der Form

$$(*) \qquad u'' + \frac{P}{EI_0} e^{(k/l)x} u = 0, \qquad u'(0) = 0, \quad u(l) = 0$$

zurückgeführt werden kann. Die Frage ist jetzt: Für welche Lasten $P$ besitzt $(*)$ eine nichttriviale Lösung? Auch diese Aufgabe kann geschlossen gelöst werden, auch wenn die Analyse weniger elementar ist. Als allgemeine Lösung der linearen Differentialgleichung zweiter Ordnung

$$u'' + \frac{P}{EI_0} e^{(k/l)x} u = 0$$

erhalten wir z. B. mit Maple

$$u(x) = c_1 J_0(\alpha\beta^{x/l}) + c_2 Y_0(\alpha\beta^{x/l}),$$

wobei wir zur Abkürzung

$$\alpha := \frac{2l}{k}\sqrt{\frac{P}{EI_0}}, \qquad \beta := e^{k/2}$$

gesetzt haben. Hierbei sind $J_0$ bzw. $Y_0$ sogenannte Besselsche Funktionen erster bzw. zweiter Art der Ordnung 0. Diese sind auch in MATLAB (und auch in Maple) verfügbar. Denn nach der Aufforderung `help bessel` erhalten wir die Auskunft:

```
BESSEL Bessel functions of various kinds.
   Bessel functions are solutions to Bessel's differential
   equation of order NU:
            2                     2    2
           x * y'' +  x * y' + (x - nu ) * y = 0

   There are several functions available to produce solutions to
   Bessel's equations.  These are:

       BESSELJ(NU,Z)    Bessel function of the first kind
       BESSELY(NU,Z)    Bessel function of the second kind
       BESSELI(NU,Z)    Modified Bessel function of the first kind
       BESSELK(NU,Z)    Modified Bessel function of the second kind
       BESSELH(NU,K,Z)  Hankel function
       AIRY(K,Z)        Airy function

   See the help for each function for more details.
```

Also stellt sich die Frage, für welche $\alpha$ das lineare Gleichungssystem

$$\begin{pmatrix} J_0(\alpha\beta) & Y_0(\alpha\beta) \\ J_0'(\alpha) & Y_0'(\alpha) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

eine nichttriviale Lösung besitzt. Dies ist genau dann der Fall, wenn $\alpha$ Lösung der Gleichung

$$J_0(\alpha\beta)Y_1(\alpha) - J_1(\alpha)Y_0(\alpha\beta) = 0$$

ist, wobei wir noch $J_0' = -J_1$, $Y_0' = -Y_0$ ausgenutzt haben ($J_1$ bzw. $Y_1$ ist Besselsche Funktion erster bzw. zweiter Art zum Index 1). Sei $\alpha_0$ ihre kleinste Lösung. Dann ist

$$P_0 := EI_0\Big(\frac{k\alpha_0}{2l}\Big)^2$$

die kleinste Last, bei der eine Deformation der Säule eintritt. Deshalb nennt man $P_0$ die *Knicklast*. Es handelt sich hier um ein Beispiel einer Eigenwertaufgabe. □

Eine Differentialgleichung zweiter Ordnung hat natürlich im allgemeinen Fall die Form

$$(*) \qquad F(x, u, u', u'') = 0.$$

Wir beschränken uns auf den Fall *expliziter* Gleichungen, bei denen die Funktion $F$ nach der höchsten Ableitung $u''$ der gesuchten Funktion $u$ aufgelöst werden kann.

**Definition 1.1** Die Differentialgleichung zweiter Ordnung $(*)$ heißt *quasilinear*, falls

$$F(x, u, u', u'') := -u'' + B(x, u)u' + C(x, u),$$

*semilinear*, falls

$$F(x, u, u', u'') := -u'' + b(x)u' + C(x, u),$$

bzw. *linear*, falls

$$F(x, u, u', u'') := -u'' + b(x)u' + c(x)u - f(x).$$

Im allgemeinen Fall sind die Randbedingungen

$$G_i(a, b, u(a), u(b), u'(a), u'(b)) = 0, \qquad i = 1, 2,$$

nichtlinear und gekoppelt. In Anwendungen ist oft ausreichend, Randbedingungen in linearer und entkoppelter Form zu betrachten. Genauer betrachten wir in diesem Abschnitt lineare *Sturmsche Randwertaufgaben*. Hierbei handelt es sich um das folgende Problem: Man bestimme ein $u \in C^2[a, b]$ mit

$$\text{(DGL)} \qquad (Lu)(x) := -[p(x)u'(x)]' + q(x)u(x) = g(x) \qquad \text{für alle } x \in [a, b]$$

und

$$\text{(RB)} \qquad \begin{array}{rcl} R_1 u & := & \alpha_1 u(a) + \alpha_2 p(a) u'(a) \;=\; \eta_1, \\ R_2 u & := & \beta_1 u(b) + \beta_2 p(b) u'(b) \;=\; \eta_2. \end{array}$$

Hierbei wird generell vorausgesetzt:

(V) Es ist $p \in C^1[a, b]$, $q, g \in C[a, b]$ und $p(x) > 0$ für alle $x \in [a, b]$. Ferner ist $\alpha_1^2 + \alpha_2^2 > 0$ und $\beta_1^2 + \beta_2^2 > 0$, und o. B. d. A. $\alpha_1 \geq 0$ und $\beta_1 \geq 0$. Schließlich seien $\eta_1, \eta_2 \in \mathbb{R}$.

Ist $g \equiv 0$, so ist die Differentialgleichung (DGL) homogen, ist $(\eta_1, \eta_2) = (0, 0)$, so sind die Randbedingungen (RB) homogen. Ist $\alpha_2 = 0$ bzw. $\beta_2 = 0$, so spricht man von einer Randbedingung erster Art (oder vom Dirichletschen Typ) in $a$ bzw. $b$, ist $\alpha_1 = 0$ bzw. $\beta_1 = 0$ so ist die entsprechende Randbedingung von zweiter Art (uder vom Neumannschen Typ), sind $\alpha_1$ und $\alpha_2$ bzw. $\beta_1$ und $\beta_2$ von Null verschieden, so spricht man schließlich von einer Randbedingung dritter Art (oder einer vom Robinschen Typ).

**Bemerkung:** Eine lineare Differentialgleichung der Form

$$-u'' + b(x)u' + c(x)u = f(x)$$

kann durch Multiplikation mit

$$p(x) := \exp\Big(-\int_a^x b(\xi)\, d\xi\Big)$$

auf die sogenannte "selbstadjungierte" Form (DGL) gebracht werden, da

$$-p(x)u'' + p(x)b(x)u' + p(x)c(x)u = -[p(x)u']' + \underbrace{p(x)c(x)}_{=:q(x)} u = \underbrace{p(x)f(x)}_{=:g(x)}.$$

Dies ist nur für theoretische Zwecke gelegentlich von Vorteil. □

### 4.1.2 Existenz, Eindeutigkeit

Im folgenden Satz werden notwendige und hinreichende Bedingungen dafür angegeben, dass die lineare Sturmsche Randwertaufgabe mit der Differentialgleichung (DGL) und der Randbedingung (RB) eindeutig lösbar ist.

**Satz 1.2** *Seien $u_1, u_2$ linear unabhängige Lösungen der homogenen Differentialgleichung $Lu = 0$. Dann ist die inhomogene Randwertaufgabe (DGL), (RB) genau dann eindeutig lösbar, wenn die Verträglichkeitsbedingung*

$$\text{(VB)} \qquad \det \begin{pmatrix} R_1u_1 & R_1u_2 \\ R_2u_1 & R_2u_2 \end{pmatrix} \neq 0$$

*erfüllt ist. Ferner gilt (VB) genau dann, wenn die Implikation*

$$\text{(VB)}^* \qquad Lu = 0, \qquad Ru = \begin{pmatrix} R_1u \\ R_2u \end{pmatrix} = 0 \Longrightarrow u = 0$$

*gilt.*

**Beweis:** Ist $u^*$ eine spezielle Lösung von (DGL) $Lu = g$, so lautet die allgemeine Lösung

$$u = c_1u_1 + c_2u_2 + u^*.$$

Die beiden Randbedingungen (RB) ergeben zwei lineare Gleichungen für die beiden Konstanten $c_1, c_2$, nämlich

$$R_iu = c_1R_iu_1 + c_2R_iu_2 + R_iu^* = \eta_i, \qquad i = 1, 2,$$

welche genau dann eindeutig lösbar sind, wenn (VB) gilt. Der Rest der Behauptung ist trivial. □

Insbesondere hängt die eindeutige Lösbarkeit von (DGL), (RB) nicht von den "Inhomogenitäten" $g$ und $(\eta_1, \eta_2)$ ab. Die Verhältnisse sind also wie bei einem linearen Gleichungssystem mit ebenso vielen Gleichungen wie Unbekannten: Dieses ist genau dann für jede rechte Seite eindeutig lösbar, wenn das homogene System nur die triviale Lösung besitzt.

**Beispiel:** Sei $Lu := -u'' - u$, $R_1u = u(0) + u'(0)$, $R_2u := u(\pi)$. Durch $u_1(x) := \cos x$, $u_2(x) := \sin x$ sind zwei linear unabhängige Lösungen von $Lu = 0$ gegeben. Ferner ist

$$\det \begin{pmatrix} R_1u_1 & R_1u_2 \\ R_2u_1 & R_2u_2 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} = 1,$$

so dass (VB) erfüllt ist. Ist dagegen $R_1u := u(0)$, $R_2u := u(\pi)$, so ist

$$\det \begin{pmatrix} R_1u_1 & R_1u_2 \\ R_2u_1 & R_2u_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} = 0,$$

also ist (VB) nicht erfüllt. In diesem Fall hat die homogene Aufgabe die Lösungen $u(x) = c \sin x$ mit $c \in \mathbb{R}$. □

**Bemerkung:** Neben den Standardvoraussetzungen (V) sei nun $q(x) \geq 0$ für alle $x \in [a, b]$. Wir wollen uns überlegen, dass dann bei geeigneten Vorzeichen der in den Randbedingungen vorkommenden Konstanten $\alpha_i, \beta_i$, $i = 1, 2$, die Bedingung $(\text{VB})^*$ in

Satz 1.2 erfüllt ist und daher die entsprechende inhomogene Sturmsche Randwertaufgabe eindeutig lösbar ist. Denn sei $u$ eine Lösung der homogenen Aufgabe $Lu = 0$, $Ru = 0$. Dann ist

$$\begin{aligned} 0 &= \int_a^b Lu(x)u(x)\,dx \\ &= \int_a^b [-[p(x)u'(x)]'u(x) + q(x)u^2(x)]\,dx \\ &= \int_a^b [p(x)u'(x)^2 + q(x)u^2(x)]\,dx - p(x)u'(x)u(x)\,\Big|_a^b \\ &= \int_a^b [p(x)u'(x)^2 + q(x)u^2(x)]\,dx - p(b)u'(b)u(b) + p(a)u'(a)u(a). \end{aligned}$$

Wir machen jetzt eine Fallunterscheidung.

- Es ist $\alpha_1 > 0$ und $\beta_1 > 0$.

Aus obiger Gleichungskette folgt dann, dass

$$0 = \int_a^b [p(x)u'(x)^2 + q(x)u^2(x)]\,dx + \frac{\beta_2}{\beta_1}[p(b)u'(b)]^2 - \frac{\alpha_2}{\alpha_1}[p(a)u'(a)]^2$$

für jedes $u$ mit $Lu = 0$, $Ru = 0$. Hieraus erhalten wir: Ist $\beta_2 \geq 0$ und $\alpha_2 \leq 0$ (dies ist also insbesondere für Randbedingungen erster Art erfüllt), so ist $u' \equiv 0$ und folglich $u$ konstant. Aus $Ru = 0$ folgt, dass $u = 0$.

- Es ist ($\alpha_1 = 0$ und $\beta_1 > 0$) oder ($\alpha_1 > 0$ und $\beta_1 = 0$).

Wir betrachten zunächst den ersten Fall, also $\alpha_1 = 0$ und $\beta_1 > 0$. Aus obiger Gleichungskette folgt dann

$$0 = \int_a^b [p(x)u'(x)^2 + q(x)u^2(x)]\,dx + \frac{\beta_2}{\beta_1}[p(b)u'(b)]^2$$

für jedes $u$ mit $Lu = 0$ und $Ru = 0$. Ist also $\beta_2 \geq 0$, so ist wieder $u$ konstant und die Randbedingung $R_2u = 0$ liefert, dass $u = 0$. Im zweiten Fall, also $\alpha_1 > 0$ und $\beta_1 = 0$, folgt aus $\alpha_2 \leq 0$, dass $u = 0$.

- Es ist $\alpha_1 = 0$ und $\beta_1 = 0$.

Für jedes $u$ mit $Lu = 0$, $Ru = 0$ ist dann

$$0 = \int_a^b [p(x)u'(x)^2 + q(x)u^2(x)]\,dx.$$

Hieraus folgt wieder, dass $u$ konstant ist. Ist also $q$ nicht nur nichtnegativ auf $[a, b]$, sondern auch $\int_a^b q(x)\,dx > 0$ bzw. $q$ nicht identisch verschwindend, so ist $u = 0$.

Damit haben wir durchdiskutiert, unter welchen Voraussetzungen für nichtnegatives $q$ aus $Lu = 0$, $Ru = 0$ folgt, dass $u = 0$. □

### 4.1.3 Die Greensche Funktion

Unter der generellen Voraussetzung (V) und der Bedingung (VB) bzw. (VB)$^*$ in Satz 1.2 besitzt die Sturmsche Randwertaufgabe $Lu = g$, $Ru = \eta$ eine eindeutige Lösung. Wir werden im folgenden $\eta = 0$ annehmen und nur noch die halbhomogene Aufgabe $Lu = g$, $Ru = 0$ betrachten. Den allgemeinen Fall kann man hierauf zurückführen, indem man zunächst ein $\phi \in C^2[a,b]$ mit $R\phi = \eta$ bestimmt[1] und anschließend die halbhomogene Aufgabe $Lv = g - L\phi$, $Rv = 0$ betrachtet.

Im folgenden suchen wir eine Darstellung der Lösung von $Lu = g$, $Ru = 0$. Wir werden zeigen, dass sich die Lösung in der Form

$$u(x) = \int_a^b G(x,\xi) g(\xi)\, d\xi$$

mit der sogenannten Greenschen Funktion $G$ darstellen lässt. Die entsprechenden Resultate sind in dem folgenden Satz zusammengestellt.

**Satz 1.3** *Gegeben sei die (halbhomogene) Sturm'sche Randwertaufgabe, also die Differentialgleichung*

$$\text{(DGL)} \qquad Lu := -(p(x)u')' + q(x)u = g(x)$$

*und die Randbedingungen*

$$\text{(RB)} \qquad Ru := \begin{pmatrix} R_1 u \\ R_2 u \end{pmatrix} := \begin{pmatrix} \alpha_1 u(a) + \alpha_2 p(a) u'(a) \\ \beta_1 u(b) + \beta_2 p(b) u'(b) \end{pmatrix} = 0.$$

*Neben der generellen Voraussetzung (V) gelte*

$$\text{(VB)}^* \qquad Lu = 0, \qquad Ru = \begin{pmatrix} R_1 u \\ R_2 u \end{pmatrix} = 0 \Longrightarrow u = 0.$$

*Dann existiert genau eine sogenannte Greensche Funktion $G\colon [a,b] \times [a,b] \longrightarrow \mathbb{R}$ zu $(L,R)$ mit*

1. *$G\colon [a,b] \times [a,b] \longrightarrow \mathbb{R}$ ist stetig.*
2. *In jedem der beiden Dreiecke*

$$\Delta_1 := \{(x,\xi) : a \le \xi \le x \le b\}, \qquad \Delta_2 := \{(x,\xi) : a \le x \le \xi \le b\}$$

*existieren die partiellen Ableitungen $G_x$, $G_{xx}$ und sind stetig, wobei natürlich auf der Diagonalen die dem Dreieck entsprechende einseitige Ableitung zu nehmen ist.*

[1] Ist $R_1 u := u(a)$ und $R_2 u := u(b)$, so kann man z. B.

$$\phi(x) := \eta_1 \frac{b-x}{b-a} + \eta_2 \frac{x-a}{b-a}$$

nehmen.

3. *Bei festem* $\xi \in [a,b]$ *ist* $LG(x,\xi) = 0$ *für* $x \in [a,b] \setminus \{\xi\}$.

4. *Es gilt die Sprungbedingung*

$$G_x(x+0,x) - G_x(x-0,x) = -\frac{1}{p(x)} \qquad \text{für } x \in (a,b).$$

5. $R_1 G(\cdot,\xi) = R_2 G(\cdot,\xi) = 0$ *für jedes* $\xi \in (a,b)$.

*Die eindeutige Lösung von (DGL), (RB) ist dann gegeben durch*

$$u(x) = \int_a^b G(x,\xi) g(\xi)\, d\xi.$$

**Beweis:** Seien $u_1, u_2$ linear unabhängige Lösungen von $Lu = 0$. Die Bedingung 3. an die Greensche Funktion ist genau dann erfüllt, wenn

$$G(x,\xi) = \begin{cases} (a_1(\xi) + b_1(\xi))u_1(x) + (a_2(\xi) + b_2(\xi))u_2(x), & (x,\xi) \in \Delta_1, \\ (a_1(\xi) - b_1(\xi))u_1(x) + (a_2(\xi) - b_2(\xi))u_2(x), & (x,\xi) \in \Delta_2 \end{cases}$$

mit noch unbestimmten $a_1, a_2, b_1, b_2$, zu deren Bestimmung man vier Bedingungen, nämlich die Stetigkeitsforderung, die Sprungbedingung und die beiden Randbedingungen, zur Verfügung hat. Die Stetigkeitsforderung führt auf

$$b_1(\xi)u_1(\xi) + b_2(\xi)u_2(\xi) = 0,$$

die Sprungbedingung auf

$$b_1(\xi)u_1'(\xi) + b_2(\xi)u_2'(\xi) = -\frac{1}{2p(\xi)}.$$

Insgesamt erhält man ein lineares Gleichungssystem für $b_1(\xi), b_2(\xi)$, dessen Koeffizientenmatrix nichtsingulär ist, da ihre Determinante als Wronski-Determinante nicht verschwindet. Für $\xi \in (a,b)$ ist

$$\begin{aligned} R_1 G(\cdot,\xi) &= (a_1(\xi) - b_1(\xi))R_1 u_1 + (a_2(\xi) - b_2(\xi))R_1 u_2, \\ R_2 G(\cdot,\xi) &= (a_1(\xi) + b_1(\xi))R_2 u_1 + (a_2(\xi) + b_2(\xi))R_2 u_2. \end{aligned}$$

Die Bedingung 5. liefert also zur Bestimmung von $a_1(\xi), a_2(\xi)$ das lineare Gleichungssystem

$$\begin{aligned} a_1(\xi)R_1 u_1 + a_2(\xi)R_1 u_2 &= b_1(\xi)R_1 u_1 + b_2(\xi)R_1 u_2, \\ a_1(\xi)R_2 u_1 + a_2(\xi)R_2 u_2 &= -b_1(\xi)R_2 u_1 - b_2(\xi)R_2 u_2 \end{aligned}$$

und dieses ist wegen $(\mathrm{VB})^*$ eindeutig lösbar. Damit ist die eindeutige Existenz einer Greenschen Funktion bewiesen.

Zu zeigen bleibt, dass durch

$$u(x) := \int_a^b G(x,\xi) g(\xi)\, d\xi$$

eine (und dann auch die einzige) Lösung von (DGL), (RB) gegeben ist. Es ist

$$u(x) = \int_a^x G(x,\xi)g(\xi)\,d\xi + \int_x^b G(x,\xi)g(\xi)\,d\xi$$

und daher

$$\begin{aligned} u'(x) &= G(x,x)g(x) + \int_a^x G_x(x,\xi)g(\xi)\,d\xi - G(x,x)g(x) + \int_x^b G_x(x,\xi)g(\xi)\,d\xi \\ &= \int_a^x G_x(x,\xi)g(\xi)\,d\xi + \int_x^b G_x(x,\xi)g(\xi)\,d\xi. \end{aligned}$$

Durch erneutes Differenzieren erhält man

$$\begin{aligned} u''(x) &= G_x(x+0,x)g(x) + \int_a^x G_{xx}(x,\xi)g(\xi)\,d\xi \\ &\qquad - G_x(x-0,x)g(x) + \int_x^b G_{xx}(x,\xi)g(\xi)\,d\xi \\ &= -\frac{g(x)}{p(x)} + \int_a^b G_{xx}(x,\xi)g(\xi)\,d\xi. \end{aligned}$$

Damit wird

$$\begin{aligned} Lu(x) &= -p(x)u''(x) - p'(x)u'(x) + q(x)u(x) \\ &= g(x) + \int_a^b LG(x,\xi)g(\xi)\,d\xi \\ &= g(x). \end{aligned}$$

Also ist die Differentialgleichung (DGL) durch $u$ erfüllt. Wegen

$$u(x) = \int_a^b G(x,\xi)g(\xi)\,d\xi, \qquad u'(x) = \int_a^b G_x(x,\xi)g(\xi)\,d\xi$$

gilt ferner

$$R_1 u = \int_a^b R_1 G(x,\xi)g(\xi)\,d\xi = 0, \qquad R_2 u = \int_a^b R_2 G(x,\xi)g(\xi)\,d\xi = 0.$$

Insgesamt ist der Satz damit bewiesen. □

**Beispiel:** Mit der beim Beweis angewandten Methode wollen wir die Greensche Funktion zu

$$Lu := -u'', \qquad Ru := \begin{pmatrix} u(a) \\ u(b) \end{pmatrix}$$

berechnen. Als linear unabhängige Lösungen von $Lu = 0$ nehme man $u_1(x) := 1$, $u_2(x) = x$. Dann berechnen sich $b_1, b_2$ aus

$$\begin{aligned} b_1(\xi) + b_2(\xi)\xi &= 0, \\ b_2(\xi) &= -\frac{1}{2}, \end{aligned}$$

was

$$b_1(\xi) = \frac{1}{2}\xi, \qquad b_2(\xi) = -\frac{1}{2}$$

ergibt. Anschließend sind $a_1, a_2$ aus

$$\begin{aligned} a_1(\xi) + a_2(\xi)a &= \frac{1}{2}\xi - \frac{1}{2}a \\ a_1(\xi) + a_2(\xi)b &= -\frac{1}{2}\xi + \frac{1}{2}b \end{aligned}$$

zu berechnen, woraus man

$$a_1(\xi) = \frac{1}{2(b-a)}[\xi(a+b) - 2ab], \qquad a_2(\xi) = \frac{1}{2(b-a)}[a+b-2\xi]$$

erhält. Insgesamt wird

$$G(x,\xi) = \frac{1}{b-a}\begin{cases} (\xi-a)(b-x), & a \le \xi \le x \le b, \\ (b-\xi)(x-a), & a \le x \le \xi \le b. \end{cases}$$

Man beachte, dass die Greensche Funktion in diesem Fall eine auf $[a,b]\times[a,b]$ nichtnegative Funktion ist. □

In der folgenden Bemerkung lernen wir eine etwas einfachere Methode zur Berechnung der Greenschen Funktion kennen, die insbesondere die "selbstadungierte" Form der Differentialgleichung (DGL) besser ausnutzt.

**Bemerkung:** Gegeben sei wieder die Sturm'sche Randwertaufgabe (DGL), (RB). Dann gilt:

- *Sind $u_1, u_2$ linear unabhängige Lösungen von $Lu = 0$, so ist*

$$p(x)[u_1(x)u_2'(x) - u_1'(x)u_2(x)] = \text{const} \neq 0.$$

Denn: $u_1, u_2$ sind linear unabhängige Lösungen der Differentialgleichung zweiter Ordnung

$$u'' + \frac{p'(x)}{p(x)}u' - \frac{q(x)}{p(x)}u = 0.$$

Für die Wronski-Determinante

$$\Phi(x) := \det\begin{pmatrix} u_1(x) & u_2(x) \\ u_1'(x) & u_2'(x) \end{pmatrix}$$

gilt aber (siehe das Beispiel einer Differentialgleichung $n$-ter Ordnung im Anschluss an Lemma 2.1 in Abschnitt 2.2)

$$\begin{aligned} \Phi(x) &= \Phi(a)\exp\Big(-\int_a^x \frac{p'(\xi)}{p(\xi)}\,d\xi\Big) \\ &= \Phi(a)\exp[-\ln p(x) + \ln p(a)] \\ &= \frac{\Phi(a)p(a)}{p(x)}, \end{aligned}$$

so daß

$$p(x)\Phi(x) = p(x)[u_1(x)u_2'(x) - u_1'(x)u_2(x)] = p(a)\Phi(a) = \text{const}.$$

Seien nun $u_1, u_2$ linear unabhängige Lösungen von $Lu = 0$ mit $R_1u_1 = 0$ und $R_2u_2 = 0$ (z. B. bestimme man $u_1$ als eindeutige Lösung von $Lu = 0$ mit den beiden Randbedingungen $R_1u = 0$, $R_2u = 1$, entsprechend $u_2$ als die Lösung von $Lu = 0$ mit $R_1u = 1$, $R_2u = 0$). Dann ist die Greensche Funktion zu $(L, R)$ gegeben durch

$$G(x,\xi) = -\frac{1}{c}\begin{cases} u_1(\xi)u_2(x), & a \le \xi \le x \le b, \\ u_2(\xi)u_1(x), & a \le x \le \xi \le b \end{cases}$$

mit

$$c := p(x)[u_1(x)u_2'(x) - u_1'(x)u_2(x)] = \text{const} \neq 0.$$

Denn die fünf eine Greensche Funktion charakterisierenden Eigenschaften aus dem vorigen Satz sind erfüllt. Weiter erkennt man, dass die Greensche Funktion zu $(L, R)$ *symmetrisch* ist, d. h. es ist $G(x, \xi) = G(\xi, x)$ für alle $(x, \xi) \in [a, b] \times [a, b]$ gilt. □

**Beispiel:** Mit den obigen Bemerkungen wollen wir die Greensche Funktion zu

$$Lu := -u'' + \lambda^2 u, \qquad Ru := \begin{pmatrix} u(a) \\ u(b) \end{pmatrix}$$

mit $\lambda > 0$ berechnen. Durch

$$u_1(x) := \sinh \lambda(x - a), \qquad u_2(x) := \sinh \lambda(b - x)$$

sind zwei linear unabhängige Lösungen von $Lu = 0$ mit $u_1(a) = 0$, $u_2(b) = 0$ gegeben. Ferner ist

$$c := u_1(x)u_2'(x) - u_1'(x)u_2(x) = -\lambda \sinh \lambda(b - a),$$

so dass die Greensche Funktion zu $(L, R)$ durch

$$G(x,\xi) = \frac{1}{\lambda \sinh \lambda(b-a)}\begin{cases} \sinh \lambda(\xi - a)\sinh \lambda(b - x), & a \le \xi \le x \le b, \\ \sinh \lambda(b - \xi)\sinh \lambda(x - a), & a \le x \le \xi \le b \end{cases}$$

gegeben ist. Wieder stellt man fest, dass die Greensche Funktion nichtnegativ ist. □

**Beispiel:** Im letzten Beispiel wussten wir von vornherein, dass $Lu = 0$, $Ru = 0$ nur die triviale Lösung $u = 0$ besitzt und daher die Greensche Funktion für beliebiges $\lambda > 0$ existiert. Das ist bei

$$Lu := -u'' - \lambda^2 u, \qquad Ru := \begin{pmatrix} u(a) \\ u(b) \end{pmatrix}$$

anders. Denn

$$u_1(x) := \sin \lambda(x - a), \qquad u_2(x) := \sin \lambda(b - x)$$

sind, allerdings nur für

$$\frac{\lambda(b-a)}{\pi} \notin \mathbb{Z},$$

zwei linear unabhängige Lösungen von $Lu = 0$. Für $\lambda(b-a)/\pi = m \in \mathbb{Z}$ besitzt die homogene Aufgabe $Lu = 0$, $Ru = 0$ dagegen die nichttrivialen Lösungen $u(x) = c\sin\lambda(x-a)$ für beliebiges $c \neq 0$. Für $\lambda(b-a)/\pi \notin \mathbb{Z}$ kann man die Greensche Funktion aber nach obigem Muster ausrechnen und erhält in diesem Falle

$$G(x,\xi) = \frac{1}{\lambda \sin\lambda(b-a)} \begin{cases} \sin\lambda(\xi-a)\sin\lambda(b-x), & a \le \xi \le x \le b, \\ \sin\lambda(b-\xi)\sin\lambda(x-a), & a \le x \le \xi \le b. \end{cases}$$

Hier stellt man fest, dass die Greensche Funktion für $\lambda(b-a) < \pi$ nichtnegativ ist. □

Als Anwendung unserer Ergebnisse über die Greensche Funktion zu Sturmschen Randwertaufgaben wollen wir einen kleinen Abstecher zu nichtlinearen Randwertaufgaben machen und den folgenden Satz[2] beweisen.

**Satz 1.4** *Die Funktion $f\colon [a,b]\times\mathbb{R} \longrightarrow \mathbb{R}$ sei stetig und genüge der Lipschitzbedingung*

$$|f(x,u) - f(x,v)| \le L\,|u-v| \qquad \text{für alle } (x,u),(x,v) \in [a,b]\times\mathbb{R}.$$

*Die Lipschitzkonstante $L$ sei so klein, dass*

$$L < \frac{\pi^2}{(b-a)^2}.$$

*Dann ist die Randwertaufgabe*

$$-u'' = f(x,u), \qquad u(a) = u(b) = 0$$

*eindeutig lösbar.*

**Beweis:** Wir geben zunächst einen naheliegenden, aber nicht ganz zum Ziel führenden "Beweis" an, danach wird dieser "Beweis" so modifiziert, dass er zum Beweis wird.

Man normiere $C[a,b]$ durch die Maximumnorm $\|u\| := \max_{x\in[a,b]} |u(x)|$ und betrachte die zu (P) äquivalente Fixpunktaufgabe für die durch

$$F(u)(x) := \int_a^b G(x,\xi) g(\xi,u(\xi))\,d\xi$$

definierte Abbildung $F\colon C[a,b] \longrightarrow C[a,b]$ mit der Green'schen Funktion

$$G(x,\xi) = \frac{1}{b-a} \begin{cases} (\xi-a)(b-x) & \text{für} \quad a \le \xi \le x \le b, \\ (b-\xi)(x-a) & \text{für} \quad a \le x \le \xi \le b. \end{cases}$$

Die Green'sche Funktion $G$ ist nichtnegativ und daher ist

$$\begin{aligned} |F(u)(x) - F(v)(x)| &\le \int_a^b G(x,\xi)\,|f(\xi,u(\xi)) - f(\xi,v(\xi))|\,d\xi \\ &\le L\int_a^b G(x,\xi)\,d\xi\,\|u-v\| \end{aligned}$$

[2]Diesen Satz findet man als Aufgabe bei W. Walter (1993, S. 221).

für alle $u, v \in C[a,b]$ und alle $x \in [a,b]$. Nun kann man $v(x) := \int_a^b G(x,\xi)\,d\xi$ entweder direkt ausrechnen, oder, etwas geschickter, beachten, daß

$$v(x) = \int_a^b G(x,\xi)\,d\xi = \int_a^b G(x,\xi)\cdot 1\,d\xi$$

die Lösung von

$$-v'' = 1, \qquad v(a) = v(b) = 0$$

ist. Hieraus erhält man $v(x) = \frac{1}{2}(x-a)(b-x)$ und damit

$$\begin{aligned} |F(u)(x) - F(v)(x)| &\leq \frac{L}{2}(x-a)(b-x)\,\|u-v\| \\ &\leq L\,\frac{(b-a)^2}{8}\,\|u-v\| \end{aligned}$$

für alle $u, v \in C[a,b]$ und alle $x \in [a,b]$. Das impliziert wiederum

$$\|F(u) - F(v)\| \leq L\,\frac{(b-a)^2}{8}\,\|u-v\| \qquad \text{für alle } u, v \in C[a,b].$$

Nun ist aber $8 < \pi^2$ und man muss sich zum Beweis der behaupteten Aussage noch etwas mehr anstrengen. Hierzu benutzen wir einen Trick, den wir schon bei Anfangswertaufgaben, genauer beim Beweis des Satzes von Picard-Lindelöf, angewandt haben. Wir definieren nämlich mit einem $\epsilon > 0$ die auf $[a,b]$ positive Funktion

$$w_\epsilon(x) := \epsilon + \sin\frac{\pi}{b-a}(x-a)$$

und anschließend auf $C[a,b]$ die gewichtete Maximumnorm

$$\|u\|_\epsilon := \max_{x\in[a,b]} \frac{|u(x)|}{w_\epsilon(x)}.$$

Klar ist, dass $\|\cdot\|_\epsilon$ zu $\|\cdot\|$ äquivalent ist, so dass auch $(C[a,b], \|\cdot\|_\epsilon)$ ein Banachraum ist. Wir betrachten wieder die Abbildung $F\colon C[a,b] \longrightarrow C[a,b]$ wie oben und berechnen ihre Lipschitzkonstante bezüglich der Norm $\|\cdot\|_\epsilon$. Für beliebige $u, v \in C[a,b]$ und $x \in [a,b]$ ist

$$\begin{aligned} \frac{|F(u)(x) - F(v)(x)|}{w_\epsilon(x)} &\leq \frac{L}{w_\epsilon(x)} \int_a^b G(x,\xi) w_\epsilon(\xi) \frac{|u(\xi) - v(\xi)|}{w_\epsilon(\xi)}\,d\xi \\ &\leq \frac{L}{w_\epsilon(x)} \int_a^b G(x,\xi) w_\epsilon(\xi)\,d\xi\,\|u-v\|_\epsilon. \end{aligned}$$

Definiert man $v_\epsilon(x) := \int_a^b G(x,\xi) w_\epsilon(\xi)\,d\xi$, so ist

$$-v_\epsilon'' = \epsilon + \sin\frac{\pi}{b-a}(x-a), \qquad v_\epsilon(a) = v_\epsilon(b) = 0$$

und daher

$$v_\epsilon(x) = \frac{\epsilon}{2}(x-a)(b-x) + \frac{(b-a)^2}{\pi^2}\sin\frac{\pi}{b-a}(x-a).$$

Nun ist

$$\frac{v_\epsilon(x)}{w_\epsilon(x)} \le \frac{v_\epsilon(\frac{1}{2}(a+b))}{w_\epsilon(\frac{1}{2}(a+b))} = \frac{\epsilon(b-a)^2/8 + (b-a)^2/\pi^2}{\epsilon+1}.$$

Also ist $F$ auf $C[a,b]$ lipschitzstetig mit der Lipschitzkonstanten

$$q_\epsilon := L\frac{\epsilon(b-a)^2/8 + (b-a)^2/\pi^2}{\epsilon+1}.$$

Es ist

$$\lim_{\epsilon\to 0+} q_\epsilon = L\frac{(b-a)^2}{\pi^2} < 1,$$

für hinreichend kleines $\epsilon > 0$ ist also $q_\epsilon < 1$ und die Behauptung folgt aus dem Kontraktionssatz. □

**Beispiel:** Die Randwertaufgabe

$$u'' + c^2 \sin u = e(x), \qquad u(a) = u(b) = 0$$

mit $e \in C[a,b]$ besitzt eine eindeutige Lösung, falls $c^2 < \pi^2/(b-a)^2$. □

**Bemerkung:** Die Aussage des Satzes ist insofern optimal, als sie für $L = \pi^2/(b-a)^2$ falsch wird. Denn

$$-u'' = \frac{\pi^2}{(b-a)^2}u, \qquad u(a) = u(b) = 0$$

besitzt die *unendlich vielen* Lösungen

$$u(x) = c\,\sin\frac{\pi}{b-a}(x-a)$$

mit $c \in \mathbb{R}$. Außerdem besitzt

$$-u'' = \frac{\pi^2}{(b-a)^2}(u+1), \qquad u(a) = u(b) = 0$$

*keine* Lösung. Denn die allgemeine Lösung der Differentialgleichung ist

$$u(x) = -1 + c_1 \sin\frac{\pi}{b-a}(x-a) + c_2\cos\frac{\pi}{b-a}(x-a).$$

Die Randbedingungen liefern

$$u(a) = -1 + c_2 = 0, \qquad u(b) = -1 - c_2 = 0,$$

und das ist nicht gleichzeitig erfüllbar. □

### 4.1.4 Aufgaben

1. Sei $p \in C^1[a,b]$ mit $p(x) > 0$ auf $[a,b]$, ferner $q \in C[a,b]$ mit $q(x) \geq 0$ für alle $x \in [a,b]$. Man zeige: Ist $u \in C[a,b] \cap C^2(a,b)$ mit

$$Lu(x) := -[p(x)u'(x)]' + q(x)u(x) \geq 0 \qquad \text{für alle } x \in (a,b)$$

und

$$Ru := \begin{pmatrix} u(a) \\ u(b) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

so ist $u(x) \geq 0$ für alle $x \in [a,b]$.

2. Sei $p \in C^1[a,b]$ mit $p(x) > 0$ auf $[a,b]$, ferner $q \in C[a,b]$ mit $q(x) > 0$ für alle $x \in [a,b]$. Für $u \in C[a,b] \cap C^2(a,b)$ gelten dann die folgenden Implikationen:

   (a) $Lu(x) \leq 0$ in $(a,b) \Longrightarrow u(x) \leq \max(0, u(a), u(b))$,

   (b) $Lu(x) \geq 0$ in $(a,b) \Longrightarrow u(x) \geq \min(0, u(a), u(b))$.

   Hinweis: Im ersten Fall nehme man im Widerspruch zur Behauptung an, $u$ besitze in $(a,b)$ ein positives Maximum. Den zweiten Fall beweise man entsprechend durch Widerspruch.

3. Gegeben seien der Differentialoperator $L\colon C_n^1[a,b] \longrightarrow C_n[a,b]$ und der Randoperator $R\colon C_n[a,b] \longrightarrow \mathbb{R}^n$ durch

$$(Lu)(x) := u'(x) - A(x)u(x), \qquad Ru := Cu(a) + Du(b).$$

   Hierbei ist $A \in C_{n\times n}[a,b]$ und $C, D \in \mathbb{R}^{n\times n}$. Sei $U(\cdot)$ ein Fundamentalsystem zu $Lu = 0$ und $R(U) := CU(a) + DU(b)$. Man zeige:

   (a) Die folgenden drei Aussagen sind äquivalent:

      (i) Die homogene Aufgabe $Lu = 0$, $Ru = 0$ besitzt nur die triviale Lösung $u = 0$.

      (ii) Die Matrix $R(U)$ ist nichtsingulär.

      (iii) Zu vorgegebenen $g \in C_n[a,b]$, $\eta \in \mathbb{R}^n$ besitzt die Randwertaufgabe

$$Lu = g(x), \qquad Ru = \eta$$

      genau eine Lösung.

   (b) Sei $R(U)$ nichtsingulär. Definiert man die sogenannte *Greensche Matrix* $G\colon [a,b] \times [a,b] \longrightarrow \mathbb{R}^{n\times n}$ durch

$$G(x,\xi) := \begin{cases} U(x)[I - R(U)^{-1}DU(b)]U^{-1}(\xi), & a \leq \xi \leq x \leq b, \\ -U(x)R(U)^{-1}DU(b)U^{-1}(\xi), & a \leq x < \xi \leq b, \end{cases}$$

   so hat diese die folgenden Eigenschaften:

      (i) Es gilt die Sprungbedingung $G(x+0,x) - G(x-0,x) = I$ für $x \in [a,b]$.

      (ii) Bei festem $\xi \in [a,b]$ ist $LG(x,\xi) = 0$ für $x \in [a,b] \setminus \{\xi\}$.

      (iii) Für festes $\xi \in (a,b)$ ist $RG(\cdot,\xi) = 0$.

(iv) Bei vorgegebenem $g \in C_n[a,b]$ lässt sich die eindeutige Lösung von $Lu = g(x)$, $Ru = 0$ durch

$$u(x) := \int_a^b G(x,\xi) g(\xi)\, d\xi$$

darstellen.

(v) Durch die Bedingungen (i)–(iii) ist die Greensche Matrix eindeutig festgelegt.

4. Man[3] definiere den Differentialoperator $L$ und den Randoperator $R$ durch

$$Lu(x) := -u''(x) - \frac{1}{4x^2} u(x), \qquad Ru := \begin{pmatrix} u(1) \\ u(2) \end{pmatrix}.$$

Man bestimme die zugehörige Greensche Funktion.

## 4.2 Das Sturm-Liouvillesche Eigenwertproblem

### 4.2.1 Problemstellung, Beispiele

Ein Beispiel einer Eigenwertaufgabe hatten wir in Unterabschnitt 4.1.1 angegeben, als wir über das Knicklastproblem informierten. Ganz kurz gehen wir hierauf noch einmal ein.

**Beispiel:** Es wird eine Säule mit Länge $l$, Elastizitätsmodul $E$ und einem Flächenträgheitsmoment $I$, welches mit wachsendem $x$ nach dem Gesetz

$$I(x) = I_0 e^{-(k/l)x} \qquad (k > 0 \text{ konstant})$$

abnimmt, betrachtet. Die Frage, bei welcher Last $P$ die Säule aus ihrer vertikalen Anfangslage seitlich ausweichen wird, führt zu dem Eigenwertproblem, diejenigen $P$ bzw. $\lambda = P/(EI_0)$ zu bestimmen, für die

$$(*) \qquad -u'' = \lambda e^{(k/l)x} u, \qquad u'(0) = 0, \quad u(l) = 0$$

eine nichttriviale Lösung besitzt. Der[4] Parameter $\lambda$ ist dank seiner *physikalischen* Bedeutung positiv. Aber auch *mathematisch* lässt sich leicht einsehen, dass $(*)$ einzig und allein im Falle $\lambda > 0$ nichttriviale Lösungen haben kann. Ist nämlich $u$ eine solche, so ist dank partieller Integration mit Hilfe der Randbedingungen

$$\begin{aligned} \lambda \int_0^l e^{(k/l)x} u(x)^2\, dx &= -\int_0^l u(x) u''(x)\, dx \\ &= -u(x)u'(x)\Big|_0^l + \int_0^l u'(x)^2\, dx \\ &= \int_0^l u'(x)^2\, dx. \end{aligned}$$

[3] Diese Aufgabe haben wir W. Walter (1993, S. 226) entnommen.

[4] Wir folgen hier fast wörtlich H. Heuser (1989, S. 386). Das gilt auch für das übernächste Beispiel.

Das erste und das letzte Integral in dieser Gleichungskette ist positiv, also muss auch $\lambda$ positiv sein. □

Jetzt wollen wir die hier zu untersuchende Problemstellung angeben. Wieder sei der Differentialoperator $L$ durch

$$Lu := -(p(x)u')' + q(x)u$$

und der Randoperator $R$ durch

$$Ru := \begin{pmatrix} R_1 u \\ R_2 u \end{pmatrix} := \begin{pmatrix} \alpha_1 u(a) + \alpha_2 p(a)u'(a) \\ \beta_1 u(b) + \beta_2 p(b)u'(b) \end{pmatrix}$$

gegeben. Die generellen Voraussetzungen des Unterabschnitts 4.1.1 seien wieder erfüllt, d. h. es seien $p \in C^1[a,b]$ mit $p(x) > 0$ für alle $x \in [a,b]$, $q \in C[a,b]$ und $\alpha_1^2 + \beta_1^2 > 0$, $\beta_1^2 + \beta_2^2 > 0$. Schließlich sei noch eine Funktion $r \in C[a,b]$ mit $r(x) > 0$ für alle $x \in [a,b]$ gegeben. Dann besteht das *Sturm-Liouvillesche Eigenwertproblem* darin, Zahlen $\lambda \in \mathbb{R}$ (Eigenwert) und zugehörige (nichttriviale) Funktionen $u \in C^2[a,b] \setminus \{0\}$ (Eigenfunktion) derart zu bestimmen, dass

$$\text{(EWA)} \qquad Lu(x) = \lambda r(x)u(x) \qquad \text{für alle } x \in [a,b], \qquad Ru = 0.$$

**Beispiel:** Gegeben sei die Eigenwertaufgabe

$$-u'' = \lambda u, \qquad u(0) = u(\pi) = 0.$$

Für $\lambda \leq 0$ existiert keine nichttriviale Lösung. Für $\lambda = \mu^2 > 0$ ist die allgemeine Lösung von $-u'' = \mu^2 u$ durch

$$u(x) = c_1 \sin \mu x + c_2 \cos \mu x$$

gegeben. Die Randbedingung $u(0) = 0$ ergibt $c_2 = 0$. Schließlich ist $u(\pi) = 0$ genau dann, wenn $\mu \in \mathbb{Z}$. Als Eigenwerte erhält man daher $\lambda_k := k^2$ mit zugehörigen Eigenfunktionen $u_k(x) := \sin kx$ (und nichttrivialen Vielfachen). □

**Beispiel:** Sei $\theta(x,t)$ die orts- und zeitabhängige Temperaturverteilung in einem von $x = 0$ bis $x = l$ reichenden dünnen Stab, dessen linkes Ende auf der konstanten Temperatur 0 gehalten wird, während am rechten Ende Wärmeabgabe an ein umgebendes Medium der Temperatur 0 zugelassen sein soll; zur Zeit $t_0 = 0$ habe der Stab an der Stelle $x$ die vorgegebene Temperatur $f(x)$. Die Temperaturverteilung ergibt sich dann als diejenige Lösung der Wärmeleitungsgleichung

$$\frac{\partial \theta}{\partial t} = a^2 \frac{\partial^2 \theta}{\partial x^2},$$

die den Randbedingungen

$$\theta(0,t) = 0, \quad \frac{\partial \theta}{\partial x}(l,t) + \sigma\theta(l,t) = 0 \quad \text{für alle } t \geq 0$$

und der Anfangsbedingung

$$\theta(x,0) = f(x) \quad \text{für } 0 \le x \le l$$

genügt; $a$ und $\sigma$ sind positive Materialkonstanten.

Geht man mit dem Separationsansatz $\theta(x,t) = u(x)v(t)$ in die Wärmeleitungsgleichung ein, so erhält man

$$\frac{\dot{v}(t)}{a^2 v(t)} = \frac{u''(x)}{u(x)} = -\lambda.$$

Links steht eine nur von $t$, rechts davon eine nur von $x$ abhängende Funktion, daher sind beide Ausdrücke konstant gleich einem $-\lambda$. Die Suche nach nichttrivialem $u$ führt daher auf die Sturm-Liouvillesche Eigenwertaufgabe

$$(*) \qquad -u'' = \lambda u, \qquad u(0) = 0, \quad \sigma u(l) + u'(l) = 0.$$

Auch diese Aufgabe kann höchstens für positive $\lambda$ nichttriviale Lösungen haben. Ist nämlich $u$ eine solche, so erhält man

$$\lambda \int_0^l u(x)^2\,dx = \sigma u(l)^2 + \int_0^l u'(x)^2\,dx$$

und hieraus $\lambda > 0$. Die allgemeine Lösung der Differentialgleichung $-u'' = \lambda u$ wird also durch

$$u(x) = A\cos\sqrt{\lambda}x + B\sin\sqrt{\lambda}x$$

gegeben. Wegen $u(0) = 0$ muss $A = 0$, also

$$u(x) = B\sin\sqrt{\lambda}x \qquad (B \neq 0)$$

sein. Mit der zweiten Randbedingung $\sigma u(l) + u'(l) = 0$ ergibt sich daraus

$$\tan\sqrt{\lambda}l = -\frac{\sqrt{\lambda}}{\sigma}.$$

Diese Gleichung für $\lambda$ besitzt abzählbar viele positive Lösungen $\lambda_1 < \lambda_2 < \cdots$, und offenbar gilt $\lim_{k\to\infty} \lambda_k = \infty$. Wir erhalten: Die Eigenwerte von $(*)$ sind die positiven Lösungen $\lambda_k$ von $\tan(\sqrt{\lambda}l) = -\sqrt{\lambda}/\sigma$, die zugehörigen Eigenfunktionen werden durch $\sin\sqrt{\lambda_k}x$, $k = 1, 2, \ldots$, gegeben. □

### 4.2.2 Der Existenzsatz und der Trennungssatz

Unter den oben genannten Bezeichnungen und generellen Voraussetzungen sei die Sturm-Liouvillesche Eigenwertaufgabe

$$\text{(EWA)} \qquad Lu = \lambda r(x)u, \qquad Ru = 0$$

gegeben. Natürlich ist ein (nichttriviales) Vielfaches einer Eigenfunktion wieder eine Eigenfunktion (zum gleichen Eigenwert). Wichtig ist die Bemerkung, dass Eigenwerte des

obigen Sturm-Liouvilleschen Eigenwertproblems notwendig *einfach* in dem Sinne sind, dass es zu einem Eigenwert nicht zwei zugehörige linear unabhängige Eigenfunktionen geben kann. Denn sind $u, v$ zwei Eigenfunktionen zum Eigenwert $\lambda$, also $Lu = \lambda r(x)u$, $Lv = \lambda r(x)v$, so ist

$$0 = uLv - vLu = -u(pv')' + v(pu')' = -[p(uv' - vu')]'.$$

Also ist

$$W := p(uv' - vu') = \text{const.}$$

Wir zeigen, dass diese Konstante verschwindet, woraus $uv' - vu' = 0$ bzw. die lineare Abhängigkeit von $u$ und $v$ folgt. Ist $\alpha_2 = 0$, so ist notwendig $\alpha_1 \neq 0$ und $u(a) = v(a) = 0$, folglich $W(a) = 0 = \text{const}$. Ist dagegen $\alpha_2 \neq 0$, so ist $p(a)u'(a) = \gamma_1 u(a)$, $p(a)v'(a) = \gamma_1 v(a)$ mit $\gamma_1 := -\alpha_1/\alpha_2$. Folglich ist $W(a) = 0$, die Einfachheit der Eigenwerte beim Sturm-Liouvilleschen Eigenwertproblem ist bewiesen.

Unser erstes Ziel in diesem Unterabschnitt ist ein Beweis des folgenden *Existenzsatzes*. Es gibt verschiedene Beweismöglichkeiten für diesen grundlegenden Satz. Wir folgen W. Walter (1993, S. 228 ff.) und schildern eine auf H. Prüfer zurückgehende elementare (aber nicht einfache) Beweismethode.

**Satz 2.1** *Zur Eigenwertaufgabe (EWA) gibt es unendlich viele Eigenwerte $\lambda_k$ mit*

$$\lambda_0 < \lambda_1 < \cdots < \lambda_k < \cdots, \qquad \lim_{k\to\infty} \lambda_k = +\infty.$$

*Die zum Eigenwert $\lambda_k$ gehörende Eigenfunktion $u_k$ hat im offenen Intervall $(a, b)$ genau $k$ Nullstellen.*

**Beweis:** Da der Beweis nicht ganz einfach ist, begleiten wir die Beweisschritte durch ein konkretes Beispiel.

- *Gegeben sei die Eigenwertaufgabe*

$$-u'' = \lambda u, \qquad u(0) = 0, \quad u(1) + u'(1) = 0.$$

*Wir wissen schon, dass diese Aufgabe die k-te positive Wurzel von* $\tan\sqrt{\lambda} = -\sqrt{\lambda}$ *als Eigenwert besitzt und eine zugehörige Eigenfunktion durch* $u_k(x) = \sin\sqrt{\lambda_k}x$ *gegeben ist.*

Die erste Idee besteht darin, die gegebene Differentialgleichung und die linke Randbedingung, also

$$-(p(x)u')' + q(x)u = \lambda r(x)u, \qquad \alpha_1 u(a) + \alpha_2 p(a)u'(a) = 0,$$

als eine Anfangswertaufgabe für ein System von zwei Differentialgleichungen erster Ordnung in der Form

$$\begin{aligned} \xi' &= [q(x) - \lambda r(x)]\eta, & \xi(a) &= \cos\alpha, \\ \eta' &= \frac{1}{p(x)}\xi, & \eta(a) &= \sin\alpha, \end{aligned}$$

zu schreiben, wobei $\alpha \in [0, \pi)$ so bestimmt ist, dass

$$\alpha_1 \sin\alpha + \alpha_2 \cos\alpha = 0.$$

Denn dann ist $u = \eta$ eine nichttriviale Lösung der gegebenen Differentialgleichung, welche der linken Randbedingung genügt.

- *In unserem Beispiel erhalten wir für $(\xi, \eta)$ die Anfangswertaufgabe*

$$\begin{array}{rclrcl} \xi' &=& -\lambda\eta, & \xi(0) &=& 1, \\ \eta' &=& \xi, & \eta(0) &=& 0. \end{array}$$

*Folglich ist (die Abhängigkeit von $\lambda$ unterdrücken wir)*

$$\xi(x) = \begin{cases} \cos(\sqrt{\lambda}x), & \lambda > 0, \\ 1, & \lambda = 0, \\ \cosh(\sqrt{-\lambda}x), & \lambda < 0, \end{cases}$$

$$\eta(x) = \begin{cases} \sin(\sqrt{\lambda}x)/\sqrt{\lambda}, & \lambda > 0, \\ x, & \lambda = 0, \\ -\sinh(\sqrt{-\lambda}x)/\sqrt{-\lambda}, & \lambda < 0. \end{cases}$$

Durch

$$\xi(x) = \rho(x)\cos\phi(x), \qquad \eta(x) = \rho(x)\sin\phi(x)$$

stellen wir (in diesem Zusammenhang spricht man auch von der *Prüfer-Transformation*) die nichttriviale Bahn $(\xi, \eta)$ in Polarkoordinaten dar. Dann ist

$$\rho(x) = \sqrt{\xi^2(x) + \eta^2(x)}, \qquad \phi(x) = \arctan\frac{\eta(x)}{\xi(x)},$$

wobei $\phi(a) = \alpha \in [0, \pi)$ und die mehrdeutige Arcusfunktion bzw. $\phi$ so festgelegt werden kann, dass sie stetig ist. Der Witz bei der obigen Transformation besteht nun darin, dass man *eine* Differentialgleichung erster Ordnung für $\phi$ herleiten und dadurch auf Eigenschaften von $\phi(\cdot) = \phi(\cdot, \lambda)$ schließen kann. Denn es ist

$$\begin{aligned} \phi' &= \frac{1}{1 + (\eta/\xi)^2} \frac{\eta'\xi - \eta\xi'^2}{\xi^2} \\ &= \frac{1}{\xi^2 + \eta^2} \Big( \frac{1}{p(x)} \xi^2 - [q(x) - \lambda r(x)]\eta^2 \Big) \\ &= \frac{1}{p(x)} \cos^2\phi + [\lambda r(x) - q(x)] \sin^2\phi \\ &= \frac{1}{p(x)} + \Big( \lambda r(x) - q(x) - \frac{1}{p(x)} \Big) \sin^2\phi. \end{aligned}$$

Sei also jetzt $\phi(\cdot, \lambda)$ bei gegebenem $\lambda \in \mathbb{R}$ die Lösung von

$$\phi' = \frac{1}{p(x)} + \Big( \lambda r(x) - q(x) - \frac{1}{p(x)} \Big) \sin^2\phi =: f(x, \phi), \qquad \phi(a, \lambda) = \alpha,$$

wobei $\alpha \in [0, \pi)$ wie oben angegeben zu bestimmen ist.

Jetzt bestimme man $\beta \in (0, \pi]$ mit

$$\beta_1 \sin\beta + \beta_2 \cos\beta = 0$$

und überlege sich, dass $u = \eta$ genau dann eine (nichttriviale) Eigenfunktion zum Eigenwert $\lambda$ der Eigenwertaufgabe (EWA) ist, wenn $\phi(b, \lambda) = \beta + k\pi$ mit $k \in \mathbb{Z}$. Denn angenommen, $u$ ist Eigenfunktion zum Eigenwert $\lambda$. Dass die rechte Randbedingung erfüllt ist, bedeutet $\beta_1\eta(b) + \beta_2\xi(b) = 0$. Folglich ist

$$\tan\beta = \frac{\sin\beta}{\cos\beta} = -\frac{\beta_2}{\beta_1} = \frac{\eta(b)}{\xi(b)}$$

und daher

$$\phi(b, \lambda) = \arctan\frac{\eta(b)}{\xi(b)} = \beta + k\pi$$

mit einem gewissen $k \in \mathbb{Z}$. Ist umgekehrt $\phi(b, \lambda) = \beta + k\pi$ mit $k \in \mathbb{Z}$, so genügt $u$ auch der rechten Randbedingung, so dass die Aussage richtig ist.

- *Wir kommen wieder auf das Beispiel zurück. Hier ist* $\beta = \frac{3}{4}\pi$*, weiter ist*

$$\phi(1, \lambda) = \begin{cases} \arctan\Big(\dfrac{\tan\sqrt{\lambda}}{\sqrt{\lambda}}\Big), & \lambda > 0, \\ \arctan 1, & \lambda = 0, \\ \arctan\Big(-\dfrac{\tanh\sqrt{-\lambda}}{\sqrt{-\lambda}}\Big), & \lambda < 0. \end{cases}$$

  *Offensichtlich hat die Gleichung* $\phi(1, \lambda) = \frac{3}{4}\pi + k\pi$ *nur für* $\lambda > 0$ *eine Lösung. Diese Gleichung ist äquivalent zu* $\tan\sqrt{\lambda}/\sqrt{\lambda} = -1$.

Wir beweisen die folgenden Aussagen über $\phi(\cdot, \lambda)$:

(a) Aus $\phi(x_0, \lambda) = k\pi$ mit $k \in \mathbb{Z}$ folgt $\phi'(x_0, \lambda) > 0$. Ferner ist $\phi(x, \lambda) > 0$ für alle $(x, \lambda) \in (a, b] \times \mathbb{R}$.

(b) Es ist

$$\frac{\partial}{\partial\lambda}\phi(x, \lambda) = \phi_\lambda(x, \lambda) > 0 \qquad \text{für alle } (x, \lambda) \in (a, b] \times \mathbb{R},$$

d. h. bei festem $x \in (a, b]$ ist $\phi(x, \cdot)$ monoton wachsend.

(c) Es ist $\lim_{\lambda\to-\infty} \phi(b, \lambda) = 0$.

(d) Es gibt positive Konstanten $\delta, D, \lambda_0$ mit

$$\delta\sqrt{\lambda} \le \phi(b, \lambda) \le D\sqrt{\lambda} \qquad \text{für alle } \lambda \ge \lambda_0.$$

Der erste Teil der Aussage in (a) ist wegen $p > 0$ trivial, der zweite folgt daraus. Denn insbesondere (mit $k = 0$) sagt der erste Teil aus, dass $y = \phi(x, \lambda)$ in der $(x, y)$-Ebene die Gerade $y = 0$ nur einmal schneiden kann, und zwar von unten nach oben. Wegen $\phi(a, \lambda) = \alpha \geq 0$ folgt die Behauptung.

Aus allgemeinen Sätzen folgt, dass $\phi$ stetig differenzierbar von $\lambda$ abhängt. Zur Abkürzung setzen wir $\psi(x) := \phi_\lambda(x, \lambda)$ mit festem $\lambda \in \mathbb{R}$ und erhalten für $\psi$ die lineare Differentialgleichung erster Ordnung

$$\psi' = \psi \underbrace{\Big(\lambda r(x) - q(x) - \frac{1}{p(x)}\Big) 2 \sin\phi(x,\lambda)\cos\phi(x,\lambda)}_{=:l(x)} + r(x)\sin^2\phi(x,\lambda), \qquad \psi(a) = 0.$$

Wegen (a) ist $h(x) := r(x)\sin^2\phi(x, \lambda)$ bis auf endlich viele Stellen positiv auf $[a, b]$. Aus der Darstellung von

$$\psi(x) = \int_a^x e^{L(x)-L(t)}\, h(t)\, dt \qquad \text{mit} \qquad L(x) := \int_a^x l(t)\, dt$$

erhält man $\psi(x) > 0$ für alle $x \in (a, b]$ und das ist die Behauptung (b).

Sei $\epsilon \in (0, \pi - \alpha)$ beliebig. Wir wollen zeigen, dass es ein $\lambda_0 = \lambda_0(\epsilon)$ mit $\phi(b, \lambda) < \epsilon$ für alle $\lambda \leq \lambda_0$ gibt. Man definiere

$$w(x) := \frac{\epsilon}{b-a}(x - a) + \frac{\pi - \epsilon}{b-a}(b - x).$$

Mit $r_0 := \min_{x\in[a,b]} r(x)$ wähle man nun $\lambda_0 < 0$ so klein, daß $\lambda_0 r_0 - q(x) - 1/p(x) < 0$ für alle $x \in [a, b]$ und

$$\frac{1}{p(x)} + \Big(\lambda_0 r_0 - q(x) - \frac{1}{p(x)}\Big)\sin^2\epsilon < \frac{2\epsilon - \pi}{b-a} \qquad \text{für alle } x \in [a, b].$$

Für $\lambda \leq \lambda_0$ und beliebiges $x \in [a, b]$ ist dann

$$\begin{aligned} f(x, w(x)) &= \frac{1}{p(x)} + \Big(\lambda r(x) - q(x) - \frac{1}{p(x)}\Big)\sin^2 w(x) \\ &\leq \frac{1}{p(x)} + \underbrace{\Big(\lambda_0 r_0 - q(x) - \frac{1}{p(x)}\Big)}_{<0}\sin^2 w(x) \\ &\leq \frac{1}{p(x)} + \Big(\lambda_0 r_0 - q(x) - \frac{1}{p(x)}\Big)\sin^2\epsilon \\ &< \frac{2\epsilon - \pi}{b-a} \\ &= w'(x). \end{aligned}$$

Ferner ist

$$\phi(a, \lambda) = \alpha < \pi - \epsilon = w(a).$$

Hieraus wollen wir schließen, dass $\phi(x, \lambda) < w(x)$ für alle $x \in [a, b]$ und alle $\lambda \leq \lambda_0$ ist, woraus insbesondere $\phi(b, \lambda) < \epsilon$ für alle $\lambda \leq \lambda_0$ folgt. Denn angenommen, $w(\cdot) - \phi(\cdot, \lambda)$

hätte für ein festes $\lambda \le \lambda_0$ eine Nullstelle in $[a,b]$. Sei $x_0 \in (a,b]$ die erste Nullstelle, also $w(x) > \phi(x,\lambda)$ für alle $x \in [a,x_0)$ und $w(x_0) = \phi(x_0,\lambda)$. Wir wollen zeigen, dass dann $w'(x_0) \le f(x_0, w(x_0))$, was ein Widerspruch ist. Denn es ist

$$\begin{aligned} w'(x_0) &= \lim_{h\to 0+} \frac{w(x_0) - w(x_0-h)}{h} \\ &\le \lim_{h\to 0+} \frac{\phi(x_0,\lambda) - \phi(x_0-h,\lambda)}{h} \\ &= \phi'(x_0,\lambda) \\ &= f(x_0, \phi(x_0,\lambda)) \\ &= f(x_0, w(x_0)). \end{aligned}$$

Damit ist auch (c) bewiesen.

Da $p$ und $r$ nach Voraussetzung positiv sind, existiert ein $\lambda_0 > 0$ und positive Konstanten $A_0, B_0, A, B$ derart, daß

$$\begin{aligned} A_0 + \lambda B_0 \sin^2\phi(x,\lambda) &\le \frac{1}{p(x)} + \Big(\lambda r(x) - q(x) - \frac{1}{p(x)}\Big)\sin^2\phi(x,\lambda) \\ &\le A + \lambda B \sin^2\phi(x,\lambda) \qquad \text{für alle } (x,\lambda) \in [a,b]\times[\lambda_0,\infty). \end{aligned}$$

Hieraus folgt

$$\frac{\phi'(x,\lambda)}{A+\lambda B\sin^2\phi(x,\lambda)} \le 1 \le \frac{\phi'(x,\lambda)}{A_0+\lambda B_0\sin^2\phi(x,\lambda)} \qquad \text{für alle } (x,\lambda)\in[a,b]\times[\lambda_0,\infty).$$

Eine Integation über $[a,b]$ liefert

$$\int_a^b \frac{\phi'(x,\lambda)}{A+\lambda B\sin^2\phi(x,\lambda)}\,dx \le b-a \le \int_a^b \frac{\phi'(x,\lambda)}{A_0+\lambda B_0\sin^2\phi(x,\lambda)}\,dx$$

für alle $\lambda \ge \lambda_0$. Die Substitution $s = \phi(x,\lambda)$ liefert

$$\int_\alpha^{\phi(b,\lambda)} \frac{ds}{A+\lambda B\sin^2 s} \le b-a \le \int_\alpha^{\phi(b,\lambda)} \frac{ds}{A_0+\lambda B_0\sin^2 s} \qquad \text{für alle } \lambda \ge \lambda_0.$$

Es sei $k \in \mathbb{N}$ mit $k\pi \le \phi(b,\lambda) < (k+1)\pi$. Verkleinert man das links stehende Integral dadurch, dass man nur über $[\pi, k\pi] \subset [\alpha, \phi(b,\lambda)]$ integriert, so erhält man

$$\begin{aligned} b-a &\ge \int_\pi^{k\pi} \frac{ds}{A+\lambda B\sin^2 s} \\ &= (k-1)\int_0^\pi \frac{ds}{A+\lambda B\sin^2 s} \\ &\ge (k-1)\int_0^\pi \frac{ds}{A+\lambda B s^2} \\ &= \frac{k-1}{\sqrt{\lambda}}\int_0^{\sqrt{\lambda}\pi} \frac{dt}{A+Bt^2} \\ &\qquad \text{(Substitution } \sqrt{\lambda}s = t) \\ &\ge \frac{\gamma(k-1)}{\sqrt{\lambda}} \end{aligned}$$

mit einer gewissen Konstanten $\gamma > 0$. Also ist

$$\phi(b,\lambda) - 2\pi \le (k-1)\pi \le \frac{\pi}{\gamma}\sqrt{\lambda}(b-a)$$

bzw. $\phi(b,\lambda) \le D\sqrt{\lambda}$ für alle hinreichend großen $\lambda$, wobei $D > 0$ eine geeignete Konstante ist. Zum Nachweis der zweiten Ungleichung geht man ähnlich vor. Hier vergrößert man die zweite Ungleichung, indem man sogar über $[0,(k+1)\pi]$ integriert. Dann ist also

$$b-a \le (k+1)\int_0^{\pi} \frac{ds}{A_0 + \lambda B_0 \sin^2 s} = 2(k+1)\int_0^{\pi/2} \frac{ds}{A_0 + \lambda B_0 \sin^2 s}.$$

Das letzte Integral schätze man nach oben ab, indem man $\frac{1}{2}s \le \sin s$ auf $[0,\frac{1}{2}\pi]$ ausnutzt. Anschließend substituiere man wieder $\sqrt{\lambda}s = t$ und erhält

$$b-a \le 2(k+1)\int_0^{\pi/2} \frac{ds}{A_0 + \lambda B_0 s^2/4} \le \frac{2(k+1)}{\sqrt{\lambda}}\int_0^{\infty} \frac{dt}{A_0 + B_0 t^2/4} = \frac{C(k+1)}{\sqrt{\lambda}}$$

mit einer Konstanten $C > 0$, woraus die gewünschte Ungleichung $\delta\sqrt{\lambda} \le \phi(b,\lambda)$ mit einer Konstanten $\delta > 0$ für alle hinreichend großen $\lambda$ folgt.

Damit sind schließlich die obigen Behauptungen (a)–(d) vollständig bewiesen.

Wegen (b) ist $\phi(b,\lambda)$ in $\lambda$ auf $\mathbb{R}$ monoton wachsend, wegen (c) und (d) hat $\phi(b,\cdot)$ den Wertebereich $(0,\infty)$. Für $k = 0,1,\ldots$ gibt es daher genau ein $\lambda_k$ mit $\phi(b,\lambda_k) = \beta + k\pi$, während die Gleichung $\phi(b,\lambda) = \beta + k\pi$ für $k = -1,-2,\ldots$ keine Lösung besitzt. Die Zahlen $\lambda_k$ sind die gesuchten Eigenwerte, die Funktionen $u_k(x) = \eta(x,\lambda_k)$ die zugehörigen Eigenfunktionen. Für große $k$ gilt nach (d) eine Aussage über das asymptotische Wachstum der Eigenwerte, dass nämlich positive Konstanten $\delta, D$ mit

$$\delta^2\lambda_k \le (\beta + k\pi)^2 \le D^2\lambda_k$$

bzw. positive Konstanten $c, C$ mit

$$ck^2 \le \lambda_k \le Ck^2 \qquad \text{für alle hinreichend großen } k \in \mathbb{N}$$

existieren. Damit ist der erste Teil des Existenzsatzes bewiesen.

Nach Konstruktion ist $x \in (a,b)$ genau dann eine Nullstelle von $u_k(\cdot) = \eta(\cdot,\lambda_k)$, wenn $\phi(x,\lambda_k) = n\pi$. Nun ist

$$0 \le \phi(a,\lambda_k) = \alpha < \pi \qquad \text{und} \qquad k\pi < \phi(b,\lambda_k) = \beta + k\pi \le (k+1)\pi.$$

Wegen (a) nimmt $\phi(\cdot,\lambda_k)$ auf $(a,b)$ den Wert $n\pi$ für $n = 1,2,\ldots,k$ genau einmal an, für andere $n \in \mathbb{Z}$ nicht. Damit ist der Existenzsatz vollständig bewiesen. □

Über die Nullstellen der $k$-ten Eigenfunktion $u_k$ kann noch etwas mehr ausgesagt werden. Hierzu beweisen wir zunächst den *Trennungssatz von Sturm*. Dabei sagen wir, daß die Nullstellen zweier Funktionen $u, v$ sich gegenseitig trennen, wenn zwischen zwei aufeinander folgenden Nullstellen von $u$ eine Nullstelle von $v$ liegt und umgekehrt.

**Satz 2.2** *Der Differentialoperator $L$ sei durch $Lu := -(p(x)u')' + q(x)u$ mit den üblichen Voraussetzungen an $p, q$ gegeben. Sind dann $u_1, u_2$ zwei linear unabhängige Lösungen von $Lu = 0$, so trennen sich ihre Nullstellen gegenseitig.*

**Beweis:** Seien $x_0, x_1$ zwei aufeinander folgende Nullstellen von $u_1$, also $u_1(x_0) = u_1(x_1) = 0$ und $u_1(x) \neq 0$ für $x \in (x_0, x_1)$. Angenommen, es wäre $u_2(x) \neq 0$ für alle $x \in (x_0, x_1)$. O. B. d. A. können wir annehmen, daß $u_1$ und $u_2$ auf $(x_0, x_1)$ positiv sind. Wir wollen hieraus schließen, daß $u_1$ und $u_2$ notwendig linear abhängig sind, was der gewünschte Widerspruch wäre.

Es ist

$$\begin{aligned} 0 &= u_1 L u_2 - u_2 L u_1 \\ &= -u_1 (p u_2')' + u_2 (p u_1')' \\ &= -[p(u_1 u_2' - u_2 u_1')]' \\ &= -(p\Phi)' \qquad \text{mit} \quad \Phi := u_1 u_2' - u_2 u_1'. \end{aligned}$$

Dann ist

$$\Phi(x_0) = \underbrace{u_1(x_0)}_{=0} u_2'(x_0) - \underbrace{u_2(x_0)}_{\geq 0} \underbrace{u_1'(x_0)}_{\geq 0} \leq 0$$

und ebenso

$$\Phi(x_1) = \underbrace{u_1(x_1)}_{=0} u_2'(x_1) - \underbrace{u_2(x_1)}_{\geq 0} \underbrace{u_1'(x_1)}_{\leq 0} \geq 0.$$

Da andererseits $p\Phi$ konstant ist, ist $p\Phi \equiv 0$ bzw. $\Phi \equiv 0$. Die Wronski-Determinante zu $u_1, u_2$ verschwindet also bzw. $u_1, u_2$ sind linear abhängig, womit der gewünschte Beweis erbracht ist. □

Nun erhalten wir leicht die folgende Aussage über die Nullstellen aufeinander folgender Eigenfunktionen zu dem oben angegebenen Sturm-Liouvilleschen Eigenwertproblem.

**Satz 2.3** *Seien $\lambda_k < \lambda_{k+1}$ zwei aufeinander folgende Eigenwerte zum obigen Sturm-Liouvilleschen Eigenwertproblem und $u_k$ bzw. $u_{k+1}$ zwei zugehörige Eigenfunktionen. Dann liegt zwischen je zwei Nullstellen von $u_k$ eine Nullstelle von $u_{k+1}$.*

**Beweis:** Die Beweisidee ist ganz ähnlich wie die zum letzten Satz. Wir nehmen an, es sei $u_k(x_0) = u_k(x_1) = 0$ und $u_k$ sowie $u_{k+1}$ seien positiv auf $(x_0, x_1)$. Auf $(x_0, x_1)$ ist dann

$$0 < (\lambda_{k+1} - \lambda_k) u_k u_{k+1} = -u_k L u_{k+1} + u_{k+1} L u_k = -[p(u_k u_{k+1}' - u_{k+1} u_k')]'.$$

Mit $\Phi_k := u_k u_{k+1}' - u_{k+1} u_k'$ ist also $p\Phi_k$ auf $(x_0, x_1)$ monoton fallend, insbesondere also $p(x_0)\Phi_k(x_0) > p(x_1)\Phi_k(x_1)$. Andererseits ist

$$\Phi_k(x_0) = \underbrace{u_k(x_0)}_{=0} u_{k+1}'(x_0) - \underbrace{u_{k+1}(x_0)}_{\geq 0} \underbrace{u_k'(x_0)}_{\geq 0} \leq 0,$$

folglich $p(x_0)\Phi_k(x_0) \leq 0$, und entsprechend

$$\Phi_k(x_1) = \underbrace{u_k(x_1)}_{=0} u_{k+1}'(x_1) - \underbrace{u_{k+1}(x_1)}_{\geq 0} \underbrace{u_k'(x_1)}_{\leq 0} \geq 0,$$

und folglich $p(x_1)\Phi_k(x_1) \geq 0$, womit wir den gewünschten Widerspruch hergestellt haben. □

### 4.2.3 Aufgaben

1. Gegeben sei das Eigenwertproblem

$$-u'' = \lambda u, \qquad u(0) = u'(0), \quad u(1) = 0.$$

Man[5] bestimme die Eigenwerte $\lambda_k$ und Eigenfunktionen $u_k$ und zeige, dass

$$\sqrt{\lambda_k} = \frac{1}{2}\pi + k\pi + \beta_k \qquad (k = 0, 1, \ldots) \quad \text{mit} \quad \beta_k \downarrow 0 \quad (k \to \infty).$$

Man skizziere die ersten beiden Eigenfunktionen $u_0$ und $u_1$.

2. Man[6] löse das Eigenwertproblem

$$-(xu')' = \frac{\lambda}{x}u, \qquad u'(1) = 0, \quad u'(e^{2\pi}) = 0.$$

Ist $\lambda = 0$ ein Eigenwert?

3. Sei $p \in C^1(\mathbb{R})$, $q \in C(\mathbb{R})$ und $p(x) > 0$ für alle $x \in \mathbb{R}$. Hiermit definiere man den Differentaloperator $L\colon C^2(\mathbb{R}) \longrightarrow C(\mathbb{R})$ durch

$$(Lu)(x) := -[p(x)u'(x)]' + q(x)u(x).$$

Man zeige: Eine nichttriviale Lösung $u$ von $Lu = 0$ hat nur einfache Nullstellen, und zwar endlich oder abzählbar viele. Im zweiten Fall haben die Nullstellen keinen Häufungspunkt.

4. Man löse das folgende Eigenwertproblem mit *periodischen* Randbedingungen:

$$-u'' = \lambda u, \qquad u(0) = u(1), \quad u'(0) = u'(1).$$

[5] Diese Aufgabe findet man bei W. Walter (1993, S. 235).
[6] Diese Aufgabe findet man bei W. Walter (1993, S. 235).

# Kapitel 5

# Lösungen zu den Aufgaben

## 5.1 Aufgaben zu Kapitel 1

### 5.1.1 Aufgaben zu Abschnitt 1.1

1. Man betrachte eine große Population von $N$ Individuen. Geburten, "natürliche Tode", Ein- und Auswanderungen mögen vernachlässigt werden. Es grassiere eine Krankheit, die sich durch Kontakt zwischen Individuen ausbreitet. Diese Krankheit sei so beschaffen, dass ein Individuum entweder durch sie stirbt oder nach einer Genesung immun gegen sie wurde. Die Population kann dann in drei Klassen eingeteilt werden.

   - In der Klasse $S$ sind die anfälligen (**s**usceptibles) zusammengefasst, also diejenigen, die die Krankheit noch nicht bekommen haben und nicht gegen sie immun sind. Ihre Zahl zur Zeit $t$ sei $S(t)$.
   - In der Klasse $I$ sind die **i**nfizierten enthalten, also diejenigen, die die Krankheit haben und andere anstecken können. Zur Zeit $t$ sei ihre Zahl $I(t)$.
   - Zur Klasse $R$ gehört der Rest (**r**emoved), genauer also diejenigen, die tot, isoliert oder immun sind. $R(t)$ sei die Anzahl der Individuen der Klasse $R$ zur Zeit $t$.

   Die Krankheit genüge der folgenden Gesetzmäßigkeit.

   (a) Die Änderungsrate der anfälligen Population ist proportional zur Anzahl der Kontakte zwischen anfälliger und infizierter Population. Wir nehmen daher an, es sei

   $$S' = -\beta SI$$

   mit einer Konstanten (der sogenannten Infektionsrate) $\beta > 0$.

   (b) Individuen werden aus der Klasse $I$ der Infizierten mit einer Rate entfernt (sie sterben, werden isoliert oder immun), die proportional zu ihrer Anzahl ist. Daher ist

   $$I' = \beta SI - \gamma I, \qquad R' = \gamma I.$$

   Mit $S_0, I_0$ seien die positiven Populationen der Klassen $S$ und $I$ zur Anfangszeit $t = 0$ bezeichnet. Zu dieser Zeit sei noch niemand an der Krankheit gestorben bzw. ihretwegen

isoliert oder immun. Man hat daher die Anfangswertaufgabe

$$\text{(P)} \qquad \begin{array}{rclrcl} S' &=& -\beta SI, & S(0) &=& S_0, \\ I' &=& \beta SI - \gamma I, & I(0) &=& I_0, \\ R' &=& \gamma I, & R(0) &=& 0. \end{array}$$

Dies ist das sogenannte Kermack-McKendrick-Modell für die Ausbreitung ansteckender Krankheiten. Wir gehen davon aus, dass obige Anfangswertaufgabe eine eindeutige Lösung $(S, I, R)$ auf $[0, \infty)$ besitzt. Man zeige (die ersten beiden Aussagen sind anschaulich völlig trivial, müssen aber trotzdem bewiesen werden):

(a) Es sind $I(\cdot)$ und $S(\cdot)$ auf $[0, \infty)$ positiv.

(b) Es ist $S(\cdot)$ auf $[0, \infty)$ monoton fallend. Daher existiert $S_\infty := \lim_{t\to\infty} S(t)$.

(c) Es ist $S(t) + I(t) - (\gamma/\beta) \ln S(t) = \text{const}$ für alle $t$.

(d) Ist $S_0 > \gamma/\beta$, so kommt es zu einer Epidemie in dem Sinne, dass es ein $t > 0$ mit $I(t) > I_0$ gibt. Weiter gibt es ein $t^* > 0$ derart, dass $I(\cdot)$ auf $[0, t^*]$ monoton wachsend und auf $[t^*, \infty)$ monoton fallend ist. Es ist $\lim_{t\to\infty} I(t) = 0$ und $S_\infty$ ist die eindeutige Lösung der transzendenten Gleichung

$$S_0 \exp\left(-\frac{(N-x)\beta}{\gamma}\right) - x = 0.$$

(e) Ist $S_0 < \gamma/\beta$, so ist $I(\cdot)$ auf $[0, \infty)$ monoton fallend und $\lim_{t\to\infty} I(t) = 0$. Es kommt also zu keiner Epidemie und die Krankheit verschwindet letztendlich.

Hinweis: Es kann zweckmäßig sein, zunächst die folgende Aussage zu beweisen:

- Sei $h\colon [0, \infty) \longrightarrow \mathbb{R}$ stetig. Dann besitzt die Anfangswertaufgabe $x' = h(t)x$, $x(0) = x_0$ die eindeutige Lösung

$$x(t) = x_0 \exp\left(\int_0^t h(\tau)\, d\tau\right).$$

**Lösung:** Wir beweisen zunächst die im Hinweis gemachte Aussage. Dass

$$x(t) := x_0 \exp(\int_0^t h(\tau)\, d\tau)$$

eine Lösung von $x' = h(t)x$, $x(0) = x_0$, ist, erkennt man einfach durch Einsetzen. Daher nehmen wir jetzt an, die auf $[0, \infty)$ stetig differenzierbare Funktion $x$ sei eine Lösung der angegebenen Anfangswertaufgabe. Dann ist

$$\frac{d}{dt}\left[\exp\left(-\int_0^t h(\tau)\, d\tau\right) x(t)\right] = \exp\left(-\int_0^t h(\tau)\, d\tau\right)[x'(t) - h(t)x(t)] = 0.$$

Daher ist

$$\exp\left(-\int_0^t h(\tau)\, d\tau\right) x(t) = x_0 \qquad \text{bzw.} \qquad x(t) = x_0 \exp\left(\int_0^t h(\tau)\, d\tau\right).$$

Nun kommen wir zu den eigentlichen Aussagen. $I(\cdot)$ ist Lösung von

$$I' = (\beta S(t) - \gamma)I, \qquad I(0) = I_0.$$

Aus der Aussage im Hinweis erhalten wir, dass

$$I(t) = I_0 \exp\left(\int_0^t (\beta S(\tau) - \gamma)\, d\tau\right) > 0.$$

Entsprechend zeigt man mit Hilfe der ersten Differentialgleichung, dass auch $S$ auf $[0,\infty)$ positiv ist. Wegen $S'(t) = -\beta S(t)I(t) < 0$, ist auch der zweite Teil bewiesen. Weiter ist

$$\begin{aligned}\frac{d}{dt}[S(t) + I(t) - (\gamma/\beta)\ln S(t)] &= S'(t) + I'(t) - (\gamma/\beta)\frac{S'(t)}{S(t)} \\ &= -\beta S(t)I(t) + \beta S(t)I(t) - \gamma I(t) + (\gamma/\beta)I(t) \\ &= 0,\end{aligned}$$

woraus auch die dritte Behauptung folgt.

In der vierten Aussage wird $S_0 > \gamma/\beta$ vorausgesetzt. Daher ist $I'(0) = I_0[\beta S_0 - \gamma] > 0$ und folglich $I(t) > I_0$ für alle hinreichend kleinen $t > 0$. Es ist $0 < S_\infty$, denn die Annahme $S_\infty = 0$ würde wegen der aus der dritten Behauptung folgenden Beziehung

$$(*) \qquad I(t) = \underbrace{I_0 + S_0}_{=N} - S(t) + (\gamma/\beta)\ln\frac{S(t)}{S_0}$$

einen Widerspruch ergeben. Wäre $S(t) > \gamma/\beta$ für alle $t \geq 0$, so wäre $I(t) \geq I_0$ für alle $t \geq 0$. Dann wäre aber $R'(t) \geq \gamma I_0$ und folglich $R(t) \geq \gamma I_0 t$ für alle $t \geq 0$, was wegen $R(t) \leq N$ natürlich nicht sein kann. Folglich existiert genau ein $t^* \in (0,\infty)$ mit $S(t^*) = \gamma/\beta$, ferner ist $S_\infty < \gamma/\beta$. Wegen

$$I'(t) = I(t)[\beta S(t) - \gamma]$$

ist $I(\cdot)$ auf $[0,t^*]$ monoton wachsend und auf $[t^*,\infty)$ monoton fallend. Weiter existiert ein $t^{**} > t^*$ mit $S(t) \leq \frac{1}{2}(S_\infty + \gamma/\beta)$ für alle $t \geq t^{**}$. Für diese $t$ ist

$$\begin{aligned}0 &< I(t) \\ &= I_0 \exp\left[-\beta\int_0^t [\rho - S(\tau)]\, d\tau\right] \\ &= I_0 \exp\left[-\beta\int_0^{t^{**}} [\rho - S(\tau)]\, d\tau\right] \exp\left[-\beta\int_{t^{**}}^t [\rho - S(\tau)]\, d\tau\right. \\ &\leq I_0 \exp\left[-\beta\int_0^{t^{**}} [\rho - S(\tau)]\, d\tau\right] \exp\left[-\frac{\beta}{2}(\rho - S_\infty)(t - t^{**})\right]\end{aligned}$$

und daher $\lim_{t\to\infty} I(t) = 0$. Aus $(*)$ erhält man

$$0 = N - S_\infty + \rho\ln\frac{S_\infty}{S_0}.$$

Hieraus erhält man sehr schnell einen Beweis der letzten Behauptung.

Nun zur fünften Aussage, in der wir annehmen, es sei $S_0 < \gamma/\beta$. Da $S(\cdot)$ monoton fallend ist, ist $S(t) < \gamma/\beta$ für alle $t$. Folglich ist

$$I'(t) = \underbrace{(\beta S(t) - \gamma)}_{<0}\,\underbrace{I(t)}_{>0} < 0,$$

also ist $I(\cdot)$ auf $[0,\infty)$ monoton fallend. Ferner ist

$$\begin{aligned} 0 &< I(t) \\ &= I_0 \exp\left(\int_0^t (\beta S(\tau) - \gamma)\, d\tau\right) \\ &\le I_0 \exp\left(\int_0^t (\beta S_0 - \gamma)\, d\tau\right) \\ &= I_0 \exp(-(\underbrace{\gamma - \beta S_0}_{>0})t) \\ &\to 0, \end{aligned}$$

womit auch die fünfte Behauptung bewiesen ist.

Wir wollen noch das Resultat unserer Bemühungen, Maple auf das angegebene System von drei Differentialgleichungen anzuwenden, mitteilen. Nach

```
eqn:=D(s)(t)=-beta*s(t)*i(t),D(i)(t)=(beta*s(t)-gamma)*i(t),
D(r)(t)=gamma*i(t):
initial:=s(0)=s_0,i(0)=i_0,r(0)=0:
dsolve({eqn,initial},{s(t),i(t),r(t)});
```

(wir verwenden hier kleine Buchstaben s,i,r, weil `I` bei Maple die imaginäre Einheit ist) erhalten wir keine Antwort, Maple findet also keine geschlossene Lösung.

2. Sei $p$ die Lösung der Anfangswertaufgabe für die logistische Differentialgleichung

$$p' = ap - bp^2, \qquad p(t_0) = p_0,$$

wobei $a, b, p_0$ positive Konstanten mit $p_0 < \frac{1}{2}(a/b)$ sind.

   (a) Seien $t_1 < t_2$ mit $t_1 > t_0$ und $t_1 - t_0 = t_2 - t_1$ gegeben. Man zeige, dass $a$ und $b$ eindeutig durch $p_0 = p(t_0), p(t_1), p(t_2)$ bestimmt sind. Dies bedeutet: Legt man das logistische Wachstumsmodell zugrunde und sind die Populationen $p_0, p_1, p_2$ zu äquidistanten Zeiten $t_0, t_1, t_2$ bekannt, so sind hierdurch die Parameter $a, b$ im Modell eindeutig festgelegt.

   (b) Man zeige, dass genau ein $t^* > t_0$ mit $p(t^*) = \frac{1}{2}(a/b)$ existiert und die Darstellung

$$p(t) = \frac{a/b}{1 + e^{-a(t-t^*)}}$$

   gilt.

   (c) Aus

| $k$ | $t_k$ | $p(t_k)$ |
|---|---|---|
| 0 | 1790 | 3 929 000 |
| 1 | 1850 | 23 192 000 |
| 2 | 1910 | 91 972 000 |

   bestimme man $a$ und $b$. Anschließend berechne man $t^*$ mit $p(t^*) = \frac{1}{2}(a/b)$.

Hinweis: Es darf Maple eingesetzt werden.

**Lösung:** Zur Abkürzung setzen wir $p_1 := p(t_1)$ und $p_2 := p(t_2)$. Es ist

$$p(t) = \frac{ap_i}{bp_i + (a - bp_i)\,\exp[-a(t - t_i)]}, \qquad i = 0, 1, 2,$$

da dies die Lösung der logistischen Differentialgleichung mit der Anfangsbedingung $p(t_i) = p_i$, $i = 0, 1, 2$, ist. Aus den Gleichungen

$$p_1[bp_0 + (a - bp_0)\exp(-a(t_1 - t_0))] = ap_0$$

und

$$p_2[bp_1 + (a - bp_1)\exp(-a(t_2 - t_1))] = ap_1$$

erhält man unter Berücksichtigung von $t_1 - t_0 = t_2 - t_1$, dass

$$\frac{p_0(a - bp_1)}{p_1(a - bp_0)} = \frac{p_1(a - bp_2)}{p_2(a - bp_1)}.$$

Hieraus kann man wenigstens in eindeutiger Weise das Verhältnis $\xi = a/b$ berechnen, denn es ist

$$\xi = \frac{p_1[p_1(p_0 + p_2) - 2p_0p_2]}{p_1^2 - p_0p_2}.$$

Da wir insbesondere $p_0 < a/b$ vorausgesetzt haben, ist $p(\cdot)$ auf $(t_0, \infty)$ monoton wachsend mit $\lim_{t\to\infty} p(t) = \xi$. Folglich ist

$$\eta := \left(\frac{p_0}{p_1}\right)\frac{\xi - p_1}{\xi - p_0} \in (0, 1),$$

so dass genau ein $a > 0$ mit

$$\exp(-a(t_1 - t_0)) = \eta$$

existiert, nämlich

$$a = -\frac{\log \eta}{t_1 - t_0}.$$

Die eindeutige Existenz von $t^* > t_0$ mit $p(t^*) = \frac{1}{2}(a/b)$ ist völlig trivial. Auch die behauptete Darstellung von $p(\cdot)$ ist einfach einzusehen, denn es ist

$$\begin{aligned} p(t) &= \frac{ap(t^*)}{bp(t^*) + (a - bp(t^*))e^{-a(t-t^*)}} \\ &= \frac{a\frac{1}{2}(a/b)}{b\frac{1}{2}(a/b) + (a - b\frac{1}{2}(a/b))e^{-a(t-t^*)}} \\ &= \frac{a/b}{1 + e^{-a(t-t^*)}}. \end{aligned}$$

Im letzten Teil der Aufgabe berechnen wir zunächst $\xi = a/b$ aus

$$\xi = \frac{p_1[p_1(p_0 + p_2) - 2p_0p_2]}{p_1^2 - p_0p_2}.$$

Wir erhalten durch

```
p_0:=3929000;p_1:=23192000;p_2:=91972000;
xi:=p_1*(p_1*(p_0+p_2)-2*p_0*p_2)/(p_1^2-p_0*p_2);
```

das Resultat

$$\xi = \frac{8705233252768000}{44127719} = 1.97273583363 \cdot 10^9,$$

letzteres nach Eingabe von `evalf(xi)`. Durch

```
a:=-ln((p_0/p_1)*(xi-p_1)/(xi-p_0))/60;
```

und `a:=evalf(a);` erhalten wir $a = 0.03133953992$, dann entsprechend $b = a/\xi = 0.158863378 \cdot 10^{-9}$. Bei der Berechnung von $t^*$ mit $p(t^*) = \frac{1}{2}(a/b)$ machen wir es uns einfach. Aus

```
p:=t->a*p_0/(b*p_0+(a-b*p_0)*exp(-a*(t-1790)));
fsolve({p(t)=0.5*xi},{t});
```

erhalten wir als Lösung $t^* = 1914.318643$. Natürlich hätte man die Berechnung von $a$ und $b$ mit Maple noch einfacher bekommen können, sozusagen ohne überhaupt ein bisschen nachzudenken:

```
> eqn:={p_1*(b*p_0+(a-b*p_0)*exp(-a*T))=a*p_0,
p_2*(b*p_1+(a-b*p_1)*exp(-a*T))=a*p_1};
```

$$eqn := \{p\_1\,(b\,p\_0 + (a - b\,p\_0)\,e^{(-a\,T)}) = a\,p\_0,$$
$$p\_2\,(b\,p\_1 + (a - b\,p\_1)\,e^{(-a\,T)}) = a\,p\_1\}$$

```
> sol:=solve(eqn,{a,b});
```

$$sol := \{b = b,\ a = 0\},$$
$$\left\{a = -\frac{\ln(-\frac{p\_0\,(p\_1 - p\_2)}{p\_2\,(p\_1 - p\_0)})}{T},\ b = -\frac{\ln(-\frac{p\_0\,(p\_1 - p\_2)}{p\_2\,(p\_1 - p\_0)})\,(-p\_0\,p\_2 + p\_1^2)}{T\,(p\_2\,p\_1 - 2\,p\_0\,p\_2 + p\_1\,p\_0)\,p\_1}\right\}$$

```
> p_0:=3929000: p_1:=23192000: p_2:=91972000: T:=60:
> evalf(sol);
```

$$\{b = b,\ a = 0.\},\ \{a = .03133953992,\ b = .1588633378\,10^{-9}\}$$

3. Gegeben sei die Anfangswertaufgabe

$$\begin{aligned} x' &= 2x - 0.01xy, & x(0) &= 300, \\ y' &= -y + 0.01xy, & y(0) &= 150. \end{aligned}$$

Aus Abbildung 1.2 kann man ablesen, dass die Lösung $(x(\cdot), y(\cdot))$ eine Periode $T \approx 5$ besitzt. Man berechne (wie auch immer) eine verbesserte Näherung.

**Lösung:** Mit Hilfe von

```
restart;
a:=2: b:=0.01: c:=1: d:=0.01: x_0:=300: y_0:=150:
eqn:=diff(x(t),t)=a*x(t)-v*x(t)*y(t),diff(y(t),t)=-c*y(t)+d*x(t)*y(t):
initial:=x(0)=x_0,y(0)=y_0:
```

```
sol:=dsolve({eqn,initial},{x(t),y(t)},type=numeric);
X:=s->subs(sol(s),x(t)):
Y:=s->subs(sol(s),y(t)):
```

stehen die beiden Lösungskomponenten zur Verfügung. Nach einem Schritt des Newton-Verfahrens

```
T:=5.0:
T:=T-(X(T)-300)/(X(T)*(a-b*Y(T)));
```

erhält man $T = 4.999920102$.

4. Das Lotka-Volterra-System

$$\begin{aligned} x' &= ax - bxy, \\ y' &= -cy + dxy \end{aligned}$$

mit positiven Konstanten $a, b, c, d$ besitzt den Gleichgewichtspunkt $(c/d, a/b)$. Man mache die Variablentransformation $u = x - c/d$, $v = y - a/b$ und stelle für $u, v$ ein Differentialgleichungssystem auf. Man löse das durch Weglassen der nichtlinearen Terme entstehende System, indem man nachweist, dass $u$ und $v$ der Differentialgleichung zweiter Ordnung $w'' + acw = 0$ genügen. Für die Anfangswertaufgabe

$$\begin{aligned} x' &= ax - bxy, & x(0) &= x_0, \\ y' &= -cy + dxy, & y(0) &= y_0 \end{aligned}$$

mit $x_0 \approx c/d$ und $y_0 \approx a/b$ berechne man hierdurch eine Näherungslösung.

**Lösung:** Es ist

$$u' = x' = ax - bxy = a(u + c/d) - b(u + c/d)(v + a/b) = -b(c/d)v - buv$$

und entsprechend

$$v' = y' = -cy + dxy = -c(v + a/b) + d(u + c/d)(v + a/b) = d(a/b)u + duv.$$

Lässt man jeweils die nichtlinearen Terme fort, so erhält man das lineare System

$$\begin{aligned} u' &= -b(c/d)v, \\ v' &= d(a/b)u. \end{aligned}$$

Dann ist

$$u'' = -b(c/d)v' = -b(c/d)d(a/b)u = -acu$$

und entsprechend $v'' = -acv$. Die Lösung von

$$\begin{aligned} u' &= -b(c/d)v, & u(0) &= x_0 - c/d, \\ v' &= d(a/b)u, & v(0) &= y_0 - a/b \end{aligned}$$

erhält man daher durch Lösen von

$$u'' + acu = 0, \qquad u(0) = x_0 - c/d, \qquad u'(0) = -b(c/d)(y_0 - a/b)$$

und

$$v'' + acv = 0, \qquad v(0) = y_0 - a/b, \qquad v'(0) = d(a/b)(x_0 - c/d).$$

Dies ergibt

$$u(t) = (x_0 - c/d)\cos(\sqrt{ac}t) - \frac{b\sqrt{c}}{d\sqrt{a}}(y_0 - a/b)\sin(\sqrt{ac}t)$$

und

$$v(t) = (y_0 - a/b)\cos(\sqrt{ac}t) + \frac{d\sqrt{a}}{b\sqrt{c}}(x_0 - c/d)\sin(\sqrt{ac}t).$$

Die sich ergebenden Näherungslösungen sind also

$$x(t) \approx c/d + (x_0 - c/d)\cos(\sqrt{ac}t) - \frac{b\sqrt{c}}{d\sqrt{a}}(y_0 - a/b)\sin(\sqrt{ac}t)$$

und

$$y(t) \approx a/b + (y_0 - a/b)\cos(\sqrt{ac}t) + \frac{d\sqrt{a}}{b\sqrt{c}}(x_0 - c/d)\sin(\sqrt{ac}t).$$

**Beispiel:** Durch

```
restart;
a:=2: b:=0.01: c:=1: d:=0.01: x_0:=300: y_0:=150:
eqn:=diff(x(t),t)=a*x(t)-v*x(t)*y(t),diff(y(t),t)=-c*y(t)+d*x(t)*y(t):
initial:=x(0)=x_0,y(0)=y_0:
sol:=dsolve({eqn,initial},{x(t),y(t)},type=numeric);
X:=s->subs(sol(s),x(t)):
Y:=s->subs(sol(s),y(t)):
plot([evaln(X(t)),evaln(Y(t))],t=0..5);
```

erhält man z.B. den in Abbildung 5.1 links stehenden Plot. Mit Hilfe von

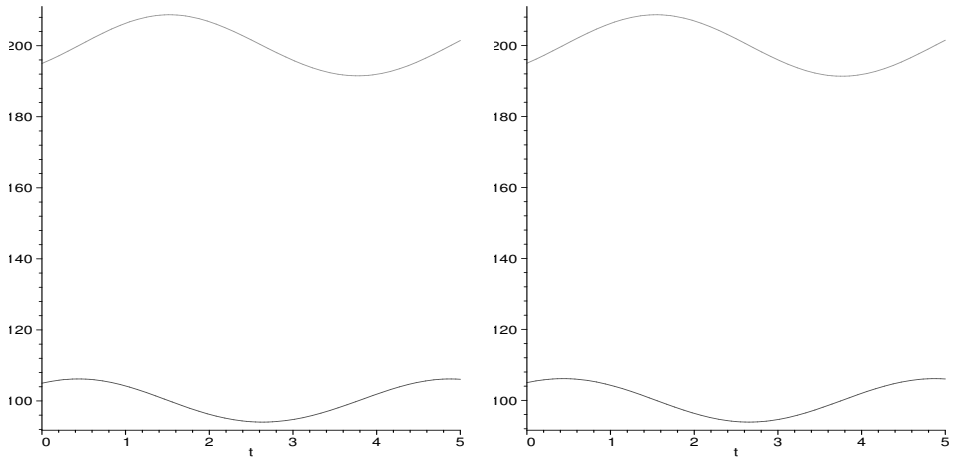


Abbildung 5.1: Exakte Lösung und Näherungslösung

```
> restart;
> a:=2: b:=0.01: c:=1: d:=0.01: x_0:=105: y_0:=195:
> u:=t->(x_0-c/d)*cos(sqrt(a*c)*t)-(b/d)*sqrt(c/a)*(y_0-a/b)*sin(sqrt(a
*c)*t):
```

```
> v:=t->(y_0-a/b)*cos(sqrt(a*c)*t)+(d/b)*sqrt(a/c)*(x_0-c/d)*sin(sqrt(a
*c)*t):
```

und anschließendes `plot([c/d+u(t),a/b+v(t)],t=0..5);` erhalten den in Abbildung 5.1 rechts stehenden Plot, der sich zumindestens qualitativ von dem links stehenden nicht unterscheidet. □

5. Das Wachstumgesetz von B. Gompertz (1779-1865) soll das Wachsen von Tumoren gut beschreiben. Es basiert auf der Anfangswertaufgabe

$$V' = -rV \ln\left(\frac{V}{K}\right), \qquad V(0) = V_0.$$

Hierbei sind $r$ und $K$ gegebene Konstanten, $V(t)$ die Größe des Tumors zur Zeit $t$ und $V_0$ der Anfangszustand. Mit Hilfe von Maple löse man diese Anfangswertaufgabe.

**Lösung:** Nach

```
sol:=dsolve({D(V)(t)=-r*V(t)*ln(V(t)/K),V(0)=V_0},V(t));
simplify(sol);
```

erhalten wir

$$V(t) = \left(\frac{V_0}{K}\right)^{e^{-tr}} K.$$

Hier haben wir einmal bei der Beschreibung der Differentialgleichung `D(V)(t)` statt `diff(V(t),t)` geschrieben.) Merkwürdigerweise gelingt es uns nicht, mit Maple zu verifizieren, dass dies wirklich eine Lösung ist.

6. Wir[1] machen über die Population $p$ einer Spezies mit $p(0) = p_0$ die folgenden Annahmen:

   - Die Population hängt nur vom Vorhandensein eines Grundstoffs $R$ ab.
   - Die Population verbraucht laufend diesen Grundstoff, genauer sei

$$R(t) = R_0 - b \int_0^t p(s)\, ds.$$

   - Die Wachstumsrate $p'(t)$ der Population ist proportional zu $p(t)$ und $R(t)$.

Mit positiven Konstanten $R_0, b, c$ ist also eine Lösung $p$ von

$$p'(t) = cp(t)\left(R_0 - b \int_0^t p(s)\, ds\right), \qquad p(0) = p_0$$

zu bestimmen. Dies ist keine Differentialgleichung, sondern eine Integro-Differentialgleichung für $p$. Man stelle für $z(t) := \int_0^t p(s)\, ds$ eine Anfangswertaufgabe erster Ordnung

[1] Diese und einige weitere Aufgaben haben wir dem Skript "Einführung in die Theorie der Differentialgleichungen" von H. Behncke (Universität Osnabrück) entnommen. Sehr viele Beispiele sind übrigens bei

H. Heuser (1989) *Gewöhnliche Differentialgleichungen*. B. G. Teubner, Stuttgart

enthalten.

auf. Für $p_0 := 1, R_0 := 1, b := 0.002$ und $c := 1$ berechne man (mit Maple) $z$ und $p$, ferner plotte man beide Funktionen über dem Intervall $[0, 10]$.

**Lösung:** Offensichlich genügt $z$ der nichtlinearen Anfangswertaufgabe zweiter Ordnung

$$z'' = cz'(R_0 - bz), \qquad z(0) = 0,\ z'(0) = p_0.$$

Eine Integration über dem Intervall $[0, t]$ liefert, dass $z$ der Anfangswertaufgabe erster Ordnung

$$z' = p_0 + cR_0 z - \frac{cb}{2} z^2, \qquad z(0) = 0$$

genügt. Für die angegebenen Parameter ist die Anfangswertaufgabe

$$z' = 1 + z - 0.01z^2, \qquad z(0) = 0$$

zu lösen. Mit Maple erhält man

```
> restart;
> dsolve({diff(z(t),t)=1+z(t)-0.01*z(t)^2,z(0)=0},z(t));
```

$$z(t) = 50 + 10\sqrt{26}\tanh(\frac{1}{10}\sqrt{26}\,t + \frac{1}{2}\ln(\frac{-5+\sqrt{26}}{5+\sqrt{26}}))$$

```
> z:=unapply(rhs(%),t);p:=D(z);
```

$$z := t \to 50 + 10\sqrt{26}\tanh(\frac{1}{10}\sqrt{26}\,t + \frac{1}{2}\ln(\frac{-5+\sqrt{26}}{5+\sqrt{26}}))$$

$$p := t \to 26 - 26\tanh(\frac{1}{10}\sqrt{26}\,t + \frac{1}{2}\ln(\frac{-5+\sqrt{26}}{5+\sqrt{26}}))^2$$

Beide Funktionen können durch `plot(z(t),t=0..10)` bzw. `plot(p(t),t=0..10)` geplottet werden. Die entsprechenden Plots (wir haben nur noch die Achsen beschriftet) sind in Abbildung 5.2 zu finden. Will man das Ergebnis überprüfen, so kann man fol-

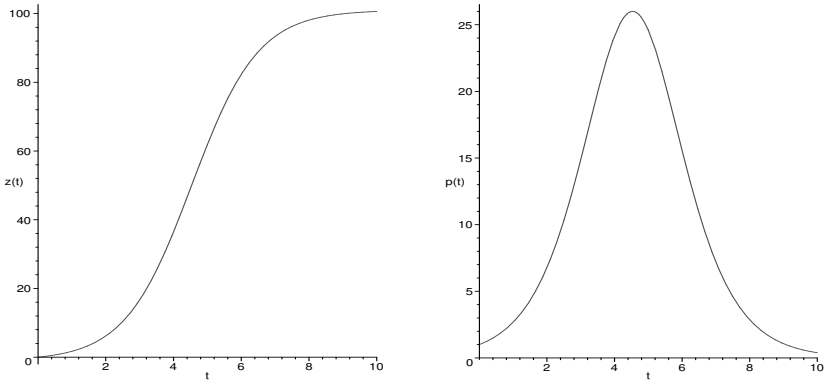


Abbildung 5.2: Die Funktionen $z$ und $p$

gendermaßen vorgehen. Nach

```
eqn:=diff(z(t),t)-1-z(t)+0.01*z(t)^2;
dsolve({eqn,z(0)=0},z(t));
subs(%,eqn);
simplify(%);
```

erhält man $0. = 0$. Ersetzt man in der Differentialgleichung 0.01 durch $\frac{1}{100}$, so ist $0 = 0$ das Resultat.

### 5.1.2 Aufgaben zu Abschnitt 1.2

1. Das vollständige elliptische Integral erster Ordnung ist mit dem Gaußschen arithmetisch-geometrischen Mittel (AGM) verwandt[2]. Insbesondere zeige man:

   (a) Gegeben seien Zahlen $a, b$ mit $0 < b \leq a$. Auf die folgende Weise erzeuge man Folgen $\{a_k\}$, $\{b_k\}$.

   - Setze $a_0 := a$, $b_0 := b$.
   - Für $k = 0, 1, \ldots$:
     - Berechne $a_{k+1} := \frac{1}{2}(a_k + b_k)$.
     - Berechne $b_{k+1} := \sqrt{a_k b_k}$.

   Man zeige: Die Folgen $\{a_k\}$ und $\{b_k\}$ konvergieren monoton nicht wachsend bzw. monoton nicht fallend gegen einen gemeinsamen Grenzwert $M(a, b)$, das sogenannte arithmetisch-geometrische Mittel von $a$ und $b$.

   (b) Für $0 < b \leq a$ und $\lambda > 0$ ist $M(\lambda a, \lambda b) = \lambda M(a, b)$.

   (c) Für $0 < b \leq a$ ist
   $$M(a, b) = M\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

   (d) Für $0 < b \leq 1$ ist
   $$M(1, b) = \frac{1+b}{2} M\left(1, \frac{2\sqrt{b}}{1+b}\right).$$

   (e) Für $0 < x \leq 1$ ist
   $$\frac{1}{M(1, x)} = \frac{2}{\pi} K(\sqrt{1-x^2}),$$
   wobei mit
   $$K(k) := \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}}$$
   das vollständige elliptische Integral erster Art bezeichnet wird.

   (f) Man zeige, dass
   $$\frac{1}{M(\sqrt{2}, 1)} = \frac{2}{\pi} \int_0^1 \frac{dt}{\sqrt{1-t^4}}.$$

   **Lösung:** Wir nehmen an, es sei $0 < b_k \leq a_k$, was für $k = 0$ richtig ist. Wegen der Ungleichung vom geometrisch-arithmetischen Mittel ist
   $$b_k \leq b_{k+1} = \sqrt{a_k b_k} \leq \frac{a_k + b_k}{2} = a_{k+1} \leq a_k.$$

[2] Hierüber kann man sich sehr gut auf den ersten Seiten von

J. M. Borwein, P. B. Borwein (1987) *Pi and the AGM*. J. Wiley, New York

informieren. Zu recht bezeichnen sie das arithmetisch-geometrische Mittel als eine der Juwelen der klassischen Analysis.

Als monotone, nach unten bzw. oben beschränkte Folgen sind $\{a_k\}$ und $\{b_k\}$ konvergent gegen $a_\infty$ bzw. $b_\infty$. Z. B. wegen $a_\infty = \frac{1}{2}(a_\infty + b_\infty)$ ist $a_\infty = b_\infty$. Die Folgen $\{a_k\}$ und $\{b_k\}$ besitzen also denselben Grenzwert.

Mit $\lambda > 0$ und $0 < b \le a$ setze man $a_0 := a$, $b_0 := b$ sowie $a_0' := \lambda a$, $b_0' := \lambda b$. Anschließend definiere man die Folgen $\{a_k\}$, $\{b_k\}$ sowie $\{a_k'\}$, $\{b_k'\}$ durch

$$a_{k+1} := \frac{a_k + b_k}{2}, \qquad b_{k+1} := \sqrt{a_k b_k}$$

bzw.

$$a_{k+1}' := \frac{a_k' + b_k'}{2}, \qquad b_{k+1}' := \sqrt{a_k' b_k'}.$$

Durch vollständige Induktion nach $k$ zeigt man, dass $a_k' = \lambda a_k$, $b_k' = \lambda b_k$. , woraus natürlich $M(\lambda a, \lambda b) = \lambda M(a, b)$ folgt.

Die Gleichung

$$M(a,b) = M\left(\frac{a+b}{2}, \sqrt{ab}\right)$$

für $0 < b \le a$ ist völlig selbstverständlich, denn ob man im obigen Algorithmus bei $a_0, b_0$ oder $a_1, b_1$ beginnt, ist für den gemeinsamen Grenzwert irrelevant.

Für $0 < b \le 1$ ist unter Berücksichtigung der beiden letzten Ergebnisse

$$M(1,b) = M\left(\frac{1+b}{2}, \sqrt{b}\right) = \frac{1+b}{2} M\left(1, \frac{2\sqrt{b}}{1+b}\right).$$

Sei $0 < b \le a$. Man definiere

$$T(a,b) := \frac{2}{\pi} \int_0^{\pi/2} \frac{d\theta}{\sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}}.$$

Mit der Substitution $t := b \tan\theta$ wird

$$\cos^2\theta = \frac{b^2}{b^2+t^2}, \qquad \sin^2\theta = \frac{t^2}{b^2+t^2}, \qquad d\theta = \frac{b}{b^2+t^2}\, dt$$

und folglich

$$T(a,b) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}}.$$

Nun ist

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}} &= \int_{-\infty}^{0} \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}} + \int_0^{\infty} \frac{dt}{\sqrt{(a^2+t^2)(b^2+t^2)}} \\
&= \int_{-\infty}^{\infty} \Bigg\{ \frac{1 - u/\sqrt{ab+u^2}}{\sqrt{[a^2 + (u - \sqrt{ab+u^2})^2][b^2 + (u - \sqrt{ab+u^2})^2]}} \\
&\qquad + \frac{1 + u/\sqrt{ab+u^2}}{\sqrt{[a^2 + (u + \sqrt{ab+u^2})^2][b^2 + (u + \sqrt{ab+u^2})^2]}} \Bigg\} du \\
&\qquad (t = u - \sqrt{ab+u^2} \text{ bzw. } t = u + \sqrt{ab+u^2}) \\
&= \int_{-\infty}^{\infty} \frac{1}{C(u)} \Bigg\{ \frac{C(u) - u}{\sqrt{[a^2 + (C(u)-u)^2][b^2 + (C(u)-u)^2]}} \\
&\qquad + \frac{C(u) + u}{\sqrt{[a^2 + (C(u)+u)^2][b^2 + (C(u)+u)^2]}} \Bigg\} du,
\end{aligned}$$

wobei wir zur Abkürzung

$$C(u) := \sqrt{ab + u^2}$$

gesetzt haben. Nun ist

$$\begin{aligned} & \frac{C(u) - u}{\sqrt{[a^2 + (C(u) - u)^2][b^2 + (C(u) - u)^2]}} + \frac{C(u) + u}{\sqrt{[a^2 + (C(u) + u)^2][b^2 + (C(u) + u)^2]}} \\ = \; & \frac{1}{2\sqrt{[(a+b)/2]^2 + u^2}} + \frac{1}{2\sqrt{[(a+b)/2]^2 + u^2}} \\ = \; & \frac{1}{2\sqrt{[(a+b)/2]^2 + u^2}}, \end{aligned}$$

wenn man den ersten Summanden mit $C(u)+u$ und den zweiten mit $C(u)-u$ erweitert. Insgesamt ist

$$\int_{-\infty}^{\infty} \frac{dt}{\sqrt{(a^2 + t^2)(b^2 + t^2)}} = \int_{-\infty}^{\infty} \frac{du}{\sqrt{\{[(a+b)/2]^2 + u^2\}\{ab + u^2\}}}$$

und daher

$$T(a, b) = T\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

Eine Fortsetzung liefert

$$T(a, b) = T(M(a, b), M(a, b)) = \frac{1}{M(a, b)}.$$

Für $0 < x \leq 1$ ist insbesondere

$$\begin{aligned} \frac{1}{M(1, x)} &= T(1, x) \\ &= \frac{2}{\pi} \int_0^{\pi/2} \frac{d\theta}{\sqrt{\cos^2 \theta + x^2 \sin^2 \theta}} \\ &= \frac{2}{\pi} \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - (1 - x^2) \sin^2 \theta}} \\ &= \frac{2}{\pi} K(\sqrt{1 - x^2}), \end{aligned}$$

was zu zeigen war.

Es ist

$$\begin{aligned} \frac{1}{M(\sqrt{2}, 1)} &= \frac{1}{\sqrt{2} M(1, 1/\sqrt{2})} \\ &= \frac{\sqrt{2}}{\pi} K(1/\sqrt{2}) \\ &\qquad \text{(wegen des gerade eben bewiesenen Teils)} \\ &= \frac{\sqrt{2}}{\pi} \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - \frac{1}{2} \sin^2 \theta}} \\ &= \frac{\sqrt{2}}{\pi} \int_0^{1} \frac{dt}{\sqrt{(1 - t^2)(1 - \frac{1}{2} t^2)}} \end{aligned}$$

$$
\begin{aligned}
& \qquad\qquad \text{(Substitution } t = \sin\theta\text{)} \\
&= \frac{2}{\pi}\int_0^1 \frac{dt}{\sqrt{(1-t^2)(2-t^2)}} \\
&= \frac{2}{\pi}\int_0^1 \frac{dx}{\sqrt{1-x^4}} \\
& \qquad\qquad \text{(Substitution } x^2 = t^2/(2-t^2)\text{)},
\end{aligned}
$$

womit die Behauptung bewiesen ist.

2. Man schreibe ein Maple-Programm, mit dem k Schritte des Gauß-Verfahrens zur Berechnung von $M(a,b)$ durchgeführt werden. Insbesondere berechne man $M(\sqrt{2},1)$ und prüfe numerisch die von Gauß gefundene Identität

$$M(\sqrt{2},1)\int_0^1 \frac{dt}{\sqrt{1-t^4}} = \frac{\pi}{2}$$

nach.

Lösung: Wir schreiben die folgende Prozedur, bei welcher Zwischenergebnisse mit 30 Dezimalen in ein file `agm` geschrieben werden:

```
> agm:=proc(a,b,k) local i,alpha,beta,c;
> alpha:=a;beta:=b;fprintf("agm", "%.30f %.30f %d\n", alpha, beta,
> 0);
> for i from 1 to k do
> c:=0.5*(alpha+beta);
> beta:=sqrt(alpha*beta);
> alpha:=c;
> fprintf("agm", "%.30f %.30f %d\n", alpha, beta, i);
> end do;
> fclose("agm");
> end proc;
```

```
agm := proc(a, b, k)
local i, α, β, c;
    α := a;
    β := b;
    fprintf("agm", "%.30f  %.30f %d\n", α, β, 0);
    for i to k do
        c := .5 * α + .5 * β;
        β := sqrt(α * β);
        α := c;
        fprintf("agm", "%.30f  %.30f %d\n", α, β, i)
    end do;
    fclose("agm")
end proc
```

```
> Digits:=30;
```

$$Digits := 30$$

```
> agm(evalf(sqrt(2)),1,5);
```

Hätte man hier `fprintf` durch `print` ersetzt und die beiden ersten String-Parameter jeweils weggelassen, so hätte man einen Output am Bildschirm erhalten. Wichtig ist beim Schreiben der Prozedur, dass die Input-Parameter nicht verändert werden. Diese werden daher sogleich lokalen Variablen übergeben. Als Ergebnis erhalten wir:

| $a_k$ | $b_k$ | $k$ |
|---|---|---|
| 1.41421356237309504880168872421 0 | 1.00000000000000000000000000000 | 0 |
| 1.20710678118654752440084436210 0 | 1.18920711500272106671749997056 0 | 1 |
| 1.19815694809463429555917216633 0 | 1.19812352149312012260658557182 0 | 2 |
| 1.19814023479387720908287886908 0 | 1.19814023467730720579838378819 0 | 3 |
| 1.19814023473559220744063132864 0 | 1.19814023473559220743921365593 0 | 4 |
| 1.19814023473559220743992249228 0 | 1.19814023473559220743992249228 0 | 5 |

Zur Verifikation der Gauß'schen Identität benutzen wir:

```
> agm:=proc(a,b,k) local i,alpha,beta,c;
> alpha:=evalf(a);beta:=evalf(b);
> for i from 1 to k do
> c:=0.5*(alpha+beta);
> beta:=sqrt(alpha*beta);
> alpha:=c;
> end do;
> c;
> end proc;
```

$$\begin{array}{l} agm := \mathbf{proc}(a,\, b,\, k) \\ \mathbf{local}\, i,\, \alpha,\, \beta,\, c; \\ \quad \alpha := \mathrm{evalf}(a)\,; \\ \quad \beta := \mathrm{evalf}(b)\,; \\ \quad \mathbf{for}\, i\, \mathbf{to}\, k\, \mathbf{do}\, c := .5 * \alpha + .5 * \beta\,;\ \beta := \mathrm{sqrt}(\alpha * \beta)\,;\ \alpha := c\, \mathbf{end\ do}\,; \\ \quad c \\ \mathbf{end\ proc} \end{array}$$

```
> Digits:=30:
> gauss:=agm(sqrt(2),1,5):
> evalf(int(1/sqrt(1-t^4),t=0..1)*gauss-Pi/2);
```

$$0.$$

3. Für einige Jahre[3] entledigte man sich in den USA eines Teils des radioaktiven Mülls, indem dieser in Fässer kam, die in die See geworfen wurden. Es wurde davon ausgegangen (nach hoffentlich sorgfältigen Tests), dass die Fässer so dicht sind, dass eine Lagerung unbedenklich ist. Es stellte sich aber die Frage, ob eine zu hohe Aufprallgeschwindigkeit

[3]Siehe M. Braun (1975, S. 68 ff.).

zu einem Leck führen könnte. Nach Tests ergab sich, dass die Fässer ab einer Aufprallgeschwindigkeit von 12.2 m/sec platzen konnten, so dass die Aufgabe darin besteht, die Aufprallgeschwindigkeit zu ermitteln.

Ein Fass wiege $m := 240\,\mathrm{kg}$, das Volumen sei $V := 0.21\,\mathrm{m}^3$. Der Wasserwiderstand $D$ sei proportional zur Geschwindigkeit $v$ des Fasses: $D = cv$, wobei durch Experimente $c = 0.12\,\mathrm{kg}\cdot\mathrm{sec/m}$ festgestellt wurde. Durch den Auftrieb erleidet das Fass einen Gewichtsverlust $B$, der gleich dem Gewicht des verdrängten Salzwassers ist (Prinzip des Archimedes). Daher ist $B$ das Produkt aus Volumen $V = 0.21\,\mathrm{m}^3$ des Fasses und der Dichte $1025\,\mathrm{kg/m}^3$ von Salzwasser, also ist $B = 215.25\,\mathrm{kg}$. Bezeichnet man mit $x(t)$ die Tiefe des Fasses zur Zeit $t$ ($x = 0$ sei die Meeresoberfläche), so lautet die Newtonsche Bewegungsgleichung daher

$$m\ddot{x} = g(m - B - cv) = g(m - B - c\dot{x}),$$

wobei $g = 9.81\,\mathrm{m/sec}^2$. Ferner sind die Anfangsbedingungen

$$x(0) = 0, \qquad \dot{x}(0) = 0$$

gegeben.

Bei welcher Wassertiefe übersteigt die Geschwindigkeit $v$ die kritische Aufprallgeschwindigkeit von 12.2 m/sec?

**Lösung:** Durch die folgenden Maple-Befehle erhalten wir die Lösung:

```
> sol:=dsolve(
> {240*diff(x(t),t,t)=9.81*(240-215.25-0.12*diff(x(t),t)),x(0)=0,D(x)(0
> )=0},x(t)):
> x:=unapply(rhs(sol),t):
> v:=D(x):
> t_A:=solve(v(t)=12.2,t);
```

$$t\_A := 12.43081904$$

```
> x_A:=x(t_A);
```

$$x\_A := 76.59854$$

Die kritische Wassertiefe ist also etwa 76.6 m.

4. Bei W. Walter (1996, S. 5)[4] findet man ein System von zwei Differentialgleichungen zweiter Ordnung, durch das die Bewegung eines Satelliten im Gravitationsfeld zweier Körper (z. B. Erde und Mond) modelliert wird, nämlich

$$\begin{aligned}
\ddot{x} &= x + 2\dot{y} - \mu'\frac{x+\mu}{[(x+\mu)^2+y^2]^{3/2}} - \mu\frac{x-\mu'}{[(x-\mu')^2+y^2]^{3/2}},\\
\ddot{y} &= y - 2\dot{x} - \mu'\frac{y}{[(x+\mu)^2+y^2]^{3/2}} - \mu\frac{y}{[(x-\mu')^2+y^2]^{3/2}}.
\end{aligned}$$

Hierbei ist $\mu$ eine gegebene Konstante und $\mu' := 1 - \mu$. Für $\mu := 0.01213$ und die Anfangsbedingungen

$$x(0) = 1.2, \quad \dot{x}(0) = 0, \quad y(0) = 0, \quad \dot{y}(0) = -1.04936$$

[4] W. Walter (1996) *Gewöhnliche Differentialgleichungen. 6. Auflage*. Springer-Verlag, Berlin-Heidelberg-New York.

plotte man die Bahn $\{(x(t), y(t)) : 0 \leq t \leq 10\}$.

**Lösung:** Wir benutzen die folgenden Maple-Befehle:

```
mu:=0.01213: must:=1-mu:
eqn1:=diff(x(t),t,t)=x(t)+2*diff(y(t),t)
    -must*(x(t)+mu)/((x(t)+mu)^2+y(t)^2)^(3/2)
    -mu*(x(t)-must)/((x(t)-must)^2+y(t)^2)^(3/2):
eqn2:=diff(y(t),t,t)=y(t)-2*diff(x(t),t)-must*y(t)/((x(t)+mu)^2+y(t)^2)^(3/2)
    -mu*y(t)/((x(t)-must)^2+y(t)^2)^(3/2):
initial1:=x(0)=1.2,D(x)(0)=0:
initial2:=y(0)=0,D(y)(0)=-1.04936;
sol:=dsolve({eqn1,eqn2,initial1,initial2},{x(t),y(t)},type=numeric):
with(plots):
odeplot(sol,[x(t),y(t)],0..10,numpoints=300,labels=["x","y"],
      title="Bahn eines Satelliten");
```

Das Resultat ist in Abbildung 5.3 angegeben. Erstaunlicherweise erhält man eine peri-

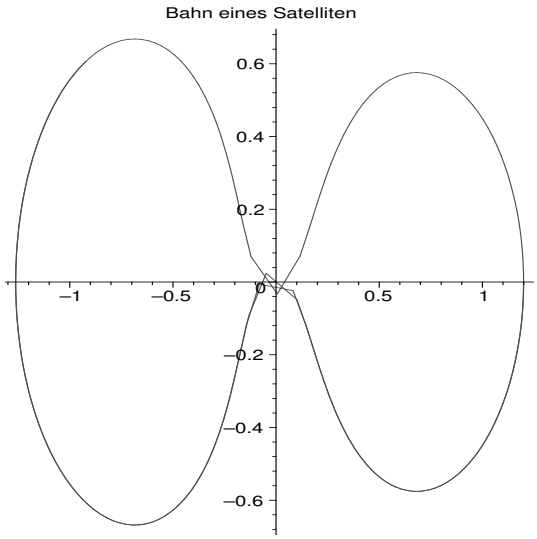


Abbildung 5.3: Bewegung im Gravitationsfeld zweier Körper

odische Bahn.

5. Mit $p > 0$, $\epsilon \in [0, 1)$ und $\alpha \in [0, 2\pi)$ sei die Ellipse $K_\alpha$ durch

$$K_\alpha := \left\{ \frac{p}{1 + \epsilon \cos(\phi - \alpha)} (\cos\phi, \sin\phi) : \phi \in [0, 2\pi] \right\}$$

gegeben. Man zeige:

   (a) Dreht man die Ellipse $K_\alpha$ um den Winkel $\alpha$ im Uhrzeigersinn, so erhält man $K_0$.

   (b) Die Ellipse $K_0$ (und damit auch $K_\alpha$) hat die Halbachsen

$$a = \frac{p}{1 - \epsilon^2}, \qquad b = \frac{p}{\sqrt{1 - \epsilon^2}}$$

   und die Brennpunkte $(0, 0)$ und $(-2p\epsilon/(1 - \epsilon^2), 0)$.

(c) Die Ellipse $K_0$ lässt sich in der Form

$$K_0 = \left\{(x,y) : \frac{(x-x_0)^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$$

darstellen, wobei

$$x_0 := -\frac{p\epsilon}{1-\epsilon^2}.$$

**Lösung:** Die Multiplikation eines Vektors im $\mathbb{R}^2$ mit der Matrix

$$G := \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

dreht diesen Vektor um den Winkel $\alpha$ im Uhrzeigersinn. Wegen

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} = \begin{pmatrix} \cos(\phi-\alpha) \\ \sin(\phi-\alpha) \end{pmatrix}$$

ist $G(K_\alpha) = K_0$, was im ersten Teil zu zeigen war.

Es ist

$$a = \frac{p}{2}\left(\frac{1}{1+\epsilon} + \frac{1}{1-\epsilon}\right) = \frac{p}{1-\epsilon^2}$$

und entsprechend (die Funktion $h(\phi) := \sin\phi/(1+\epsilon\cos\phi)$ nimmt ihre Extrema für $\cos\phi + \epsilon = 0$ an)

$$b = \frac{p}{2}\left(\frac{\sqrt{1-\epsilon^2}}{1-\epsilon^2} + \frac{\sqrt{1-\epsilon^2}}{1-\epsilon^2}\right) = \frac{p}{\sqrt{1-\epsilon^2}}.$$

Offenbar ist

$$(x_0, y_0) := \left(-\frac{p\epsilon}{1-\epsilon^2}, 0\right)$$

der Mittelpunkt der Ellipse, ihre Brennpunkte sind wie behauptet $(x_0 \pm \sqrt{a^2-b^2}, 0)$.

Der letzte Teil der Aufgabe ist jetzt völlig klar.

6. Wir betrachten den Fall eines Körpers der Masse $m$, der unter dem Einfluss der Schwerkraft sich senkrecht nach unten bewegt, wobei der zur Geschwindigkeit proportionale Luftwiderstand berücksichtigt werde. Mit einer Konstanten $\rho > 0$ und (konstantem) $g = 9.81\,\mathrm{m/sec^2}$ hat man die Anfangswertaufgabe

$$m\ddot{x} = mg - \rho\dot{x}, \qquad x(0) = 0, \quad \dot{x}(0) = v_0$$

zu lösen. Man zeige, dass $\lim_{t\to\infty} \dot{x}(t)$ existiert, der Körper also eine endliche Endgeschwindigkeit erreicht[5].

**Lösung:** Die Geschwindigkeit $v := \dot{x}$ ist Lösung der Anfangswertaufgabe

$$m\dot{v} = mg - \rho v, \qquad v(0) = v_0.$$

Nach

```
> dsolve({m*diff(v(t),t)=m*g-rho*v(t),v(0)=v_0},v(t));
```

[5]Bei H. Heuser (1989, S. 30) findet man hierzu die Bemerkung: Von dieser Tatsache profitiert der Fallschirmspringer immer dann, wenn sein Schirm überhaupt aufgeht.

$$\mathrm{v}(t) = \frac{g\,m}{\rho} - \frac{e^{(-\frac{\rho\,t}{m})}\,(m\,g - v\_0\,\rho)}{\rho}$$

erkennt man, dass

$$\lim_{t\to\infty} v(t) = \frac{gm}{\rho}.$$

### 5.1.3 Aufgaben zu Abschnitt 1.3

1. Man löse[6], "per hand" und mit Maple, die Anfangswertaufgabe

$$x' - x + e^{-2t}x^2 = 0, \qquad x(0) = 1.$$

Wo ist die Lösung erklärt?

**Lösung:** Es handelt sich hier um eine Bernoullische Differentialgleichung. Es ist $x(t) = z(t)^{-1}$, wobei $z$ die Lösung der Anfangswertaufgabe

$$z' + z - e^{-2t} = 0, \qquad z(0) = 1,$$

ist. Hieraus erhält man $z(t) = e^{-t}(2 - e^{-t})$ und dann

$$x(t) = \frac{e^{2t}}{-1 + 2e^t}.$$

Diese Lösung ist offenbar auf $(-\ln 2, \infty)$ erklärt. Mit Maple erhält man natürlich dieselbe Lösung.

2. Man bestimme, "per hand" und mit Maple, die allgemeine Lösung der Differentialgleichungen

   (a) $x' = e^x \sin t$,

   (b) $x' = (t - x + 3)^2$,

   (c) $(1 + t^2)x' + tx = t\sqrt{1 + t^2}$.

**Lösung:** Die Differentialgleichung $x' = e^x \sin t$ ist eine Differentialgleichung mit getrennten Veränderlichen. Über $-e^{-x(t)} = -\cos t + C$ erhält man $x(t) = -\ln(\cos t - C)$ als Lösung. Praktisch dieselbe Form, nämlich

$$x(t) = \ln\left(-\frac{1}{-\cos t + C}\right),$$

wird von Maple ausgegeben.

Die allgemeine Lösung der Differentialgleichung $x' = (t - x + 3)^2$ erhält man aus $x(t) = t + 3 - y(t)$, wobei $y$ die allgemeine Lösung von $y' = 1 - y^2$ ist. Dies ist eine Differentialgleichung mit getrennten Veränderlichen und man erhält $\operatorname{arctanh} y(t) = t + C$ bzw. $y(t) = \tanh(t + C)$ als allgemeine Lösung. Daher ist $x(t) = t + 3 - \tanh(t + C)$ die allgemeine Lösung der gegebenen Differentialgleichung. Nach

```
dsolve(diff(x(t),t)=(t-x(t)+3)^2,x(t));
```

[6] Diese Aufgabe wurde in der Staatsexamensklausur September 2001 gestellt.

erhält man

$$\begin{aligned} x(t) &= \frac{-4+te^{2t}C-t+2e^{2t}C}{-1+e^{2t}C} \\ &= t+\frac{e^{2t}C-4}{-1+e^{2t}C} \\ &= t+3-\frac{e^{2t}C+1}{e^{2t}C-1} \\ &= t+3-\tanh(t+D) \end{aligned}$$

mit $\tanh D = (C+1)/(C-1)$, also im Prinzip dieselbe Lösung.

Die Differentialgleichung $(1+t^2)x' + tx = t\sqrt{1+t^2}$ geht nach Division mit $1+t^2$ in (die lineare Differentialgleichung)

$$x' + \frac{t}{1+t^2}x = \frac{t}{\sqrt{1+t^2}}$$

über. Es ist

$$\exp\left(-\int_0^t \frac{\tau}{1+\tau^2}\,d\tau\right) = \exp\left(-\frac{1}{2}\ln(1+t^2)\right) = \frac{1}{\sqrt{1+t^2}}$$

und folglich

$$x(t) = \frac{c}{\sqrt{1+t^2}} + \frac{1}{\sqrt{1+t^2}}\int_0^t \sqrt{1+\tau^2}\frac{\tau}{\sqrt{1+\tau^2}}\,d\tau = \frac{c}{\sqrt{1+t^2}} + \frac{t^2}{2\sqrt{1+t^2}}$$

die gesuchte allgemeine Lösung. Dieselbe Lösung erhält man sofort nach

```
dsolve((1+t^2)*diff(x(t),t)+t*x(t)=t*sqrt(1+t^2),x(t));
```

3. Man gebe für die Differentialgleichung

$$x' = x^2 + 1 - t^2$$

das Richtungsfeld an und plotte Isoklinen. Ferner bestimme man sämtliche Lösungen (eine Lösung ist aus dem Richtungsfeld ersichtlich).

**Lösung:** Nach

```
with(DEtools):
dfieldplot(diff(x(t),t)=x(t)^2+1-t^2,x(t),t=-1..1,x=-1..1,
         arrows=medium,title="Richtungsfeld",scene=[t,x]);
```

erhält man den Plot in Abbildung 5.4 links. Die Isoklinen haben wir in Abbildung 5.4 rechts dargestellt. Sie sind durch

```
with(plots):
contourplot(x^2+1-t^2,t=-1..1,x=-1..1,color=black);
```

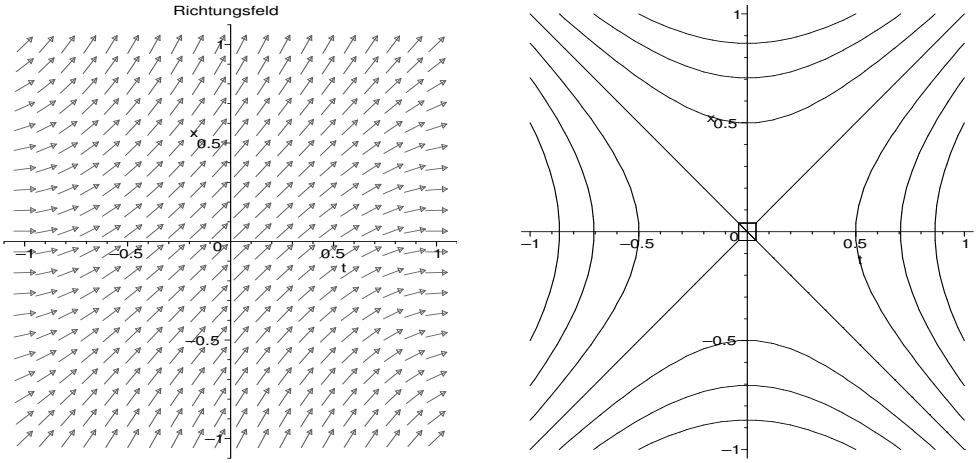


Abbildung 5.4: Richtungsfeld und Isoklinen zu $x' = x^2 + 1 - t^2$

hergestellt. Offensichtlich ist $\phi(t) := t$ eine spezielle Lösung. Die allgemeine Lösung hat die Form $x = \phi + z$, wobei $z$ die allgemeine Lösung der Bernoullischen Differentialgleichung $z' - 2tz - z^2 = 0$ ist. Mittels der Transformation $y = 1/z$ erhält man die lineare Differentialgleichung $y' + 2ty + 1 = 0$ mit der allgemeinen Lösung

$$y(t) = e^{-t^2}\left(C + \int_0^t e^{-\tau^2}\,d\tau\right).$$

Die allgemeine Lösung von $x' = x^2 + 1 - t^2$ ist daher

$$x(t) = t + \frac{e^{t^2}}{C + \int_0^t e^{-\tau^2}\,d\tau}.$$

Dasselbe Ergebnis erhält man auch mit Maple.

4. Man bestimme, "per hand" und mit Maple, die Lösung der Anfangswertaufgaben

   (a) $x' = t^2\sqrt{1-x^2}$, $x(1) = 0$,

   (b) $x' = t^2/(e^x + \cos x)$, $x(-1) = 0$,

   (c) $x' = (2/t)x + t^4$, $x(1) = -6$,

   (d) $x' = 2(x/t)^3 + x/t$, $x(1) = 2$.

   **Lösung:** Die Differentialgleichung $x' = t^2\sqrt{1-x^2}$ ist eine Differentialgleichung mit getrennten Veränderlichen. Wir erhalten $\arcsin x(t) = t^3/3 + c$ mit konstantem $c$ bzw. $x(t) = \sin(t^3/3 + c)$ als allgemeine Lösung. Die Anfangsbedingung $x(1) = 0$ liefert $c = -1/3$ und damit $x(t) = \sin(t^3/3 - 1/3)$ als gesuchte Lösung. Diese findet natürlich auch Maple.

   Die Differentialgleichung $x' = t^2/(e^x + \cos x)$ ist eine Differentialgleichung mit getrennten Veränderlichen. Mit einer Konstanten $c$ erhalten wir $e^{x(t)} + \sin x(t) = \frac{1}{3}t^3 + c$, die Anfangsbedingung $x(-1) = 0$ liefert $c = \frac{4}{3}$, so dass $x(\cdot)$ implizit durch $e^{x(t)} + \sin x(t) = \frac{4}{3} + \frac{1}{3}t^3$ gegeben. Um sich über den Verlauf von $x(\cdot)$ zu informieren, gibt es in Maple verschiedene Möglichkeiten. Nach

```
with(plots);implicitplot(exp(x)+sin(x)=4/3+t^3/3,t=-1..1,x=-2..2);
```

erhält man z. B. den in Abbildung 5.5 links stehenden Plot. Man kann aber auch mittels

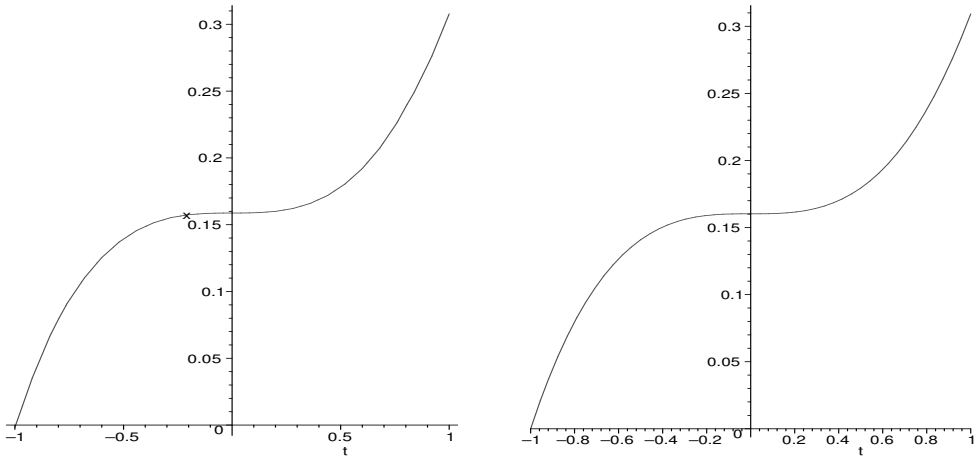


Abbildung 5.5: Lösung von $x' = t^2/(e^x + cosx)$, $x(-1) = 0$

```
x:=t->fsolve(exp(y)+sin(y)=4/3+t^3/3,y,y=-2..2);
```

eine Funktion $x(\cdot)$ definieren, die mit Hilfe von

```
plot(evaln(x(t)),t=-1..1);
```

geplottet werden kann. Das Ergebnis findet man in Abbildung 5.5 rechts, wobei man natürlich keinen wesentlichen Unterschied zum links stehenden Counterpart feststellen wird. Der Vorteil ist jetzt, dass auch numerische Werte von $x(t)$ für spezielle $t$ berechnet werden können.

Die Differentialgleichung $x' = (2/t)x + t^4$ ist eine lineare Differentialgleichung, ihre allgemeine Lösung ist $x(t) = ct^2 + t^5/3$. Die Anfangsbedingung liefert $c = -19/3$, so dass $x(t) = -\frac{19}{3}t^2 + \frac{1}{3}t^5$ die gesuchte Lösung ist.

Die Differentialgleichung $x' = 2(x/t)^3 + x/t$ ist homogen. Die gesuchte Lösung ist $x(t) = ty(t)$, wobei $y$ die Lösung von $y' = 2y^3/t$, $y(1) = 2$ ist. Dies ist eine Anfangsaufgabe für eine Differentialgleichung mit getrennten Veränderlichen. Aus

$$\frac{d}{dt}\left(-\frac{1}{4y(t)^2}\right) = \frac{1}{t}$$

erhält man durch Integration über $[1, t]$ unter Berücksichtigung von $y(1) = 2$, dass

$$-\frac{1}{4y(t)^2} + \frac{1}{16} = \ln t,$$

woraus $y(t) = (\frac{1}{4} - 4\ln t)^{-1/2}$ folgt. Daher ist $x(t) = t(\frac{1}{4} - 4\ln t)^{-1/2}$ die gesuchte Lösung, die natürlich auch Maple findet.

5. Sei $F$ eine auf einem Gebiet $D$ der $(t,x)$-Ebene definierte reellwertige Funktion, die dort stetig partiell differenzierbar sei. Ist $x(\cdot)$ eine auf einem Intervall $I$ stetig differenzierbare Funktion mit $(t,x(t)) \in D$ und $F(t,x(t)) = \text{const}$ für alle $t \in I$, so genügt $x$ auf $I$ der Differentialgleichung $F_t(t,x) + F_x(t,x)x' = 0$. Umgekehrt nennen wir eine Differentialgleichung $g(t,x) + h(t,x)x' = 0$ *exakt*, wenn ein (hinreichend glattes) $F$ mit $F_t(t,x) = g(t,x)$, $F_x(t,x) = h(t,x)$ in $D$ existiert. Notwendig und hinreichend hierfür ist $g_x(t,x) = h_t(t,x)$, $F$ nennt man die zugehörige Stammfunktion.

   Man löse die folgenden Anfangswertaufgaben für eine exakte Differentialgleichung. Auch die Möglichkeiten von Maple können getestet werden.

   (a) $2tx + t^2x' = 0$, $x(2) = -3$,

   (b) $(x^2 + \cos t) + 2txx' = 0$, $x(\pi) = -3$,

   (c) $x + (t - \sin x)x' = 0$, $x(0) = \pi/2$.

   **Lösung:** Die angegebenen Differentialgleichungen sind offenbar exakt, es kommt nur darauf an, eine zugehörige Stammfunktion zu bestimmen. Im ersten Fall kann $F(t,x) = t^2x$ genommen werden. Aus $x(t) = Ct^{-2}$ und der Anfangsbedingung $x(2) = -3$ erhalten wir $C = -12$, so dass $x(t) = -12t^{-2}$ die gesuchte Lösung der ersten Anfangswertaufgabe ist. Auch Maple findet diese.

   Bei der zweiten Differentialgleichung gehen wir folgendermaßen vor, um eine Stammfunktion $F$ zu finden. Aus $F_t(t,x) = x^2 + \cos t$, $F_x(t,x) = 2tx$ erhalten wir durch Integration über $t$ bzw. $x$, dass

$$F(t,x) = tx^2 + \sin t + k(x), \qquad F(t,x) = tx^2 + m(t).$$

   Es liegt nahe, $k(x) = 0$ und $m(t) = \sin t$ zu wählen. Aus $tx(t)^2 + \sin t = C$ und der Anfangsbedingung erhalten wir $x(t) = -\sqrt{(9\pi - \sin t)/t}$. Auch Maple findet diese Lösung, nur in der Darstellung $x(t) = -\sqrt{t(-\sin t + 9\pi)}/t$.

   Bei der dritten Differentialgleichung gewinnt man durch $F(t,x) = tx + \cos x$ leicht eine Stammfunktion. Aus $tx(t) + \cos x(t) = C$ und der Anfangsbedingung erhält man, dass $tx(t) + \cos x(t) = 0$. Diese Gleichung kann nun nicht einfach nach $x$ aufgelöst werden. Dies äußert sich auch in Maple:

```
> dsolve({x(t)+(t-sin(x(t)))*diff(x(t),t)=0,x(0)=Pi/2},x(t));
```

$$\mathrm{x}(t) = \mathrm{RootOf}(\_Z\, t + \cos(\_Z) - \cos(\frac{1}{2}\,\pi))$$

```
> simplify(%);
```

$$\mathrm{x}(t) = \mathrm{RootOf}(\_Z\, t + \cos(\_Z))$$

   Mit Hilfe von Maple wollen wir die implizite Gleichung $tx + \cos(x) = 0$ visualisieren. Mit

```
with(plots);
implicitplot(t*x+cos(x)=0,t=-1..1,x=-2..4);
```

   erhalten wir den in Abbildung 5.6 angegebenen Plot.

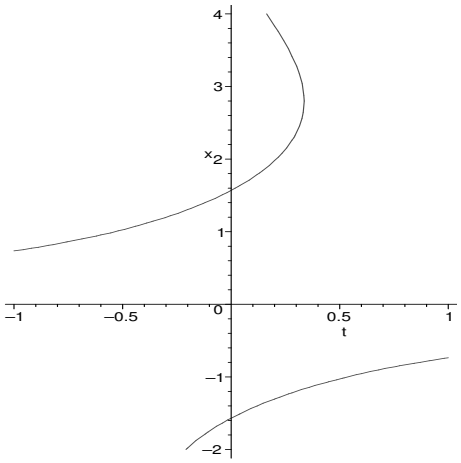


Abbildung 5.6: Lösung von $tx + \cos(x) = 0$

6. Gegeben sei die Differentialgleichung $g(t,x) + h(t,x)x' = 0$, wobei $g$ und $h$ "hinreichend glatt" sind. Man nennt eine nicht verschwindende Funktion $k$ einen *integrierenden Faktor* für diese Differentialgleichung, wenn $k(t,x)g(t,x) + k(t,x)h(t,x)x'$ eine exakte Differentialgleichung ist. Es liegt nahe, als integrienden Faktor eine Funktion $k$ anzusetzen, die alleine von $t$ oder $x$ abhängt. Z. B. ist eine nicht verschwindende Funktion $k = k(t)$ genau dann ein integrierender Faktor für obige Differentialgleichung, wenn

$$\frac{g_x(t,x) - h_t(t,x)}{h(t,x)} = \frac{k'(t)}{k(t)} = \frac{d}{dt}\ln k(t),$$

insbesondere muss hier die linke Seite von $x$ unabhängig sein.

Durch die Bestimmung eines integrierenden Faktors löse man die folgenden Anfangswertaufgaben. Ferner plotte man die Lösungen auf einem geeigneten Intervall.

(a) $\cos x + (t\sin x)x' = 1$, $x(-1) = \pi/2$,

(b) $(2t^2 + 2tx^2 + 1)x + (3x^2 + t)x' = 0$, $x(0) = 1$.

**Lösung:** Für die erste Differentialgleichung machen wir für den integrierenden Faktor den Ansatz $k(t,x) = k(t)$. Aus

$$-\frac{2}{t} = \frac{d}{dt}\log k(t)$$

erhalten wir $k(t) := 1/t^2$ als einen integrierenden Faktor. Eine Stammfunktion aus der resultierenden Gleichung ist $F(t,x) := (1 - \cos x)/t$. Die Anfangsbedingung ergibt, dass eine Lösung implizit durch $\cos x(t) - t = 1$ bzw. explizit durch $x(t) = \arccos(1 + t)$ gegeben ist. Die Lösung ist also auf $[-2, 0]$ erklärt, ein Plot (erzeugt durch `plot(arccos(1+t),t=-2..0);`) findet man in Abbildung 5.7 links. Wir überprüfen dies (Maple war übrigens, auch nach Angabe des integrierenden Faktors, nicht in der Lage, die Anfangswertaufgabe geschlossen zu lösen), indem wir durch

```
sol:=dsolve({cos(x(t))+(t*sin(x(t)))*diff(x(t),t)=1,x(-1)=Pi/2},
      type=numeric);
plots[odeplot](sol,[t,x(t)],-2..0);
```

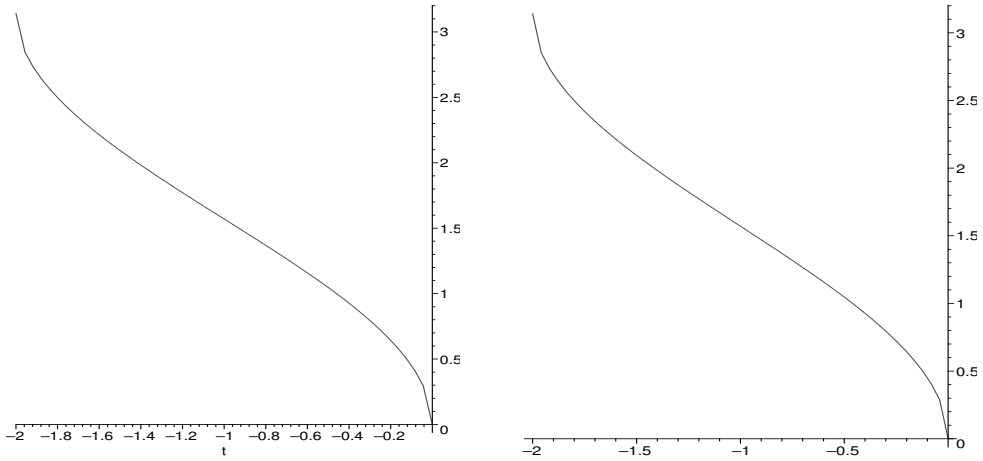


Abbildung 5.7: Lösung von $\cos x + (t \sin x)x' = 1$, $x(-1) = \pi/2$

die Anfangswertaufgabe zunächst numerisch lösen und dann die Lösung auf dem Intervall $[-2, 0]$ plotten. Das Ergebnis sieht man in Abbildung 5.7 rechts.

Nun zur zweiten Aufgabe. Wir machen wieder den Ansatz $k(t, x) = k(t)$ für einen integrierenden Faktor. Aus

$$2t = \frac{d}{dt} \log k(t)$$

erhalten wir $k(t) = e^{t^2}$, als Stammfunktion der resultierenden Aufgabe erhält man sehr leicht $F(t, x) = xe^{t^2}(t + x^2)$. Aus der Anfangsbedingung folgt, dass die Lösung implizit durch

$$x(t)e^{t^2}(t + x(t)) = 1$$

gegeben ist. Eine explizite Darstellung ist zwar möglich (kubische Gleichung!), aber sehr umständlich. Durch

```
with(plots);
implicitplot(x*exp(t^2)*(t+x^2)=1,t=-0.5..3,x=-2..2);
```

erhalten wir den Plot in Abbildung 5.8 links. Indem wir, wie oben, die Anfangswertaufgabe erst numerisch lösen und dann plotten, erhalten wir das (glattere) Bild in Abbildung 5.8 rechts.

7. Eine implizite Differentialgleichung der Form $x = tx' + g(x')$ nennt man eine *Clairautsche Differentialgleichung*. Man zeige:

   (a) Ist $g(\cdot)$ an der Stelle $a$ definiert, so ist $x(t) := at + g(a)$ eine Lösung der Differentialgleichung.

   (b) Ist $g$ stetig differenzierbar und $g'(s) \neq 0$ auf einem Intervall $I$, so ist eine Lösung in Parameterdarstellung durch

   $$t = -\dot{g}(s), \qquad x = -s\dot{g}(s) + g(s)$$

   gegeben.

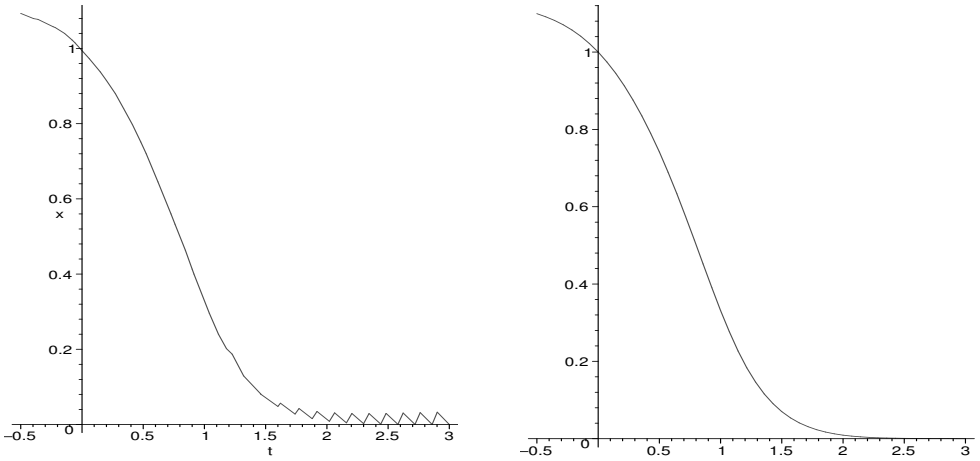


Abbildung 5.8: Lösung von $(2t^2 + 2tx^2 + 1)x + (3x^2 + t)x' = 0$, $x(0) = 1$

Schließlich bestimme man Lösungen der Clairautschen Differentialgleichung $x = tx' + \exp(x')$ und plotte sie.

**Lösung:** Der erste Teil der Aufgabe ist völlig trivial. Da $-\dot{g}(s) \neq 0$ auf dem Intervall $I$, existiert die Umkehrfunktion $(-\dot{g})^{-1}$. Wir haben nachzuweisen, dass

$$x(t) := t(-\dot{g})^{-1}(t) + g((-\dot{g})^{-1}(t))$$

eine Lösung der Clairautschen Differentialgleichung ist. Hierzu berechnen wir zunächst

$$x'(t) = (-\dot{g})^{-1}(t) + t\frac{d}{dt}(-\dot{g})^{-1}(t) - \underbrace{(-\dot{g})((-\dot{g})^{-1}(t))}_{=t}\frac{d}{dt}(-\dot{g})^{-1}(t) = (-\dot{g})^{-1}(t).$$

Daher ist

$$x(t) - tx'(t) - g(x'(t)) = t(-\dot{g})^{-1}(t) + g((-\dot{g})^{-1}(t)) - t(-\dot{g})^{-1}(t) - g((-\dot{g})^{-1}(t)) = 0,$$

also $x$ eine Lösung der Clairautschen Differentialgleichung.

Die Differentialgleichung $x = tx' + \exp(x')$ ordnet sich mit $g(s) := \exp(s)$ der Clairautschen Differentialgleichung unter. Für alle $a \in \mathbb{R}$ ist $x(t) := at + e^a$ eine Lösung. Weiter ist $(-\dot{g})^{-1} = \ln(-t)$ und daher

$$x(t) := t(-\dot{g})^{-1}(t) + g((-\dot{g})^{-1}(t)) = t\ln(-t) + \exp(\ln(-t)) = t[\ln(-t) - 1]$$

eine Lösung. Diese ist natürlich nur für $t < 0$ erklärt. Diese Lösung stellen wir in Abbildung 5.9 auf zwei Weisen dar. Links haben wir den Befehl

```
plot(t*(ln(-t)-1),t=-4..0);
```

gegeben, wir haben also ausgenutzt, dass wir eine explizite Darstellung der Lösung haben. Das Bild rechts ist durch den parametrischen Plotbefehl

```
plot([-exp(s),-s*exp(s)+exp(s),s=-10..ln(4)]);
```

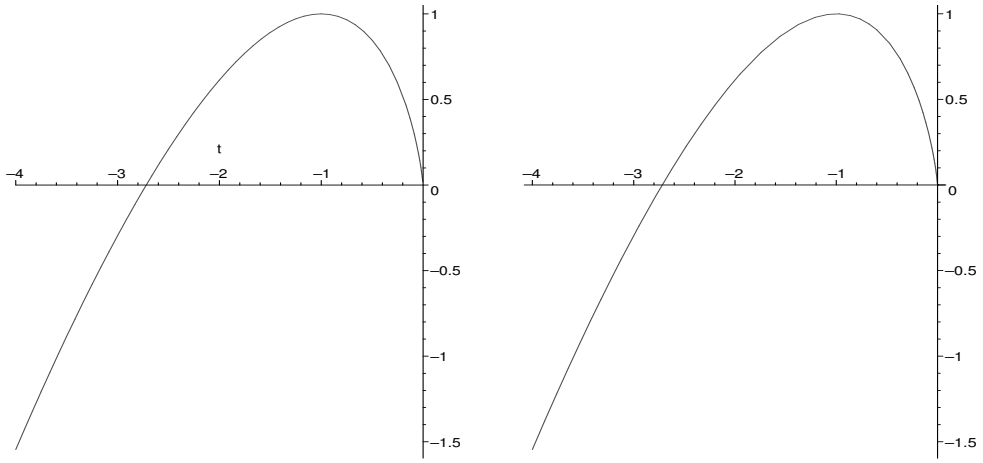


Abbildung 5.9: Lösung von $x = tx' + \exp(x')$

entstanden. Man stellt natürlich keinen Unterschied fest. Übrigens können (zumindestens gewisse) Clairautsche Differentialgleichungen auch von Maple "geknackt" werden, siehe:

```
> dsolve(x(t)=t*diff(x(t),t)+exp(diff(x(t),t)),x(t));
```

$$\mathrm{x}(t) = \ln(-t)\,t - t,\ \mathrm{x}(t) = t\ \_C1 + e^{\_C1}$$

8. Man bestimme für die Clairautschen Differentialgleichungen

   (a) $x = tx' - \sqrt{x' - 1}$,

   (b) $x = tx' + x'^2$

   Lösungen in expliziter Form.

   **Lösung:** Für $a \geq 1$ ist $x(t) := at - \sqrt{a-1}$ Lösung der ersten Differentialgleichung, für alle $a \in \mathbb{R}$ ist $x(t) := at + a^2$ eine Lösung der zweiten. Die erste Differentialgleichung ordnet sich mit $g(s) := -\sqrt{s-1}$ einer Clairautschen Differentialgleichung unter. Es ist

$$-\dot{g}(s) = \frac{1}{2\sqrt{s-1}}, \qquad (-\dot{g})^{-1}(t) = 1 + \frac{1}{4t^2}.$$

   Für $t \neq 0$ ist daher auch

$$x(t) = t(-\dot{g})^{-1}(t) + g((-\dot{g})^{-1}(t)) = t\left(1 + \frac{1}{4t^2}\right) - \sqrt{1 + \frac{1}{4t^2} - 1} = t - \frac{1}{4t}$$

   eine Lösung (die Maple übrigens nicht findet).

   Entsprechend gehen wir bei der zweiten Differentialgleichung vor. Hier ist

$$g(s) = s^2, \qquad (-\dot{g})^{-1}(t) = -\frac{t}{2}.$$

   Daher ist auch

$$x(t) = t(-\dot{g})^{-1}(t) + g((-\dot{g})^{-1}(t)) = t\left(-\frac{t}{2}\right) + \left(-\frac{t}{2}\right)^2 = -\frac{t^2}{4},$$

   eine Lösung, die auch Maple findet.

9. Zur Lösung von $x' = t^2 + x^2$, $x(0) = 1$, mache man einen Potenzreihenansatz $x(t) = \sum_{k=0}^{\infty} a_k t^k$. Man stelle eine Rekursionsformel für die Koeffizienten auf und zeige, dass $a_k \geq 1$, $k = 0, 1, \ldots$. Man berechne die ersten 15 Koeffizienten mit Maple.

   **Lösung:** Der Ansatz $x(t) = \sum_{k=0}^{\infty} a_k t^k$ führt auf die Identität

$$\sum_{k=0}^{\infty}(k+1)a_{k+1}t^k = t^2 + \left(\sum_{k=0} a_k t^k\right)^2 = t^2 + \sum_{k=0}^{\infty} t^k \sum_{j=0}^{k} a_j a_{k-j},$$

   woraus man durch Koeffizientenvergleich

$$(k+1)a_{k+1} = \begin{cases} \displaystyle\sum_{j=0}^{k} a_j a_{k-j}, & k \neq 2, \\ \displaystyle 1 + \sum_{j=0}^{k} a_j a_{k-j}, & k = 2. \end{cases}$$

   Aus der Anfangsbedingung erhält man $a_0 = 1$. Die Behauptung, dass alle $a_k$ größer oder gleich 1 sind, sind also für den Induktionsanfang $k = 0$ richtig. Weiter ist $a_1 = a_2 = 1$. Angenommen $a_j \geq 1$, $j = 0, \ldots, k$, wobei $k \geq 2$. Dann ist

$$(k+1)a_{k+1} = \sum_{j=0}^{k} a_j a_{k-j} \geq k+1,$$

   also auch $a_{k+1} \geq 1$. Die ersten 15 Koeffizienten sind leicht mit Maple berechenbar:

```
> Order:=15;
```

$$Order := 15$$

```
> dsolve({diff(x(t),t)=t^2+x(t)^2,x(0)=1},x(t),type=series);
```

$$\mathrm{x}(t) = 1 + t + t^2 + \frac{4}{3}t^3 + \frac{7}{6}t^4 + \frac{6}{5}t^5 + \frac{37}{30}t^6 + \frac{404}{315}t^7 + \frac{369}{280}t^8 + \frac{428}{315}t^9 + \frac{1961}{1400}t^{10} + \frac{75092}{51975}t^{11} + \frac{1238759}{831600}t^{12} + \frac{9884}{6435}t^{13} + \frac{17121817}{10810800}t^{14} + \mathrm{O}(t^{15})$$

```
> evalf(%);
```

$$\mathrm{x}(t) = 1. + 1.\,t + 1.\,t^2 + 1.333333333\,t^3 + 1.166666667\,t^4 + 1.200000000\,t^5 + 1.233333333\,t^6 + 1.282539683\,t^7 + 1.317857143\,t^8 + 1.358730159\,t^9 + 1.400714286\,t^{10} + 1.444771525\,t^{11} + 1.489609187\,t^{12} + 1.535975136\,t^{13} + 1.583769656\,t^{14} + \mathrm{O}(t^{15})$$

10. Sei $x \in C^1(0,a)$ auf $(0,a)$ positiv und $\lim_{t\to a} x(t) = 0$. Für jedes $t \in (0,a)$ sei der Abstand des Punktes $P := (t, x(t))$ vom Schnittpunkt $T$ der Tangente an $x(t)$ in $P$ mit der $x$-Achse gleich dem Abstand von $T$ zum Nullpunkt. Man zeige[7], dass $x$ einer Differentialgleichung erster Ordnung genügt und löse diese.

    **Lösung:** Die Tangente an $x$ in $P := (t, x(t))$ ist $y(s) = x(t) + x'(t)(s-t)$, der Schnittpunkt der Tangente mit der $x$-Achse ist also $T = (0, x(t) - x'(t)t)$. Wegen

[7] Diese Aufgabe wurde, mit geringfügig anderer Notation, in der Staatsexamensklausur September 2001 gestellt.

$\|P - T\|_2 = \|T\|_2$ genügt $x$ der Differentialgleichung $\sqrt{t^2 + t^2 x'^2} = |x - tx'|$, weiter ist $x(a) = 0$. Daher ist die Lösung der Anfangswertaufgabe

$$t^2 = x^2 - 2txx', \qquad x(a) = 0$$

zu bestimmen. Mit Hilfe von

```
dsolve({t^2=x(t)^2-2*t*x(t)*diff(x(t),t),x(a)=0},x(t));
```

erhält man

$$x(t) = \pm\sqrt{t(a-t)},$$

die gesuchte Lösung ist also $x(t) = \sqrt{t(a-t)}$. Wie erhält man dieses Ergebnis aber, wenn man Maple (oder ein anderes mathematisches Anwendersystem) nicht zur Verfügung hat? Hierzu beachte man, dass man die Anfangswertaufgabe auch in der Form

$$x' = \frac{1}{2}\left(\frac{x}{t} - \frac{t}{x}\right), \qquad x(a) = 0$$

schreiben kann, aus der man ersieht, dass es sich hier um eine homogene Differentialgleichung handelt. Daher ist $x(t) = ty(t)$, wobei $y$ Lösung von

$$y' = -\frac{y + 1/y}{2t}, \qquad y(a) = 0.$$

Mit

$$H(y) := \int^y \frac{1}{y + 1/y}\, dy = \frac{1}{2}\ln(1 + y^2)$$

ist

$$H(y(t)) = -\int_a^t \frac{1}{2t}\, dt + H(0)$$

bzw.

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

Hieraus erhält man (wir suchen eine auf $(0, a)$ positive Funktion) die gesuchte Lösung $x(t) = \sqrt{t(a-t)}$.

### 5.1.4 Aufgaben zu Abschnitt 1.4

1. Auf $C[a, b]$ ist durch

$$\|x\|_\infty := \max_{t\in[a,b]} |x(t)|, \qquad \|x\|_2 := \left(\int_a^b x(t)^2\, dt\right)^{1/2}$$

jeweils eine Norm gegeben. Man zeige, dass *keine* Konstante $C > 0$ mit $\|x\|_\infty \le C\,\|x\|_2$ für alle $x \in C[a, b]$ existiert, d. h. in $C[a, b]$ sind nicht je zwei Normen äquivalent.

Hinweis: Man konstruiere eine Folge $\{x_k\} \subset C[a, b]$ mit $\|x_k\|_\infty = 1$ für alle $k$ und $\lim_{k\to\infty} \|x_k\|_2 = 0$.

**Lösung:** Für $k \in \mathbb{N}$ definiere man $x_k$ durch

$$x_k(t) := \begin{cases} 1 - \dfrac{t}{a + (b-a)/k}, & t \in [a, a + (b-a)/k], \\ 0, & t \in [a + (b-a)/k, b]. \end{cases}$$

Offenbar ist $\{x_k\} \subset C[a,b]$ und $\|x_k\|_\infty = 1$, ferner ist

$$\begin{aligned}\|x_k\|_2^2 &= \int_a^{a+(b-a)/k} \left(1 - \frac{t}{a+(b-a)/k}\right)^2 dt \\ &= \frac{1}{3}\frac{(a+(b-a)/k)^3 - a^3}{(a+(b-a)/k)^2} - \frac{(a+(b-a)/k)^2 - a^2}{a+(b-a)/k} + (b-a)/k \\ &= -\frac{1}{3}\frac{(-b+a)^3}{(ak+b-a)^2 k}.\end{aligned}$$

Die letzten beiden Gleichungen haben wir (der Bequemlichkeit halber) durch

```
int((1-t/(a+(b-a)/k))^2,t=a..a+(b-a)/k);
factor(%);
```

erhalten. Offenbar ist $\lim_{k\to\infty} \|x_k\| = 0$. Die Aufgabe ist damit gelöst.

2. Der lineare Raum $C_n[a,b]$ aller stetigen Abbildungen $x\colon [a,b] \longrightarrow \mathbb{R}^n$, versehen mit der Norm

$$\|x\| := \max_{t\in[a,b]} \|x(t)\|,$$

wobei auf der rechten Seite $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist, ist ein Banach-Raum.

**Lösung:** Sei $\{x_k\}$ eine Cauchy-Folge in $C_n[a,b]$. Zu vorgegebenem $\epsilon > 0$ gibt es dann ein $K(\epsilon) \in \mathbb{N}$ mit

$$(*) \qquad \|x_k(t) - x_l(t)\| \le \|x_k - x_l\| \le \epsilon \qquad \text{für alle } k,l \ge K(\epsilon) \text{ und alle } t \in [a,b].$$

Für jedes $t \in [a,b]$ ist daher $\{x_k(t)\} \subset \mathbb{R}^n$ eine Cauchy-Folge. Da der $\mathbb{R}^n$ versehen mit einer beliebigen Norm ein Banach-Raum ist, konvergiert die Folge $\{x_k(t)\}$ für jedes $t \in [a,b]$, es existiert also eine Abbildung $x\colon [a,b] \longrightarrow \mathbb{R}^n$ mit $\lim_{k\to\infty} x_k(t) = x(t)$. Aus $(*)$ folgt mit $l \to \infty$, dass

$$\|x_k(t) - x(t)\| \le \epsilon \qquad \text{für alle } k \ge K(\epsilon) \text{ und alle } t \in [a,b].$$

Folglich ist $\|x_k - x\| \le \epsilon$ für alle $k \ge K(\epsilon)$, d. h. die Folge $\{x_k\} \subset C_n[a,b]$ konvergiert gleichmäßig gegen $x$. Der gleichmäßige Limes stetiger Funktionen ist bekanntlich eine stetige Funktion, wie man leicht mit einem $\epsilon/3$-Argument nachweist. Also ist $x \in C_n[a,b]$, die Behauptung ist bewiesen.

3. Der lineare Raum $C_n^1[a,b]$ aller stetig differenzierbaren Abbildungen $x\colon [a,b] \longrightarrow \mathbb{R}^n$, versehen mit der Norm

$$\|x\| := \max(\max_{t\in[a,b]} \|x(t)\|, \max_{t\in[a,b]} \|x'(t)\|),$$

wobei auf der rechten Seite $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist, ist ein Banach-Raum.

**Lösung:** Sei $\{x_k\}$ eine Cauchy-Folge in $C_n^1[a,b]$ bezüglich der angegebenen Norm. Dann sind $\{x_k\}, \{x_k'\}$ Cauchy-Folgen in $C_n[a,b]$, versehen mit der Norm

$$\|x\| := \max_{t\in[a,b]} \|x(t)\|.$$

Folglich konvergieren $\{x_k\}$ und $\{x_k'\}$ gleichmäßig gegen $x \in C_n[a,b]$ bzw. $y \in C_n[a,b]$. Zunächst zeigen wir, dass

$$x(t) = x(a) + \int_a^t y(s)\,ds \qquad \text{für alle } t \in [a,b],$$

woraus $x \in C_n^1[a,b]$ und $x' = y$ folgt. Denn für beliebiges $t \in [a,b]$ ist

$$x(t) - x(a) - \int_a^t y(s)\,ds = [x(t) - x_k(t)] - [x(a) - x_k(a)] - \int_a^t [y(s) - x_k'(s)]\,ds$$

und folglich

$$\begin{aligned}
\left\| x(t) - x(a) - \int_a^t y(s)\,ds \right\| &\le \|x(t) - x_k(t)\| + \|x(a) - x_k(a)\| \\
&\quad + \left\| \int_a^t [y(s) - x_k'(s)]\,ds \right\| \\
&\le \|x(t) - x_k(t)\| + \|x(a) - x_k(a)\| \\
&\quad + \int_a^t \|y(s) - x_k'(s)\|\,ds \\
&\le \underbrace{\|x(t) - x_k(t)\|}_{\to 0} + \underbrace{\|x(a) - x_k(a)\|}_{\to 0} \\
&\quad + (t-a) \underbrace{\max_{s\in[a,b]} \|y(s) - x_k'(s)\|}_{\to 0},
\end{aligned}$$

woraus $x(t) = x(a) + \int_a^t y(s)\,ds$ für alle $t \in [a,b]$ folgt. Dann konvergiert $\{x_k\}$ bezüglich der auf $C_n^1[a,b]$ gegebenen Norm gegen $x$, so dass $C_n^1[a,b]$ bezüglich dieser Norm ein Banach-Raum ist.

4. Auf $C[0,1]$, dem linearen Raum der auf dem Intervall $[0,1]$ stetigen, reellwertigen Funktionen definiere man die reellwertige Abbildung $\|\cdot\|$ durch $\|x\| := \max_{t\in[0,1]} t^2|x(t)|$. Man zeige[8], dass $(C[0,1], \|\cdot\|)$ ein linearer normierter Raum, aber kein Banach-Raum ist.

   Hinweis: Man betrachte die Folge $\{x_k\} \subset C[0,1]$, die durch

$$x_k(t) := \begin{cases} k, & t \in [0, 1/k], \\ 1/t, & t \in [1/k, 1] \end{cases}$$

   definiert ist.

   **Lösung:** Dass $(C[0,1], \|\cdot\|)$ ein linearer normierter Raum ist, ist fast trivial. Nun zum Nachweis der Definitheit muss man sich ein bisschen überlegen. Aus $x \in C[0,1]$ und $\|x\| = 0$ folgt $t^2 x(t) = 0$ für alle $t \in [0,1]$, hieraus zunächst $x(t) = 0$ für alle $t \in (0,1]$ und dann wegen der Stetigkeit von $x$ auch $x(t) = 0$ für alle $t \in [0,1]$. Wir zeigen, dass die im Hinweis angegebene Folge $\{x_k\}$ eine Cauchy-Folge (natürlich bezüglich der Norm $\|\cdot\|$) ist, aber nicht konvergiert. Die ersten vier Glieder dieser Folge stellen wir

[8] Diese Aufgabe haben wir W. Walter (1993, S. 55) entnommen.

in Abbildung 5.10 dar. Für $k \geq l$ ist

$$\begin{aligned}
\|x_k - x_l\| &= \max_{t\in[0,1]} t^2|x_k(t) - x_l(t)| \\
&= \max\left(\max_{t\in[0,1/k]} t^2|k-l|, \max_{t\in[1/k,1/l]} t^2|1/t - l|\right) \\
&= \begin{cases} \dfrac{k-l}{k^2}, & 2l \leq k \\ \max\left(\dfrac{k-l}{k^2}, \dfrac{1}{4l}\right), & l \leq k \leq 2l. \end{cases}
\end{aligned}$$

Hieraus liest man ab, dass $\{x_k\}$ eine Cauchy-Folge ist. Auf jedem kompakten Teilin-

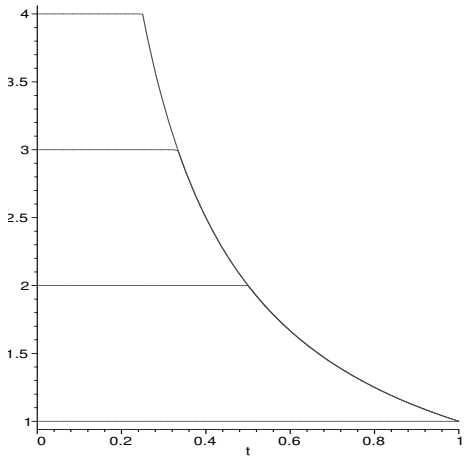


Abbildung 5.10: Die ersten vier Glieder der Folge $x_k$

tervall $I$ von $(0, 1]$ konvergiert die Folge $\{x_k\}$ gleichmäßig, also bezüglich der Norm $\|x\|_I := \max_{t\in I} |x(t)|$, gegen $x(t) := 1/t$. Als potentieller Limes der Folge $\{x_k\}$ kommt also nur die Funktion $x(t) = 1/t$ in Frage, welche aber nicht auf $[0, 1]$ stetig ist. Damit ist die Aufgabe gelöst.

5. Man beweise den Brouwerschen Fixpunktsatz im eindimensionalen Fall. Man zeige also: Sei $[a, b] \subset \mathbb{R}$ ein kompaktes Intervall, $F\colon [a, b] \longrightarrow \mathbb{R}$ eine stetige Funktion mit $F([a, b]) \subset [a, b]$. Dann existiert ein $x \in [a, b]$ mit $F(x) = x$.

   **Lösung:** Wir beweisen die Aussage durch Widerspruch. Angenommen, es ist $F(x) \neq x$ für alle $x \in [a, b]$. Da die durch $g(x) := x - F(x)$ definierte Funktion $g\colon [a, b] \longrightarrow \mathbb{R}$ stetig ist und keine Nullstelle besitzt. ist $g$ auf $[a, b]$ positiv oder negativ. Ist $g$ positiv, so ist insbesondere $g(a) > 0$ und daher $F(a) < a$, folglich $F(a) \notin [a, b]$. Einen entsprechenden Widerspruch zu der Voraussetzung, dass $F$ das Intervall $[a, b]$ in sich abbildet, erhält man für den Fall, dass $g$ auf $[a, b]$ negativ ist.

6. Sei $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ lipschitzstetig und gleichmäßig monoton auf dem $\mathbb{R}^n$, d. h. es existieren positive Konstanten $L$ und $c$ mit

$$\begin{aligned}
\|f(x) - f(y)\| &\leq L\,\|x - y\| \\
(f(x) - f(y))^T(x - y) &\geq c\,\|x - y\|^2
\end{aligned} \qquad \text{für alle } x, y \in \mathbb{R}^n.$$

Hierbei sei $\|\cdot\|$ die euklidische Norm im $\mathbb{R}^n$. Dann besitzt die Gleichung $f(x) = u$ für jedes $u \in \mathbb{R}^n$ genau eine Lösung.

**Lösung:** Sei $u \in \mathbb{R}^n$ beliebig vorgegeben. Mit noch unbestimmtem $\alpha > 0$ definieren wir $F\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ durch $F(x) := x - \alpha(f(x) - u)$. Für beliebige $x, y \in \mathbb{R}^n$ ist dann

$$\begin{aligned} \|F(x) - F(y)\|^2 &= \|x - y\|^2 - 2\alpha[f(x) - f(y)]^T(x - y) + \alpha^2 \|f(x) - f(y)\|^2 \\ &\leq (1 - 2\alpha c + \alpha^2 L)\|x - y\|^2. \end{aligned}$$

Setzt man hier speziell $\alpha := c/L$, so erhält man, dass

$$\|F(x) - F(y)\| \leq \sqrt{1 - c^2/L}\,\|x - y\| \qquad \text{für alle } x, y \in \mathbb{R}^n$$

Die so definierte Abbildung $F$ kontrahiert also auf dem $\mathbb{R}^n$, besitzt also wegen des Fixpunktsatzes für kontrahierende Abbildungen genau einen Fixpunkt. Da Fixpunkte von $F$ und Lösungen von $f(x) = u$ übereinstimmen, folgt die Gültigkeit der behaupteten Aussage.

7. Man beweise die folgende Variante zum Kontraktionssatz: Sei $X$ ein Banach-Raum (mit einer Norm $\|\cdot\|$) und $F\colon X \longrightarrow X$ eine Abbildung.. Es existiere ein $x_0 \in X$ und ein $r > 0$ derart, dass $F$ auf der Kugel

$$B[x_0; r] := \{x \in X : \|x - x_0\| \leq r\}$$

kontrahierend mit einer Lipschitzkonstanten $q < 1$ ist. Ferner sei $\|F(x_0) - x_0\| \leq (1-q)r$. Dann besitzt $F$ in $B[x_0; r]$ genau einen Fixpunkt.

**Lösung:** Man wende den Kontraktionssatz mit $K := B[x_0; r]$ an. Zu zeigen bleibt lediglich, dass die (offensichtlich abgeschlossene) Kugel $B[x_0; r]$ durch $F$ in sich abgebildet wird. Das ist aber einfach, denn ist $x \in B[x_0; r]$, so ist

$$\|F(x) - x_0\| \leq \|F(x) - F(x_0)\| + \|F(x_0) - x_0\| \leq q \underbrace{\|x - x_0\|}_{\leq r} + (1 - q)r \leq r,$$

also $F(B[x_0; r]) \subset B[x_0; r]$.

8. Man zeige:

   (a) Die durch die Iterationsvorschrift $x_{k+1} := \exp(-x_k)$ gewonnene Folge $\{x_k\}$ konvergiert für jedes $x_0 \in \mathbb{R}$ gegen die eindeutige Lösung von $x = e^{-x}$.

   (b) Die durch die Iterationsvorschrift

   $$x_{k+1} := \frac{\exp(-x_k)(1 + x_k)}{1 + \exp(-x_k)}$$

   gewonnene Folge $\{x_k\}$ konvergiert für jedes $x_0 \in [0, 1]$ gegen die eindeutige Lösung von $x = e^{-x}$.

   Ausgehend von $x_0 := 0.3$ berechne man mit Maple für beide Iterationsvorschriften $x_1, \ldots, x_{10}$ und vergleiche die Ergebnisse.

   **Lösung:** Zunächst betrachten wir die Iterationsvorschrift $x_{k+1} := \exp(-x_k)$. Für ein beliebiges $x_0 \in \mathbb{R}$ ist $x_k \in [e^{-1}, 1]$, $k = 3, \ldots$. Da $F(x) := e^{-x}$ das abgeschlossene

Intervall $[e^{-1}, 1]$ kontrahierend in sich abbildet, folgt aus dem Kontraktionssatz die Konvergenz der Folge $\{x_k\}$ gegen den eindeutige Fixpunkt von $F$ in $[e^{-1}, 1]$. Da ein Fixpunkt von $F$ aber notwendigerweise in $[e^{-1}, 1]$ liegt, ist der erste Teil der Aussage bewiesen.

Nun zur zweiten Iterationsvorschrift. Diesmal definiert man die Iterationsfunktion $F$ durch

$$F(x) := \frac{\exp(-x)(1+x)}{1+\exp(-x)}.$$

Es kann leicht gezeigt werden, dass $F$ das Intervall $[0, 1]$ kontrahierend in sich abbildet. Statt eines genauen Argumentes geben wir in Abbildung 5.11 einen Plot von $F$ und $F'$

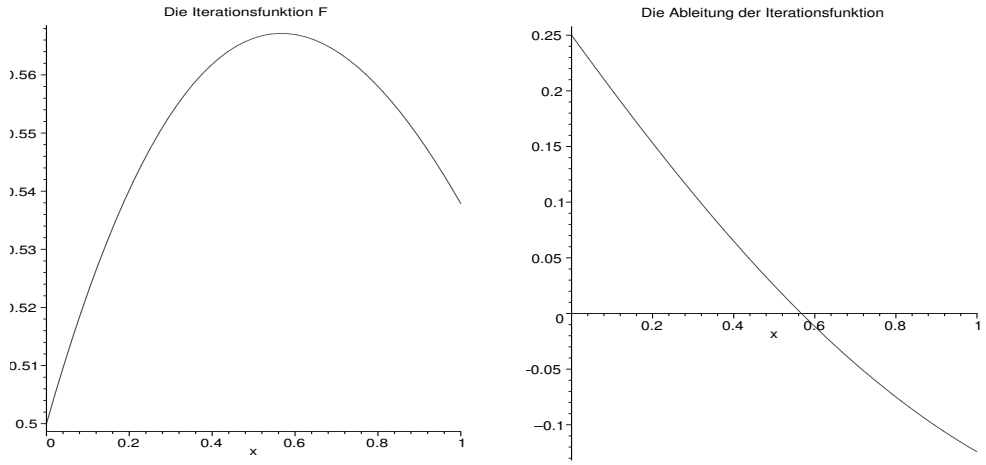


Abbildung 5.11: Die Iterationsfunktion $F$ und ihre Ableitung

über dem Intervall $[0, 1]$ an. Diese haben wir durch

```
F:=x->exp(-x)*(1+x)/(1+exp(-x));
plot(F(x),x=0..1,title="Die Iterationsfunktion F");
plot(D(F)(x),x=0..1,title="Die Ableitung der Iterationsfunktion");
```

gewonnen. Bezeichnet man die durch die erste Vorschrift gewonnene Folge mit $\{x_k\}$, die zweite mit $\{z_k\}$, so erhält man die folgende Tabelle:

| $x_k$ | $z_k$ | $k$ |
|---|---|---|
| .74081822068171786606687377931 8 | .55322472814484331670230735519 5 | 1 |
| .47672369071459406821335165923 5 | .56710808756833433742566773203 2 | 2 |
| .62081404228142328172629808930 3 | .56714329018554330032683406273 4 | 3 |
| .53750670626749250316415907482 5 | .56714329040978387299086989648 4 | 4 |
| .58420302777682462826983279591 4 | .56714329040978387299996866221 0 | 5 |
| .55755003668341333877909576705 2 | .56714329040978387299996866221 1 | 6 |
| .57261022079205282553859915340 0 | .56714329040978387299996866221 0 | 7 |
| .56405121729970387190376242699 7 | .56714329040978387299996866221 1 | 8 |
| .56889965292891139509859022256 3 | .56714329040978387299996866221 0 | 9 |
| .56614805544407171737193557278 7 | .56714329040978387299996866221 1 | 10 |

Man erkennt, das das zweite Verfahren (Newton-Verfahren, angewandt auf die Nullstellenaufgabe $x - \exp(-x) = 0$) um ein Vielfaches besser ist.

9. Wir[9] definieren in $C(I)$, $I := [0, a]$, drei Normen, die Maximumnorm

$$\|x\|_0 := \max_{t\in I} |x(t)|$$

sowie die Normen

$$\|x\|_1 := \max_{t\in I} |x(t)|e^{-at}, \qquad \|x\|_2 := \max_{t\in I} |x(t)|e^{-t^2}.$$

Man berechne für den durch

$$T(x)(t) := \int_0^t \tau x(\tau)\, d\tau$$

definierten Operator $T\colon C(I) \longrightarrow C(I)$ die entsprechenden Operatornormen $\|T\|_0$, $\|T\|_1$ und $\|T\|_2$. Hierbei ist die Operatornorm $\|T\|_j$, $j = 0, 1, 2$, gegeben durch

$$\|T\|_j := \sup_{x\in C(I)\setminus\{0\}} \frac{\|T(x)\|_j}{\|x\|_j}.$$

**Lösung:** Für beliebiges $x \in C(I)$ und beliebiges $t \in I$ ist

$$|T(x)(t)| = \left|\int_0^t \tau x(\tau)\, d\tau\right| \le \int_0^t \tau\, |x(\tau)|\, d\tau \le \int_0^t \tau\, d\tau\, \|x\|_0 = \frac{1}{2}t^2\, \|x\|_0 \le \frac{1}{2}a^2\, \|x\|_0$$

und daher

$$\|T(x)\|_0 \le \frac{1}{2}a^2\, \|x\|_0,$$

folglich $\|T\|_0 \le \frac{1}{2}a^2$. Hier gilt sogar Gleichheit, denn für die konstante Funktion $x^*(t) := 1$ gilt

$$\|T\|_0 \ge \frac{\|T(x^*)\|_0}{\|x^*\|_0} = \frac{1}{2}a^2.$$

Nach dem selben Muster gehen wir auch für die beiden anderen Normen vor. Für beliebiges $x \in C(I)$ und beliebiges $t \in I$ ist

$$\begin{aligned} e^{-at}|T(x)(t)| &= e^{-at}\left|\int_0^t \tau e^{a\tau} e^{-a\tau} x(\tau)\, d\tau\right| \\ &\le e^{-at}\int_0^t \tau e^{a\tau}\, d\tau\, \|x\|_1 \\ &= e^{-at}\left[\frac{e^{at}(at-1)+1}{a^2}\right]\|x\|_1 \\ &\le \frac{a^2 - 1 + e^{-a^2}}{a^2}\, \|x\|_1 \end{aligned}$$

und daher

$$\|T\|_1 \le \frac{a^2 - 1 + e^{-a^2}}{a^2}.$$

Diesmal definiere man $x^*(t) := e^{at}$. Dann ist $\|x^*\|_1 = 1$ und

$$\|T(x^*)\|_1 = \max_{t\in I} e^{-at}\left|\int_0^t \tau e^{a\tau}\, d\tau\right| = \frac{a^2 - 1 + e^{-a^2}}{a^2},$$

[9] Diese Aufgabe haben wir W. Walter (1993, S. 56) entnommen.

folglich

$$\|T\|_1 \le \frac{a^2 - 1 + e^{-a^2}}{a^2}.$$

Für beliebige $x \in C(I)$ und $t \in I$ ist

$$\begin{aligned} e^{-t^2}|T(x)(t)| &= e^{-t^2}\left|\int_0^t \tau e^{\tau^2} e^{-\tau^2} x(\tau)\,d\tau\right| \\ &\le e^{-t^2}\int_0^t \tau e^{\tau^2}\,d\tau\,\|x\|_2 \\ &= e^{-t^2}\frac{1}{2}\left[e^{t^2} - 1\right]\|x\|_2 \\ &\le \frac{1}{2}[1 - e^{-a^2}]\,\|x\|_2 \end{aligned}$$

und daher

$$\|T\|_2 \le \frac{1}{2}[1 - e^{-a^2}].$$

Definiert man $x^*(t) := e^{t^2}$ und argumentiert man wie oben, so erkennt man, dass hier sogar Gleichheit gilt. Insgesamt haben wir also erhalten, dass

$$\|T\|_j = \begin{cases} \dfrac{1}{2}a^2, & j = 0, \\ \dfrac{a^2 - 1 + e^{-a^2}}{a^2}, & j = 1, \\ \dfrac{1}{2}[1 - e^{-a^2}], & j = 2. \end{cases}$$

10. Man zeige[10], dass die Integralgleichung

$$x(t) = \frac{1}{2}t^2 + \int_0^t \tau x(\tau)\,d\tau, \qquad t \in I := [0, a]$$

genau eine Lösung besitzt und bestimme diese durch Zurückführung auf ein Anfangswertproblem bzw. durch explizite Berechnung der sukzessiven Approximationen unter Benutzung der Aufgabe 9, beginnend etwa mit $x_0 := 0$.

**Lösung:** Ist $x \in C(I)$ eine Lösung der Integralgleichung, so ist $x \in C^1(I)$ und $x$ genügt der Anfangswertaufgabe $x' = t(1 + x)$, $x(0) = 0$, ist daher durch $x(t) = -1 + e^{t^2/2}$ gegeben. Definiert man den Operator $\hat{T}\colon C(I) \longrightarrow C(I)$ durch

$$\hat{T}(x)(t) := \frac{1}{2}t^2 + \int_0^t \tau x(\tau)\,d\tau,$$

so ist dieser bezüglich der in Aufgabe 9 definierten Norm $\|\cdot\|_2$ kontrahierend. Daher konvergiert die Folge $\{x_k\}$ mit $x_{k+1} := \hat{T}(x_k)$ für jedes $x_0 \in C(I)$ gegen die eindeutige Lösung der Integralgleichung. Setzt man $x_0(t) := 0$, so ist

$$x_k(t) = \sum_{j=1}^{k} \frac{(t^2/2)^j}{j!}.$$

[10]Diese Aufgabe haben wir W. Walter (1993, S. 56) entnommen.

Dies zeigen wir durch vollständige Induktion nach $k$. Für $k = 0$ ist es richtig (die leere Summe ist 0). Wir machen den Schluss von $k$ nach $k + 1$. Es ist

$$\begin{aligned} x_{k+1}(t) &= \frac{1}{2}t^2 + \int_0^t \tau x_k(\tau)\, d\tau \\ &= \frac{1}{2}t^2 + \int_0^t \tau \sum_{j=1}^{k} \frac{(\tau^2/2)^j}{j!}\, d\tau \\ &= \frac{1}{2}t^2 + \sum_{j=1}^{k} \frac{1}{2^j j!} \int_0^t \tau^{2j+1}\, d\tau \\ &= \frac{1}{2}t^2 + \sum_{j=1}^{k} \frac{t^{2(j+1)}}{2^{j+1}(j+1)!} \\ &= \sum_{j=1}^{k+1} \frac{(t^2/2)^j}{j!}, \end{aligned}$$

womit die Darstellung für $x_k$ bewiesen ist.

11. Man zeige: Ist $A \in \mathbb{R}^{n\times n}$ eine positive Matrix, also alle Einträge von $A$ positiv, so besitzt $A$ einen positiven Eigenwert $\lambda^*$ mit zugehörigem positiven Eigenvektor $x^*$, d. h. alle Komponenten von $x^*$ sind positiv. Ferner ist $|\lambda| \le \lambda^*$ für alle Eigenwerte $\lambda$ von $A$, d. h. $\lambda^*$ ist der *Spektralradius* von $A$.

    Hinweis: Für den ersten Teil setze man $K := \{x \in \mathbb{R}^n : x \ge 0,\ e^T x = 1\}$ (hierbei ist $e \in \mathbb{R}^n$ der Vektor, dessen Komponenten alle gleich 1 sind), definiere $F\colon K \longrightarrow \mathbb{R}$ durch $F(x) := Ax/e^T Ax$ und wende den Brouwerschen Fixpunktsatz an. Für den zweiten Teil kann man benutzen, dass jeder Eigenwert von $A$ auch Eigenwert von $A^T$ ist.

    **Lösung:** Die im Hinweis definierte Menge $K$ ist offensichtlich nichtleer, konvex und kompakt. Für jedes $x \in K$ (dann ist wenigstens eine Komponente von $x$ positiv) ist $e^T Ax > 0$. Daher ist $F$ auf $K$ stetig, ferner offensichtlich $F(K) \subset K$. Der Brouwersche Fixpunktsatz liefert die Existenz eines $x^* \in K$ mit $F(x^*) = x^*$ bzw. $Ax^* = (e^T Ax^*)x^*$. Dann ist $x^*$ ein Eigenvektor zum Eigenwert $\lambda^* := e^T Ax^* > 0$. Da $A$ eine positive Matrix und $x^*$ ein nichtnegativer, von Null verschiedener Vektor ist, ist $Ax^*$ und dann auch $x^*$ ein positiver Vektor. Damit ist der erste Teil der Aussage bewiesen.

    Für den Beweis des zweiten Teils sei $\lambda \in \mathbb{C}$ ein beliebiger Eigenwert von $A$. Da $\lambda$ auch ein Eigenwert von $A^T$ ist, gibt es ein $y \in \mathbb{C}^n$ mit $A^T y = \lambda y$ und $\|y\|_1 = e^T|y| = 1$. Hierbei sei $|z| \in \mathbb{R}^n$ bei gegebenem $z \in \mathbb{C}^n$ der Vektor, dessen Komponenten $|z_j|$, $j = 1, \ldots, n$, sind. Dann ist

$$|\lambda|\,|y| = |\lambda y| = |A^T y| \le A^T|y|,$$

    wobei die Dreiecksungleichung und die Positivität von $A$ bzw. $A^T$ benutzt wurde. Mit dem Eigenvektor $x^* > 0$ zum Eigenwert $\lambda^* > 0$ ist daher

$$|\lambda|\,(x^*)^T|y| \le (x^*)^T A^T|y| = (Ax^*)^T|y| = (\lambda^* x^*)^T|y| = \lambda^*\,(x^*)^T|y|.$$

    Wegen $x^* > 0$ und $y \ne 0$ ist $(x^*)^T|y| > 0$ und daher $|\lambda| \le \lambda^*$. Auch der zweite Teil der Aussage ist bewiesen.

12. Man zeige: Sei $F\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ stetig. Es mögen Konstanten $\alpha \in (0,1)$, $\beta > 0$ existieren derart, dass $\|F(x)\| \le \alpha \|x\| + \beta$ für alle $x \in \mathbb{R}^n$. Hierbei sei $\|\cdot\|$ eine beliebige Norm im $\mathbb{R}^n$. Dann besitzt $F$ mindestens einen Fixpunkt.

    **Lösung:** Sei $K := \{x \in \mathbb{R}^n : \|x\| \le \beta/(1-\alpha)\}$. Dann ist $F(K) \subset K$, denn für beliebiges $x \in K$ ist
    $$\|F(x)\| \le \alpha \|x\| + \beta \le \frac{\alpha\beta}{1-\alpha} + \beta = \frac{\beta}{1-\alpha},$$
    also $F(x) \in K$. Der Brouwersche Fixpunktsatz liefert die Behauptung.

13. Man zeige[11]: Sei $K := \{x \in \mathbb{R}^n : \|x\| \le r\}$ mit $r > 0$ die abgeschlossene Kugel um den Nullpunkt mit dem Radius $r$, wobei $\|\cdot\|$ eine beliebige Norm auf dem $\mathbb{R}^n$ ist. Sei $F\colon K \longrightarrow \mathbb{R}^n$ stetig. Gilt dann die Implikation
    $$\lambda > 1,\ \|x\| = r \Longrightarrow F(x) \neq \lambda x,$$
    so besitzt $F$ einen Fixpunkt in $K$.

    Hinweis: Man mache einen Widerspruchsbeweis und wende den Brouwerschen Fixpunktsatz auf die durch $G(x) := r[F(x) - x]/\|F(x) - x\|$ definierte Abbildung an.

    **Lösung:** Der Beweis erfolgt durch Widerspruch. Wir nehmen also an, $F$ habe keinen Fixpunkt in $K$. Dann ist die Abbildung $G\colon K \longrightarrow \mathbb{R}^n$ mit
    $$G(x) := r\frac{F(x) - x}{\|F(x) - x\|}$$
    wohldefiniert und stetig, ferner $\|G(x)\| = r$ für alle $x \in K$. Der Brouwersche Fixpunktsatz liefert die Existenz eines Fixpunktes $x^*$ von $G$ in $K$. Offensichtlich ist $\|x^*\| = 1$, ferner ist $F(x^*) = \lambda^* x^*$ mit
    $$\lambda^* := 1 + (1/r)\|F(x^*) - x^*\| > 1,$$
    ein Widerspruch zur Voraussetzung.

14. Man beweise die folgende direkte Verallgemeinerung des Brouwerschen Fixpunktsatzes.

    Sei $(X, \|\cdot\|)$ ein linearer normierter Raum und $K \subset X$ nichtleer, konvex und kompakt. Die stetige Abbildung $F\colon K \longrightarrow X$ bilde $K$ in sich ab, d. h. es sei $F(K) \subset K$. Dann besitzt $F$ mindestens einen Fixpunkt in $K$.

    **Lösung:** Diese Aussage ist offenbar ein einfacher Spezialfall des Schauderschen Fixpunktsatzes. Als kompakte Menge ist $K$ nämlich auch abgeschlossen. Als stetiges Bild der kompakten Menge $K$ ist $F(K)$ selbst kompakt, insbesondere daher relativ kompakt. Alle Voraussetzungen des Schauderschen Fixpunktsatzes sind daher erfüllt.

15. Durch das folgende Gegenbeispiel[12] zeige man, dass die Aussage in Aufgabe 14 falsch wird, wenn man "kompakt" durch "abgeschlossen und beschränkt" ersetzt. Sei $X := l^2$ der Hilbertsche Folgenraum aller Folgen $x := \{x_j\}$ reeller Zahlen mit $\sum_{j=1}^{\infty} x_j^2 < \infty$, versehen mit der Norm $\|x\| := (\sum_{j=1}^{\infty} x_j^2)^{1/2}$. Ferner sei $K := \{x \in l^2 : \|x\| \le 1\}$ die abgeschlossene Einheitskugel und $F\colon K \longrightarrow l^2$ definiert durch $y = F(x)$ mit
    $$y_1 := (1 - \|x\|^2)^{1/2}, \quad y_j := x_{j-1} \quad (j = 2, 3, \ldots).$$
    Man zeige:

[11] Siehe J. M. Ortega, W. C. Rheinboldt (1970, S. 163).
[12] Siehe z. B. J. Franklin (1980, S. 275).

(a) $(l^2, \|\cdot\|)$ ist ein Banach-Raum.

(b) Die Menge $K \subset l^2$ ist nichtleer, abgeschlossen, beschränkt und konvex.

(c) Die Abbildung $F$ bildet $K$ stetig in sich ab.

(d) Die Abbildung $F$ besitzt keinen Fixpunkt in $K$.

**Lösung:** Dass $\|\cdot\|$ eine Norm ist, folgt leicht daraus, dass $\|\cdot\|$ mittels $\|x\| = (x,x)^{1/2}$ durch das innere Produkt $(x,y) := \sum_{j=1}^{\infty} x_j y_j$ erzeugt ist. Sei nun $\{x_k\} \subset l^2$ mit $x_k = \{(x_k)_j\}_{j\in\mathbb{N}}$ eine Cauchyfolge. Zu vorgebenem $\epsilon > 0$ gibt es daher ein $K(\epsilon) \in \mathbb{N}$ mit

$$\|x_k - x_l\| = \left(\sum_{j=1}^{\infty}((x_k)_j - (x_l)_j)^2\right)^{1/2} \leq \epsilon \qquad \text{für alle } k,l \geq K(\epsilon).$$

Hieraus liest man ab, dass die Folge $\{(x_k)_j\}_{k\in\mathbb{N}}$ eine Cauchyfolge in $\mathbb{R}$ und daher , insbesondere konvergent ist. Daher existiert

$$x_j^* := \lim_{k\to\infty}(x_k)_j, \qquad j = 1,2,\ldots$$

und damit die Folge $x^* = \{x_j^*\}$ reeller Zahlen. Für beliebiges endliches $N \in \mathbb{N}$ folgt aus

$$\left(\sum_{j=1}^{N}((x_k)_j - (x_l)_j)^2\right)^{1/2} \leq \|x_k - x_l\| \leq \epsilon$$

für alle $k,l \geq K(\epsilon)$ mit $l \to \infty$, dass

$$\left(\sum_{j=1}^{N}((x_k)_j - x_j^*)^2\right)^{1/2} \leq \epsilon$$

für alle $k \geq K(\epsilon)$ und alle $N \in \mathbb{N}$. Mit $N \to \infty$ folgt $\|x_k - x^*\| \leq \epsilon$ für alle $k \geq K(\epsilon)$. Daher ist $x^* \in l^2$ und die Cauchyfolge $\{x_k\}$ konvergent gegen $x^*$. Daher ist $(l^2, \|\cdot\|)$ ein Banach-Raum.

Dass $K$ nichtleer, abgeschlossen, beschränkt und konvex ist, ist völlig trivial.

Für alle $x \in K$ ist $\|F(x)\| = 1$, also $F(K) \subset K$. Ferner ist $F$ offensichtlich stetig. Wäre $F(x) = x$ mit einem $x \in K$, so wäre einerseits $x_{k-1} = x_k$, $k = 2,\ldots$, also alle Komponenten von $x$ gleich und daher $x = 0$. Andererseits ist $F(0) = \{1,0,\ldots\}$, ein Widerspruch. Insgesamt ist die Aussage bewiesen.

16. Mit positiven Konstanten $c_0, c_1$ sei

$$K := \{x \in C^1[a,b] : \|x\|_\infty \leq c_0,\ \|x'\|_\infty \leq c_1\}.$$

Man zeige, dass aus jeder Folge $\{x_k\} \subset K$ eine gleichmäßig konvergente Teilfolge ausgewählt werden kann.

Hinweis: Man wende den Satz von Arzela-Ascoli an.

**Lösung:** Wir zeigen mit Hilfe des Satzes von Arzela-Ascoli, dass $K$ als Teilmenge von $C[a,b]$, versehen mit der Maximumnorm, relativ kompakt ist. Hierzu ist die Beschränktheit und gleichgradige Stetigkeit von $K$ nachzuweisen. Wegen $\|x\|_\infty \leq c_0$ für alle $x \in K$ ist ersteres klar. Letzteres erhält man aus $-c_1 \leq x'(\tau) \leq c_1$ für alle $\tau \in [a,b]$ und alle

$x \in K$. Seien $s, t \in [a, b]$ beliebig. Ist $s \leq t$, so integriere man obige Ungleichungen über $[s, t]$, im anderen Fall über $[t, s]$. In jedem Fall erhält man $|x(t) - x(s)| \leq c_1 |t - s|$ für alle $t, s \in [a, b]$ und alle $x \in K$. Zu jedem $\epsilon > 0$ gibt es dann ein $\delta(\epsilon) := \epsilon / c_1$ mit

$$t, s \in [a, b],\ |t - s| \leq \delta(\epsilon),\ x \in K \Longrightarrow |x(t) - x(s)| \leq \epsilon.$$

Damit ist bewiesen, dass $K$ auch gleichgradig stetig ist. Die Behauptung folgt aus dem Satz von Arzela-Ascoli.

17. Man zeige die Umkehrung im Satz von Arzela-Ascoli, also:

    Sei $C_n[a, b]$ versehen mit der Maximumnorm $\|x\| := \max_{t \in [a,b]} \|x(t)\|$ für $x \in C_n[a, b]$, wobei $\|\cdot\|$ rechts eine beliebige Norm auf dem $\mathbb{R}^n$ ist. Eine relativ kompakte Menge $K \subset C_n[a, b]$ ist beschränkt und gleichgradig stetig.

    Hinweis: Zum Nachweis der gleichgradigen Stetigkeit von $K$ zeige man zunächst, dass es zu vorgegebenem $\epsilon > 0$ endlich viele $\{z_1, \ldots, z_p\} \subset K$ mit $\min_{i=1,\ldots,p} \|x - z_i\| \leq \epsilon/3$ für alle $x \in K$ gibt. Mit Hilfe der gleichmäßigen Stetigkeit der $z_i$, $i = 1, \ldots, p$, schließe man auf die gleichgradige Stetigkeit von $K$.

    **Lösung:** Eine relativ kompakte Teilmenge $K$ eines beliebigen linearen normierten Raumes $(X, \|\cdot\|)$ ist beschränkt. Zu zeigen bleibt die gleichgradige Stetigkeit von $K$. Da $K$ relativ kompakt ist, gibt es zu gegebenem $\epsilon > 0$ gi endlich viele $\{z_1, \ldots, z_p\} \subset K$ mit

    $$\min_{i=1,\ldots,p} \|x - z_i\| \leq \frac{\epsilon}{3} \qquad \text{für alle } x \in K,$$

    d. h. jedes $x \in K$ hat zu wenigstem einem der $z_i$, $i = 1, \ldots, p$ einen Abstand, der nicht größer als $\epsilon/3$ ist. Denn andernfalls existiert eine Folge $\{z_k\} \subset K$ mit $\|z_k - z_l\| > \epsilon/3$ für alle $k \neq l$, was ein Widerspruch dazu wäre, dass man aus $\{z_k\}$ eine konvergente Teilfolge auswählen kann. Die $z_i$, $i = 1, \ldots, p$, sind auf der kompakten Menge $[a, b]$ gleichmäßig stetig, d. h. es existieren $\delta_i = \delta_i(\epsilon, z_i) > 0$, $i = 1, \ldots, p$, mit

    $$t, s \in [a, b],\ |t - s| \leq \delta_i \Longrightarrow \|z_i(t) - z_i(s)\| < \frac{\epsilon}{3}, \quad i = 1, \ldots, p.$$

    Nun setze man $\delta := \min_{i=1,\ldots,p} \delta_i$ und wähle $x \in K$ beliebig. Es existiert dann ein $i \in \{1, \ldots, p\}$ mit $\|x - z_i\| \leq \epsilon/3$. Mit beliebigen $t, s \in [a, b]$ mit $|t - s| \leq \delta$ ist dann

    $$\begin{aligned} \|x(t) - x(s)\| &\leq \|x(t) - z_i(t)\| + \|z_i(t) - z_i(s)\| + \|z_i(s) - x(s)\| \\ &\leq \|x - z_i\| + \|z_i(t) - z_i(s)\| + \|z_i - x\| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon, \end{aligned}$$

    womit die gleichgradige Stetigkeit von $K$ gezeigt ist.

# 5.2 Aufgaben zu Kapitel 2

## 5.2.1 Aufgaben zu Abschnitt 2.1

1. Die Anfangswertaufgabe

$$x'' = f(t, x), \qquad x(t_0) = x_0,\ x'(t_0) = x_0'$$

für eine Differentialgleichung zweiter Ordnung kann natürlich als eine Anfangswertaufgabe für zwei Differentialgleichungen erster Ordnung geschieben werden. Man zeige, dass sie auch äquivalent ist zu der Integralgleichung

$$x(t) = x_0 + x_0'(t - t_0) + \int_{t_0}^{t} (t - s) f(s, x(s))\, ds.$$

Genauer: Sei $I$ ein Intervall mit $t_0 \in I$. Ist $x \in C^2(I)$ eine Lösung der Anfangswertaufgabe, so ist $x$ auch eine Lösung der Integralgleichung. Ist umgekehrt $x \in C(I)$ eine Lösung der Integralgleichung, so ist sogar $x \in C^2(I)$ und $x$ ist eine Lösung der Anfangswertaufgabe.

**Lösung:** Sei $x \in C^2(I)$ zunächst eine Lösung der Anfangswertaufgabe. Für beliebiges $t \in I$ ist dann

$$\begin{aligned} \int_{t_0}^{t} (t - s) f(s, x(s))\, ds &= \int_{t_0}^{t} (t - s) x''(s)\, ds \\ &= x(t) - [x(t_0) + x'(t_0)(t - t_0)] \\ &= x(t) - [x_0 + x_0'(t - t_0)], \end{aligned}$$

also $x$ eine Lösung der Integralgleichung. Sei nun umgekehrt $x \in C(I)$ eine Lösung der Integralgleichung. Offensichtlich ist dann $x(t_0) = x_0$ und $x \in C^1(I)$, ferner

$$x'(t) = x_0' + \int_{t_0}^{t} f(s, x(s))\, ds.$$

Hierus liest man $x'(t_0) = x_0'$, $x \in C^2(I)$ und $x''(t) = f(t, x(t))$ ab, d. h. $x$ ist eine Lösung der Anfangswertaufgabe.

2. Bei gegebenem $a > 0$ sei die Funktion $f = f(t, s, x)$ auf $D := \{(t, s, x) \in \mathbb{R}^3 : 0 \le s \le t \le a\}$ stetig und dort bezüglich der letzten Variablen $x$ global lipschitzstetig, d. h. es existiere $L > 0$ mit

$$|f(t, s, x) - f(t, s, y)| \le L\,|x - y| \qquad \text{für alle } (t, s, x),\ (t, s, y) \in D.$$

Dann besitzt die Volterrasche Integralgleichung

$$x(t) = g(t) + \int_{0}^{t} f(t, s, x(s))\, ds$$

für jedes $g \in C[0, a]$ genau eine auf $[0, a]$ stetige Lösung.

**Lösung:** Wir definieren die Abbildung $F\colon C[0, a] \longrightarrow C[0, a]$ durch

$$F(x)(t) := g(t) + \int_{0}^{t} f(t, s, x(s))\, ds$$

und auf $C[0, a]$ die (zur Maximumnorm äquivalente) Norm

$$\|x\|_* := \max_{t \in [0,a]} e^{-2Lt} |x(t)|.$$

Wir zeigen, dass $F$ bezüglich dieser Norm kontrahierend ist, so dass aus dem Fixpunktsatz für kontrahierende Abbildungen die Behauptung folgen wird. Für $x, y \in C[0,a]$ und beliebiges $t \in [0,a]$ ist

$$\begin{aligned} e^{-2Lt}|F(x)(t) - F(y)(t)| &= e^{-2Lt}\left|\int_0^t [f(t,s,x(s)) - f(t,s,y(s))]\right| \\ &\le e^{-2Lt} L \int_0^t |x(s) - y(s)|\, ds \\ &= e^{-2Lt} L \int_0^t e^{2Ls} \underbrace{e^{-2Ls}|x(s) - y(s)|}_{\le \|x-y\|_*}\, ds \\ &\le \frac{1}{2} e^{-2Lt}(e^{2Lt} - 1)\|x - y\|_* \\ &\le \frac{1}{2}\|x - y\|_*. \end{aligned}$$

Indem man links zum Maximum über $[0,a]$ übergeht, erhält man

$$\|F(x) - F(y)\|_* \le \frac{1}{2}\|x - y\|_* \qquad \text{für alle } x, y \in C[0,a].$$

Die Aussage ist damit bewiesen.

3. Gegeben sei die Anfangswertaufgabe

$$x' = tx^2, \qquad x(0) = 1.$$

Mit $x_0 := 1$ und

$$x_{k+1}(t) := 1 + \int_0^t s x_k(s)^2\, ds$$

berechne man $x_1, x_2, x_3$. Man bestimme ein Intervall $[0,\alpha]$ mit $\alpha > 0$, auf dem eine Lösung eindeutig existiert und mache eine Fehlerabschätzung.

**Lösung:** Mit Hilfe von Maple erhalten wir

$$\begin{aligned} x_1(t) &= 1 + \frac{1}{2}t^2, \\ x_2(t) &= 1 + \frac{1}{2}t^2 + \frac{1}{4}t^4 + \frac{1}{24}t^6, \\ x_3(t) &= 1 + \frac{1}{2}t^2 + \frac{1}{4}t^4 + \frac{1}{8}t^6 + \frac{1}{24}t^8 + \frac{1}{96}t^{10} + \frac{1}{576}t^{12} + \frac{1}{8064}t^{14}. \end{aligned}$$

Für eine Fehlerabschätzung untersuchen wir, unter welchen Voraussetzungen an noch unbestimmte Parameter $\alpha >$ und $\beta \in (0,1]$ die Abbildung

$$F(x)(t) := 1 + \int_0^t s x(s)^2\, ds$$

die abgeschlossene Menge

$$K := \{x \in C[0,\alpha] : \|x - x_0\| \le \beta\}$$

kontrahierend in sich abbildet, wobei $x_0(t) := 1$ und als Norm in $C[0,\alpha]$ die (ungewichtete) Maximumnorm zugrunde gelegt sei. Seien $t \in [0,\alpha]$, $x \in K$ beliebig. Es ist

$$\begin{aligned} |F(x)(t) - x_0(t)| &= \left| \int_0^t s x(s)^2 \, ds \right| \\ &\le (1+\beta)^2 \int_0^t s \, ds \\ &\le \frac{1}{2}\alpha^2 (1+\beta)^2. \end{aligned}$$

Daher bildet die Abbildung $F$ die Menge $K$ in sich ab, wenn

$$\frac{1}{2}\alpha^2(1+\beta)^2 \le \beta.$$

Für beliebige $x, y \in K$ und $t \in [0,\alpha]$ ist

$$\begin{aligned} |F(x)(t) - F(y)(t)| &= \left| \int_0^t s[x(s)^2 - y(s)^2] \, ds \right| \\ &\le \int_0^t s|x(s)+y(s)| \, |x(s)-y(s)| \, ds \\ &\le 2(1+\beta) \int_0^t s \, ds \, \|x-y\| \\ &\le \alpha^2(1+\beta) \, \|x-y\|. \end{aligned}$$

Insgesamt bildet $F$ die (abgeschlossene) Menge $K$ kontrahierend in sich ab, wenn

$$\frac{1}{2}\alpha^2(1+\beta)^2 \le \beta, \qquad \alpha^2(1+\beta) < 1.$$

Setzt man z. B. $\beta := 1$, so sind beide Ungleichungen erfüllt, wenn $0 < \alpha < \sqrt{2}/2$. Die exakte Lösung der gegebenen Anfangswertaufgabe ist übrigens $x(t) = 2/(2-t^2)$. Eine Lösung existiert also nur auf $(-\sqrt{2}, \sqrt{2})$.

4. Die lineare Anfangswertaufgabe erster Ordnung

$$x' = 2tx + t, \qquad x(0) = x_0$$

besitzt die Lösung

$$x(t) = x_0 e^{t^2} + \frac{1}{2}(e^{t^2} - 1),$$

wie man durch `dsolve({diff(x(t),t)=2*t*x(t)+t,x(0)=x_0},x(t));` oder eigene Rechnung feststellt. Mit $x_0(t) := x_0$ sei

$$x_{k+1}(t) := x_0 + \int_0^t (2s x_k(s) + s) \, ds.$$

Man stelle die Iterierten $x_k$ geschlossen dar und begründe, weshalb die Folge $\{x_k\}$ auf jedem kompakten Intervall in $\mathbb{R}$ gleichmäßig gegen die Lösung der gegebenen Anfangswertaufgabe konvergiert.

**Lösung:** Wir zeigen durch vollständige Induktion nach $k$, dass

$$x_k(t) = x_0\left(1 + t^2 + \frac{t^4}{2!} + \cdots + \frac{t^{2k}}{k!}\right) + \frac{1}{2}\left(t^2 + \frac{t^4}{2!} + \cdots + \frac{t^{2k}}{k!}\right).$$

Hierzu genügt es, den Induktionsschluss durchzuführen. Nun ist

$$\begin{aligned}
x_{k+1}(t) &= x_0 + \int_0^t [2sx_k(s) + s]\,ds \\
&= x_0 + \frac{t^2}{2} + 2\int_0^t \left[x_0 s\left(1 + s^2 + \frac{s^4}{2!} + \cdots + \frac{s^{2k}}{k!}\right)\right. \\
&\qquad \left. + \frac{s}{2}\left(s^2 + \frac{s^4}{2!} + \cdots + \frac{s^{2k}}{k!}\right)\right] ds \\
&= x_0 + \frac{t^2}{2} + x_0\left(t^2 + \frac{t^4}{2!} + \frac{t^6}{3!} + \cdots + \frac{t^{2(k+1)}}{(k+1)!}\right) \\
&\qquad + \frac{1}{2}\left(\frac{t^4}{2!} + \frac{t^6}{3!} + \cdots + \frac{t^{2(k+1)}}{(k+1)!}\right) \\
&= x_0\left(1 + t^2 + \frac{t^4}{2!} + \cdots + \frac{t^{2(k+1)}}{(k+1)!}\right) + \frac{1}{2}\left(t^2 + \frac{t^4}{2!} + \cdots + \frac{t^{2(k+1)}}{(k+1)!}\right),
\end{aligned}$$

womit die Behauptung bewiesen ist. Unter Beachtung der Potenzreihenentwicklung

$$e^{t^2} = \sum_{j=0}^{\infty} \frac{t^{2j}}{j!}$$

und der Tatsache, dass Partialsummen auf kompakten Teilmengen gleichmäßig gegen den Limes konvergieren, folgt der Rest der Behauptung. Natürlich hätte aber auch mit der Bemerkung im Anschluss an den Satz von Picard-Lindelöf argumentiert werden können.

5. Man zeige, dass die Anfangswertaufgabe für das mathematische Pendel, also

$$x'' + \omega_0^2 \sin x = 0, \quad x(0) = x_0, \quad x'(0) = 0,$$

für beliebige $\omega_0$ und $x_0$ genau eine Lösung besitzt. Diese existiert auf ganz $\mathbb{R}$ und ist gerade, also $x(t) = x(-t)$ für alle $t$. Für $\omega_0 := 2$ und $x_0 := 1$ berechne man mit Hilfe des Gaußschen Verfahrens vom arithmetisch-geometrischen Mittel die Periodenlänge $T = (4/\omega_0)K(\sin \frac{1}{2}x_0)$. Schließlich plotte man die Lösung auf $[0, 2T]$.

**Lösung:** Schreibt man die Anfangswertaufgabe für die Differentialgleichung zweiter Ordnung als eine Anfangswertaufgabe für ein System von zwei Differentialgleichungen erster Ordnung, so erhält man

$$\begin{aligned}
x_1' &= x_2, & x_1(0) &= x_0, \\
x_2' &= -\omega_0^2 \sin x_1, & x_2(0) &= 0.
\end{aligned}$$

Die rechte Seite

$$f(t, x) := \begin{pmatrix} x_2 \\ -\omega_0^2 \sin x_1 \end{pmatrix}$$

ist global lipschitzstetig. Denn ist $\|\cdot\|$ die Maximumnorm auf dem $\mathbb{R}^2$, so ist für beliebige $x, y \in \mathbb{R}^2$ offenbar

$$\|f(t,x) - f(t,y)\| = \max(|x_2 - y_2|, \omega_0^2 |\sin x_1 - \sin y_1|) \leq \max(1, \omega_0^2)\|x - y\|.$$

Aus dem Korollar 1.3 zum Satz von Picard-Lindelöf folgt die Existenz genauer einer Lösung $x$ auf $\mathbb{R}$. Um zu zeigen, dass diese gerade ist, definieren wir $y(t) := x(-t)$. Dann genügt $y$ denselben Anfangsbedingungen wie $x$, da $y(0) = x(0) = x_0$ und $y'(0) = -x'(0) = 0$, aber auch derselben Differentialgleichung, da

$$y''(t) + \omega_0^2 \sin y(t) = x''(-t) + \omega_0^2 \sin x(-t) = 0.$$

Da gerade gezeigt wurde, dass diese Anfangswertaufgabe genau eine Lösung besitzt, ist $x(t) = y(t)$ für alle $t$ bzw. $x$ gerade. Für $\omega_0 = 2$ und $x_0 = 1$ ist die Periodenlänge

$$T = \frac{4}{\omega_0} K(\sin \tfrac{1}{2} x_0) = 2K(\sin \tfrac{1}{2}) = \frac{\pi}{M(1, \cos \frac{1}{2})}.$$

Bei der Berechnung von $M(1, \cos \frac{1}{2})$ erhalten wir die folgenden Werte

| $k$ | $b_k$ | $a_k$ |
|---|---|---|
| 0 | 0.877582561890373 | 1.000000000000000 |
| 1 | 0.936793767000172 | 0.938791280945186 |
| 2 | 0.937791992130215 | 0.937792523972679 |
| 3 | 0.937792258051410 | 0.937792258051447 |
| 4 | 0.937792258051429 | 0.937792258051429 |

Daher erhalten wir als Periodenlänge

$$T = 3.34998783218523.$$

In Abbildung 5.12 findet man einen Plot über dem Intervall $[0, 2T]$.

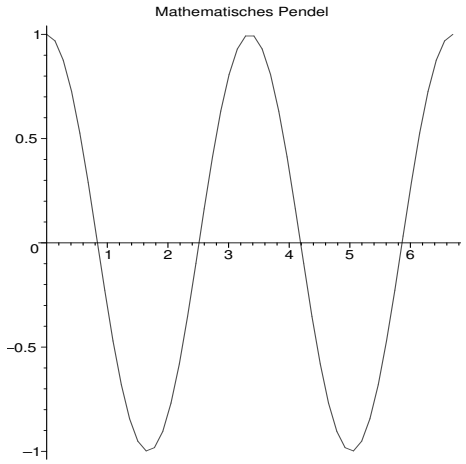


Abbildung 5.12: Die Lösung von $x'' + 4 \sin x = 0,\ x(0) = 1,\ x'(0) = 0$

6. Gegeben sei die Anfangswertaufgabe

(P) $$x' = t + \sin x, \qquad x(0) = 0.$$

(a) Man zeige, dass (P) auf jedem kompakten Teilintervall $I$ von $\mathbb{R}$ mit $0 \in I$ genau eine Lösung besitzt.

(b) Mit Hilfe von Maple-Befehlen plotte man die Lösung von (P) auf dem Intervall $[-1, 1]$.

(c) Man zeige, dass die Lösung $x$ von (P) nichtnegativ ist.

**Lösung:** Die rechte Seite von (P) ist in $x$ global lipschitzstetig. Denn mit $f(t, x) := t + \sin x$ ist

$$|f(t, x) - f(t, y)| \leq |x - y| \qquad \text{für alle } (t, x)\,(t, y) \in \mathbb{R}^2.$$

Aus dem Korollar 1.3 zum Satz von Picard-Lindelöf folgt die Existenz genau einer Lösung von (P) auf jedem kompakten Intervall, das 0 enthält.

Mit Hilfe von

```
kor:=dsolve({diff(x(t),t)=t+sin(x(t)),x(0)=0},x(t),type=numeric);
plots[odeplot](kor,[t,x(t)],-1..1,labels=["t","x"]);
```

erhalten wir den in Abbildung 5.13 dargestellten Plot.

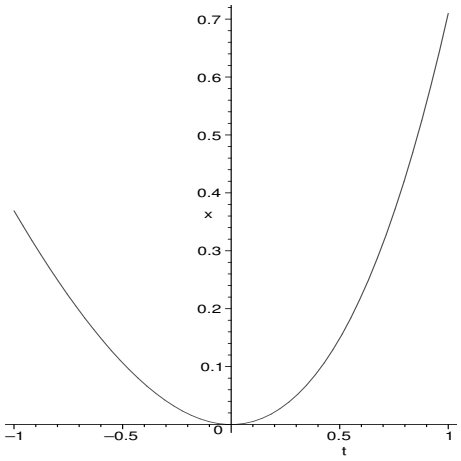


Abbildung 5.13: Lösung von $x' = t + \sin x$, $x(0) = 0$

Für $t \geq 1$ ist $x'(t) = 1 + \sin x(t) \geq 0$, folglich nach Integration $x(t) \geq 0$. Entsprechend kann man für $t \leq 1$ argumentieren. Wir betrachten auf dem mit der Maximumnorm versehenen Banach-Raum $C[-1, 1]$ die Abbildung, $F\colon C[-1, 1] \longrightarrow C[-1, 1]$, definiert durch

$$F(x)(t) := \frac{1}{2}t^2 + \int_0^t \sin x(s)\, ds.$$

Da Fixpunkte von $F$ und Lösungen von (P) übereinstimmen, wissen wir schon, dass $F$ genau einen Fixpunkt besitzt. Wir definieren die abgeschlossene, konvexe Menge

$$K := \{x \in C[-1, 1] : 0 \leq x(t) \leq 1 \ \text{ für alle } t \in [-1, 1]\}.$$

Es ist $F(K) \subset K$, denn für $x \in K$ und $t \in [-1,1]$ ist

$$0 \le \frac{1}{2}t^2 = F(0)(t) \le F(x)(t) \le F(1)(t) = \frac{1}{2}(1+\sin 1)t^2 \le 1.$$

Da $F(K)$ relativ kompakt ist, besitzt $F$ einen Fixpunkt in $K$, womit die Behauptung schließlich bewiesen[13] ist.

7. Man beweise[14], dass die Volterra-Integralgleichung

$$x(t) = g(t) + \int_0^t k(t,s,x(s))\,ds$$

mindestens eine in $[0,a]$ stetige Lösung besitzt, wenn $g \in C[0,a]$ und der „Kern“ $k(t,s,z)$ für $0 \le s \le t \le a$, $z \in \mathbb{R}$, stetig ist und einer Wachstumsbedingung $|k(t,s,z)| \le L(1+|z|)$ mit eine Konstanten $L > 0$ genügt.

Hinweis: Man wende den Schauderschen Fixpunktsatz an mit den folgenden Daten: Sei $X := C[0,a]$ der mit der Maximumnorm $\|x\| := \max_{t\in[0,a]} |x(t)|$ versehene Banach-Raum. Sei

$$K := \{x \in C[0,a] : |x(t)| \le \rho(t) \ \text{ für alle } t \in [0,a]\},$$

wobei $\rho(\cdot)$ die Lösung der Anfangswertaufgabe

$$\rho' = L(1+\rho), \qquad \rho(0) = \|g\|$$

ist und $F\colon K \longrightarrow C[0,a]$ durch

$$F(x)(t) := g(t) + \int_0^t k(t,s,x(s))\,ds$$

definiert ist. Man zeige also, dass mit diesen Daten die Voraussetzungen des Schauderschen Fixpunktsatzes erfüllt sind.

**Lösung:** Die Menge $K$ ist offensichtlich konvex und abgeschlossen. Sie ist nichtleer, da $\rho(\cdot)$ zumindestens nichtnegativ ist[15]. Wir zeigen nun weiter:

(a) $F\colon K \longrightarrow C[0,a]$ ist stetig.
Denn: Sei ein $\epsilon > 0$ vorgegeben. Man definiere

$$C := \{(t,s,x) \in \mathbb{R}^3 : 0 \le s \le t \le a,\ |x| \le \|\rho\|\}, \qquad M := \max_{(t,s,x)\in C} |k(t,s,x)|.$$

Auf der kompakten Menge $C$ ist der Kern $k$ gleichmäßig stetig. Daher existiert $\delta = \delta(\epsilon) > 0$ mit

$$(t,s,x),\ (\tau,s,y) \in C, \quad |t-\tau| + |x-y| \le \delta \Longrightarrow |k(t,s,x) - k(\tau,s,y)| \le \frac{\epsilon}{a}.$$

[13] Dieser Beweis müsste vereinfacht werden können. Aber wie?
[14] Diese Aufgabe ist dem Buch von W. Walter über Gewöhnliche Differentialgleichungen entnommen.
[15] Es ist

$$\rho(t) = -1 + (1+\|g\|)e^{Lt}.$$

Seien nun $x, y \in K$ mit $\|x - y\| \leq \delta$ gegeben. Mit $t \in [0, a]$ ist

$$\begin{aligned} |F(x)(t) - F(y)(t)| &= \left| \int_0^t [k(t,s,x(s)) - k(t,s,y(s))]\, ds \right| \\ &\leq \int_0^t |k(t,s,x(s)) - k(t,s,y(s))|\, ds \\ &\leq t\frac{\epsilon}{a} \end{aligned}$$

und folglich $\|F(x) - F(y)\| \leq \epsilon$. Damit ist die Stetigkeit von $F$ auf $K$ bewiesen.

(b) Es ist $F(K) \subset K$.
Denn: Sei $x \in K$ und $t \in [0, a]$. Dann ist

$$\begin{aligned} |F(x)(t)| &= \left| g(t) + \int_0^t k(t,s,x(s))\, ds \right| \\ &\leq |g(t)| + \int_0^t |k(t,s,x(s))|\, ds \\ &\leq \|g\| + \int_0^T L(1 + |x(s)|)\, ds \\ &\leq \|g\| + \int_0^t L(1 + \rho(s))\, ds \\ &= \rho(t), \end{aligned}$$

also $F(x) \in K$ bzw. $F(K) \subset K$.

(c) $F(K)$ ist relativ kompakt.
Zunächst ist $F(K)$ offensichtlich beschränkt, da $F(K) \subset K$ und $K$ beschränkt. Um den Satz von Arzela-Ascoli anwenden zu können, muss noch die gleichgradige Stetigkeit von $F(K)$ nachgewiesen werden. Seien hierzu $t, \tau \in [a, b]$ mit $|t - \tau| \leq \delta$ und ein beliebiges $x \in K$ gegeben. O. B. d. A. sei $\tau \leq t$. Dann ist

$$\begin{aligned} |F(x)(t) - F(x)(\tau)| &\leq |g(t) - g(\tau)| + \int_0^\tau |k(t,s,x(s)) - k(\tau,s,x(s))|\, ds \\ &\qquad \int_\tau^t |k(t,s,x(s))|\, ds \\ &\leq |g(t) - g(\tau)| + \epsilon + M\,|t - \tau|, \end{aligned}$$

woraus unter Berücksichtigung der gleichmäßigen Stetigkeit von $g$ auf $[0, a]$ die Behauptung folgt.

8. Man beweise die folgende Aussage:

Sei $D \subset \mathbb{R}^{n+1}$ offen, $f\colon D \longrightarrow \mathbb{R}^n$ stetig und bezüglich des zweiten Arguments (global) lipschitzstetig mit einer Lipschitzkonstanten $L$. Sei $(t_0, x_0) \in D$ und $x$ eine Lösung von $x' = f(t,x)$, $x(t_0) = x_0$, auf $I := \{t \in \mathbb{R} : |t - t_0| \leq \alpha\}$ mit $(t, x(t)) \in D$ für alle $t \in I$. Entsprechend sei auch $(\hat{t}_0, \hat{x}_0) \in D$ mit $\hat{t}_0 \in I$ und $\hat{x}$ eine Lösung von $x' = f(t,x)$, $x(\hat{t}_0) = \hat{x}_0$, auf $\hat{I} := \{t \in \mathbb{R} : |t - t_0| \leq \hat{\alpha}\}$ mit $(t, \hat{x}(t)) \in D$ für alle $t \in \hat{I}$. Dann ist

$$\|x(t) - \hat{x}(t)\| \leq (M\,|t_0 - \hat{t}_0| + \|x_0 - \hat{x}_0\|)\, e^{L\,|t - \hat{t}_0|} \qquad \text{für alle } t \in I \cap \hat{I},$$

wobei $M := \max_{t\in I} \|f(t, x(t))\|$.

**Lösung:** Zunächst ist

$$\|x(t_0) - x(\hat{t}_0)\| = \left\| x(t_0) - x(t_0) - \int_{t_0}^{\hat{t}_0} f(s, x(s))\, ds \right\| \le M\,|t_0 - \hat{t}_0|,$$

wobei wir benutzt haben, dass wegen $t_0, \hat{t}_0 \in I$ das gesamte Intervall von $t_0$ nach $\hat{t}_0$ zu $I$ gehört. Dann ist

$$\|x(t) - \hat{x}(t)\| \le \|x(\hat{t}_0) - \hat{x}_0\| + \int_{\hat{t}_0}^{t} L\,\|x(s) - \hat{x}(s)\|\, ds$$

für alle $t \in I \cap \hat{I}$ mit $t \ge \hat{t}_0$. Für diese $t$ folgt wegen der speziellen Gronwallschen Ungleichung

$$\|x(t) - \hat{x}(t)\| \le \|x(\hat{t}_0) - \hat{x}_0\|\, e^{L(t-\hat{t}_0)}.$$

Entsprechend ist für $t \in I \cap \hat{I}$ mit $t \le \hat{t}_0$ offenbar

$$\|x(t) - \hat{x}(t)\| \le \|x(\hat{t}_0) - \hat{x}_0\|\, e^{-L(t-\hat{t}_0)},$$

und daher gilt für alle $t \in I \cap \hat{I}$ die folgende Ungleichungskette:

$$\begin{aligned} \|x(t) - \hat{x}(t)\| &\le \|x(\hat{t}_0) - \hat{x}_0\|\, e^{L\,|t-\hat{t}_0|} \\ &\le (\|x(\hat{t}_0) - x(t_0)\| + \|x_0 - \hat{x}_0\|)\, e^{L\,|t-\hat{t}_0|} \\ &\le (M\,|t_0 - \hat{t}_0| + \|x_0 - \hat{x}_0\|)\, e^{L\,|t-\hat{t}_0|} \end{aligned}$$

und das ist die Behauptung.

9. Die Abbildung $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ sei stetig und genüge einer einseitigen Lipschitzbedingung, d. h. es existiere ein $l \in \mathbb{R}$ (welches auch negativ sein kann) mit

$$[f(x) - f(y)]^T (x - y) \le l\,\|x - y\|^2 \qquad \text{für alle } x, y \in \mathbb{R}^n,$$

wobei hier $\|\cdot\|$ die euklidische Norm bedeute. Man zeige, dass die Anfangswertaufgabe

$$x' = f(x), \qquad x(0) = x_0$$

für jedes $x_0 \in \mathbb{R}^n$ höchstens eine[16] Lösung auf $[0, \infty)$ besitzt.

**Lösung:** Sei $x$ eine Lösung von $x' = f(x)$, $x(0) = x_0$ und $y$ eine Lösung derselben Differentialgleichung mit dem Anfangswert $y(0) = y_0$. Dann ist

$$\frac{d}{dt}\|x(t) - y(t)\|^2 = 2[x'(t) - y'(t)]^T [x(t) - y(t)] \le 2l\,\|x(t) - y(t)\|^2.$$

Also ist

$$\frac{d}{dt}\|x(t) - y(t)\|^2 = 2l\,\|x(t) - y(t)\|^2 - r(t)$$

[16] Es kann auch die Existenz einer Lösung nachgewiesen werden, siehe Theorem 1.4.1 bei

K. Strehmel, R. Weiner (1992) *Linear-implizite Runge-Kutta-Methoden und ihre Anwendung*. B. G. Teubner, Stuttgart-Leipzig.

mit einer nichtnegativen, stetigen Funktion $r$. Also genügt $v(t) := \|x(t) - y(t)\|^2$ einer linearen Differentialgleichung erster Ordnung, nämlich $v' = 2lv - r$. Folglich ist

$$\begin{aligned}\|x(t) - y(t)\|^2 &= e^{2lt}\left[\|x(0) - y(0)\| - \int_0^t e^{-2l\tau} r(\tau)\, d\tau\right] \\ &\le e^{2lt}\,\|x(0) - y(0)\|^2 \\ &= e^{2lt}\,\|x_0 - y_0\|^2.\end{aligned}$$

Ist nun $x_0 = y_0$, so erhält man die Eindeutigkeitsaussage.

10. Mit Hilfe von Maple[17] bestimme man das maximale Existenzintervall für die Anfangswertaufgaben:

    (a) $x' = (1 - 2t)/\cos x$, $x(1) = 2$,

    (b) $x' = x/t + 4t^2x^2$, $x(1) = 1/15$,

    (c) $x' = x(1 - x)$, $x(0) = 2$.

    **Lösung:** Lösung von $x' = (1 - 2t)/\cos x$, $x(1) = 2$ ist $x(t) = \arcsin(t - t^2 + \sin 2)$. Durch

    ```
    solve({-1<=t-t^2+sin(2),t-t^2+sin(2)<=1},t);
    ```

    erhalten wir, dass das maximale 1-enthaltende Intervall in

    $$\left[\frac{1}{2} + \frac{1}{2}\sqrt{-3 + 4\sin 2}, \frac{1}{2} + \frac{1}{2}\sqrt{5 + 4\sin(2)}\right] = [0.8991208173, 1.969454806]$$

    An den Endpunkten dieses Intervalles hat $x$ den Wert $-\pi/2$ bzw. $\pi/2$, in beiden Fällen verschwindet dort $\cos x(t)$ bzw. ist $x$ nicht differenzierbar. Das maximale Existenzintervall ist also

    $$(t_{\min}, t_{\max}) := \left(\frac{1}{2} + \frac{1}{2}\sqrt{-3 + 4\sin 2}, \frac{1}{2} + \frac{1}{2}\sqrt{5 + 4\sin(2)}\right).$$

    Die Anfangswertaufgabe $x' = x/t + 4t^2x^2$, $x(1) = 1/15$, hat $x(t) = t/(16 - t^4)$ als Lösung. Das maximale Existenzintervall ist also $(t_{\min}, t_{\max}) = (-4, 4)$.

    Die Anfangswertaufgabe $x' = x(1 - x)$, $x(0) = 2$, hat die Lösung $x(t) = 1/(1 - \frac{1}{2}e^{-t})$. Das maximale Existenzintervall ist daher $(t_{\min}, t_{\max}) = (-\ln 2, \infty)$.

## 5.2.2 Aufgaben zu Abschnitt 2.2

1. Man bestimme mit Hilfe von Maple sämtliche Lösungen von

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} (3t-1) & -(1-t) \\ -(t+2) & (t-2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} te^{t^2} \\ -e^{t^2} \end{pmatrix}.$$

[17] Die Aufgabe haben wir dem Buch

D. Betounes (2001) *Differential Equations. Theory and Applications with Maple*. Springer-Verlag, Berlin-New York-Heidelberg

entnommen.

**Lösung:** Als ein durch $X(0) = I$ normiertes Fundamentalsystem berechnet man

$$X(t) = \frac{1}{9}e^{(-3+t)t}\begin{pmatrix} 2+7e^{3t}-3t & 2-2e^{3t}-3t \\ 7-7e^{3t}+3t & 7+2e^{3t}+3t \end{pmatrix}.$$

Dies geschah durch

```
dsolve({diff(x_1(t),t)=(3*t-1)*x_1(t)-(1-t)*x_2(t),
diff(x_2(t),t)=-(t+2)*x_1(t)+(t-2)*x_2(t),x_1(0)=1,x_2(0)=0},
{x_1(t),x_2(t)});
dsolve({diff(x_1(t),t)=(3*t-1)*x_1(t)-(1-t)*x_2(t),
diff(x_2(t),t)=-(t+2)*x_1(t)+(t-2)*x_2(t),x_1(0)=0,x_2(0)=1},
{x_1(t),x_2(t)});
```

Als spezielle Lösung der inhomogenen Gleichung (mit zur Zeit $t = 0$ verschwindendem Anfangswert) erhält man

$$x^*(t) = \frac{1}{81}e^{(-3+t)t}\begin{pmatrix} 8-12t+e^{3t}(-8+36t+9t^2+9t^3) \\ 28+12t+e^{3t}(-28-9t-9t^2-9t^3) \end{pmatrix}.$$

Alle Lösungen sind durch $X(t)c + x^*(t)$ mit beliebigem $c \in \mathbb{R}^2$ gegeben.

2. Sei

$$A := \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix}$$

mit reellem, positiven $\omega$. Man berechne (wie auch immer) $e^{At}$, ferner gebe man eine Darstellung der Lösung von

$$\text{(P)} \qquad x'' + \omega^2 x = g(t), \qquad x(0) = x_0, \quad x'(0) = x_0'$$

an, wobei $x_0, x_0'$ gegebene reelle Zahlen sind und $g(\cdot)$ auf $\mathbb{R}$ stetig ist.

**Lösung:** Offenbar ist

$$e^{At} = \begin{pmatrix} \cos\omega t & \frac{1}{\omega}\sin\omega t \\ -\omega\sin\omega t & \cos\omega t \end{pmatrix}.$$

Umscheiben von (P) in ein System ergibt

$$z' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ g(t) \end{pmatrix}, \qquad z(0) = \begin{pmatrix} x_0 \\ x_0' \end{pmatrix}.$$

Für die Lösung von (P) samt ihrer Ableitung erhält man daher die Darstellung

$$\begin{aligned} \begin{pmatrix} x(t) \\ x'(t) \end{pmatrix} &= e^{At}\begin{pmatrix} x_0 \\ x_0' \end{pmatrix} + \int_0^t e^{A(t-s)}\begin{pmatrix} 0 \\ g(s) \end{pmatrix} ds \\ &= \begin{pmatrix} \cos\omega t & \frac{1}{\omega}\sin\omega t \\ -\omega\sin\omega t & \cos\omega t \end{pmatrix}\begin{pmatrix} x_0 \\ x_0' \end{pmatrix} \\ &\qquad + \int_0^t \begin{pmatrix} \cos\omega(t-s) & \frac{1}{\omega}\sin\omega(t-s) \\ -\omega\sin\omega(t-s) & \cos\omega(t-s) \end{pmatrix}\begin{pmatrix} 0 \\ g(s) \end{pmatrix} ds. \end{aligned}$$

Daher ist

$$x(t) = x_0\cos\omega t + \frac{x_0'}{\omega}\sin\omega t + \frac{1}{\omega}\int_0^t \sin\omega(t-s)g(s)\,ds.$$

3. Für[18] jede der folgenden Matrizen $A$ berechne man das Fundamentalsystem $e^{At}$ zu $x' = Ax$, wobei Maple benutzt werden darf.

   (a)
   $$A := \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix},$$

   (b)
   $$A := \begin{pmatrix} -1 & 1 \\ 0 & -2 \end{pmatrix},$$

   (c)
   $$A := \begin{pmatrix} -1 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

   (d)
   $$A := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

   **Lösung:** Wir erhalten:

   (a) Die angegebene Matrix ist eine Diagonalmatrix, es ist
   $$e^{At} = \begin{pmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{pmatrix},$$

   (b) Die angegebene Matrix ist einer Diagonalmatrix ähnlich. Es ist
   $$e^{At} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} e^{-2t} & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{pmatrix}.$$

   (c) Auch hier ist $A$ einer Diagonalmatrix ähnlich. Es ist
   $$\begin{aligned} e^{At} &= \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-2t} * 0 & 0 & \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{5t} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} e^{-t} & e^{-t} - e^{-2t} & 0 \\ 0 & e^{-2t} & 0 \\ 0 & 0 & e^{5t} \end{pmatrix}. \end{aligned}$$

   (d) Hier hat $A$ die Form eines Jordan-Blocks. Aus Teil 3 von Lemma 2.2 erhalten wir, dass
   $$e^{At} = \begin{pmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$$

[18] Diese sehr einfache Übungsaufgabe haben wir

D. Betounes (2001) *Differential Equations: Theory and Applications with Maple.* Springer-Verlag, New York-Berlin-Heidelberg

entnommen.

4. Für zwei durch eine Feder gekoppelte Pendel gleicher Masse $m = 1$ und gleicher Länge $l$ lauten die Bewegungsgleichungen

$$\begin{aligned} \ddot{x} &= -\alpha x - k(x-y) \\ \ddot{y} &= -\alpha y - k(y-x), \end{aligned}$$

wobei $g$ die Erdbeschleunigung und $k$ die (positive) Federkonstante bedeuten und $\alpha := g/l$ gesetzt ist. Schreibt man die beiden Differentialgleichungen zweiter Ordnung als ein System von vier Differentialgleichungen erster Ordnung, so erhält man ein homogenes System mit der Koeffizientenmatrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -(\alpha+k) & 0 & k & 0 \\ 0 & 0 & 0 & 1 \\ k & 0 & -(\alpha+k) & 0 \end{pmatrix}.$$

Man bestimme ein zu $x' = Ax$ gehörendes (nicht notwendig normiertes) Fundamentalsystem. Ferner löse man die Anfangswertaufgabe für den Fall, dass zur Zeit $t = 0$ ein Pendel angestoßen wird bzw. die Anfangswerte $x(0) = y(0) = y'(0) = 0$, $x'(0) = 1$ vorgegeben werden.

**Lösung:** Durch die Eingabe

```
A:=Matrix([[0,1,0,0],[-(alpha+k),0,k,0],[0,0,0,1],[k,0,-(alpha+k),0]]);
(lambda,C):=LinearAlgebra[Eigenvectors](A);
```

(bzw. `(lambda,C):=Eigenvectors(A);` wenn vorher durch `with(LinearAlgebra):` das `LinearAlgebra`-Paket geladen ist) erhält man die Eigenwerte

$$\lambda_{1,2} = \pm i\sqrt{\alpha}, \qquad \lambda_{3,4} = \pm i\sqrt{\alpha+2k}$$

mit zugehörigen Eigenvektoren

$$c_{1,2} = \begin{pmatrix} 1 \\ \pm i\sqrt{\alpha} \\ 1 \\ \pm i\sqrt{\alpha} \end{pmatrix}, \qquad c_{3,4} = \begin{pmatrix} 1 \\ \pm i\sqrt{\alpha+2k} \\ -1 \\ \mp i\sqrt{\alpha+2k} \end{pmatrix}.$$

Daher erhält man durch

$$\begin{aligned} x_1(t) &= \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \cos\sqrt{\alpha}t - \begin{pmatrix} 0 \\ \sqrt{\alpha} \\ 0 \\ \sqrt{\alpha} \end{pmatrix} \sin\sqrt{\alpha}t, \\ x_2(t) &= \begin{pmatrix} 0 \\ \sqrt{\alpha} \\ 0 \\ \sqrt{\alpha} \end{pmatrix} \cos\sqrt{\alpha}t + \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \sin\sqrt{\alpha}t, \\ x_3(t) &= \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \cos\sqrt{\alpha+2k}t - \begin{pmatrix} 0 \\ \sqrt{\alpha+2k} \\ 0 \\ -\sqrt{\alpha+2k} \end{pmatrix} \sin\sqrt{\alpha+2k}t, \end{aligned}$$

$$x_4(t) = \begin{pmatrix} 0 \\ \sqrt{\alpha+2k} \\ 0 \\ -\sqrt{\alpha+2k} \end{pmatrix} \cos\sqrt{\alpha+2k}t + \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \sin\sqrt{\alpha+2k}t,$$

Spalten eines Fundamentalsystems $X(t)$ zu $x' = Ax$. Die allgemeine Lösung ist eine Linearkombination dieser Spalten. Wir lösen das lineare Gleichungssystem mit der Koeffizientenmatrix $X(0)$ und der rechten Seite $(0,1,0,0)^T$. Mit

```
X_0:=Matrix([[1,0,1,0],[0,sqrt(alpha),0,sqrt(alpha+2*k)],[1,0,-1,0],
[0,sqrt(alpha),0,-sqrt(alpha+2*k)]]);
b:=Vector([0,1,0,0]);
LinearAlgebra[LinearSolve](X_0,b);
```

erhält man, dass die Lösung der Anfangswertaufgabe

$$\begin{aligned} \ddot{x} &= -\alpha x - k(x-y) \\ \ddot{y} &= -\alpha y - k(y-x), \end{aligned} \qquad \begin{pmatrix} x(0) \\ x'(0) \\ y(0) \\ y'(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

durch

$$\begin{pmatrix} x(t) \\ x'(t) \\ y(t) \\ y'(t) \end{pmatrix} = \frac{1}{2\sqrt{\alpha}} x_2(t) + \frac{1}{2\sqrt{\alpha+2k}} x_4(t)$$

gegeben ist. Daher ist

$$\begin{aligned} x(t) &= \frac{1}{2\sqrt{\alpha}} \sin\sqrt{\alpha}t + \frac{1}{2\sqrt{\alpha+2k}} \sin\sqrt{\alpha+2k}t, \\ y(t) &= \frac{1}{2\sqrt{\alpha}} \sin\sqrt{\alpha}t - \frac{1}{2\sqrt{\alpha+2k}} \sin\sqrt{\alpha+2k}t. \end{aligned}$$

5. Man bestimme (wie auch immer)[19] ein reelles Fundamentalsystem von Lösungen der Differentialgleichungssysteme $x' = Ax$ mit

$$A := \begin{pmatrix} 3 & 6 \\ -2 & -3 \end{pmatrix}, \qquad A := \begin{pmatrix} 8 & 1 \\ -4 & 4 \end{pmatrix}.$$

**Lösung:** Im ersten Fall erhält man als Eigenwerte und zugehörige Eigenvektoren

$$\lambda_{1,2} = \pm i\sqrt{3}, \qquad c_{1,2} = \begin{pmatrix} \frac{1}{2}(-3 \mp \sqrt{3}) \\ 1 \end{pmatrix}.$$

Als Spalten eines Fundamentalsystems erhält man

$$\begin{aligned} x_1(t) &= \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} \cos\sqrt{3}t + \begin{pmatrix} \frac{\sqrt{3}}{2} \\ 0 \end{pmatrix} \sin\sqrt{3}t, \\ x_2(t) &= \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ 0 \end{pmatrix} \cos\sqrt{3}t + \begin{pmatrix} -\frac{3}{2} \\ 1 \end{pmatrix} \sin\sqrt{3}t. \end{aligned}$$

[19] Die Aufgabe ist W. Walter (1996, S. 159) entnommen.

Im zweiten Fall hat $A$ den doppelten Eigenwert $\lambda_{1,2} = 6$ und ist nicht diagonalisierbar. Bei Maple wird dies nach dem Aufruf von `Eigenvectors(A)` dadurch deutlich gemacht, dass neben dem Eigenvektor $(1, -2)^T$ noch ein Nullvektor ausgegeben wird. Durch eine Ähnlichkeitstransformation mit der Matrix

$$C := \begin{pmatrix} 2 & 1 \\ -4 & 0 \end{pmatrix}$$

erhält man die Jordansche Normalform von $A$, d. h. es ist

$$\begin{pmatrix} 2 & 1 \\ -4 & 0 \end{pmatrix}^{-1} A \begin{pmatrix} 2 & 1 \\ -4 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 0 & 6 \end{pmatrix}.$$

Dies kann man z. B. durch den Maple-Befehl `JordanForm` erhalten. Daher ist

$$e^{At} = \begin{pmatrix} 2 & 1 \\ -4 & 0 \end{pmatrix} e^{6t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -4 & 0 \end{pmatrix}^{-1} = e^{6t} \begin{pmatrix} 1+2t & t \\ -4t & 1-2t \end{pmatrix}.$$

Etwas schneller hätten wir dies erhalten können, indem wir

```
dsolve({diff(x(t),t)=8*x(t)+y(t),diff(y(t),t)=-4*x(t)+4*y(t),
    x(0)=1,y(0)=0},{x(t),y(t)});
```

und

```
dsolve({diff(x(t),t)=8*x(t)+y(t),diff(y(t),t)=-4*x(t)+4*y(t),
    x(0)=0,y(0)=1},{x(t),y(t)});
```

eingegeben hätten.

6. Sei

$$A := \begin{pmatrix} 0 & 1 & 0 \\ 4 & 3 & -4 \\ 1 & 2 & -1 \end{pmatrix}.$$

Man berechne $e^A$ und vergleiche mit $\sum_{j=0}^{10} A^j/j!$. Man mache hierbei möglichst viele der Rechnungen mit Maple oder MATLAB.

**Lösung:** Wir wenden zunächst MATLAB an. Nach

```
A=[0 1 0;4 3 -4;1 2 -1];format long
expm(A)
```

erhalten wir

$$e^A = \begin{pmatrix} 5.00000000000001 & 3.71828182845906 & -4.00000000000000 \\ 10.87312731383622 & 8.15484548537718 & -10.87312731383620 \\ 7.71828182845905 & 6.43656365691810 & -6.71828182845904 \end{pmatrix}.$$

Als Resultat des kleinen MATLAB-Programms

```
A=[0 1 0;4 3 -4;1 2 -1];format long
E=eye(size(A));
F=eye(size(A));
for j=1:10
      F=F*A/j;
      E=E+F;
end;
E
```

erhalten wir die Matrix

$$\begin{pmatrix} 4.99999889770723 & 3.71828125000000 & -3.99999889770723 \\ 10.87312610229277 & 8.15484485229277 & -10.87312610229277 \\ 7.71828014770723 & 6.43656277557319 & -6.71828014770723 \end{pmatrix}.$$

In MATLAB wird $e^A$ nicht mit Hilfe einer Reihenentwicklung berechnet. Trotzdem findet sich ein Function-File `expm2`, das mit Hilfe der Reihenentwicklung arbeitet. Wir modifizieren es geringfügig, indem wir auch die Anzahl der benötigten Terme ausgeben.

```
function [E,number] = expm4(A)
%EXPM2  Matrix exponential via Taylor series.
%   E = expm4(A) illustrates the classic definition for the
%   matrix exponential.  As a practical numerical method,
%   this is often slow and inaccurate. number is number of
%   terms.
%
%   See also EXPM, EXPM1, EXPM2, EXPM3.

E = zeros(size(A));
F = eye(size(A));
k = 1;
while norm(E+F-E,1) > 0
   E = E + F;
   F = A*F/k;
   k = k+1;
end;
number=k;
```

Als Resultat von `[E,k]=expm4(A)` erhalten wir

$$E = \begin{pmatrix} 5.00000000000000 & 3.71828182845905 & -4.00000000000000 \\ 10.87312731383618 & 8.15484548537713 & -10.87312731383618 \\ 7.71828182845904 & 6.43656365691809 & -6.71828182845904 \end{pmatrix}, \qquad k = 21.$$

Nun arbeiten wir mit Maple. Mit Hilfe von `JordanForm` berechnet man die Jordansche Normalform

$$A = CJC^{-1} \qquad \text{mit} \quad C := \begin{pmatrix} 5 & 4 & -4 \\ 0 & 4 & 0 \\ 5 & 6 & -5 \end{pmatrix}, \quad J := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Mit

```
X:=t->C.Matrix([[1,0,0],[0,exp(t),t*exp(t)],[0,0,exp(t)]]).C^(-1);
X(t);
```

erhalten wir (wir übernehmen den Output von Maple, daher die eckigen Klammern)

$$\begin{bmatrix} 5+4\,t\,e^t-4\,e^t & 1-e^t+2\,t\,e^t & -4-4\,t\,e^t+4\,e^t \\ 4\,t\,e^t & e^t+2\,t\,e^t & -4\,t\,e^t \\ 5+6\,t\,e^t-5\,e^t & 1-e^t+3\,t\,e^t & -4-6\,t\,e^t+5\,e^t \end{bmatrix}$$

Nach `Digits:=15: evalf(X(1));` erhalten wir $e^A$, nämlich

$$\begin{bmatrix} 5. & 3.71828182845905 & -4. \\ 10.8731273138362 & 8.15484548537715 & -10.8731273138362 \\ 7.71828182845905 & 6.43656365691810 & -6.71828182845905 \end{bmatrix}$$

Noch einfacher ist es allerdings, wenn man das `linalg`-Paket ladet. Nach

```
with(linalg):
A:=Matrix([[0,1,0],[4,3,-4],[1,2,-1]]);
exponential(A,t);
```

ist der Output

$$\begin{bmatrix} 5+4\,t\,e^t-4\,e^t & -e^t+1+2\,t\,e^t & -4\,t\,e^t+4\,e^t-4 \\ 4\,t\,e^t & 2\,t\,e^t+e^t & -4\,t\,e^t \\ -5\,e^t+5+6\,t\,e^t & 3\,t\,e^t-e^t+1 & -4-6\,t\,e^t+5\,e^t \end{bmatrix}$$

Bis auf die Reihenfolge von Summanden in den Einträgen ist das dieselbe Lösung wie oben.

Nun noch ein kleines Maple-Programm. Nach ($A$ ist schon besetzt, ferner ist das `LinearAlgebra`-Paket geladen)

```
E:=IdentityMatrix(3): F:=E;
for j from 1 to 10 do
   F:=F.A/j:
   E:=E+F:
end do:
E;
```

erhalten wir

$$\begin{bmatrix} \frac{4535999}{907200} & \frac{23797}{6400} & \frac{-3628799}{907200} \\ \frac{98641}{9072} & \frac{29592301}{3628800} & \frac{-98641}{9072} \\ \frac{5601619}{725760} & \frac{23356999}{3628800} & \frac{-4875859}{725760} \end{bmatrix}$$

Nach anschließendem `evalf(E)` erhalten wir

$$\begin{bmatrix} 4.99999889770723 & 3.71828125000000 & -3.99999889770723 \\ 10.8731261022928 & 8.15484485229277 & -10.8731261022928 \\ 7.71828014770723 & 6.43656277557319 & -6.71828014770723 \end{bmatrix}$$

7. Man bestimme die allgemeine Lösung von

$$x'' - 6x' + 25x = e^{2t}.$$

Anschließend bestimme man die Lösung zu den Anfangswerten $x(0) = 1$, $x'(0) = 0$.

**Lösung:** Eine spezielle Lösung ist $\frac{1}{17}e^{2t}$. Die Nullstellen von $\lambda^2 - 6\lambda + 25 = 0$ sind $\lambda_{1,2} = 3 \pm 4i$. Die allgemeine Lösung der homogenen Aufgabe ist daher $e^{3t}(\alpha \cos 4t + \beta \sin 4t)$, folglich die allgemeine Lösung der inhomogenen Aufgabe

$$x(t) = e^{3t}(\alpha \cos 4t + \beta \sin 4t) + \frac{1}{17}e^{2t}$$

mit beliebigen $\alpha, \beta$. Die gewünschte Lösung der Anfangswertaufgabe ist

$$x(t) = \frac{1}{17}e^{2t} + e^{3t}\left(\frac{16}{17}\cos 4t - \frac{25}{34}\sin 4t\right).$$

Mit Maple hätte man natürlich dieselben Ergebnisse erzielt.

8. Sei

$$A(t) := \begin{pmatrix} 0 & 1 \\ -\sin t & \cos t \end{pmatrix}.$$

Man definiere $\phi(t) := \int_0^t e^{-\sin s}\, ds$ und zeige, dass durch

$$X(t) := \begin{pmatrix} e^{\sin t} & e^{\sin t}\phi(t) \\ \cos t e^{\sin t} & 1 + \cos t e^{\sin t}\phi(t) \end{pmatrix}$$

ein Fundamentalsystem zu $x' = A(t)x$ gegeben ist. Anschließend bestimme man die charakteristischen Multiplikatoren, d. h. die Eigenwerte von $C := X(0)^{-1}X(2\pi)$.

**Lösung:** Es ist

$$\begin{aligned} X'(t) &= \begin{pmatrix} \cos t e^{\sin t} & 1 + \cos t e^{\sin t}\phi(t) \\ (-\sin t + \cos^2 t)e^{\sin t} & \cos t + (-\sin t + \cos^2 t)e^{\sin t}\phi(t) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} e^{\sin t} & e^{\sin t}\phi(t) \\ \cos t e^{\sin t} & 1 + \cos t e^{\sin t}\phi(t) \end{pmatrix} \\ &= A(t)X(t), \end{aligned}$$

also ist $X(t)$ ein Lösungssystem. Es ist

$$\det X(t) = \det \begin{pmatrix} e^{\sin t} & e^{\sin t}\phi(t) \\ \cos t e^{\sin t} & 1 + \cos t e^{\sin t}\phi(t) \end{pmatrix} = e^{\sin t} \neq 0,$$

also ist $X(\cdot)$ ein Fundamentalsystem. Die charakterischen Multiplikatoren sind Eigenwerte der zu $X$ gehörenden Monodromie-Matrix $C := X(0)^{-1}X(2\pi)$. Nun ist

$$\begin{aligned} C &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & \phi(2\pi) \\ 1 & 1 + \phi(2\pi) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & \phi(2\pi) \\ 1 & 1 + \phi(2\pi) \end{pmatrix} \\ &= \begin{pmatrix} 1 & \phi(2\pi) \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Die Monodromie-Matrix hat also den doppelten Eigenwert 1, dies ist auch der charakteristische Multiplikator.

### 5.2.3 Aufgaben zu Abschnitt 2.3

1. Sei
$$A := \begin{pmatrix} 3 & 6 \\ -2 & -3 \end{pmatrix} \qquad \text{bzw.} \qquad A := \begin{pmatrix} -2 & -1 \\ 4 & -1 \end{pmatrix}.$$
Man untersuche die Stabilität der Nulllösung des Differentialgleichungssystems $x' = Ax$ und veranschauliche beide Fälle durch Phasenportraits.

**Lösung:** Im ersten Fall hat $A$ die Eigenwerte $\lambda_{1,2} = \pm\sqrt{3}i$. Nach Satz 3.2 ist die Nulllösung stabil. In Abbildung 5.14 links geben wir zwei Phasenbahnen an. Im zweiten Fall sind die Eigenwerte $\lambda_{1,2} = -\frac{3}{2} \pm \frac{1}{2}\sqrt{15}i$, die Nulllösung ist also asymptotisch stabil. In Abbildung 5.14 rechts findet man zwei Phasenbahnen.

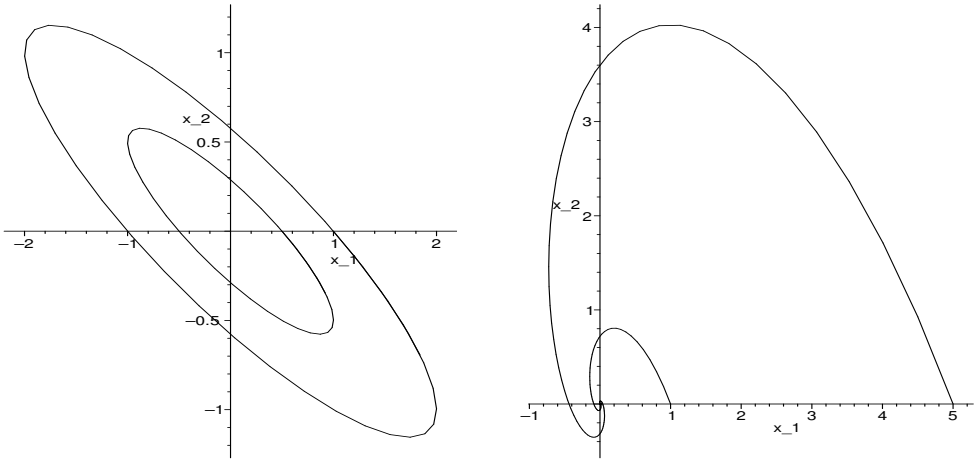


Abbildung 5.14: Phasenbahnen

2. Sei $A := \text{tridiag}(1, -2, 1) \in \mathbb{R}^{n\times n}$ die Tridiagonalmatrix, bei der alle Hauptdiagonaleinträge gleich $-2$ und alle Nebendiagonaleinträge gleich 1 sind. Man zeige, dass die Nulllösung von $x' = Ax$ asymptotisch stabil ist.

**Lösung:** Zu zeigen ist, dass alle Eigenwerte von $A$ negativen Realteil haben. Nun ist $A$ symmetrisch, also alle Eigenwerte reell. Sei $\lambda$ ein Eigenwert von $A$ und $x$ ein zugehöriger, durch $\|x\|_\infty = 1$ normierter Eigenwert. Sei $|x_i| = 1$. Wegen
$$x_{i-1} - 2x_i + x_{i+1} = \lambda x_i$$
(ist $i = 1$, so ist $x_{i-1} = 0$, ist $i = n$, so ist $x_{i+1} = 0$) ist
$$|2 + \lambda| = |2 + \lambda|\,|x_i| = |x_{i-1} + x_{i+1}| \le |x_{i-1}| + |x_{i+1}| \le 2\,|x_i| = 2,$$
also $\lambda \in [-4, 0]$. Wäre $\lambda = 0$ ein Eigenwert, so folgt aus obiger Ungleichung, dass auch $|x_{i-1}| = |x_i| = |x_{i+1}| = 1$. Also sind alle Komponenten von $x$ gleich 1. Die erste (oder die letzte) Gleichung ergibt einen Widerspruch.

Einen alternativen Beweis kann man dadurch bringen, dass man die Eigenwerte explizit berechnet. Sie sind durch
$$\lambda_j = -4\sin^2\left(\frac{j\pi}{2(n+1)}\right), \qquad j = 1, \ldots, n,$$
gegeben.

3. Man zeige: Hat das Polynom

$$p(z) := z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

mit den reellen Koeffizienten $a_0, \ldots, a_{n-1}$ nur Nullstellen mit negativem Realteil, so sind die Koeffizienten $a_0, \ldots, a_{n-1}$ von $p$ notwendigerweise positiv.

**Lösung:** Seien $-\alpha_j$, $j = 1, \ldots, k$, die reellen Nullstellen und $-\beta_j \pm i\gamma_j$, $j = 1, \ldots, m$, die komplexen Nullstellen von $p$, also $\alpha_j > 0$, $j = 1, \ldots, k$, und $\beta_j > 0$, $j = 1, \ldots, m$. Dann ist

$$\begin{aligned} p(z) &= \prod_{j=1}^{k}(z+\alpha_j)\prod_{j=1}^{m}(z+\beta_j - i\gamma_j)(z+\beta_j+i\gamma_j) \\ &= \prod_{j=1}^{k}(z+\alpha_j)\prod_{j=1}^{m}(z^2+2\beta_j z+\beta_j^2+\gamma_j^2). \end{aligned}$$

Daher ist $p$ Produkt von Polynomen mit positiven Koeffizienten, besitzt also selbst nur positive Koeffizienten. Die Aussage ist bewiesen.

4. Man betrachte die sogenannte Hillsche Differentialgleichung

$$x'' + b(t)x = 0,$$

wobei $b(\cdot)$ stetig und $T$-periodisch. Man zeige, dass die Nulllösung zum äquivalenten System

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b(t) & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

*nicht* asymptotisch stabil ist, da das Produkt der beiden charakteristischen Multiplikatoren 1 ist.

**Lösung:** Wir wenden Satz 3.5 an. Hiernach ist die Nulllösung zu $x' = A(t)x$, wobei die Matrixfunktion stetig und $T$-periodisch ist, genau dann asymptotisch stabil, wenn alle charakteristischen Multiplikatoren betragsmäßig kleiner als 1 sind. Sei

$$X(t) = \begin{pmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{pmatrix}$$

das durch $X(0) = I$ normierte Fundamentalsystem, $X(T)$ ist die zugehörige Monodromiematrix. Die charakteristischen Multiplikatoren $\rho$ sind die Eigenwerte von $X(T)$. Wegen der Wronski-Beziehung (siehe Lemma 2.1) ist

$$\det X(T) = \exp\left(\int_0^T \operatorname{tr} A(s)\, ds\right) = 1.$$

Die Determinante einer Matrix ist das Produkt ihrer Eigenwerte. Das Produkt der beiden charakteristischen Multiolikatoren ist also 1, daher kann bei der Hill'schen Differentialgleichung keine asymptotische Stabilität vorliegen.

5. Wie in Aufgabe 4 betrachte man die Hillsche Differentialgleichung

$$x'' + b(t)x = 0,$$

wobei $b(\cdot)$ stetig und $T$-periodisch. Ferner sei $b(\cdot)$ eine gerade Funktion. Sei

$$X(t) = \begin{pmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{pmatrix}$$

das durch $X(0) = I$ normierte Fundamentalsystem zu dem äquivalenten System

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b(t) & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Man zeige:

(a) Es ist $x_1(\cdot)$ gerade und $x_2(\cdot)$ ungerade.

(b) Es ist $X(T)^{-1} = X(-T)$.

(c) Es ist $x_1(T) = x_2'(T)$.

(d) Die beiden chrakteristischen Multiplikatoren sind Nullstellen der quadratischen Gleichung $\mu^2 - 2x_1(T)\mu + 1 = 0$. Das Stabilitätsverhalten der Nulllösung der Hillschen Gleichung wird also im wesentlichen nur durch $x_1(T)$ bestimmt. Genauer gilt: Ist $|x_1(T)| < 1$, so ist die Nulllösung stabil, ist $|x_1(T)| > 1$ so ist sie instabil.

(e) Gegeben sei die spezielle Mathieusche Differentialgleichung

$$x'' + (\delta + \gamma \cos 2t)x = 0,$$

hier ist also $T = \pi$. Man bestimme numerisch das Stabilitätsverhalten der Nulllösung für $(\delta, \gamma) = (1, 2),\ (\frac{1}{4}, 1)$ und illustriere dies durch Bahnen in der Phasenebene.

**Lösung:** $x_1$ ist die Lösung von

$$x'' + b(t)x = 0, \qquad x(0) = 1,\ x'(0) = 0.$$

Wir definieren $y_1(t) := x_1(-t)$. Dann ist

$$y_1''(t) + b(t)y_1(t) = x_1''(-t) + b(t)x_1(-t) = x_1''(-t) + b(-t)x_1(-t) = 0,$$

ferner genügt $y_1$ auch den Anfangsbedingungen $y_1(0) = 1$, $y_1'(0) = 0$. Daher ist $y_1(t) = x_1(t)$ bzw. $x_1$ gerade. Entsprechend zeigt man, dass $x_2$ ungerade.

Wegen $X(t + T) = X(t)X(T)$ für alle $t$ ist (setze $t := -T$) $X(T)^{-1} = X(-T)$.

Wegen Lemma 2.1 ist $\det X(t) = 1$ für alle $t$ und daher wegen der gerade eben bewiesenen Aussage

$$\begin{aligned} \begin{pmatrix} x_2'(T) & -x_2(T) \\ -x_1'(T) & x_1(T) \end{pmatrix} &= X(T)^{-1} \\ &= X(-T) \\ &= \begin{pmatrix} x_1(-T) & x_2(-T) \\ x_1'(-T) & x_2'(-T) \end{pmatrix} \\ &= \begin{pmatrix} x_1(T) & -x_2(T) \\ -x_1'(T) & x_2'(T) \end{pmatrix}. \end{aligned}$$

Hieraus folgt $x_1(T) = x_2'(T)$.

Die charakterischen Multiplikatoren sind Eigenwerte von $X(T)$ bzw. Lösungen von

$$0 = \mu^2 - \mathrm{tr}X(T)\mu + \det X(T) = \mu^2 - 2x_1(T)\mu + 1.$$

Diese sind also gegeben durch

$$\mu_1 = x_1(T) + \sqrt{x_1(T)^2 - 1}, \qquad \mu_2 = x_1(T) - \sqrt{x_1(T)^2 - 1}.$$

Ist $|x_1(T)| < 1$, so ist $|\mu_1| = |\mu_2| = 1$, die Nulllösung also stabil. Ist dagegen $|x_1(T)| > 1$, so ist $\mu_1 > 1$ für $x_1(T) > 1$ bzw. $\mu_2 < -1$ für $x_1(T) < -1$, die Nulllösung ist also nicht stabil.

Nun lösen wir die Aufgabe

$$x'' + (\delta + \gamma\cos t)x = 0, \qquad x(0) = 1, x'(0) = 0$$

für $(\delta, \gamma) = (1, 2)$ bzw. $(\delta, \gamma) = (\frac{1}{4}, 1)$ numerisch und werten diese Näherung in $T = \pi$ aus. Nach

```
sol:=dsolve({diff(x(t),t$2)+(1+2*cos(2*t))*x(t),
x(0)=1,D(x)(0)=0},x(t),numeric);
subs(sol(Pi),x(t));
```

erhält man $x_1(\pi) \approx -2.198$, es liegt also keine Stabilität vor. Im zweiten Fall ist $x_1(\pi) \approx -0.521$, die Nulllösung wird also stabil sein. Nun veranschaulichen wir uns Bahnen in der Phasenebene. Nach

```
with(DEtools):
phaseportrait([diff(x_1(t),t)=x_2(t),
diff(x_2(t),t)=-(1+2*cos(2*t))*x_1(t)],[x_1(t),x_2(t)],
t=0..10,[[x_1(0)=1,x_2(0)=0]],
arrows=NONE,linecolor=black,stepsize=0.1,thickness=0);
```

erhalten wir die Bahn in Abbildung 5.15 links. Hier ist also $(\delta, \gamma) = (1, 2)$, ein instabiler Fall. Ganz anders sieht das Bild für $(\delta, \gamma) = (\frac{1}{4}, 1)$ aus, siehe Abbildung 5.15 rechts. Dieses haben wir durch

```
with(DEtools):
phaseportrait([diff(x_1(t),t)=x_2(t),
diff(x_2(t),t)=-(1/4+cos(2*t))*x_1(t)],[x_1(t),x_2(t)],
t=0..10,[[x_1(0)=1,x_2(0)=0]],
arrows=NONE,linecolor=black,stepsize=0.1,thickness=0);
```

gewonnen. Eigentlich dürfen Phasenbahnen sich wegen der eindeutigen Lösbarkeit von Anfangswertaufgaben nicht schneiden, dies muss daher auf Rundungsfehler zurückzuführen sein.

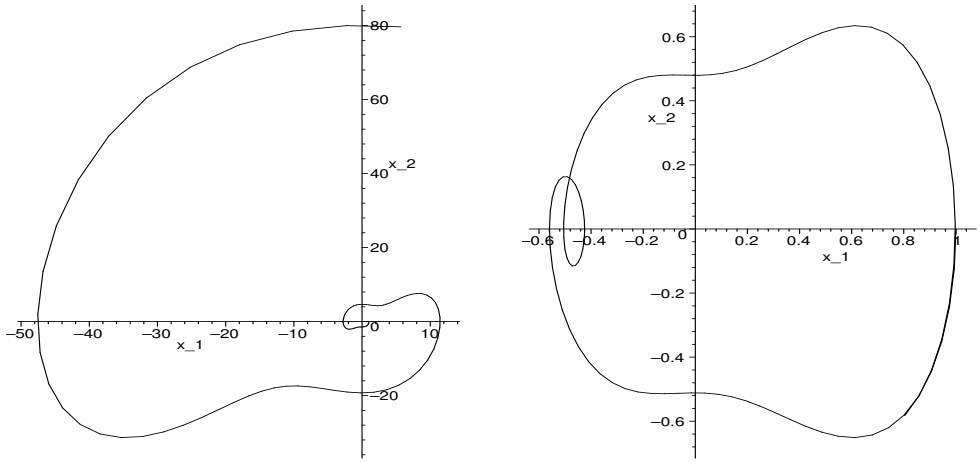


Abbildung 5.15: Phasenbahnen bei der Hillschen Differentialgleichung

6. Man zeige in Lemma 3.9 die Umkehrung. Genauer beweise man: Gibt es zu $A \in \mathbb{R}^{n\times n}$ eine symmetrische, positiv definite Matrix $B \in \mathbb{R}^{n\times n}$ derart, dass $A^TB + BA$ negativ definit ist, so ist $\Re(\lambda) < 0$ für alle Eigenwerte $\lambda$ von $A$.

   **Lösung:** Wir betrachten das Differentialgleichungssystem $x' = Ax$ und hierzu die durch $V(x) := x^TBx$ definierte positiv definite Funktion $V\colon \mathbb{R}^n \longrightarrow \mathbb{R}$. Es ist

   $$\dot{V}(x) = x^T(A^TB + BA)x < 0 \qquad \text{für alle } x \in \mathbb{R}^n \setminus \{0\},$$

   nach Satz 3.8 ist die Nulllösung eine asymptotisch stabile Lösung von $x' = Ax$. Wegen des zweiten Teils von Satz 3.2 ist $\Re(\lambda) < 0$ für jeden Eigenwert $\lambda$ von $A$.

7. Man betrachte die Differentialgleichung zweiter Ordnung

   $$x'' + h(x) = 0,$$

   wobei $h \in C(\mathbb{R})$ und $xh(x) > 0$ für alle $x \neq 0$ (woraus $h(0) = 0$ folgt). Man zeige, dass die Nulllösung zu dieser Differentialgleichung bzw. dem äquivalenten Differentialgleichungssystem

   $$\begin{aligned} x_1' &= x_2 \\ x_2' &= -h(x_1) \end{aligned}$$

   stabil ist.

   **Lösung:** Wir definieren $V\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ durch

   $$V(x_1, x_2) := \int_0^{x_1} h(s)\,ds + \frac{1}{2}x_2^2.$$

   Offensichtlich ist $V$ positiv definit auf $\mathbb{R}^2$. Dann ist

   $$\dot{V}(x_1, x_2) = \nabla V(x_1, x_2)^T \begin{pmatrix} x_2 \\ -h(x_1) \end{pmatrix} = \begin{pmatrix} h(x_1) \\ x_2 \end{pmatrix}^T \begin{pmatrix} x_2 \\ -h(x_1) \end{pmatrix} = 0,$$

   aus dem ersten Teil von Satz 3.2 folgt die Behauptung.

8. Man zeige: Die Nulllösungen der Differentialgleichungssysteme

$$\begin{aligned} x_1' &= -x_2\sqrt{x_1^2+x_2^2} \\ x_2' &= x_1\sqrt{x_1^2+x_2^2}, \end{aligned} \qquad \text{bzw.} \qquad \begin{aligned} x_1' &= (x_1^2+x_2^2-1)x_1 - x_2 \\ x_2' &= x_1 + (x_1^2+x_2^2-1)x_2 \end{aligned}$$

sind stabil bzw. asymptotisch stabil. Weiter zeige man, dass die Nulllösung von

$$\begin{aligned} x_1' &= (1-x_1^2-x_2^2)x_1 - x_2 \\ x_2' &= x_1 + (1-x_1^2-x_2^2)x_2 \end{aligned}$$

nicht stabil ist.

**Lösung:** Man definiere die positiv definite Funktion $V\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ durch $V(x_1,x_2) := \frac{1}{2}(x_1^2+x_2^2)$. Im ersten Fall ist

$$\dot{V}(x_1,x_2) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} -x_2\sqrt{x_1^2+x_2^2} \\ x_1\sqrt{x_1^2+x_2^2} \end{pmatrix} = 0,$$

aus Satz 3.8 folgt die Stabilität der Nulllösung. Im zweiten Fall ist

$$\dot{V}(x_1,x_2) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} (x_1^2+x_2^2-1)x_1 - x_2 \\ x_1 + (x_1^2+x_2^2-1)x_2 \end{pmatrix} = (x_1^2+x_2^2)(x_1^2+x_2^2-1),$$

also ist $\dot{V}$ auf der offenen (euklidischen) Einheitskugel negativ definit. Wiederum aus Satz 3.8, diesmal aber dem zweiten Teil, folgt die Behauptung.

Auch für den Beweis des letzten Teils der Aufgabe benutzen wir die gerade definierte Funktion $V$. Mit einer Lösung $x(t) = (x_1(t), x_2(t))$ ist dann

$$\frac{d}{dt}V(x_1(t),x_2(t)) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}^T \begin{pmatrix} x_1'(t) \\ x_2'(t) \end{pmatrix} = (x_1(t)^2+x_2(t)^2)[1-(x_1(t)^2+x_2(t)^2)].$$

Daher genügt

$$V(t) := V(x_1(t),x_2(t))$$

der Anfangswertaufgabe

$$V' = 2V(1-2V), \qquad V(0) = V_0 := \frac{1}{2}[x_1(0)^2+x_2(0)^2].$$

Als Lösung erhält man

$$V(t) = \frac{V_0}{2V_0 + e^{-2t}(1-2V_0)}.$$

Daher ist $\lim_{t\to\infty} V(t) = \frac{1}{2}$ bzw. $\lim_{t\to\infty} \|x(t)\|^2 = 1$, ein Widerspruch zur Stabilität.

In Abbildung 5.16 tragen wir Bahnen zu den Anfangswerten $x_1(0) = 1$, $x_2(0) = 0$ bzw. $x_1(0) = x_2(0) = \frac{1}{2}$ für die beiden letzten Fälle ein. Im zweiten Fall sieht man, wie sich die Bahn einem Kreis annähert.

9. Man zeige, dass $(\frac{1}{2},\frac{1}{2})$ eine asymptotisch stabile Lösung des Differentialgleichungssystems

$$\begin{aligned} x_1' &= x_1(1-x_1-x_2) \\ x_2' &= x_2(\tfrac{3}{4}-x_2-\tfrac{1}{2}x_1) \end{aligned}$$

ist.

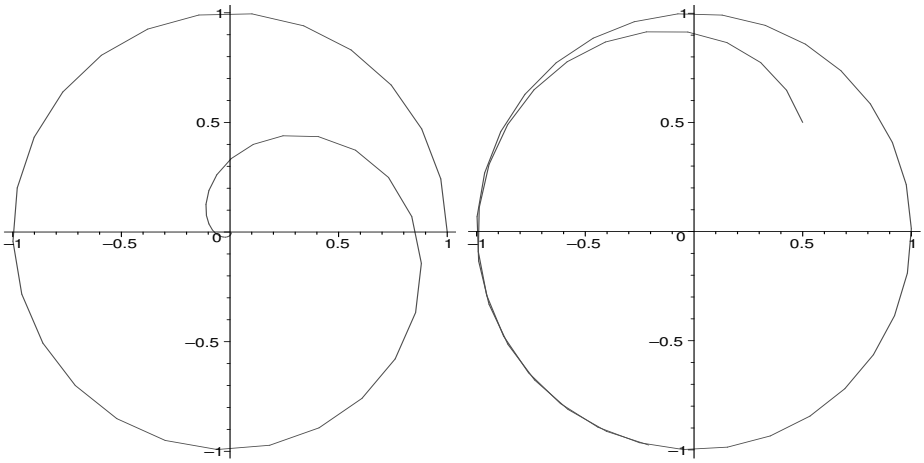


Abbildung 5.16: Asymptotisch stabile und instabile Nulllösung

**Lösung:** Nach der Variablentransformation $x_1 = \frac{1}{2} + y_1$, $x_2 = \frac{1}{2} + y_2$ hat man die asymptotische Stabilität der Nulllösung des Systems

$$\begin{aligned} y_1' &= -\tfrac{1}{2}y_1 - \tfrac{1}{2}y_2 - y_1^2 - y_1y_2 \\ y_2' &= -\tfrac{1}{4}y_1 - \tfrac{1}{2}y_2 - \tfrac{1}{2}y_1y_2 - y_2^2 \end{aligned}$$

bzw. von

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} -y_1^2 - y_1y_2 \\ -\frac{1}{2}y_1y_2 - y_2^2 \end{pmatrix}.$$

Die Matrix

$$A := \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{4} & -\frac{1}{2} \end{pmatrix}$$

hat die beiden reellen negativen Eigenwerte $\lambda_{1,2} = -\frac{1}{2} \pm \frac{1}{4}\sqrt{2}$. Wir zeigen, dass mit

$$g(y) := \begin{pmatrix} -y_1^2 - y_1y_2 \\ -\frac{1}{2}y_1y_2 - y_2^2 \end{pmatrix}$$

gilt, dass $g(y) = o(\|y\|)$, woraus dann die Behauptung folgt (siehe z. B. die Bemerkung im Anschluss an Lemma 3.9). Dies ist aber klar, denn

$$g'(y) = \begin{pmatrix} -2y_1 - y_2 & -y_1 \\ -\frac{1}{2}y_1 & -\frac{1}{2}y_1 - 2y_2 \end{pmatrix}$$

und folglich $g'(0) = 0$.

10. Gegeben sei das Differentialgleichungssystem (Lorenz-Attraktor)

$$\begin{aligned} x_1' &= -\sigma x_1 + \sigma x_2 \\ x_2' &= rx_1 - x_2 - x_1x_3 \\ x_3' &= -bx_3 + x_1x_2, \end{aligned}$$

wobei $b, r, \sigma$ positive Konstanten sind. Man zeige, dass die triviale Lösung asymptotisch stabil ist, wenn $r \in (0, 1)$. Mit Hilfe des im Anschluss von Satz 3.6 (ohne Beweis) angegebenen Instabilitätssatzes begründe man, dass die Nullösung für $r > 1$ nicht stabil ist.

**Lösung:** Wir schreiben das gegebene Differentialgleichungssystem in der Form $x' = Ax + g(x)$, wobei

$$A := \begin{pmatrix} -\sigma & \sigma & 0 \\ r & -1 & 0 \\ 0 & 0 & -b \end{pmatrix}, \qquad g(x) := \begin{pmatrix} 0 \\ -x_1 x_3 \\ x_1 x_2 \end{pmatrix}.$$

Die Eigenwerte von $A$ sind

$$\lambda_1 := -b, \qquad \lambda_{2,3} := -\frac{1}{2}\sigma - \frac{1}{2} \pm \frac{1}{2}\sqrt{\sigma^2 - 2\sigma + 1 + 4r\sigma}.$$

Alle drei Eigenwerte sind reell. Sie sind alle drei negativ genau dann, wenn $-(\sigma+1) + \sqrt{(\sigma-1)^2 + 4r\sigma} < 0$ bzw. $r \in (0,1)$. Da $g(0) = 0$ und $g'(0) = 0$ folgt die asymptotische Stabilität der trivialen Lösung, siehe Satz 3.6 bzw. die Bemerkung im Anschluss an Lemma 3.9. Für $r > 1$ ist $\lambda_2 > 0$, aus dem Instabilitättssatz folgt, dass die triviale Lösung nicht stabil ist.

11. Gegeben sei das Differentialgleichungssystem

$$\begin{aligned} x' &= ax - bxy - ex^2, \\ y' &= -cy + dxy - fy^2 \end{aligned}$$

mit positiven Konstanten $a, b, c, d, e, f$. Dieses hat den Gleichgewichtspunkt bzw. die konstante Lösung

$$(\hat{x}, \hat{y}) = \frac{1}{ef + bd}(af + bc, ad - ce).$$

Man zeige, dass dies für $(a, b, c, d, e, f) := (2, 0.01, 1, 0.01, 0.001, 0.001)$ (siehe Abbildung 1.6) eine asymptotisch stabile Lösung ist.

**Lösung:** Wir machen die Variablentransformation

$$x = \hat{x} + u, \qquad y = \hat{y} + v$$

und haben wir die Stabilität der Nulllösung des Differentialgleichungssystems

$$\begin{aligned} u' &= (a - b\hat{y} - 2e\hat{x})u - b\hat{x}v - buv - eu^2, \\ v' &= d\hat{y}u + (-c + d\hat{x} - 2f\hat{y})v + duv - fv^2 \end{aligned}$$

zu untersuchen. Einsetzen von $(\hat{x}, \hat{y})$ liefert das System

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \frac{1}{ef + bd} \begin{pmatrix} -e(af + bc) & -b(af + bc) \\ d(ad - ce) & -f(ad - ce) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} -buv - eu^2 \\ duv - fv^2 \end{pmatrix}.$$

Die Eigenwerte von

$$A := \frac{1}{ef + bd} \begin{pmatrix} -e(af + bc) & -b(af + bc) \\ d(ad - ce) & -f(ad - ce) \end{pmatrix} \approx \begin{pmatrix} -0.1188 & -1.1881 \\ 1.8812 & -0.1181 \end{pmatrix}$$

sind $\lambda_{1,2} = -0.1184 \pm 1.4950\, i$. Damit folgt die Behauptung wieder aus dem Satz 3.6 bzw. der Bemerkung im Anschluss an Lemma 3.9.

# 5.3 Aufgaben zu Kapitel 3

## 5.3.1 Aufgaben zu Abschnitt 3.1

1. Man zeige: Ist $p \in \Pi_3$ (kubisches Polynom), so ist

$$\int_a^b p(t)\,dt = \frac{b-a}{6}\,[p(a) + 4\,p(\tfrac{1}{2}\,(a+b)) + p(b)].$$

**Lösung:** Es bilden 1, $t - \frac{1}{2}(a+b)$, $(t - \frac{1}{2}(a+b))^2$ und $(t - \frac{1}{2}(a+b))^3$ eine Basis von $\Pi_3$. Es genügt, die behauptete Gleichung für die Basiselemente nachzuweisen. Wegen

$$b - a = \int_a^b 1\,dt = \frac{b-a}{6}[1+4+1] = b-a$$

und

$$\tfrac{1}{12}(b-a)^3 = \int_a^b [t - \tfrac{1}{2}(a+b)]^2\,dt = \frac{b-a}{6}[\tfrac{1}{4}(b-a)^2 + \tfrac{1}{4}(b-a)^2] = \tfrac{1}{12}(b-a)^3$$

ist dies für das erste und dritte Basiselement richtig, was trivialerweise auch für das zweite und vierte gilt.

2. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := a_1 f(t,u) + a_2 f(t + b_1 h, u + b_2 h f(t,u))$$

und zeige, dass dieses die Ordnung 2 besitzt, falls

$$a_1 + a_2 = 1, \qquad a_2 b_1 = \frac{1}{2}, \qquad a_2 b_2 = \frac{1}{2}.$$

Spezialfälle erhält man übrigens für $a_1 = 0$, $a_2 = 1$, $b_1 = b_2 = \frac{1}{2}$ (modifiziertes Euler-Verfahren) und für $a_1 = a_2 = \frac{1}{2}$, $b_1 = b_2 = 1$ (Heun-Verfahren).

**Lösung:** Bei gegebenem $(t,u) \in [t_0, T] \times \mathbb{R}^n$ sei $z$ die Lösung von $z' = f(s,z)$, $z(t) = u$. Der lokale Diskretisierungsfehler ist gegeben durch

$$\begin{aligned}
\Delta(h,f)(t,u)) &= \frac{z(t+h) - z(t)}{h} - \Phi(h,f)(t,z(t)) \\
&= z'(t) + \frac{h}{2} z''(t) + O(h)^2 \\
&\qquad - [a_1 f(t,z(t)) + a_2 f(t + b_1 h, z(t) + b_2 h f(t,z(t))] \\
&= f(t,z(t)) + \frac{h}{2}[f_t(t,z(t)) + f_x(t,z(t)) f(t,z(t))] + O(h^2) \\
&\qquad - \{(a_1+a_2) f(t,z(t)) \\
&\qquad + a_2[b_1 h f_t(t,z(t)) + b_2 h f_x(t,z(t)) f(t,z(t))]\} \\
&= [1 - (a_1+a_2)] f(t,z(t)) + h f_t(t,z(t)) \left[\frac{1}{2} - a_2 b_1\right] \\
&\qquad + h f_x(t,z(t)) f(t,x(t)) \left[\frac{1}{2} - a_2 b_2\right] + O(h^2),
\end{aligned}$$

woraus man die Behauptung abliest.

3. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{4}k_1 + \frac{3}{4}k_3,$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t+\tfrac{1}{3}h, u+\tfrac{1}{3}hk_1), \qquad k_3 := f(t+\tfrac{2}{3}h, u+\tfrac{2}{3}hk_2).$$

Man zeige, dass dies ein Verfahren der Ordnung 3 ist. Hierbei darf man sich auf den Fall einer Differentialgleichung erster Ordnung, also $n = 1$, beschränken. Anschließend löse man diese Aufgabe mit Maple.

**Lösung:** Bei gegebenem $(t,u) \in [t_0,T] \times \mathbb{R}^n$ sei $z$ die Lösung von $z' = f(s,z)$, $z(t) = u$. Statt des Argumentes $(t, z(t))$ schreiben wir im folgenden nur $t$. Es ist also $z'(t) := f(t,z(t)) = f(t)$, daher

$$z''(t) = f_t(t) + f_x(t)f(t)$$

und folglich

$$\begin{aligned} z'''(t) &= f_{tt}(t) + f_{tx}(t)f(t) + [f_{xt}(t) + f_{xx}(t)f(t)]f(t) + f_x(t)[f_t(t) + f_x(t)f(t)] \\ &= f_{tt}(t) + 2f_{xt}(t)f(t) + f_{xx}(t)f^2(t) + f_x(t)[f_t(t) + f_x(t)f(t)]. \end{aligned}$$

Daher ist der lokale Diskretisationsfehler gegeben durch

$$\begin{aligned} \Delta(h,f)(t,u) &= \frac{z(t+h)-z(t)}{h} - \Phi(h,f)(t,z(t)) \\ &= z'(t) + \frac{h}{2}z''(t) + \frac{h^2}{6}z'''(t) + O(h^3) - \Phi(h,f)(t,z(t)) \\ &= f(t) + \frac{h}{2}[f_t(t) + f_x(t)f(t)] \\ &\qquad + \frac{h^2}{6}\{f_{tt}(t) + 2f_{xt}(t)f(t) + f_{xx}(t)f^2(t) \\ &\qquad + f_x(t)[f_t(t) + f_x(t)f(t)]\} + O(h^3) - \Phi(h,f)(t,z(t)). \end{aligned}$$

Andererseits ist

$$\begin{aligned} k_2(t) &= f(t+\tfrac{1}{3}h, z(t)+\tfrac{1}{3}hk_1(t)) \\ &= f(t) + \frac{h}{3}\begin{pmatrix} f_t(t) \\ f_x(t) \end{pmatrix}^T \begin{pmatrix} 1 \\ f(t) \end{pmatrix} \\ &\qquad + \frac{h^2}{18}\begin{pmatrix} 1 \\ f(t) \end{pmatrix}^T \begin{pmatrix} f_{tt}(t) & f_{tx}(t) \\ f_{xt}(t) & f_{xx}(t) \end{pmatrix} \begin{pmatrix} 1 \\ f(t) \end{pmatrix} + O(h^3) \\ &= f(t) + \frac{h}{3}[f_t(t) + f_x(t)f(t,u)] \\ &\qquad + \frac{h^2}{18}[f_{tt}(t) + 2f_{xt}(t)f(t) + f_{xx}(t)f^2(t)] + O(h^3). \end{aligned}$$

Entsprechend ist

$$\begin{aligned} k_3(t) &= f(t+\tfrac{2}{3}h, z(t)+\tfrac{2}{3}hk_2(t)) \\ &= f(t) + \frac{2h}{3}\begin{pmatrix} f_t(t) \\ f_x(t) \end{pmatrix}^T \begin{pmatrix} 1 \\ k_2(t) \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{2h^2}{9} \begin{pmatrix} 1 \\ k_2(t) \end{pmatrix}^T \begin{pmatrix} f_{tt}(t) & f_{tx}(t) \\ f_{xt}(t) & f_{xx}(t) \end{pmatrix} \begin{pmatrix} 1 \\ k_2(t) \end{pmatrix} + O(h^3) \\
&= f(t) + \frac{2h}{3}[f_t(t) + f_x(t)k_2(t)] \\
&\quad + \frac{2h^2}{9}[f_{tt}(t) + 2k_2 f_{xt}(t) + k_2^2 f_{xx}(t)] + O(h^3) \\
&= f(t) + \frac{2h}{3}[f_t(t) + f_x(t)f(t)] + \frac{2h^2}{9} f_x(t)[f_t(t) + f_x(t)f(t)] \\
&\quad + \frac{2h^2}{9}[f_{tt}(t,u) + 2f_{xt}(t,u)f(t,u) + f_{xx}(t,u)f^2(t,u)] + O(h^3).
\end{aligned}
$$

Folglich ist

$$
\begin{aligned}
\Phi(h,f)(t,z(t)) &= \frac{1}{4}k_1(t) + \frac{3}{4}k_3(t) \\
&= f(t) + \frac{h}{2}[f_t(t) + f_x(t)f(t)] + \frac{h^2}{6}\{f_x(t)[f_t(t) + f_x(t)f(t)] \\
&\quad + f_{tt}(t) + 2f_{xt}(t)f(t) + f_{xx}(t)f^2(t)\} + O(h^3).
\end{aligned}
$$

Wir erkennen also, dass in der Tat

$$\Delta(h,f)(t,u) = O(h^3),$$

die Konsistenzordnung des Verfahrens also 3 ist. Genau das war zu zeigen. Mit Maple geht es wesentlich einfacher:

```
> restart;
> g:=t->f(t,z(t)):
> k_1:=h->f(t,z(t)):
> k_2:=h->f(t+(h/3),z(t)+(h/3)*k_1(h)):
> k_3:=h->f(t+(2*h/3),z(t)+(2*h/3)*k_2(h)):
> Delta:=h->(z(t+h)-z(t))/h-(1/4)*k_1(h)-(3/4)*k_3(h):
> s:=series(Delta(h),h,4):
> s1:=subs((D@@3)(z)(t)=(D@@2)(g)(t),s):
> s2:=subs((D@@2)(z)(t)=D(g)(t),s1):
> s3:=subs(D(z)(t)=g(t),s2):
> simplify(%);
```

$$\mathrm{O}(h^3)$$

4. Man betrachte ein Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{6}(k_1 + 4k_2 + k_3),$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t + \tfrac{1}{2}h, u + \tfrac{1}{2}hk_1), \qquad k_3 := f(t+h, u - hk_1 + 2hk_2).$$

Man zeige, dass dies ein Verfahren der Ordnung 3 ist. Hierbei darf man sich auf den Fall einer Differentialgleichung erster Ordnung, also $n = 1$, beschränken und die Aufgabe mit Maple lösen.

**Lösung:** Wir geben fast genau wie zur Bearbeitung der vorigen Aufgabe ein:

```
g:=t->f(t,z(t)):
k_1:=h->f(t,z(t)):
k_2:=h->f(t+(h/2),z(t)+(h/2)*k_1(h)):
k_3:=h->f(t+h,z(t)-h*k_1(h)+2*h*k_2(h)):
Delta:=h->(z(t+h)-z(t))/h-(1/6)*(k_1(h)+4*k_2(h)+k_3(h)):
s:=series(Delta(h),h,4):
s_1:=subs((D@@3)(z)(t)=(D@@2)(g)(t),s):
s_2:=subs((D@@2)(z)(t)=D(g)(t),s_1):
s_3:=subs(D(z)(t)=g(t),s_2);
```

Anschließendes `simplify(%);` liefert $O(h^3)$, die Konsistenzordnung ist also 3.

5. Man erläutere, weshalb das Eulersche Polygonzugverfahren, das Verfahren von Heun und auch das klassische Runge-Kutta-Verfahren nicht gut zur numerischen Lösung der einfachen Anfangswertaufgabe $x' = \lambda x$, $x(0) = x_0$, mit kleinem *negativen* $\lambda$ geeignet sind. Man überlege sich, dass dies für das *implizite* Euler-Verfahren

$$u(t+h) = u(t) + hf(t+h, u(t+h))$$

wesentlich besser aussieht. Für $\lambda = -1000$, $x_0 = 1$ und eine Schrittweite $h = 0.01$ mache man sich klar, mit welchen Werten man beim Runge-Kutta-Verfahren bzw. dem impliziten Euler-Verfahren zu rechnen hat.

**Lösung:** Die Lösung von $x' = \lambda x$, $x(0) = x_0$, mit kleinem negativen $\lambda$ geht mit wachsendem $t$ sehr schnell gegen Null und man erwartet, dass dies auch für die durch ein Verfahren gewonnenen Näherungswerte gilt. Beim expliziten Euler-Verfahren ist $u(t+h) = (1+\lambda h)u(t)$, es sollte also $|1+\lambda h| < 1$ bzw. $0 < h < (-2/\lambda)$ sein. Beim Verfahren von Heun hat man die Vorschrift

$$u(t+h) = u(t) + \tfrac{1}{2}h[f(t,u(t)) + f(t+h, u(t) + hf(t,u(t)))].$$

Nach Einsetzen von $f(t,x) := \lambda x$ erhält man

$$u(t+h) = [1 + \lambda h + \tfrac{1}{2}(\lambda h)^2]u(t)$$

bzw. $u(t+h) = R(h\lambda)u(t)$ mit $R(\zeta) := 1+\zeta+\frac{1}{2}\zeta^2$. Offenbar ist $|R(\zeta)| < 1$ für ein reelles $\zeta$ genau dann, wenn $\zeta \in (-2, 0)$. Also sollte auch hier die Schrittweite $h$ so klein gewählt werden, dass $-2 < h\lambda$ bzw. $h < (-2/\lambda)$. Beim klassischen Runge-Kutta-Verfahren, angewandt auf die obige Anfangswertaufgabe, erhält man nach leichter Rechnung, dass $u(t+h) = R(h\lambda)u(t)$, wobei

$$R(\zeta) := 1 + \zeta + \tfrac{1}{2}\zeta^2 + \tfrac{1}{6}\zeta^3 + \tfrac{1}{24}\zeta^4.$$

Hier ist $(-2.78529, 0)$ (die Endpunkte dieses Intervalls sind die beiden reellen Nullstellen von $R$) das "reelle Stabilitätsgebiet", es ist nur unwesentlich größer als beim Euler- bzw. Heun-Verfahren. Beim impliziten Euler-Verfahren ist $u(t+h) = u(t) + \lambda h u(t+h)$ bzw. $u(t+h) = R(h\lambda)u(t)$ mit

$$R(\zeta) := \frac{1}{1-\zeta}.$$

Offensichtlich ist hier $|R(\zeta)| < 1$ für ein reelles $\zeta$ genau dann, wenn $\zeta \in (-\infty, 0)$. Die Situation ist also unvergleichlich viel besser als bei den beiden expliziten Verfahren. Für

$\lambda = -1000$, $h = 0.01$ ist $\lambda h = -10$. Dann rechnet man leicht nach, dass beim Runge-Kutta-Verfahren $u_{i+1} = 291u_i$, d. h. das Verfahren explodiert nach kurzer Zeit. Beim impliziten Euler-Verfahren ist $u_{i+1} = \frac{1}{11}u_i$. Also erhält man z. B. $u_{10} = (1/11)^{10} \approx 3.85 \cdot 10^{-11}$ als Näherung für die Lösung $x(0.1) = e^{-100} \approx 3.72 \cdot 10^{-44}$, was natürlich relativ immer noch ein ziemlicher Fehler ist.

6. Man bestimme die exakte Lösung der Anfangswertaufgabe $x' = -x^2$, $x(0) = 1$, und vergleiche diese Werte mit dem durch das Runge-Kutta-Verfahren mit den Schrittweite $h = 0.1$ und $h = 0.05$ auf dem Intervall $[0, 1]$ erhaltenen Werte.

   **Lösung:** Durch

   ```
   dsolve({D(x)(t)=-x(t)^2,x(0)=1},x(t));
   ```

   erhalten wir die Lösung $x(t) = 1/(t+1)$. Wir benutzen die MATLAB-Funktion `FixedRK` und erhalten die Werte in der folgenden Tabelle (diesmal haben wir `format long` gewählt und geben nicht alle Werte an):

| $t$ | RK ($h = 0.1$) | RK ($h = 0.05$) | $1/(1+t)$ |
|---|---|---|---|
| 0.0 | 1.00000000000000 | 1.00000000000000 | 1.00000000000000 |
| 0.2 | 0.83333372884307 | 0.83333335857227 | 0.83333333333333 |
| 0.4 | 0.71428615389276 | 0.71428574227711 | 0.71428571428571 |
| 0.6 | 0.62500040094917 | 0.62500002549664 | 0.62500000000000 |
| 0.8 | 0.55555590318321 | 0.55555557764325 | 0.55555555555556 |
| 1.0 | 0.50000029758023 | 0.50000001889745 | 0.50000000000000 |

7. Man bestimme[20] die exakte Lösung der Anfangswertaufgabe $x' = (2/t)x$, $x(1) = 1$. Anschließend bestimme man einen analytischen Ausdruck für die durch das Eulersche Polygonzugverfahren erhaltene Näherung und gebe den globalen und den lokalen Diskretisierungsfehler an.

   **Lösung:** Durch

   ```
   dsolve({D(x)(t)=(2/t)*x(t),x(1)=1},x(t));
   ```

   erhält man die Lösung $x(t) = t^2$. Beim Euler-Verfahren ist $u(t + h) = [1 + 2h/t]u(t)$ und natürlich $u(1) = 1$. Sei $t > 1$ fest und $h = (t - 1)/m$ mit $m \in \mathbb{N}$. Die durch das Euler-Verfahren gewonnene Näherung $u(t; h)$ ist offenbar gegeben durch

$$u(t;h) = \prod_{i=0}^{m-1} \left(1 + \frac{2h}{1+ih}\right).$$

   Nun ist (für beliebiges $h$)

$$\prod_{i=0}^{m-1} \left(1 + \frac{2h}{1+ih}\right) = \frac{(1+mh)(1+(m+1)h)}{1+h}.$$

[20] Diese Aufgabe ist dem Lehrbuch

R. Kress (1998) *Numerical Analysis*. Springer-Verlag, New York-Berlin-Heidelberg.

entnommen.

Dies zeigt man natürlich durch vollständige Induktion nach $m$. Für $m = 1$ (oder noch einfacher, für $m = 0$) ist die Behauptung richtig. Im Induktionsschritt ist

$$\begin{aligned}\prod_{i=0}^{m}\left(1+\frac{2h}{1+ih}\right) &= \frac{(1+mh)(1+(m+1)h)}{1+h}\left(1+\frac{2h}{1+mh}\right)\\ &= \frac{(1+(m+1)h)(1+(m+2)h)}{1+h},\end{aligned}$$

das war im Induktionsschritt zu zeigen. Folglich ist

$$u(t;h)-x(t)=\frac{t(t+h)}{1+h}-t^2=t(1-t)\frac{h}{1+h}$$

der globale Diskretisierungsfehler.

Nun sei $z$ die Lösung von $z' = (1/s)z$, $z(t) = u$. Wir erhalten $z(s) = us^2/t^2$. Daher ist der lokale Diskretisierungsfehler

$$\begin{aligned}\Delta(h,f)(t,u) &= \frac{z(t+h)-z(t)}{h}-\Phi(h,f)(t,u)\\ &= \frac{u[(t+h)^2/t^2-1]}{h}-\frac{2u}{t}\\ &= \frac{u}{t^2}h.\end{aligned}$$

8. Man schreibe ein MATLAB-Programm für das klassische Runge-Kutta-Verfahren mit automatischer Schrittweitensteuerung. Anschließend teste man das Programm an der Anfangswertaufgabe (siehe Stoer-Bulirsch)

$$x' = -200tx^2, \qquad x(-3)=\frac{1}{901},$$

welche auf $[-3, 0]$ zu lösen sei.

**Lösung:** Die exakte Lösung ist

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

Ein Plot (siehe Abbildung 5.17) zeigt, wo die Schwierigkeiten sind. Etwa auf $[-3, \frac{1}{2}]$ kann mit einer verhältnismäßig großen Schrittweite gerechnet werden, danach muss zunehmend verfeinert werden. Wir haben die MATLAB-Funktion `RKauto.m` geschrieben, welche die dort enthaltene Funktion `RKstep` und eine Funktion für die rechte Seite der Differentialgleichung benutzt (die man eventuell in einem eigenen Function-File ablegen sollte):

```
function  [tvals,xvals,steps]=RKauto(fname,x_0,t_0,t_max,H,epsilon);
%Pre:     fname   string that names a function of the form f(t,x)
%                 where t is a scalar and x is a column n-vector
%         x_0     initial condition vector
%         t_0     initial time
%         t_max   final time
%         H       initial step size guess
```

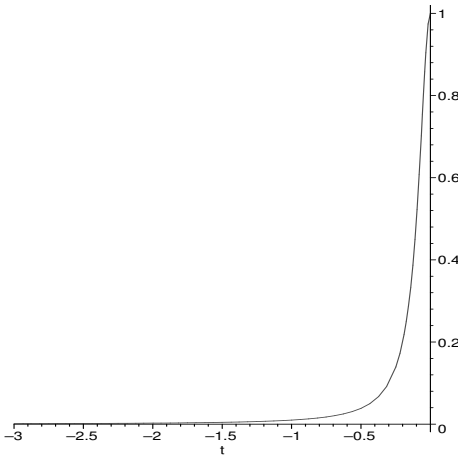


Abbildung 5.17: Die Lösung von $x' = -200tx^2$, $x(-3) = 1/901$ auf $[-3, 0]$

```
%         epsilon small number, the error is approximately
%                 equal to epsilon
%Post:    tvals(k)(k-1)-th time
%         xvals(:,k)=approximate solution at t=tvals(k)
%
t_c=t_0; x_c=x_0; tvals=t_c; xvals=x_c; f_c=feval(fname,t_c,x_c);
steps=0;
while (t_c<t_max)
     step_small=0;
     while (step_small==0)
          [xH,fH]=RKstep(fname,t_c,x_c,f_c,H);
          [xH2,fH2]=RKstep(fname,t_c,x_c,f_c,H/2);
          [xH2,fH2]=RKstep(fname,t_c+H/2,xH2,fH2,H/2);
          h=H*(15*epsilon/(16*norm(xH-xH2,inf)))^(1/5);
          steps=steps+1;
          if (H>=3*h)
              H=2*h;
          else
              step_small=1;
              t_c=t_c+H;
              x_c=xH2;
              f_c=fH2;
              H=2*h;
          end;
     end;
     xvals=[xvals x_c];
     tvals=[tvals t_c];
end;

function [x_new,f_new]=RKstep(fname,t_c,x_c,f_c,h);
%Pre:      fname   string that names a function of the form f(t,x)
%                  where t is a scalar and x is a column n-vector
```

```
%          x_c     an approximate to x'=f(t,x) at t=t_c
%          f_c     f(t_c,x_c)
%          h       the time step
%Post:     x_new   x_new is an approximate solution at t_new=t_c+h, and
%          f_new   f(t_new,x_nrw)

   k_1=f_c;
   k_2=feval(fname,t_c+(h/2),x_c+(h/2)*k_1);
   k_3=feval(fname,t_c+(h/2),x_c+(h/2)*k_2);
   k_4=feval(fname,t_c+h,x_c+h*k_3);
   x_new=x_c+(h/6)*(k_1+2*k_2+2*k_3+k_4);
   f_new=feval(fname,t_c+h,x_new);

function out=f(t,x);
   out=-200*t*x^2;
```

Mit dem Aufruf

```
[t,x,steps]=RKauto('f',1/901,-3,0,0.1,1.e-14);
```

erhalten wir `steps=687`, es wurden also $3 \cdot 687 = 2061$ Runge-Kutta-Schritte durchgeführt. Nach dem Aufruf sind `t` und `x` jeweils Zeilenvektoren mit 687 Komponenten, so viele Zeitschritte wurden also durchgeführt. Am Schluss ist

```
t(687)=2.353705998324984e-04
x(687)-1/(1+100*t(687)^2)=-4.300241132071392e-07
```

Die kleinste auftretende Schrittweite ist $h = 6.246407336988336 \cdot 10^{-4}$, die größte $h = 0.03265113115694$.

### 5.3.2 Aufgaben zu Abschnitt 3.2

1. Die Folge $\{f_k\}$ (die sogenannte Fibonacci-Folge) sei gegeben durch

$$f_0 := 1, \quad f_1 := 1, \quad f_{k+2} := f_{k+1} + f_k.$$

Man zeige, dass $\lim_{k\to\infty}(f_{k+1} - \tau f_k) = 0$, wobei $\tau := (1 + \sqrt{5})/2$. .

**Lösung:** Für die allgemeine Lösung der Gleichung $f_{k+2} = f_{k+1} + f_k$, $k = 0, 1, \ldots$, (mit noch nicht festgelegten $f_0$, $f_1$) machen wir den Ansatz $f_k = \lambda^k$. Einsetzen liefert, dass $\{f_k\}$ eine nichttriviale Lösung der Differenzengleichung ist, wenn $\lambda^2 = \lambda + 1$, was

$$\lambda_{1,2} = \frac{1}{2}(1 \pm \sqrt{5})$$

ergibt. Man beachte, dass

$$\lambda_1 = \frac{1}{2}(1 + \sqrt{5}) = \tau, \qquad \lambda_2 = \frac{1}{2}(1 - \sqrt{5}) = (-\tau)^{-1}.$$

Die allgemeine Lösung der Gleichung $f_{k+2} = f_{k+1} + f_k$ ist daher

$$f_k = \alpha\tau^k + \beta(-\tau)^{-k}$$

mit beliebigen $\alpha$, $\beta$. Anpassen der Anfangsbedingungen $f_0 = 1$ und $f_1 = 1$ liefert die Gleichungen

$$\alpha + \beta = 1, \qquad \alpha\tau + \beta(-\tau)^{-1} = 1$$

mit der Lösung

$$\alpha = \frac{1}{2} + \frac{1}{10}\sqrt{5} = \frac{\tau}{\sqrt{5}}, \qquad \beta = \frac{1}{2} - \frac{1}{10}\sqrt{5} = -\frac{(-\tau)^{-1}}{\sqrt{5}}.$$

Daher ist das $k$-te Folgenglied der gegebenen Folge $\{f_k\}$ gerade

$$f_k = \frac{1}{\sqrt{5}}[\tau^{k+1} - (-\tau)^{-(k+1)}].$$

Hierzu hätten wir übrigens auch Maple einsetzen können:

```
rsolve({f(k+2)=f(k+1)+f(k),f(0)=1,f(1)=1},f(k));
```

liefert dasselbe Ergebnis in anderer Gestalt. Folglich ist

$$\begin{aligned} f_{k+1} - \tau f_k &= \frac{1}{\sqrt{5}}[\tau^{k+2} - (-\tau)^{-(k+2)} - \tau^{k+2} + (-1)^{k+1}\tau^{-k}] \\ &= \frac{(-1)^{k+1}}{\sqrt{5}}[\tau^{-(k+2)} + \tau^{-k}] \\ &\to 0 \quad \text{für } k \to \infty. \end{aligned}$$

Damit ist die Behauptung bewiesen.

2. Man[21] löse folgende Differenzengleichungen:

   (a) $u_{j+2} - 2u_{j+1} - 3u_j = 0$, $u_0 = 0$, $u_1 = 1$,
   (b) $u_{j+1} - u_j = 2^j$, $u_0 = 0$,
   (c) $u_{j+2} - 2u_{j+1} - 3u_j = 1$, $u_0 = 0$, $u_1 = 0$,
   (d) $u_{j+1} - u_j = j$, $u_0 = 0$.

   Hierbei kann auch der Maple-Befehl `rsolve` benutzt werden.

   **Lösung:** Nach

   ```
   rsolve({u(j+2)-2*u(j+1)-3*u(j)=0,u(0)=0,u(1)=1},u(j));
   ```

   erhalten wir sofort

   $$u_j = -\frac{1}{4}(-1)^j + \frac{1}{4}3^j,$$

   was wir natürlich auch leicht auf "konventionellem" Wege erhalten hätten.

   Die anderen drei Aufgaben kann man ebenfalls sehr leicht mit Maple lösen (und zur Not das Ergebnis verifizieren). Die Lösungen sind

   $$u_j = -1 + 2^j, \qquad u_j = \frac{1}{8}(-1)^j + \frac{1}{8}3^j - \frac{1}{4}, \qquad u_j = \frac{1}{2}j(j-1).$$

[21] Diese Aufgabe haben wir aus

K. Strehmel, R. Weiner (1995) *Numerik gewöhnlicher Differentialgleichungen*. B. G. Teubner, Stuttgart.

3. Man[22] löse die Differenzengleichung

$$u_{j+4} - 6u_{j+3} + 14u_{j+2} - 16u_{j+1} + 8u_j = j$$

mit den Anfangsbedingungen

$$u_0 = 1, \quad u_1 = 2, \quad u_2 = 3, \quad u_3 = 4.$$

**Lösung:** Auch mit Maple ist es diesmal ein wenig komplizierter, zum Ergebnis zu kommen. Wir geben ein:

```
rsolve({u(j+4)-6*u(j+3)+14*u(j+2)-16*u(j+1)+8*u(j)=j,
       u(0)=1,u(1)=2,u(2)=3,u(3)=4},u(j));
```

und erhalten nach `simplify(%);` die Ausgabe

$$u_j = -2^j + \frac{1}{4}i(1-i)^j - \frac{1}{4}i(1+i)^j + \frac{1}{4}2^j j + j + 2.$$

Anschließendes `evalc(%);` liefert

$$u_j = -2^j + \frac{1}{2}e^{j\ln(2)/2}\sin(j\pi/4) + \frac{1}{4}2^j j + j + 2.$$

Erneutes `simplify(%);` ergibt

$$u_j = -2^j + 2^{j/2-1}\sin(j\pi/4) + \frac{1}{4}2^j j + j + 2.$$

Es ist also

$$u_j = 2^j(j/4 - 1) + 2^{j/2-1}\sin(j\pi/4) + j + 2.$$

4. Man[23] bestimme die allgemeine Lösung der Differenzengleichung

$$u_{j+2} - 2au_{j+1} + au_j = 1$$

mit $a \in (0,1)$ im Komplexen und im Reellen. Man zeige, dass $\lim_{j\to\infty} u_j = 1/(1-a)$.

**Lösung:** Eine spezielle Lösung der inhomogenen Aufgabe ist $u_j := 1/(1-a)$ für alle $j$. Wir haben daher noch nach der allgemeinen Lösung der homogenen Aufgabe $u_{j+2} - 2au_{j+1} + au_j = 0$, $j = 0, 1, \ldots$, zu suchen. Der Ansatz $u_j = \lambda^j$ führt für nichttriviale Lösungen auf die Gleichung $\lambda^2 - 2a\lambda + a = 0$ mit den beiden (komplexen) Lösungen

$$\lambda_{1,2} = a \pm \sqrt{a(1-a)}i.$$

Man beachte, dass $|\lambda_{1,2}| = \sqrt{a} < 1$. Die allgemeine Lösung der gegebenen Differenzengleichung im Komplexen ist

$$u_j = \frac{1}{1-a} + \alpha[a + \sqrt{a(1-a)}i]^k + \beta[a - \sqrt{a(1-a)}i]^k$$

[22] Diese Aufgabe haben wir aus

A. Quarteroni, R. Sacco, F. Saleri (2000) *Numerical Mathematics*. Springer, New York-Berlin-Heidelberg.

[23] Diese Aufgabe haben wir aus

R. Kress (1998) *Numerical Analysis*. Springer, New York-Berlin-Heidelberg.

mit beliebigen komplexen $\alpha$, $\beta$. Wegen $|\lambda_{1,2}| < 1$ folgt $\lim_{j\to\infty} u_j = 1/(1-a)$. Nun wollen wir auch noch die allgemeine Lösung im Reellen bestimmen, wozu es natürlich genügt, die allgemeine Lösung der homogenen Differenzengleichung im Reellen zu finden. Wegen

$$a + \sqrt{a(1-a)}i = \sqrt{a}e^{i\phi} \quad \text{mit} \quad \phi := \arctan\sqrt{(1-a)/a}$$

ist die allgemeine Lösung im Reellen

$$u_j = \frac{1}{1-a} + \alpha a^{j/2}[\alpha\cos(\phi j) + \beta\sin(\phi j)].$$

Damit ist die Aufgabe gelöst.

5. Man[24] bestimme $\alpha$, $\beta$ und $\gamma$ so, dass das lineare Mehrschrittverfahren

$$u_{j+3} - u_{j+1} + \alpha(u_{j+2} - u_j) = h\{\beta[f(t_{j+2}, u_{j+2}) - f(t_j, u_j)] + \gamma f(t_{j+1}, u_{j+1})\}$$

die Konsistenzordnung 3 hat. Ist das so gewonnene Verfahren stabil?

**Lösung:** Wir benutzen Maple. Nach

```
Delta:=h->(z(t+3*h)-z(t+h)+alpha*(z(t+2*h)-z(t)))/h-
     (beta*(D(z)(t+2*h)-D(z)(t))+gamma*D(z)(t+h));
s:=series(Delta(h),h,5);
factor(%);
```

erhalten wir, dass der lokale Diskretisierungsfehler durch

$$\begin{aligned}\Delta(h,f)(t,z(t)) \;=\; & -z'(t)(-2-2\alpha+\gamma) - z''(t)(-2\alpha-4+\gamma+2\beta)h \\ & - \frac{1}{6}z'''(t)(3\gamma-26-8\alpha+12\beta)h^2 \\ & - \frac{1}{6}z^{(4)}(t)(-4\alpha-20+\gamma+8\beta)h^3 + O(h^4)\end{aligned}$$

gegeben ist. Nach

```
eqn:={-2-2*alpha+gamm=0,-2*alpha-4+gamm+2*beta=0,
      3*gamm-26-8*alpha+12*beta=0};
solve(eqn,{alpha,beta,gamm});
```

(wir benutzen hier `gamm` statt `gamma`, weil letzteres in Maple schon ein besetzter Name ist) erhalten wir

$$\alpha = -4, \qquad \beta = 1, \qquad \gamma = -6.$$

Das Polynom

$$\psi(\lambda) := \lambda^3 - \lambda - 4(\lambda^2 - 1)$$

hat die Nullstellen $\pm 1$ und 4. Das Mehrschrittverfahren erfüllt also die Stabilitätsbedingung nicht.

[24] Diese und die nächsten beiden Aufgaben haben wir J. Stoer, R. Bulirsch (1990, S. 246) entnommen.

6. Es werde das durch

$$u_{j+2} + a_1 u_{j+1} + a_0 u_j = h[b_0 f(t_j, u_j) + b_1 f(t_{j+1}, u_{j+1})]$$

gegebene explizite Zweischrittverfahren betrachtet.

(a) Man bestimme $a_0$, $b_0$ und $b_1$ in Abhängigkeit von $a_1$ so, dass man ein Verfahren mindestens zweiter Konsistenzordnung hat.

(b) Für welche $a_1$-Werte ist das so gewonnene Verfahren stabil?

(c) Welche speziellen Verfahren erhält man für $a_1 = 0$ und $a_1 = -1$?

(d) Lässt sich $a_1$ so wählen, dass man ein stabiles Verfahren der Konsistenzordnung 3 erhält?

**Lösung:** Mit Maple erhalten wir sehr leicht, dass der lokale Diskretisierungsfehler durch

$$\begin{aligned}\Delta(h,f)(t,z(t)) &= z(t)(1+a_1+a_0)h^{-1} - z'(t)(-2-a_1+b_0+b_1) \\ &\quad - \frac{1}{2}z''(t)(2b_1-4-a_1)h - \frac{1}{6}z'''(t)(3b_1-a_1-8)h^2 \\ &\quad - \frac{1}{24}z^{(4)}(t)(-a_1-16+4b_1)h^3 + O(h^4)\end{aligned}$$

gegeben ist. Nach

```
eqn:={1+a_1+a_0=0,-2-a_1+b_0+b_1=0,2*b_1-4-a_1=0}:
solve(eqn,{a_0,b_0,b_1});
```

erhält man

$$b_0 = \frac{1}{2}a_1, \qquad a_0 = -1 - a_1, \qquad b_1 = 2 + \frac{1}{2}a_1.$$

Für beliebiges $a_1$ hat das entsprechende Mehrschrittverfahren eine Konsistenzordnung von mindestens zwei. Die Stabilität des entsprechenden Verfahrens wird durch die Nullstellen von

$$\psi(\lambda) := \lambda^2 + a_1\lambda - 1 - a_1$$

bestimmt. Dies sind 1 und $-1-a_1$. Die Stabilitätsbedingung ist genau dann erfüllt, wenn $a_1 \in (-2, 0]$ (hier musste $a_1 = -2$ ausgeschlossen werden, weil sonst 1 eine doppelte Nullstelle wäre). Für $a_1 = 0$ hat man das Verfahren

$$u_{j+2} - u_j = 2hf(t_{j+1}, u_{j+1}),$$

während man für $a_1 = -1$ das Verfahren

$$u_{j+2} - u_{j+1} = \frac{h}{2}[3f(t_{j+1}, u_{j+1}) - f(t_j, u_j)]$$

erhält. Will man ein Verfahren der Konsistenzordnung 3 erhalten, so müssen die Gleichungen

$$1 + a_1 + a_0 = 0, \quad -2 - a_1 + b_0 + b_1 = 0, \quad 2b_1 - 4 - a_1 = 0, \quad 3b_1 - a_1 - 8 = 0$$

erfüllt sein. Eindeutige Lösung ist

$$a_0 = -5, \quad a_1 = 4, \quad b_0 = 2, \quad b_1 = 4.$$

Da zugehörige Verfahren ist nicht stabil, da $a_1 \notin (-2, 0]$.

7. Man prüfe, ob das lineare Mehrschrittverfahren

$$u_{j+4} - u_j = \frac{h}{3}[8f(t_{j+3}, u_{j+3}) - 4f(t_{j+2}, u_{j+2}) + 8f(t_{j+1}, u_{j+1})]$$

konvergent ist.

**Lösung:** Der lokale Diskretisierungsfehler ist

$$\Delta(h, f)(t, z(t)) = \frac{14}{45} z^{(5)}(t) h^4 + O(h^5),$$

die Konsistenzordnung des Verfahrens ist also vier. Die Stabilitätsbedingung ist aber leider verletzt, da $\lambda = 1$ und $\lambda = -1$ jeweils doppelte Nullstellen von $\psi(\mu) := \mu^4 - 1$ sind. Also ist das angegebene Mehrschrittverfahren nicht konvergent.

8. Man[25] bestimme das $\alpha$-Intervall, für das das explizite lineare 3-Schrittverfahren

$$u_{j+3} + \alpha(u_{j+2} - u_{j+1}) - u_j = \frac{h}{2}(3 + \alpha)[f(t_{j+2}, u_{j+2}) + f(t_{j+1}, u_{j+1})], \quad \alpha \in \mathbb{R},$$

der Stabilitätsbedingung genügt. Ferner zeige man, dass ein $\alpha$ existiert, für das das Verfahren die Konsistenzordnung 4 hat, dass aber für ein stabiles Verfahren die Konsistenzordnung höchstens 2 sein kann.

**Lösung:** Nullstellen von

$$\psi(\lambda) := \lambda^3 + \alpha(\lambda^2 - \lambda) - 1$$

sind 1 und

$$\lambda_{1,2} = -\frac{1}{2}(\alpha + 1) \pm \frac{1}{2}\sqrt{\alpha^2 + 2\alpha - 3}.$$

Für $|\alpha + 1| < 2$ sind $\lambda_{1,2}$ echt komplex und

$$|\lambda_{1,2}| = \sqrt{\frac{1}{2}\alpha^2 + \alpha - \frac{1}{2}}.$$

Es ist $|\lambda_{1,2}| \leq 1$ für $\alpha \in [-3, 1]$. Insgesamt stellt man fest, dass das Verfahren für $\alpha \in (-3, 1]$ der Stabilitätsbedingung genügt. Für $\alpha \notin (-3, 1]$ sind $\lambda_1$ und $\lambda_2$ reell und $\lambda_1 \lambda_2 = -1$. Hieraus folgt, dass das Verfahren für $\alpha \notin (-3, 1]$ nicht stabil ist, insgesamt also genau für $\alpha \in (-3, 1]$ stabil ist.

Zur Berechnung der Konsistenzordnung wenden wir Satz 2.7 mit

$$\psi(\mu) := \mu^3 + \alpha(\mu^2 - \mu) - 1, \qquad \chi(\mu) := \frac{1}{2}(3 + \alpha)(\mu^2 + \mu)$$

an. Aus

```
series((mu^3+alpha*(mu^2-mu)-1)/ln(mu)-(1/2)*(3+alpha)*(mu^2+mu),mu=1,6);
```

[25] Diese Aufgabe haben wir aus

K. Strehmel, R. Weiner (1995) *Numerik gewöhnlicher Differentialgleichungen*. B. G. Teubner, Stuttgart.

erhalten wir

$$\begin{aligned}\frac{\psi(\mu)}{\ln\mu} - \chi(\mu) &= \left(\frac{3}{4} - \frac{1}{12}\alpha\right)(\mu-1)^2 + \left(\frac{3}{8} - \frac{1}{24}\alpha\right)(\mu-1)^3 \\ &\quad + \left(-\frac{3}{80} + \frac{11}{720}\alpha\right)(\mu-1)^4 + O((\mu-1)^5).\end{aligned}$$

Für alle $\alpha \in (-3, 1]$ ist das Verfahren stabil und hat die Konsistenzordnung 2, während das Verfahren genau für $\alpha = 9$ eine höhere Konstenzordnung als 2 hat, nämlich 4, aber hierfür nicht stabil ist.

9. Eine Anfangswertaufgabe

$$\text{(P)} \qquad x'' = f(t,x), \qquad x(t_0) = x_0,\ x'(t_0) = x_0'$$

für eine Differentialgleichung zweiter Ordnung kann man natürlich dadurch numerisch lösen, dass man die Aufgabe als ein System von zwei Differentialgleichungen erster Ordnung schreibt und dieses mit einem Ein- oder Mehrschrittverfahren löst. Die folgenden zu beweisenden Aussagen sollen Hinweise dafür geben, wie man Mehrschrittverfahren zur Lösung von (P) konstruieren kann, die diesen Umweg nicht gehen.

(a) Eine Lösung $x(\cdot)$ von (P) genügt der Identität

$$x(t+h)-2x(t)+x(t-h) = h^2 \int_0^1 (1-s)[f(t+sh, x(t+sh)) + f(t-sh, x(t-sh))]\, ds.$$

(b) Welche Mehrschrittverfahren zur Lösung von (P) suggeriert Teil (a) dieser Aufgabe? Man gebe ein explizites und ein implizites Verfahren an.

**Lösung:** In Aufgabe 1 in Abschnitt 2.1 wurde gezeigt: Ist $x(\cdot)$ eine Lösung von (P), so ist

$$x(t) = x_0 + x_0'(t-t_0) + \int_{t_0}^t (t-s) f(s, x(s))\, ds.$$

Daher ist

$$\begin{aligned}x(t+h) - 2x(t) + x(t-h) &= \int_{t_0}^{t+h} (t+h-s) f(s,x(s))\, ds \\ &\quad - 2\int_{t_0}^{t} (t-s) f(s,x(s))\, ds \\ &\quad + \int_{t_0}^{t-h} (t-h-s) f(s,x(s))\, ds \\ &= \int_{t}^{t+h} (t+h-s)\, f(s,x(s))\, ds \\ &\quad - \int_{t-h}^{t} (t-h-s) f(s,x(s))\, ds \\ &= h^2 \int_0^1 (1-\tau)[f(t+\tau h, x(t+\tau h)) \\ &\quad + f(t-\tau h, x(t-\tau h))]\, d\tau.\end{aligned}$$

Hierbei haben wir die Variablentransformation $s = t + \tau h$ bzw. $s = t - \tau h$ benutzt. Damit ist der erste Teil bewiesen.

Man setze

$$g(s) := f(t + sh, x(t + sh)) + f(t - sh, x(t - sh)).$$

Ersetzt man $g(\cdot)$ durch $g(0)$, also das konstante Interpolationspolynom zur Stützstelle $s = 0$, so erhält man

$$x(t + h) - 2x(t) + x(t - h) \approx 2h^2 f(t, x(t)) \int_0^1 (1 - s)\, ds = h^2 f(t, x(t)).$$

Das entsprechende explizite Mehrschrittverfahren ist also

$$u_{j+1} - 2u_j + u_{j-1} = h^2 f(t_j, u_j),$$

das einfachste *Störmer-Verfahren*. Das quadratische Interpolationspolynom für $g(\cdot)$ zu den Stützstellen $s = -1, 0, 1$ ist

$$p_2(s) = g(1) + \frac{1}{2}(g(1) - g(-1))(s - 1) + \frac{1}{2}[g(1) - 2g(0) + g(-1)](s^2 - 1).$$

Weiter ist

$$\begin{aligned} \int_0^1 (1 - s) p_2(s)\, ds &= \frac{1}{2} g(1) - \frac{1}{6}[g(1) - g(-1)] - \frac{5}{24}[g(1) - 2g(0) + g(-1)] \\ &= \frac{1}{12}[f(t + h, x(t + h)) + 10 f(t, x(t)) + f(t - h, x(t - h))]. \end{aligned}$$

Das entsprechende implizite Mehrschrittverfahren ist daher

$$u_{j+1} - 2u_j + u_{j-1} = \frac{h^2}{12}[f(t_{j+1}, u_{j+1}) + 10 f(t_j, u_j) + f(t_{j-1}, u_{j-1})],$$

das einfachste *Cowell-Verfahren*.

10. Man schreibe eine MATLAB-Funktion zur Lösung der Anfangswertaufgabe $x' = f(t, x)$, $x(t_0) = x_0$, welche ein Prädiktor-Korrektor-Verfahren (mit Adams-Bashforth-Formeln als Prädiktor und Adams-Moulton-Formeln als Korrektor) benutzt. Diese Funktion[26] könnte so deklariert werden:

```
function [tvals,xvals]=FixedPC(fname,t_0,x_0,h,p,m);
```

Hierbei seien die Eingabe Daten:

```
fname   string that names the function f.
  t_0   initial time.
  x_0   initial condition vector.
    h   stepsize.
    p   order of method.  (1<=p<=4).
    m   numberof steps to be taken.
```

[26] Wir halten uns eng an

C. F. van Loan (1997) *Introduction to Scientific Computing. A Matrix-Vector Approach using MATLAB*. Prentice Hall, Upper Saddle River.

Ausgabedaten seien

```
tvals   tvals(j)=t_0+(j-1)h, j=1:m+1.
xvals   approximate solution at t=tvals(j), j=1:m+1.
```

Hierzu sollten die folgenden Funktionen bereitgestellt werden (wir geben Input- und Ouputparameter an, ihre Bedeutung sollte sich fast von alleine erschließen):

```
function [tvals,xvals,fvals]=StartAB(fname,t_0,x_0,h,p);
function [t_new,x_new,f_new]=PCstep(fname,t_c,x_c,fvals,h,p);
```

Für die Funktion `StartAB` kann es zweckmäßig sein, noch eine Funktion

```
function [t_new,x_new,f_new]=RKstep(fname,t_c,x_c,f_c,h,p);
```

bereitzustellen. Als Test löse man das folgende Zweikörperproblem

$$\ddot{x} = -\frac{x}{(x^2+y^2)^{3/2}}, \qquad x(0) = 0.4, \qquad \dot{x}(0) = 0,$$
$$\ddot{y} = -\frac{y}{(x^2+y^2)^{3/2}}, \qquad y(0) = 0, \qquad \dot{y}(0) = 2$$

über dem Zeitintervall $[0, 2\pi]$ und plotte die Bahn $\{(x(t), y(t)) : t \in [0, 2\pi]\}$.

**Lösung:** Wir haben ein File `FixedPC` geschrieben (die benötigten Hilfsfunktionen sind einfach angehängt, damit natürlich für keine andere Funktion nutzbar), das den folgenden Inhalt hat:

```
function [tvals,xvals]=FixedPC(fname,t_0,x_0,h,p,m);
%
%Produces an approximate solution to the initial value problem
%x'=f(t,x), x(t_0)=x_0, using a strategy that is based on a p-th
%order Adams PC method. More exactly: Use RK of order p starting the
%computation, use Adams-Bashforth of order p as predictor
%and Adams-Moulton of order p+1 as corrector. Stepsize is fixed.
%
%Pre:  fname  string that names the function f.
%        t_0  initial time.
%        x_0  initial condition vector
%          h  stepsize
%          p  order of method  (1<=p<=4).
%          m  number of steps to be taken.
%
%Post: tvals  tvals(j)=t_0+(j-1)h,   j=1:m+1.
%      xvals  xvals(:,j) approximate solution at t=tvals(j), j=1:m+1.

[tvals,xvals,fvals]=StartAB(fname,t_0,x_0,h,p);
t_c=tvals(p);  x_c=xvals(:,p);  f_c=fvals(:,p);

for j=p:m
```

```
   [t_c,x_c,f_c]=PCstep(fname,t_c,x_c,fvals,h,p);
   tvals=[tvals t_c];  xvals=[xvals x_c]; fvals=[f_c fvals(:,1:p-1)];
end;

function  [tvals,xvals,fvals]=StartAB(fname,t_0,x_0,h,p);
%
%Uses p-th order Runge-Kutta to generate approximate solutions to
%x'=f(t,x), x(t_0)=x_0 at t=t_0, t_0+h,...,t_0+(p-1)h.
%
%Pre:   as above.
%Post:  tvals=[t_0,t_0+h,...,t_0+(p-1)h].
%       For j=1:p, xvals(:,j) approximates x(tvals(j)).
%       For j=1:p, fvals(:,j)=f(tvals(j),xvals(:,j)).

t_c=t_0;  x_c=x_0;  f_c=feval(fname,t_c,x_c);
tvals=t_c; xvals=x_c; fvals=f_c;

for j=1:p-1
   [t_c,x_c,f_c]=RKstep(fname,t_c,x_c,f_c,h,p);
   tvals=[tvals t_c];
   xvals=[xvals x_c];
   fvals=[f_c fvals];
end;

function  [t_new,x_new,f_new]=RKstep(fname,t_c,x_c,f_c,h,p);
%Pre and Post:   obvious.
%
k_1=f_c;
if p==1
   x_new=x_c+h*k_1;
elseif p==2
   k_2=feval(fname,t_c+(h/2),x_c+(h/2)*k_1);
   x_new=x_c+(h/2)*(k_1+k_2);
elseif p==3
   k_2=feval(fname,t_c+(h/2),x_c+(h/2)*k_1);
   k_3=feval(fname,t_c+h,x_c+h*(-k_1+2*k_2));
   x_new=x_c+(h/6)*(k_1+4*k_2+k_3);
elseif p==4
   k_2=feval(fname,t_c+(h/2),x_c+(h/2)*k_1);
   k_3=feval(fname,t_c+(h/2),x_c+(h/2)*k_2);
   k_4=feval(fname,t_c+h,x_c+h*k_3);
   x_new=x_c+(h/6)*(k_1+2*k_2+2*k_3+k_4);
end;
t_new=t_c+h;
f_new=feval(fname,t_new,x_new);

function [t_new,x_new,f_new]=PCstep(fname,t_c,x_c,fvals,h,p);
%As  predictor we use Adams-Bashforth of order p, as  corrector
```

```
%we use Adams-Moulton of order p+1.
%Pre and Post:   nearly obvious.
%      fvals   a n-by-p matrix where fvals(:,i) is an approximation
%              to f(t,x(t)) at t=t_c+(i-1)h, i=1:p
t_new=t_c+h;
if p==1
   x_P=x_c+h*fvals;
   f_P=feval(fname,t_new,x_P);
   x_new=x_c+(h/2)*([f_P f_vals]*[1;1]);
elseif p==2
   x_P=x_c+(h/2)*(fvals*[3;-1]);
   f_P=feval(fname,t_new,x_P);
   x_new=x_c+(h/12)*([f_P fvals]*[5;8;-1]);
elseif p==3
   x_P=x_c+(h/12)*(fvals*[23;-16;5]);
   f_P=feval(fname,t_new,x_P);
   x_new=x_c+(h/24)*([f_P fvals]*[9;19;-5;1]);
elseif p==4
   x_P=x_c+(h/24)*(fvals*[55;-59;37;-9]);
   f_P=feval(fname,t_new,x_P);
   x_new=x_c+(h/720)*([f_P fvals]*[251;646;-264;106;-19]);
end;
f_new=feval(fname,t_new,x_new);
```

Weiter haben wir ein File `Kepler.m` geschrieben, in dem die rechte Seite des Differentialgleichungssystems spezifiziert ist:

```
function up=Kepler(t,u);
%
%Pre:  t (time) is a scalar and u is a 4-vector whose components
%      have the property that
%            u(1)=x   u(2)=(d/dt)x   u(3)=y   u(4)=(d/dt)y
%Post: up is a 4-vector with the property that it is the
%      derivative of u at time t.
%
r3=(u(1)^2+u(3)^2)^1.5;
up(1)=u(2);  up(2)=-u(1)/r3;  up(3)=u(4);  up(4)=-u(3)/r3;
up=up';
```

Schließlich legen wir die übrigen Inputparameter fest, starten das Programm und plotten die Planetenbahn durch

```
x_0=[0.4;0;0;2];
t_0=0;h=2*pi/1000;m=1000; p=4;
[t,x]=FixedPC('Kepler',t_0,x_0,h,p,m);
plot(x(1,:),x(3,:))
axis('square','equal')
title('Planetenbahn ermittelt mit FixedPC')
```

```
xlabel('x')
ylabel('y')
```

In Abbildung 5.18 links geben wir den erzeugten Plot an. Zum Vergleich wollen wir

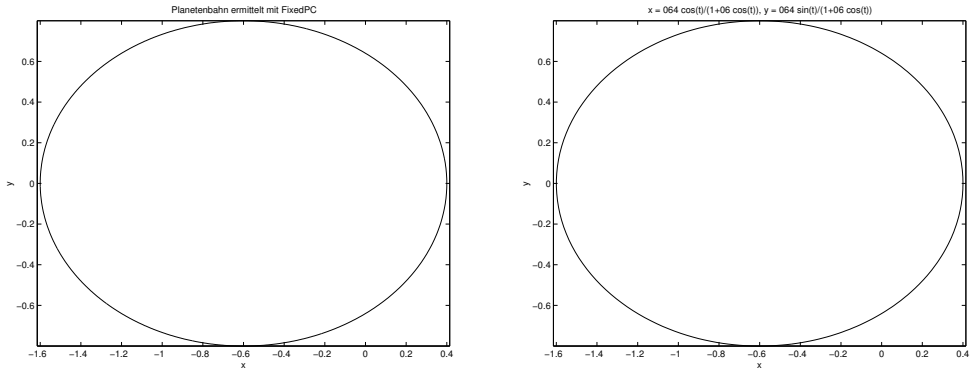


Abbildung 5.18: Planetenbahn beim Zweikörperproblem

die elliptische Planetenbahn bzw. genauer die erzeugte Ellipse auch noch mit Hilfe der Untersuchungen in Satz 2.2 berechnen. Mit den Bezeichnungen dort ist $R = 0.4$, $v_1 = 0$, $v_2 = 2$ und $\gamma M = 1$. Die erzeugte Ellipse ist

$$E = \left\{ \frac{p}{1 + \epsilon \cos(\phi - \alpha)} (\cos\phi, \sin\phi) : \phi \in [0, 2\pi] \right\},$$

wobei

$$p = \frac{R^2 v_2^2}{\gamma M}, \qquad \epsilon \cos\alpha = \frac{R v_2^2 - \gamma M}{\gamma M}, \qquad \epsilon \sin\alpha = -\frac{R v_1 v_2}{\gamma M}.$$

Dies führt auf

$$p = 0.64, \qquad \epsilon = 0.6, \qquad \alpha = 0.$$

Die wahrscheinlich einfachste Methode, die Ellipse $E$ mit MATLAB zu plotten besteht darin, den folgenden Befehl zu geben:

```
ezplot('0.64*cos(t)/(1+0.6*cos(t))','0.64*sin(t)/(1+0.6*cos(t))',[0,2*pi]);
```

In Abbildung 5.18 rechts geben wir das Resultat an. Man wird keinen Unterschied zum Bild links feststellen.

### 5.3.3 Aufgaben zu Abschnitt 3.3

1. Bei E. Hairer et al. (1993) und W. Walter (1993) findet man folgendes System von zwei Differentialgleichungen erster Ordnung (siehe auch Aufgabe 4 in Abschnitt 1.2):

$$\begin{aligned} \ddot{x} &= x + 2\dot{y} - \mu' \frac{x + \mu}{[(x+\mu)^2 + y^2]^{3/2}} - \mu \frac{x - \mu'}{[(x-\mu')^2 + y^2]^{3/2}}, \\ \ddot{y} &= y - 2\dot{x} - \mu' \frac{y}{[(x+\mu)^2 + y^2]^{3/2}} - \mu \frac{y}{[(x-\mu')^2 + y^2]^{3/2}}. \end{aligned}$$

Hierbei ist $\mu$ eine gegebene Konstante und $\mu' := 1 - \mu$. Für $\mu := 0.01212277471$ und die Anfangsbedingungen

$$x(0) = 1.2, \quad \dot{x}(0) = 0, \quad y(0) = 0, \quad \dot{y}(0) = -1.04936$$

sowie

$$x(0) = 0.994, \quad \dot{x}(0) = 0, \quad y(0) = 0, \quad \dot{y}(0) = -2.0015851063791$$

berechne man mit Hilfe der MATLAB-Funktion `ode45` jeweils eine Lösung und plotte die Phasenbahn $\{(x(t), y(t)) : t \in I\}$, wobei das Intervall $I$ einmal $[0, 7]$ und einmal $[0, 17.1]$ ist.

**Lösung:** Zunächst schreiben wir ein File `Sattel.m`, in welchem die rechte Seite (nach Übertragung in ein System von 4 Differentialgleichungen erster Ordnung) spezifiziert wird:

```
function dxdt=Sattel(t,x);

mu=0.012277471;mu_strich=1-mu;
D_1=((x(1)+mu)^2+x(3)^2)^(1.5);D_2=((x(1)-mu_strich)^2+x(3)^2)^(1.5);
dxdt=[x(2);x(1)+2*x(4)-mu_strich*(x(1)+mu)/D_1-mu*(x(1)-mu_strich)/D_2;
       x(4);x(3)-2*x(2)-mu_strich*x(3)/D_1-mu*x(3)/D_2];
```

Nach

```
[t,x]=ode45('Sattel',[0 7],[1.2;0;0;-1.04936]);
plot(x(:,1),x(:,3));
```

erhält man Abbildung 5.19 links. Das Bild ist eigentlich nicht befriedigend, weil eine

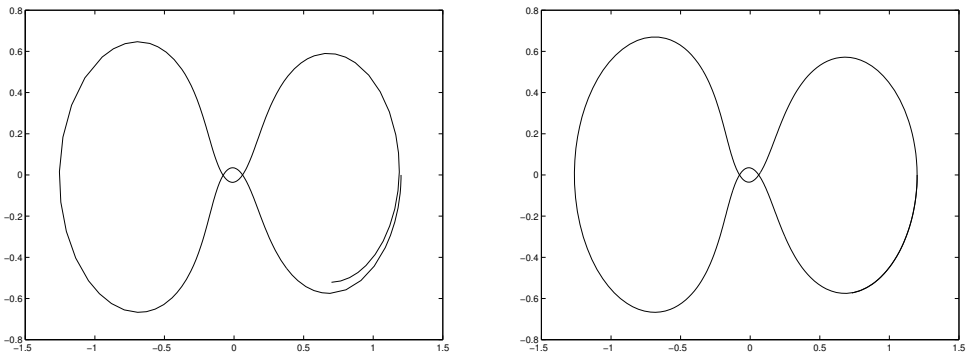


Abbildung 5.19: Satellitenbahn

periodische Phasenbahn (nach W. Walter mit der Periode $T \approx 6.19217$) entstehen müsste. Dies könnte daran liegen, dass die default-Werte für die geforderte relative und absolute Genauigkeit (`1e-3` bzw. `1e-6`) zu groß sind. Wir ändern diese folgendermaßen:

```
options=odeset('RelTol',1e-5,'AbsTol',1e-8);
```

Anschließend erfolgt der Aufruf durch

```
[t,x]=ode45('Sattel',[0 7],[1.2;0;0;-1.04936],options);
```

Man erhält den Plot in Abbildung 5.19 rechts, und das sieht schon besser aus. Entsprechend erhalten wir in Abbildung 5.20 Satellitenbahnen für den zweiten Satz von Anfangsbedingungen. Auch diesmal ist der rechts stehende Plot etwas gefälliger. Nach

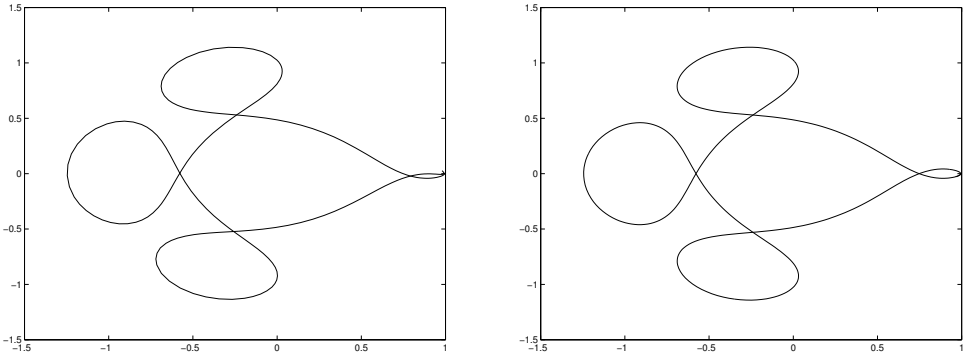

Abbildung 5.20: Weitere Satellitenbahn

E. Hairer et al. erhält man eine periodische Lösung der Periode $T \approx 17.065216560$.

2. Man löse die folgende Anfangswertaufgabe (Euler-Gleichungen für die Bewegung eines Festkörpers ohne äußere Kräfte, siehe z. B. L. F. Shampine, M. K. Gordon (1975, S. 243)[27])

$$\begin{array}{rclcrcl} x_1' &=& x_2x_3 & & x_1(0) &=& 0, \\ x_2' &=& -x_1x_3 & & x_2(0) &=& 1, \\ x_3' &=& -0.51x_1x_2 & & x_3(0) &=& 1 \end{array}$$

auf dem Zeitintervall $[0, 12]$ mit Hilfe der MATLAB-Funktion `ode45`. Ferner plotte man die Lösungskomponenten auf diesem Intervall.

**Lösung:** In ein File `Fest.m` schreiben wir:

```
function dx=Fest(t,x);
dx=[x(2)*x(3);-x(1)*x(3);-0.51*x(1)*x(2)];
```

Danach machen wir den Aufruf

```
[t,x]=ode45('Fest',[0,12],[0;1;1]);
```

Nach `plot(t,x(:,1),t,x(:,2),t,x(:,3));` hat man alle drei Komponenten geplottet, siehe Abbildung 5.21. Die Lösungen lassen sich übrigens durch elliptische Funktionen ausdrücken.

[27] L. F. Shampine, M. K. Gordon (1975) *Computer Solution of Ordinary Differential Equations. The Initial Value Problem.* W. H. Freeman and Company, San Francisco.

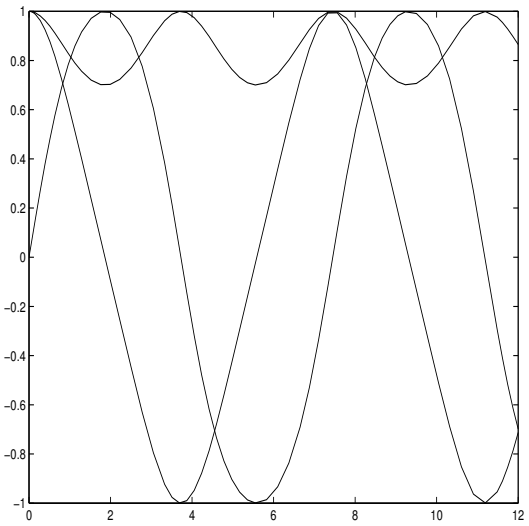


Abbildung 5.21: Die Lösung obiger Anfangswertaufgabe

### 5.3.4 Aufgaben zu Abschnitt 3.4

1. Gegeben sei das Einschrittverfahren mit der Verfahrensfunktion

$$\Phi(h,f)(t,u) := \frac{1}{6}(k_1 + 4k_2 + k_3),$$

wobei

$$k_1 := f(t,u), \qquad k_2 := f(t+\tfrac{1}{2}h, u+\tfrac{1}{2}hk_1), \qquad k_3 := f(t+h, u-hk_1+2hk_2).$$

Man berechne die zugehörige Stabilitätsfunktion und plotte mit Hilfe von Maple das Stabilitätsgebiet.

**Lösung:** Die Stabilitätsfunktion ist

$$R(z) = 1 + z + \frac{z^2}{2} + \frac{z^3}{6}.$$

Dies kann man entweder direkt nachrechnen (wobei auch hier Maple helfen kann) oder benutzen, dass es sich hier um ein 3-stufiges Runge-Kutta-Verfahren der Konsistenzordnung drei handelt. Wir plotten den Rand $\{z \in \mathbb{C} : |R(z)| = 1\}$ des Stabilitätsgebietes mit Hilfe der folgenden Sequenz:

```
with(plots):
z:=x+I*y;
implicitplot(abs(1+z+z^2/2+z^3/6)=1,x=-3..1,y=-3..3,scaling=constrained);
```

Hierbei sorgt die Option `scaling=constrained` dafür, dass auf beiden Achsen derselbe Maßstab angelegt wird. Das Ergebnis sieht man in Abbildung 5.22.

2. Man bestimme das Stabilitätsgebiet zum 2-Schrittverfahren

$$u(t+2h) - u(t) = 2hf(t+h, u(t+h)).$$

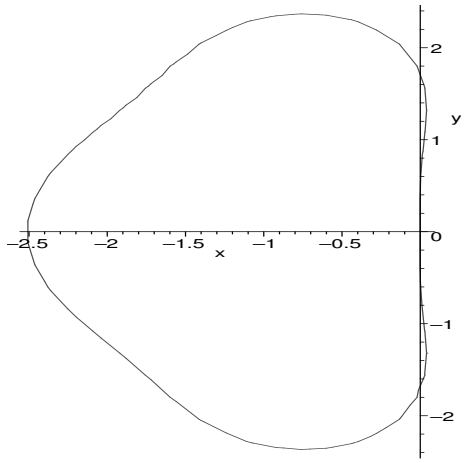


Abbildung 5.22: Das Stabilitätsgebiet $\{z \in \mathbb{C} : |1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3| \leq 1\}$

**Lösung:** Das zum Verfahren gehörende Stabilitätspolynom ist

$$\rho(\mu, z) = \mu^2 - 2\mu z - 1,$$

das Stabilitätsgebiet sei mit $S$ bezeichnet. Wurzeln des Stabilitätspolynoms sind

$$\mu_{1,2}(z) = z \pm \sqrt{1 + z^2}.$$

Wegen $\mu_1(z)\mu_2(z) = -1$ liegen für $z \in S$ beide Wurzeln auf dem Rande des Einheitskreises, es ist also etwa $\mu_1(z) = e^{i\phi}$ mit $\phi \in [0, 2\pi)$. Aus $z + \sqrt{z^2 + 1} = e^{i\phi}$ erhalten wir $z = \frac{1}{2}(e^{i\phi} - e^{-i\phi}) = i \sin \phi$. Also ist notwendigerweise

$$S \subset \{iy : y \in [-1, 1]\}.$$

Für $z = \pm i$ ist aber $\mu_1(z) = \mu_2(z) = \pm i$ eine doppelte Nullstelle des Stabilitätspolynoms, diese beiden Punkte gehören also nicht zum Stabilitätsgebiet. Insgesamt erkennen wir, dass

$$S = \{iy : y \in (-1, 1)\}$$

das gesuchte Stabilitätsgebiet ist.

3. Auf die Anfangswertaufgabe

$$x' = -2000(x - \cos t), \qquad x(0) = 0$$

wende man die (implizite) Trapezregel und das implizite Euler-Verfahren mit Schrittweite $h = 1.5/40$ an. Man plotte die erhaltenen Ergebnisse über dem Intervall $[0, 1.5]$ und vergleiche diese mit der exakten Lösung.

**Lösung:** Die exakte Lösung erhalten wir durch:

```
> dsolve({D(x)(t)=-2000*(x(t)-cos(t)),x(0)=0},x(t));
```

$$\mathrm{x}(t) = \frac{4000000}{4000001}\cos(t) + \frac{2000}{4000001}\sin(t) - \frac{4000000}{4000001}\,e^{(-2000\,t)}$$

In Abbildung 5.23 links geben wir die durch die Trapezregel (gezackte Linie) bzw. das implizite Euler-Verfahren (glatte Kurve) erhaltenenen Ergebnisse wieder. Rechts haben

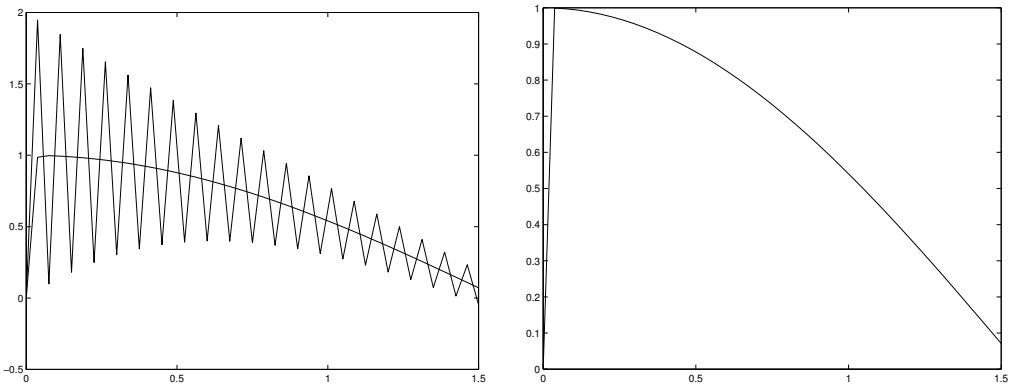

Abbildung 5.23: Lösung von $x' = -2000(x - \cos t)$, $x(0) = 0$

wir die Lösung geplottet. Frappierend sind die Unterschiede zwischen Trapezregel und implizitem Euler-Verfahren, welches ein wesentlich glatteres Ergebnis liefert. Den linken Plot haben wir ganz einfach durch das folgende Script-File erzeugt:

```
t=linspace(0,1.5,41); u=zeros(1,41);w=zeros(1,41);
h=1.5/40;
for j=1:40
   u(j+1)=((1-h*1000)*u(j)+h*1000*(cos(t(j))+cos(t(j+1))))/(1+h*1000);
   w(j+1)=(w(j)+h*2000*cos(t(j+1)))/(1+h*2000);
end;
plot(t,u,t,w);
```

4. Man zeige, dass das Stabilitätsgebiet des 2-Schrittverfahrens

$$u(t+2h) - u(t) = \frac{1}{2}h[f(t+h, u(t+h)) + 3f(t, u(t))]$$

im Kreis um $(-\frac{2}{3}, 0)$ mit dem Radius $\frac{2}{3}$ enthalten ist. Ferner zeige man, dass das reelle Intervall $[-\frac{4}{3}, 0]$ im Stabilitätsgebiet enthalten ist[28].

**Lösung:** Das zum Verfahren gehörende Stabilitätspolynom ist

$$\rho(\mu, z) = \mu^2 - \frac{z}{2}\mu - \left(1 + \frac{3z}{2}\right).$$

Lösungen von $\rho(\mu, z) = 0$ sind

$$\mu_{1,2}(z) = \frac{1}{4}(z \pm \sqrt{z^2 + 24z + 16}).$$

Mit $S$ sei das Stabilitätsgebiet bezeichnet. Ist $z \in S$, so ist

$$1 \geq |\mu_1(z)|\,|\mu_2(x)| = \left|1 + \frac{3z}{2}\right|,$$

[28] In einer Aufgabe bei K. Strehmel, R. Weiner (1995, S. 348) wird behauptet, dass das Stabilitätsgebiet des obigen 2-Schrittverfahrens ein Kreis mit dem Mittelpunkt $(-\frac{2}{3}, 0)$ und dem Radius $\frac{2}{3}$ ist. Wer kann das beweisen?

d. h. das Stabilitätsgebiet ist im Kreis um $(-\frac{2}{3}, 0)$ mit dem Radius $\frac{2}{3}$ enthalten. Nun sei $z \in [-\frac{4}{3}, 0]$. Für $z \in [-\frac{4}{3}, -12 + 8\sqrt{2})$ sind $\mu_{1,2}(z)$ konjugiert komplex:

$$\mu_{1,2}(z) = \frac{1}{4}[z \pm i\sqrt{-z^2 - 24z - 16}]$$

und daher

$$|\mu_{1,2}(z)|^2 = \frac{1}{16}[-24z - 16] \leq 1.$$

Daher gehört das Intervall $[-4/3, -12+8\sqrt{2})$ zum Stabilitätsgebiet. Für $z = -12+8\sqrt{2}$ ist zwar $\mu_1(z) = \mu_2(z) = -3+2\sqrt{2}$ eine doppelte Nullstelle, aber $|\mu_1(z)| = |\mu_2(z)| < 1$. Für $z \in (-12 + 8\sqrt{2}, 0]$ sind $\mu_{1,2}(z)$ reell. Offenbar ist nachzuweisen, dass

$$-1 \leq \frac{1}{4}[z - \sqrt{z^2 + 24z + 16}].$$

Dies ist aber richtig, wie man elementar nachweist oder aus dem Plot in Abbildung 5.24

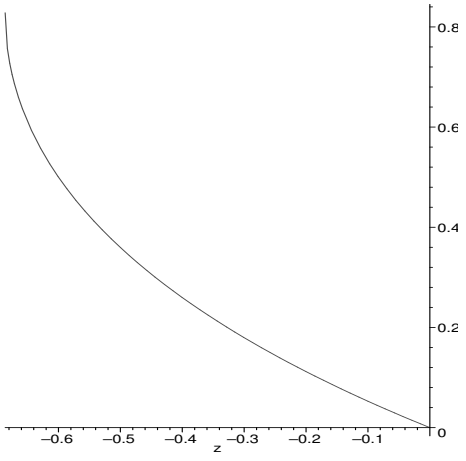


Abbildung 5.24: Plot von $\frac{1}{4}[z - \sqrt{z^2 + 24z + 16}] + 1$ auf $[-12 + 8\sqrt{2}, 0]$

entnimmt.

5. Gegeben sei das 2-stufige implizite Runge-Kutta-Verfahren mit

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

Man zeige, dass dieses die Konsistenzordnung 3 besitzt, berechne die Stabilitätsfunktion und beweise, dass das Verfahren A-stabil ist. So weit wie möglich sollte zur Bearbeitung dieser Aufgabe Maple eingesetzt werden.

**Lösung:** Es handelt sich hier um ein sogenanntes Radau-Verfahren, siehe E. Hairer, G. Wanner (1991, S. 78). Zunächst weisen wir nach, dass das Verfahren die Konsistenzordnung 3 besitzt. Der lokale Diskretisierungsfehler ist

$$\Delta(h, f)(t, u) = \frac{z(t+h) - z(t)}{h} - \frac{1}{4}[3k_1(t, u) + k_2(t, u)],$$

wobei $z$ die Lösung von $z' = f(s,z)$, $z(t) = u$, ist und $k_1(t,u), k_2(t,u)$ implizit durch

$$\begin{aligned} k_1(t,u) &= f(t+\tfrac{1}{3}h, u+\tfrac{1}{12}h(5k_1(t,u)-k_2(t,u))), \\ k_2(t,u) &= f(t+h, u+\tfrac{1}{4}h(3k_1(t,u)+k_2(t,u))) \end{aligned}$$

gegeben sind. Für die Entwicklung von $k_1$ und $k_2$ machen wir den Ansatz

$$\begin{aligned} k_1 &= A_1 + B_1 h + C_1 h^2 + O(h^3), \\ k_2 &= A_2 + B_2 h + C_2 h^2 + O(h^3). \end{aligned}$$

Geht man hiermit in obiges nichtlineares Gleichungssystem wein, so erhält man (wir haben hierzu in naheliegenderweise Maple benutzt)

$$\begin{aligned} A_1 + B_1 h + C_1 h^2 + O(h^3) &= f + [\tfrac{1}{3}f_t + \tfrac{1}{12}(5A_1 - A_2)f_x]h + [\tfrac{1}{18}f_{tt} + \tfrac{1}{12}(5B_1 - B_2)f_x \\ &\quad + \tfrac{1}{36}(5A_1 - A_2)f_{tx} + \tfrac{1}{288}(5A_1 - A_2)^2 f_{xx}]h^2 + O(h^3), \\ A_2 + B_2 h + C_2 h^2 + O(h^3) &= f + [f_t + \tfrac{1}{4}(3A_1 + A_2)f_x]h + [\tfrac{1}{2}f_{tt} + \tfrac{1}{4}(3B_1 + B_2)f_x \\ &\quad + \tfrac{1}{4}(3A_1 + A_2)f_{tx} + \tfrac{1}{32}(3A_1 + A_2)^2 f_{xx}]h^2 + O(h^3). \end{aligned}$$

Koeffizientenvergleich der Potenzen von $h$ ergibt das Gleichungssystem

$$\begin{aligned} A_1 &= f, \\ A_2 &= f, \\ B_1 &= \tfrac{1}{3}f_t + \tfrac{1}{12}(5A_1 - A_2)f_x, \\ B_2 &= f_t + \tfrac{1}{4}(3A_1 + A_2)f_x, \\ C_1 &= \tfrac{1}{18}f_{tt} + \tfrac{1}{12}(5B_1 - B_2)f_x + \tfrac{1}{36}(5A_1 - A_2)f_{tx} + \tfrac{1}{288}(5A_1 - A_2)^2 f_{xx}, \\ C_2 &= \frac{1}{2}f_{tt} + \frac{1}{4}(3B_1 + B_2)f_x + \frac{1}{4}(3A_1 + A_2)f_{tx} + \frac{1}{32}(3A_1 + A_2)^2 f_{xx}. \end{aligned}$$

Dieses kann sukzessive gelöst werden und wir erhalten (es ist $A_1 = A_2 = f$)

$$\begin{aligned} B_1 &= \tfrac{1}{3}(f_t + f_x f), \\ B_2 &= f_t + f_x f, \\ C_1 &= \tfrac{1}{18}[f_{tt} + (f_t + f_x f)f_x] + \tfrac{1}{9}f_{tx}f + \tfrac{1}{18}f_{xx}f^2, \\ C_2 &= \tfrac{1}{2}[f_{tt} + (f_t + f_x f)f_x] + f_{tx}f + \tfrac{1}{2}f_{xx}f^2. \end{aligned}$$

Damit kann der lokale Diskretisierungsfehler entwickelt werden in

$$\begin{aligned} \Delta(h,f) &= f + \tfrac{1}{2}(f_t + f_x f)h + \tfrac{1}{6}[f_{tt} + (f_t + f_x f)f_x + 2f_{tx}f + f_{xx}f^2]h^2 \\ &\qquad - \{[\tfrac{3}{4}A_1 + \tfrac{1}{4}A_2] + [\tfrac{3}{4}B_1 + \tfrac{1}{4}B_2]h^2 + [\tfrac{3}{4}C_1 + \tfrac{1}{4}C_2]h^2\} + O(h^3) \\ &= O(h^3), \end{aligned}$$

wie man durch Einsetzen feststellt. Das angegebene Verfahren ist daher ein Verfahren der Ordnung 3. Die Stabilitätsfunktion ist nach leichter Rechnung

$$R(z) = 1 + zb^T(I - zA)^{-1}e = \frac{6 + 2z}{6 - 4z + z^2}.$$

Mit

```
z:=x+I*y;
evalc(abs((6+2*z)/(6-4*z+z^2))^2);
```

und einigen Vereinfachungen erhalten wir

$$R(z) = \frac{36 + 24x + 4x^2 + 4y^2}{36 - 48x + 28x^2 - 8x^3 + x^4 + 4y^2 - 8xy^2 + 2x^2y^2 + y^4},$$

wobei $x = \Re(z)$, $y = \Im(z)$. Daher gehört $z$ genau dann zum Stabilitätsgebiet, wenn $|R(z)|^2 \leq 1$ bzw.

$$(*) \qquad x(72 - 24x + 8x^2 - x^3) \leq y^2(-8x + 2x^2 + y^2).$$

Hieraus liest man ab, dass die linke komplexe Halbebene im Stabilitätsgebiet enthalten ist, denn für $x \leq 0$ ist $(*)$ erfüllt, da

$$\underbrace{x(\underbrace{72 - 24x + 8x^2 - x^3}_{>0})}_{\leq 0} \leq \underbrace{y^2(4 - 8x + 2x^2 + y^2)}_{\geq 0}.$$

Daher ist das Verfahren A-stabil. Das Stabilitätsgebiet ist sogar noch wesentlich größer. Nach

```
with(plots):
implicitplot(x*(72-24*x+8*x^2-x^3)=y^2*(-8*x+2*x^2+y^2),x=-1..1,y=-5..5);
```

erhalten wir Abbildung 5.25 links. Alles, was sich "links" von dieser Kurve befindet,

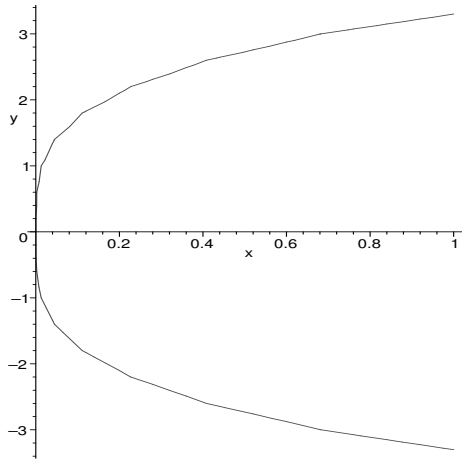


Abbildung 5.25: Stabilitätsgebiet eines Radau-Verfahrens der Ordnung 3

gehört zum Stabilitätsgebiet, also etwa auch der Punkt $z = 1 + 4i$ (in der Tat ist für diesen Punkt $|R(z)| = \frac{8}{233}\sqrt{466} \approx 0.741186$).

6. Ein implizites Runge-Kutta-Verfahren heißt *L-stabil*, wenn es A-stabil ist und zusätzlich $\lim_{z\to\infty} R(z) = 0$, wobei $R$ die zum Verfahren gehörende Stabilitätsfunktion ist.

Gegeben sei ein A-stabiles implizites $s$-stufiges Runge-Kutta-Verfahren mit den Daten

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

Hierbei sei $A$ nichtsingulär und

$$a_{sj} = b_j, \qquad j = 1, \ldots, s,$$

oder

$$a_{i1} = b_1, \qquad i = 1, \ldots, s.$$

Man zeige, dass das zugehörige Runge-Kutta-Verfahren L-stabil ist. Weiter zeige man, dass das implizite Euler-Verfahren L-stabil ist, nicht aber die Trapezregel.

**Lösung:** Die Stabilitätsfunktion zum Runge-Kutta-Verfahren mit den Daten $(A, b, c)$ ist

$$R(z) = 1 + zb^T(I - zA)^{-1}e.$$

Wegen $R(z) = 1 + b^T((1/z)I - A)^{-1}$ und der vorausgesetzten Nichtsingularität von $A$ ist

$$\lim_{z\to\infty} R(z) = 1 - b^T A^{-1} e.$$

Die Voraussetzung $a_{sj} = b_j$, $j = 1, \ldots, s$, besagt gerade, dass $e_s^T A = b^T$ (hierbei ist $e_s$ der $s$-te Einheitsvektor im $\mathbb{R}^s$). Daher ist

$$1 - b^T A^{-1} e = 1 - b^T A^{-1} e = 1 - e_s^T e = 0.$$

Ist dagegen $a_{i1} = b_1$, $i = 1, \ldots, s$, so ist $Ae_1 = b_1 e$ und daher $A^{-1}e = (1/b_1)e_1$, folglich ist auch in diesem Falle

$$1 - b^T A^{-1} e = 1 - b^T (1/b_1) e_1 = 0.$$

In beiden Fällen ist das Verfahren L-stabil. Das implizite Euler-Verfahren gehört zu den Daten

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array},$$

die Stabilitätsfunktion ist

$$R(z) = \frac{1}{1 - z}.$$

Hieraus liest man $\lim_{z\to\infty} R(z) = 0$ ab, man kann natürlich auch mit dem ersten Teil der Aufgabe argumentieren. Das Stabilitätsgebiet ist das Komplement des (offenen) Kreises um $(1, 0)$ mit dem Radius 1. Daher ist das implizite Euler-Verfahren L-stabil. Die Trapezregel gehört zu den Daten

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array} = \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2}. \end{array}$$

In diesem Falle ist die Matrix $A$ singulär, so dass der erste Teil der Aufgabe nicht angewandt werden kann. Die Stabilitätsfunktion ist

$$R(z) = \frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}.$$

Wegen $\lim_{z\to\infty} R(z) = -1$ ist die Trapezregel auch nicht L-stabil.

7. Bei einem $s$-stufigen impliziten Runge-Kutta-Verfahren für ein System von $n$ Differentialgleichungen erster Ordnung muss in jedem Zeitschritt zur Berechnung von $k_1, \ldots, k_s$ ein nichtlineares Gleichungssystem mit $ns$ Gleichungen und ebenso vielen Unbekannten gelöst werden. Ist die Verfahrensmatrix $A \in \mathbb{R}^{s\times s}$ eine untere Dreiecksmatrix, so spricht man von einem *diagonal-impliziten Runge-Kutta-Verfahren*. Jetzt können $k_1, \ldots, k_s$ sukzessive nach einander berechnet werden, so dass jetzt $s$ nichtlineare Gleichungssysteme mit $n$ Gleichungen und Unbekannten zu lösen sind. Man spricht von einem *einfach-diagonal-impliziten Runge-Kutta-Verfahren*, wenn die Diagonalelemente zusätzlich alle gleich sind, wenn also $a_{ii} = \gamma$, $i = 1, \ldots, s$.

   Man betrachte das einfach-diagonal-implizite Runge-Kutta-Verfahren zu den Daten

$$\begin{array}{c|cc} \gamma & \gamma & \\ c_2 & c_2 - \gamma & \gamma \\ \hline & b_1 & b_2 \end{array}$$

   mit $b_2 \neq 0$. Man gebe Bedingungen dafür, dass dieses Verfahren die Konsistenzordnung 2 oder sogar 3 besitzt. Schließlich gebe man ein 2-stufiges einfach-diagonal-implizites Runge-Kutta-Verfahren der Konsistenzordnung 3 an, welches $A$ stabil ist.

   **Lösung:** Es ist

$$\begin{aligned} k_1 &= f(t + \gamma h, u + h\gamma k_1), \\ k_2 &= f(t + c_2 h, u + h(c_2 - \gamma)k_1 + \gamma k_2). \end{aligned}$$

   Mit dem Ansatz

$$k_i = A_i + B_i h + C_i h^2 + O(h^3), \qquad i = 1, 2,$$

   gehen wir in das nichtlineare Gleichungssystem ein und machen einen Koeffizientenvergleich. Man erhält zunächst $A_1 = A_2 = f$, dann

$$B_1 = \gamma(f_t + f_x f), \qquad B_2 = c_2(f_t + f_x f)$$

   und

$$C_1 = \gamma^2\left[f_x(f_t + f_x f) + \frac{1}{2}(f_{tt} + 2f_{tx}f + f_{xx}f^2)\right]$$

   sowie

$$C_2 = (2c_2 - \gamma)\gamma f_x(f_t + f_x f) + \frac{1}{2}c_2^2(f_{tt} + 2f_{tx}f + f_{xx}f^2).$$

   Der lokale Diskretisierungsfehler ist folglich

$$\Delta(h, f) = \frac{z(t + h) - z(t)}{h} - (b_1 k_1 + b_2 k_2)$$

$$
\begin{aligned}
= \quad & f + \frac{h}{2}(f_t + f_x f) + \frac{h^2}{6}[f_x(f_t + f_x f) + (f_{tt} + 2f_{tx}f + f_{xx}f^2)] + O(h^3) \\
& - (b_1 + b_2)f + h[b_1\gamma + b_2 c_2)](f_t + f_x f) \\
& - h^2\{[b_1\gamma^2 + b_2(2c_2 - \gamma)\gamma]f_x(f_t + f_x f) \\
& + \tfrac{1}{2}(b_1\gamma^2 + b_2 c_2^2)(f_{tt} + 2f_{tx}f + f_{xx}f^2)\}.
\end{aligned}
$$

Folglich hat das Verfahren die Konsistenzordnung 2, wenn

$$b_1 + b_2 = 1, \qquad b_1\gamma + b_2 c_2 = \frac{1}{2}.$$

Dies ist der Fall, wenn

$$b_1 = \frac{c_2 - \gamma}{c_2 - \gamma}, \qquad b_2 = \frac{\frac{1}{2} - \gamma}{c_2 - \gamma}, \qquad c_2 \neq \gamma.$$

Damit das Verfahren die Konsistenzordnung 3 hat, muss zusätzlich gelten:

$$b_1\gamma^2 + b_2(2c_2 - \gamma)\gamma = \frac{1}{6}, \qquad \frac{1}{2}(b_1\gamma^2 + b_2 c_2^2) = \frac{1}{6}.$$

Mit Maple erhält man sehr leicht, dass

$$\gamma = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}, \qquad c_2 = \frac{1}{2} \mp \frac{1}{6}\sqrt{3}$$

und

$$b_1 = b_2 = \frac{1}{2}.$$

Die Daten dieses Verfahrens sind also gegeben durch

$$
\begin{array}{c|cc}
\frac{1}{2} \pm \frac{1}{6}\sqrt{3} & \frac{1}{2} \pm \frac{1}{6}\sqrt{3} & \\
\frac{1}{2} \mp \frac{1}{6}\sqrt{3} & \mp\frac{1}{3}\sqrt{3} & \frac{1}{2} \pm \frac{1}{6}\sqrt{3} \\
\hline
 & \frac{1}{2} & \frac{1}{2}
\end{array}
$$

Die zugehörige Stabilitätsfunktion ist

$$R(z) = \frac{1 + (1 - 2\gamma)z + (\frac{1}{2} - 2\gamma + \gamma^2)z^2}{(1 - \gamma z)^2}$$

mit

$$\gamma = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}.$$

Wir wollen uns überlegen, dass mit $\gamma := \frac{1}{2} + \frac{1}{6}\sqrt{3}$ das Verfahren A-stabil ist. Die Stabilitätsfunktion $R$ ist analytisch in der linken Halbebene. Wegen des Maximumprinzips für analytische Funktionen ist daher $|R(z)| \leq 1$ für alle $z$ mit $\Re(z) \leq 0$, wenn $|R(iy)| \leq 1$ für alle reellen $y$. Mit Hilfe von Maple erhalten wir

$$|R(iy)|^2 = \frac{36 + (24 + 12\sqrt{3})y^2 + (4 + 2\sqrt{3})y^4}{36 + (12\sqrt{3})y^2 + (7 + 4\sqrt{3})y^4} \leq 1$$

für alle $y \in \mathbb{R}$. Daher ist das entsprechende Verfahren A-stabil. Ist dagegen $\gamma := \frac{1}{2} - \frac{1}{6}\sqrt{3}$, so ist

$$|R(iy)|^2 = \frac{36 + (24 - 12\sqrt{3})y^2 + (4 - 2\sqrt{3})y^4}{36 + (24 - 12\sqrt{3})y^2 + (7 - 4\sqrt{3})y^4} > 1$$

für alle $y \in \mathbb{R}$, das Verfahren also nicht A-stabil. Das einzige A-stabile 2-stufige einfach-diagonal-implizite Runge-Kutta-Verfahren der Konsistenzordnung 3 ist also durch die Daten

$$\begin{array}{c|cc} \frac{1}{2} + \frac{1}{6}\sqrt{3} & \frac{1}{2} + \frac{1}{6}\sqrt{3} & \\ \frac{1}{2} - \frac{1}{6}\sqrt{3} & -\frac{1}{3}\sqrt{3} & \frac{1}{2} + \frac{1}{6}\sqrt{3} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

gegeben.

8. Man löse die Anfangswertaufgabe

$$x' = -100(x - \sin t), \qquad x(0) = 1,$$

auf dem Zeitintervall $[0, 10]$ mit den MATLAB-Funktionen `ode45` und `ode15s`. Man vergleiche die Anzahl der benötigten Zeitschritte.

**Lösung:** Wir schreiben das File `f.m` mit dem Inhalt

```
function out=f(t,x);
out=-100*(x-sin(t));
```

Nach dem Aufruf

```
[t,x]=ode45('f',[0,10],1);
```

und `lengt(t)` wissen wir, dass 1293 Zeitschritte durchgeführt wurden. Dagegen liefern

```
[t,x]=ode15s('f',[0,10],1);
```

und `lengt(t)` die Information, dass hier nur 78 Zeitschritte nötig waren.

9. Die folgende Definition spielt in einem vertieften Studium steifer Differentialgleichungen eine wichtige Rolle (siehe z. B. K. Strehmel, R. Weiner (1995, S. 199) und E. Hairer et al. (1993, S. 61)).

Sei $\|\cdot\|$ eine beliebige Vektornorm im $\mathbb{R}^n$. Ist dann $\|\cdot\|$ die zugeordnete Matixnorm, so heißt

$$\mu(A) := \lim_{\delta \to 0+} \frac{\|I + \delta A\| - 1}{\delta}$$

die zugeordnete *logarithmische Norm*. Man zeige:

(a) Die logarithmische Norm existiert, da die Abbildung $\delta \mapsto (\|I + \delta A\| - 1)/\delta$ auf $(0, 1]$ nach unten beschränkt und monoton nicht fallend ist.

(b) Die der euklidischen Norm $\|\cdot\|$ zugeordnete Matrixnorm ist bekanntlich die sogenannte Spektralnorm $\|\cdot\|_2$, wobei $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$. Man berechne die zugehörige logarithmische Norm.

(c) Die der Maximumnorm $\|\cdot\|_\infty$ zugeordnete Matrixnorm ist bekanntlich die maximale Betragssummennorm $\|\cdot\|_\infty$, wobei $\|A\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^n |a_{ij}|$. Man berechne die zugehörige logarithmische Norm.

(d) Man beweise die folgenden Eigenschaften der logarithmischen Norm:

i. $\mu(\alpha A) = \alpha\mu(A)$ für $\alpha \geq 0$.
ii. $|\mu(A)| \leq \|A\|$.
iii. $\mu(A+B) \leq \mu(A) + \mu(B)$.
iv. $|\mu(A) - \mu(B)| \leq \|A - B\|$.

**Lösung:** Wir definieren $\phi\colon (0,1] \longrightarrow \mathbb{R}$ durch

$$\phi(\delta) := \frac{\|I + \delta A\| - 1}{\delta}.$$

Offensichtlich ist $\phi(\delta) \geq -\|A\|$ für alle $\delta \in (0,1]$, also ist $\phi(\cdot)$ auf $(0,1]$ nach unten beschränkt. Sei nun $0 < \eta \leq \delta \leq 1$. Dann ist

$$\begin{aligned}
\|I + \eta A\| - 1 &= \left\| \frac{\eta}{\delta}(I + \delta A) + \left(1 - \frac{\eta}{\delta}\right) I \right\| - 1 \\
&\leq \frac{\eta}{\delta}\|I + \delta A\| + \left(1 - \frac{\eta}{\delta}\right)\|I\| - 1 \\
&= \frac{\eta}{\delta}(\|I + \delta A\| - 1)
\end{aligned}$$

bzw. $\phi(\eta) \leq \phi(\delta)$. Folglich ist die logarithmische Norm wohldefiniert.

Für $\delta > 0$ ist

$$\begin{aligned}
\|I + \delta A\|_2 &= \sqrt{\lambda_{\max}((I + \delta A)^T (I + \delta A))} \\
&= \sqrt{\lambda_{\max}(I + \delta(A^T + A) + \delta^2 A^T A)} \\
&= 1 + \frac{\delta}{2}\lambda_{\max}(A^T + A) + O(\delta^2)
\end{aligned}$$

und daher

$$\mu_2(A) = \frac{1}{2}\lambda_{\max}(A^T + A).$$

Wir zeigen, dass

$$\mu_\infty(A) = \max_{i=1,\ldots,n} \left( a_{ii} + \sum_{j \neq i} |a_{ij}| \right).$$

Denn für alle hinreichend kleinen $\delta > 0$ ist

$$\begin{aligned}
\|I + \delta A\|_\infty - 1 &= \max_{i=1,\ldots,n} \left( |1 + \delta a_{ii}| + \delta \sum_{j \neq i} |a_{ij}| \right) - 1 \\
&= \max_{i=1,\ldots,n} \left( 1 + \delta a_{ii} + \delta \sum_{j \neq i} |a_{ij}| \right) - 1 \\
&= \delta \max_{i=1,\ldots,n} \left( a_{ii} + \sum_{j \neq i} |a_{ij}| \right),
\end{aligned}$$

woraus die Behauptung folgt.

Die Aussage $\mu(\alpha A) = \alpha\mu(A)$ ist für $\alpha = 0$ richtig, so dass $\alpha > 0$ vorausgesetzt werden kann. Wegen

$$\mu(A) \leftarrow \frac{\|I + \delta\alpha A\| - 1}{\delta\alpha} = \frac{1}{\alpha}\frac{\|I + \delta(\alpha)A\| - 1}{\delta} \rightarrow \frac{1}{\alpha}\mu(\alpha A)$$

folgt die erste Aussage.

Wegen

$$\frac{|\|I + \delta A\| - 1|}{\delta} \leq \|A\|$$

für alle $\delta > 0$ folgt mit $\delta \rightarrow 0+$, dass $|\mu(A)| \leq \|A\|$.

Mit $\delta \rightarrow 0+$ gilt

$$\begin{aligned}
\mu(A+B) &\leftarrow \frac{\|I + \frac{1}{2}\delta(A+B)\| - 1}{\frac{1}{2}\delta} \\
&\leq \frac{\|\frac{1}{2}I + \frac{1}{2}\delta A\| - \frac{1}{2}}{\frac{1}{2}\delta} + \frac{\|\frac{1}{2}I + \frac{1}{2}\delta B\| - \frac{1}{2}}{\frac{1}{2}\delta} \\
&= \frac{\|I + \delta A\| - 1}{\delta} + \frac{\|I + \delta B\| - 1}{\delta} \\
&\rightarrow \mu(A) + \mu(B),
\end{aligned}$$

womit auch $\mu(A + B) \leq \mu(A) + \mu(B)$ bewiesen ist.

Es ist unter Benutzung der letzten beiden Teile

$$\mu(A) - \mu(B) = \mu(A - B + B) - \mu(B) \leq \mu(A - B) \leq \|A - B\|.$$

Zusammen mit einer Vertauschung der Rollen von $A$ und $B$ folgt die Behauptung.

# 5.4 Aufgaben zu Kapitel 4

## 5.4.1 Aufgaben zu Abschnitt 4.1

1. Sei $p \in C^1[a, b]$ mit $p(x) > 0$ auf $[a, b]$, ferner $q \in C[a, b]$ mit $q(x) \geq 0$ für alle $x \in [a, b]$. Man zeige: Ist $u \in C[a, b] \cap C^2(a, b)$ mit

$$Lu(x) := -[p(x)u'(x)]' + q(x)u(x) \geq 0 \qquad \text{für alle } x \in (a, b)$$

und

$$Ru := \begin{pmatrix} u(a) \\ u(b) \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

so ist $u(x) \geq 0$ für alle $x \in [a, b]$.

**Lösung:** Wir nehmen im Gegensatz zur Behauptung an, es sei $\min_{x\in[a,b]} u(x) < 0$. Nach Voraussetzung ist $u$ an den Intervallgrenzen $a$ und $b$ nichtnegativ, so dass das Minimum in einem Punkt $x^*$ im offenen Intervall $(a, b)$ angenommen wird. Ferner sei $I^*$ ein offenes, $x^*$ enthaltendes Intervall mit $v(x) < 0$ für alle $x \in I^*$. Für alle $x \in I^*$ ist dann

$$-[p(x)u'(x)]' \geq -q(x)u(x) \geq 0.$$

Wegen $u'(x^*) = 0$ folgt hieraus, dass $u'(x) \geq 0$ für alle $x \in I^*$ mit $x < x^*$ und $u'(x) \leq 0$ für alle $x \in I^*$ mit $x > x^*$. Da $u$ in $x^*$ ein Minimum annimmt, ist $u(x) = u(x^*)$ für alle $x \in I^*$. Dann ist aber $u$ auf ganz $[a, b]$ konstant und negativ, was ein Widerspruch zu $u(a) \geq 0$ bzw. $u(b) \geq 0$ ergibt.

2. Sei $p \in C^1[a, b]$ mit $p(x) > 0$ auf $[a, b]$, ferner $q \in C[a, b]$ mit $q(x) > 0$ für alle $x \in [a, b]$. Für $u \in C[a, b] \cap C^2(a, b)$ gelten dann die folgenden Implikationen:

   (a) $Lu(x) \leq 0$ in $(a, b) \Longrightarrow u(x) \leq \max(0, u(a), u(b))$,

   (b) $Lu(x) \geq 0$ in $(a, b) \Longrightarrow u(x) \geq \min(0, u(a), u(b))$.

   Hinweis: Im ersten Fall nehme man im Widerspruch zur Behauptung an, $u$ besitze in $(a, b)$ ein positives Maximum. Den zweiten Fall beweise man entsprechend durch Widerspruch.

   **Lösung:** Sei $Lu(x) \leq 0$ für alle $x \in (a, b)$. Die Funktion $u$ nehme auf $[a, b]$ in $x^*$ ihr Maximum an. Im Widerspruch zur Behauptung sei $u(x^*) > \max(0, u(a), u(b))$. Dann ist also $x^* \in (a, b)$ und folglich $u'(x^*) = 0$, $u''(x^*) \leq 0$ sowie $u(x^*) > 0$. Da wir auch die Positivität von $q$ auf $[a, b]$ vorausgesetzt haben (die Nichtnegativität hätte gelangt), ist

$$0 \geq Lu(x^*) = \underbrace{-p(x^*)u''(x^*)}_{\geq 0} + p'(x^*)\underbrace{u'(x^*)}_{=0} + \underbrace{q(x^*)u(x^*)}_{>0} > 0,$$

   ein Widerspruch. Den zweiten Fall beweist man entsprechend.

3. Gegeben seien der Differentialoperator $L\colon C_n^1[a, b] \longrightarrow C_n[a, b]$ und der Randoperator $R\colon C_n[a, b] \longrightarrow \mathbb{R}^n$ durch

$$(Lu)(x) := u'(x) - A(x)u(x), \qquad Ru := Cu(a) + Du(b).$$

   Hierbei ist $A \in C_{n\times n}[a, b]$ und $C, D \in \mathbb{R}^{n\times n}$. Sei $U(\cdot)$ ein Fundamentalsystem zu $Lu = 0$ und $R(U) := CU(a) + DU(b)$. Man zeige:

   (a) Die folgenden drei Aussagen sind äquivalent:

      (i) Die homogene Aufgabe $Lu = 0$, $Ru = 0$ besitzt nur die triviale Lösung $u = 0$.

      (ii) Die Matrix $R(U)$ ist nichtsingulär.

      (iii) Zu vorgegebenen $g \in C_n[a, b]$, $\eta \in \mathbb{R}^n$ besitzt die Randwertaufgabe

$$Lu = g(x), \qquad Ru = \eta$$

      genau eine Lösung.

   (b) Sei $R(U)$ nichtsingulär. Definiert man die sogenannte *Greensche Matrix* $G\colon [a, b] \times [a, b] \longrightarrow \mathbb{R}^{n\times n}$ durch

$$G(x, \xi) := \begin{cases} U(x)[I - R(U)^{-1}DU(b)]U^{-1}(\xi), & a \leq \xi \leq x \leq b, \\ -U(x)R(U)^{-1}DU(b)U^{-1}(\xi), & a \leq x < \xi \leq b, \end{cases}$$

   so hat diese die folgenden Eigenschaften:

      (i) Es gilt die Sprungbedingung $G(x + 0, x) - G(x - 0, x) = I$ für $x \in [a, b]$.

      (ii) Bei festem $\xi \in [a, b]$ ist $LG(x, \xi) = 0$ für $x \in [a, b] \setminus \{\xi\}$.

(iii) Für festes $\xi \in (a,b)$ ist $RG(\cdot,\xi) = 0$.

(iv) Bei vorgegebenem $g \in C_n[a,b]$ lässt sich die eindeutige Lösung von $Lu = g(x)$, $Ru = 0$ durch

$$u(x) := \int_a^b G(x,\xi)g(\xi)\,d\xi$$

darstellen.

(v) Durch die Bedingungen (i)–(iii) ist die Greensche Matrix eindeutig festgelegt.

**Lösung:** Es gelte (i). Im Widerspruch zu (i) nehmen wir an, es gäbe ein $c \neq 0$ mit $R(U)c = 0$, ferner definiere man $u(x) := U(x)c$. Im Widersoruch zu (i) wäre $u$ eine nichttriviale Lösung von $Lu = 0$, $Ru = 0$.

Es gelte (ii). Seien $g \in C_n[a,b]$ und $\eta \in \mathbb{R}^n$ beliebig vorgegeben. Die allgemeine Lösung von $Lu = g(x)$ ist $u(x) = u^*(x)+U(x)c$ mit einer speziellen Lösung $u^*$ der inhomogenen Differentialgleichung und $c \in \mathbb{R}^n$. Es ist

$$Ru = Ru^* + R(U)c,$$

so dass schlie'slich

$$u(x) := u^*(x) + U(x)R(U)^1(\eta - Ru^*)$$

die eindeutige L'osung von $Lu = g(x)$, $Ru = \eta$ ist.

Aus (iii) folgt trivialerweise (i) (setze $g := 0$ und $\eta := 0$).

Es ist

$$\begin{aligned} G(x+0,x) - G(x-0,x) &= U(x)[I - R(U)^{-1}DU(b)]U^{-1}(x) \\ &\quad + U(x)R(U)^{-1}DU(b)U^{-1}(x) \\ &= I, \end{aligned}$$

womit schon die erste Eigenschaft (i) bewiesen ist. Die zweite Eigenschaft (ii) ist selbstverständlich, da $G$ die Form

$$G(x,\xi) = \begin{cases} U(x)C(\xi), & a \le \xi \le x \le b, \\ U(x)D(\xi), & a \le x < \xi \le b \end{cases}$$

hat. Für festes $\xi \in (a,b)$ ist

$$\begin{aligned} RG(\cdot,\xi) &= CG(a,\xi) + DG(b,\xi) \\ &= -CU(a)R(U)^{-1}DU(b)U^{-1}(\xi) + DU(b)[I - R(U)^{-1}DU(b)]U^{-1}(\xi) \\ &= DU(b)U^{-1}(\xi) - [\underbrace{CU(a) + DU(b)}_{=R(U)}R(U)^{-1}DU(b)U^{-1}(\xi) \\ &= 0, \end{aligned}$$

so dass auch die dritte Eigenschaft bewiesen ist. In (iv) haben wir nachzuweisen, dass

$$u(x) := \int_a^b G(x,\xi)g(\xi)\,d\xi$$

Lösung von $Lu = g(x)$, $Ru = \eta$ ist. Es ist

$$\begin{aligned} u(x) &= U(x)[I - R(U)^{-1}DU(b)] \int_a^x U^{-1}(\xi)g(\xi)\,d\xi \\ &\quad - U(x)R(U)^{-1}DU(b) \int_x^b U^{-1}(\xi)g(\xi)\,d\xi \end{aligned}$$

und daher

$$\begin{aligned} Lu(x) &= U(x)[I - R(U)^{-1}DU(b)]U^{-1}(x)g(x) \\ &\quad + U(x)R(U)^{-1}DU(b) \int_x^b U^{-1}(x)g(x) \\ &= g(x). \end{aligned}$$

Weiter ist

$$Ru = \int_a^b \underbrace{RG(\cdot,\xi)}_{=0} g(\xi)\,d\xi = 0,$$

so dass (iv) nachgewiesen ist. In (v) weisen wir nach, dass die Greensche Matrix durch die Eigenschaften (i)–(iii) festgelegt ist. Wegen (ii) hat die Greensche Matrix die Form

$$G(x,\xi) = \begin{cases} U(x)C(\xi), & a \le \xi \le x \le b, \\ U(x)D(\xi), & a \le x < \xi \le b. \end{cases}$$

Die Sprungbeziehung (i) liefert $U(x)[C(x) - D(x)] = I$. Eine die Bedingungen (i) und (ii) erfüllende Matrixfunktion $G$ hat also notwendigerweise die Form

$$G(x,\xi) = \begin{cases} U(x)[D(\xi) + U^{-1}(\xi)], & a \le \xi \le x \le b, \\ U(x)D(\xi), & a \le x < \xi \le b. \end{cases}$$

Dann ist aber

$$RG(\cdot,\xi) = CU(a)D(\xi) + DU(b)[D(\xi) + U^{-1}(\xi)] = R(U)D(\xi) + DU(b)U^{-1}(\xi) = 0$$

genau dann, wenn

$$D(\xi) = -R(U)^{-1}DU(b)U^{-1}(\xi).$$

Dies liefert dann genau die angegebene Greensche Matrix.

4. Man[29] definiere den Differentialoperator $L$ und den Randoperator $R$ durch

$$Lu(x) := -u''(x) - \frac{1}{4x^2}u(x), \qquad Ru := \begin{pmatrix} u(1) \\ u(2) \end{pmatrix}.$$

Man bestimme die zugehörige Greensche Funktion.

**Lösung:** Bei der Bearbeitung der Aufgabe benutzen wir Maple. Als Lösung $u_1$ von $Lu = 0$, $u(1) = 0$, $u(2) = 1$ erhalten wir

$$u_1(x) := \frac{\sqrt{2}}{2}\,\frac{\sqrt{x}\ln(x)}{\ln(2)}.$$

[29] Diese Aufgabe haben wir W. Walter (1993, S. 226) entnommen.

Entsprechend ist die Lösung $u_2$ von $Lu = 0$, $u(1) = 1$, $u(2) = 0$ gegeben durch

$$u_2(x) = \sqrt{x} - \frac{\sqrt{x}\ln(x)}{\ln(2)}.$$

Weiter ist

$$c := u_1(x)u_2'(x) - u_1'(x)u_2(x) = -\frac{\sqrt{2}}{2\ln(2)}.$$

Daher ist die Greensche Funktion zu $(L, R)$ gegeben durch

$$G(x,\xi) = \begin{cases} \sqrt{\xi}\ln(\xi)\sqrt{x}[\ln(x)/\ln(2) - 1], & 1 \le \xi \le x \le 2, \\ \sqrt{\xi}[\ln(\xi)/\ln(2) - 1]\sqrt{x}\ln(x), & 1 \le x \le \xi \le 2. \end{cases}$$

### 5.4.2 Aufgaben zu Abschnitt 4.2

1. Gegeben sei das Eigenwertproblem

$$-u'' = \lambda u, \qquad u(0) = u'(0), \quad u(1) = 0.$$

Man[30] bestimme die Eigenwerte $\lambda_k$ und Eigenfunktionen $u_k$ und zeige, dass

$$\sqrt{\lambda_k} = \frac{1}{2}\pi + k\pi + \beta_k \qquad (k = 0, 1, \ldots) \quad \text{mit} \quad \beta_k \downarrow 0 \quad (k \to \infty).$$

Man skizziere die ersten beiden Eigenfunktionen $u_0$ und $u_1$.

**Lösung:** Jeder Eigenwert ist notwendigerweise positiv. Denn ist $\lambda$ ein Eigenwert und $u$ eine zugehörige Eigenfunktion, so ist

$$\lambda \int_0^1 u(x)^2\,dx = -\int_0^1 u''(x)u(x)\,dx = u(0)^2 + \int_0^1 u'(x)^2\,dx.$$

Hieraus folgt offenbar $\lambda > 0$. Für eine Eigenfunktion machen wir den Ansatz

$$u(x) = A\cos\sqrt{\lambda}x + B\sin\sqrt{\lambda}x.$$

Die Randbedingungen $u(0) = u'(0)$ und $u(1) = 0$ sind genau dann erfüllt, wenn

$$A - B\sqrt{\lambda} = 0, \qquad A\cos\sqrt{\lambda} + B\sin\sqrt{\lambda} = 0.$$

Folglich ist $\lambda > 0$ genau dann ein Eigenwert, wenn

$$\det\begin{pmatrix} 1 & -\sqrt{\lambda} \\ \cos\sqrt{\lambda} & \sin\sqrt{\lambda} \end{pmatrix} = \sin\sqrt{\lambda} + \sqrt{\lambda}\cos\sqrt{\lambda} = 0$$

bzw. $\tan\sqrt{\lambda} = -\sqrt{\lambda}$. Die Gleichung $\tan x = -x$ besitzt in $(\frac{1}{2}\pi + k\pi, (k+1)\pi)$ genau eine Lösung $x_k$, $k = 0, 1, \ldots$. Definiert man $\beta_k := x_k - \frac{1}{2}\pi - k\pi$, so ist $\beta_k > 0$. Weiter konvergiert $\{\beta_k\}$ monoton fallend gegen Null. Man erhält sehr leicht (etwa mit dem Maple-Befehl `fsolve`), dass

$$\sqrt{\lambda_0} = 2.028757838, \qquad \sqrt{\lambda_1} = 4.913180439.$$

Die Eigenfunktionen

$$u_k(x) = \sqrt{\lambda_k}\cos\sqrt{\lambda_k}x + \sin\sqrt{\lambda_k}x$$

werden für $k = 0, 1$ in Abbildung 5.26 dargestellt.

[30] Diese Aufgabe findet man bei W. Walter (1993, S. 235).

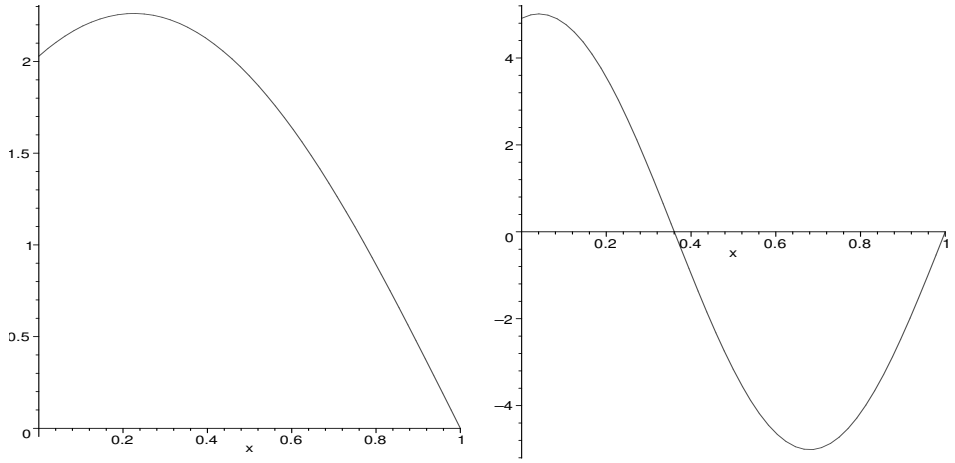


Abbildung 5.26: Die ersten beiden Eigenfunktionen

2. Man[31] löse das Eigenwertproblem

$$-(xu')' = \frac{\lambda}{x}u, \qquad u'(1) = 0, \quad u'(e^{2\pi}) = 0.$$

Ist $\lambda = 0$ ein Eigenwert?

**Lösung:** Wir beantworten zuerst die letzte Frage. Natürlich ist $\lambda = 0$ ein Eigenwert, z. B. mit der zugehörigen Eigenfunktion $u_0(x) := 1$. Ist $\lambda$ ein Eigenwert mit zugehörigem Eigenvektor $u$, so ist

$$\int_1^{e^{2\pi}} \frac{1}{x}u(x)^2\,dx = -\int_1^{e^{2\pi}} (xu'(x))'u(x)\,dx = \int_1^{e^{2\pi}} xu'(x)^2\,dx,$$

woraus wir schließen, dass alle Eigenwerte nichtnegativ sind. Nach

```
assume(lambda,positive);
dsolve(x*(D@@2)(u)(x)+D(u)(x)+(lambda/x)*u(x)=0,u(x));
```

erhalten wir

$$u(x) = A\cos(\sqrt{\lambda}\ln x) + B\sin(\sqrt{\lambda}\ln x)$$

als potentielle Eigenfunktionen. Wegen

$$u'(x) = -\frac{A\sqrt{\lambda}\sin(\sqrt{\lambda}\ln x)}{x} + \frac{B\sqrt{\lambda}\cos(\sqrt{\lambda}\ln x)}{x}$$

erhalten wir durch die Randbedingung $u'(1) = 0$, dass $B = 0$, während $u'(e^{2\pi}) = 0$ die Information $\lambda_k = (k/2)^2$ nach sich zieht. Eine zugehörige Eigenfunktion ist

$$u_k(x) = \cos((k/2)\ln x), \qquad k = 1, 2, \ldots.$$

[31] Diese Aufgabe findet man bei W. Walter (1993, S. 235).

3. Sei $p \in C^1(\mathbb{R})$, $q \in C(\mathbb{R})$ und $p(x) > 0$ für alle $x \in \mathbb{R}$. Hiermit definiere man den Differentaloperator $L\colon C^2(\mathbb{R}) \longrightarrow C(\mathbb{R})$ durch

$$(Lu)(x) := -[p(x)u'(x)]' + q(x)u(x).$$

Man zeige: Eine nichttriviale Lösung $u$ von $Lu = 0$ hat nur einfache Nullstellen, und zwar endlich oder abzählbar viele. Im zweiten Fall haben die Nullstellen keinen Häufungspunkt.

**Lösung:** Wäre $u(x_0) = u'(x_0) = 0$, so folgt $u = 0$ aud dem Eindeutigkeitssatz. Ist $u(x_k) = 0$, $k \in \mathbb{N}$, und ist $\lim_{k\to\infty} x_k = \xi$, so ist natürlich $u(\xi) = 0$. Wegen

$$u'(\xi) = \lim_{k\to\infty} \frac{u(x_k) - u(\xi)}{x_k - \xi} = 0$$

folgt wieder aus dem Eindeutigkeitssatz, dass $u = 0$.

4. Man löse das folgende Eigenwertproblem mit *periodischen* Randbedingungen:

$$-u'' = \lambda u, \qquad u(0) = u(1), \quad u'(0) = u'(1).$$

**Lösung:** Eigenwerte sind mit einem Routineargument notwendigerweise nichtnegativ. Offenbar ist $\lambda_0 = 0$ ein Eigenwert mit der Eigenfunktion $u_0(x) = 1$. Für einen Eigenwert $\lambda > 0$ hat eine zugehörige Eigenfunktion $u$ die Form

$$u(x) = A\cos\sqrt{\lambda}x + B\sin\sqrt{\lambda}x.$$

Die periodischen Randbedingungen liefern die Gleichungen

$$A(1 - \cos\sqrt{\lambda}) - B\sin\sqrt{\lambda} = 0, \qquad A\sin\sqrt{\lambda} + B(1 - \cos\sqrt{\lambda}) = 0.$$

Daher ist $\lambda > 0$ genau dann ein Eigenwert, wenn

$$\det\begin{pmatrix} 1 - \cos\sqrt{\lambda} & -\sin\sqrt{\lambda} \\ \sin\sqrt{\lambda} & 1 - \cos\sqrt{\lambda} \end{pmatrix} = (1 - \cos\sqrt{\lambda})^2 + \sin^2\sqrt{\lambda} = 0$$

bzw. $\sqrt{\lambda} = 2k\pi$ mit $k \in \mathbb{N}$. Daher hat man noch die Eigenwerte $\lambda_k = 4k^2\pi^2$, $k = 1, 2, \ldots$.