Logbuch zur Vorlesung

Gewöhnliche Differentialgleichungen im SS 21

Christian Schmeiser[1]

Contents

[1]Institut für Mathematik, Universität Wien, Nordbergstraße 15, 1090 Wien, Austria.
http://homepage.univie.ac.at/christian.schmeiser/

1 Einleitung

Die Vorlesung folgt dem Lehrbuch

[T] G. Teschl, *Ordinary Differential Equations and Dynamical Systems,* Graduate Studies in Mathematics, Volume 140, Amer. Math. Soc., Providence, 2012.

Der Vorlesungsstoff ist diesem Logbuch zu entnehmen, das die besprochenen Teile von [T] angibt, sowie nicht in [T] enthaltenes Material.

1.1 Warum Differentialgleichungen? – Newtonsche Mechanik

Newtonsche Gesetze (*Isaac Newton,* 1643–1727):

1. **Trägheit:** Ein Objekt bewegt sich mit konstanter Geschwindigkeit, solange keine Kraft auf es wirkt.
2. **Kraft = Masse × Beschleunigung**
3. **Aktion-Reaktion:** Übt ein Objekt auf ein zweites Objekt eine Kraft aus, dann übt auch das zweite Objekt auf das erste eine Kraft aus, die dieselbe Größe und die entgegengesetzte Richtung hat.

Mathematische Formulierung: (der *Teilchenmechanik*, d.h. Objekte ohne räumliche Ausdehnung)

Zeit: $t \in \mathbb{R}$

Position des Objekts zum Zeitpunkt t: $x(t) \in \mathbb{R}^3$

Geschwindigkeit ist Änderung der Position pro Zeiteinheit:

$$v_{\Delta t}(t) := \frac{x(t+\Delta t) - x(t)}{\Delta t} \in \mathbb{R}^3 .$$

Streng genommen: *Durchschnittsgeschwindigkeit* im Zeitintervall $[t, t+\Delta t]$.

Beschleunigung ist Änderung der Geschwindigkeit pro Zeiteinheit:

$$a_{\Delta t}(t) := \frac{v(t+\Delta t) - v(t)}{\Delta t} \in \mathbb{R}^3 .$$

Daher folgt eigentlich das 1. Gesetz aus dem 2. (wenn die Masse nicht Null ist).

Masse des Objekts: $m \in \mathbb{R}$, $m > 0$,
zum Zeitpunkt t konzentriert an der Position $x(t)$ (im Unterschied zu im Raum verteilter Masse → *Kontinuumsmechanik*).

2. Gesetz: $m\, a_{\Delta t}(t) = F(t) \in \mathbb{R}^3$.
Die Kraft ist daher ein Vektor mit *Größe* ($|F(t)|$) und *Richtung* (angegeben z.B. durch $F(t)/|F(t)|$ oder durch Längen- und Breitengrad).

Wählt man einen festen Anfangszeitpunkt t_0, $t > t_0$, $N \in \mathbb{N}$ und $\Delta t = \frac{t-t_0}{N}$, dann ergibt sich

$$\begin{aligned} v_{\Delta t}(t_0 + \Delta t) &= v_{\Delta t}(t_0) + \frac{\Delta t}{m} F(t_0)\,, \qquad v_{\Delta t}(t_0 + 2\Delta t) = v_{\Delta t}(t_0) + \frac{\Delta t}{m}(F(t_0) + F(t_0 + \Delta t))\,, \quad \text{usw.} \\ v_{\Delta t}(t) &= v_{\Delta t}(t_0 + N\Delta t) = v_{\Delta t}(t_0) + \frac{1}{m}\sum_{k=0}^{N-1} F(t_0 + k\Delta t)\Delta t\,. \end{aligned}$$

Das ist unbefriedigend und offensichtlich nicht konsistent, weil alles von der Wahl von Δt abhängt, das offensichtlich klein gewählt werden muss. Allerdings wie klein ist klein genug? Ein Fehler bleibt immer, wenn sich die Kraft kontinuierlich mit der Zeit verändert. Daher musste Newton die *Differential- und Integralrechnung* erfinden. Also $\Delta t \to 0$ und

$$v(t) = \dot{x}(t) = \frac{dx}{dt}(t)\,, \qquad a(t) = \dot{v}(t) = \ddot{x}(t) = \frac{d^2x}{dt^2}(t)\,,$$

sowie

$$v(t) = v(t_0) + \frac{1}{m}\int_{t_0}^{t} F(s)ds\,.$$

Die heute übliche Notation stammt übrigens von *Gottfried Wilhelm Leibniz* (1646–1716), mit dem Newton einen erbitterten Prioritätsstreit um die Erfindung der Infinitesimalrechnung austrug.

Ist die Kraft als Funktion der Zeit gegeben, lässt sich also die Geschwindigkeit durch Integration ermitteln. Analog ergibt eine weitere Integration die Bahn $x(t)$ des Objekts. Man beachte, dass man $x(t_0)$ und $v(t_0)$ vorgeben muss, um $x(t)$ für $t > t_0$ berechnen zu können. In der Teilchenmechanik wird daher das Paar (x, v) als *Zustand* des Objekts bezeichnet.

Unser erstes Modell mit einem festen $\Delta t > 0$ ist ein *diskretes Modell*, weil es den Zustand (x, v) zu den diskreten Zeitpunkten $t_k = t_0 + k\Delta t$, $k \geq 0$, liefert (ausgehend vom Anfangszeitpunkt t_0). Im Unterschied dazu nennt man das Modell mit Ableitungen und Integralen, das man mit $\Delta t \to 0$ erhält, ein *kontinuierliches Modell*. Kann man die zur Lösung notwendigen Integrale nicht explizit berechnen, dann kann man als Approximation zum diskreten Modell zurückkehren. In diesem Zusammenhang wird das diskrete Modell das *explizite Eulerverfahren* genannt (Leonhard Euler, 1707–1783).

Die wohl populärste Leistung Newtons war die Beschreibung der *Gravitation* (Apfelgeschichte). In diesem Fall ergibt sich eine Kraft, die von der Position des Objekts abhängt, also $F = F(x(t))$. Damit ergibt sich aus dem 2. Gesetz

$$m\,\ddot{x}(t) = F(x(t))\,, \tag{1}$$

ein erstes Beispiel für eine interessante *gewöhnliche Differentialgleichung*, d.h. eine Gleichung, in der Werte und Ableitungen einer unbekannten Funktion von einer Veränderlichen vorkommen. Hier ist es nicht mehr offensichtlich, wie $x(t)$ durch Integrationen berechnet werden kann.

Allerdings können wir wieder einen Schritt zurück zum diskreten Modell machen, das man in der Form

$$v(t + \Delta t) = v(t) + \Delta t \frac{F(x(t))}{m}\,, \qquad x(t + \Delta t) = x(t) + \Delta t v(t)\,,$$

schreiben kann. Der Zustand zum Zeitpunkt $t + \Delta t$ kann also aus dem Zustand zum Zeitpunkt t berechnet werden. Das zeigt, dass das diskrete Modell bei Angabe des Zustandes $(x(t_0), v(t_0))$ zum Anfangszeitpunkt gelöst werden kann in dem Sinne, das die Lösung zu jedem diskreten Zeitpunkt

$t_k = t + k\Delta t$, $k \geq 0$, berechnet werden kann. Das legt die Vermutung nahe, dass auch das *Anfangswertproblem,* das aus der Gleichung (1) und der *Anfangsbedingung*

$$x(t_0) = x_0\,, \qquad \dot{x}(t_0) = v_0\,,$$

besteht, eindeutig gelöst werden kann. Ein derartiges Resultat wird eines der Hauptresultate dieser Vorlesung sein.

$\longrightarrow$ [T, 1.1]

1.2 Klassifikation von gewöhnlichen Differentialgleichungen

$\longrightarrow$ [T, 1.2]

Nicht all Arten von Gleichungen für Funktionen von einer Veränderlichen sind gewöhnliche Differentialgleichungen. Es gibt auch *Funktionalgleichungen,* z.B.

$$x(t)x(s) = x(t+s)\,,$$

retardierte Differentialgleichungen, z.B.

$$\dot{x}(t) = f(t, x(t), x(t-\tau))\,,$$

mit festem $\tau > 0$; *Integralgleichungen,* z.B.

$$\int_a^b f(t, s, x(s))ds = 0\,;$$

Integro-Differentialgleichungen, z.B.

$$\dot{x}(t) = \int_a^t f(t, s, x(s))ds\,.$$

Auch diese treten in Anwendungen auf.

1.4 Explizit lösbare Gleichungstypen

Separable Gleichungen:

$$\dot{x} = f(x)g(t) \quad \Rightarrow \quad \frac{\dot{x}}{f(x)} = g(t)$$

Jetzt Integration bezüglich t. Linke Seite:

$$\int \frac{\dot{x}(t)}{f(x(t))}dt = \left| \begin{array}{c} y = x(t) \\ dy = \dot{x}(t)dt \end{array} \right| = \int \frac{dy}{f(y)} = F(y) + c = F(x(t)) + c$$

mit einer Stammfunktion F von $1/f$, d.h. $F'(x) = 1/f(x)$. Sei G eine Stammfunktion von g. Dann gilt

$$F(x(t)) = G(t) + c$$

mit einer Integrationskonstanten $c \in \mathbb{R}$. Die Menge aller Lösungen (man spricht auch von der *allgemeinen Lösung*) ist also eindimensional und wird parametrisiert durch die Integrationskonstante. Durch eine Anfangsbedingung $x(t_0) = x_0$ wird die Integrationskonstante eindeutig bestimmt:

$$F(x(t)) = G(t) - G(t_0) + F(x_0)\,.$$

Das ist eine implizite Darstellung der Lösung $x(t)$. Zur Bestimmung von $x(t)$ brauchen wir F^{-1}. Mit *explizit lösbar* meinen wir also lösbar bis auf Grundaufgaben wie die Bestimmung von Stammfunktionen und die Lösung algebraischer Gleichungen.
Beispiele: *Autonome Gleichung* ($g(t) = 1$):

$$\dot{x} = f(x)\,, \qquad x(0) = x_0 \tag{2}$$

Implizite Lösung: $F(x(t)) = t + F(x_0)$
a) *Linear:*

$$f(x) = \lambda x \quad \Rightarrow \quad \log(x/x_0) = \lambda t \quad \Rightarrow \quad x(t) = e^{\lambda t} x_0$$

Man beachte, dass zwar die Rechnung nur für $x_0 \neq 0$ funktioniert, das Endresultat aber auch für $x_0 = 0$.
b) *Quadratisch:*

$$f(x) = x^2 \quad \Rightarrow \quad -\frac{1}{x(t)} = t - \frac{1}{x_0} \quad \Rightarrow \quad x(t) = \frac{x_0}{1 - x_0 t}$$

Die Lösung existiert nicht für alle Zeiten. Das *maximale Existenzintervall* für $x_0 > 0$ ist $(-\infty, 1/x_0)$.
c) *Wurzel:* $x_0 > 0$

$$f(x) = \sqrt{x} \quad \Rightarrow \quad 2\sqrt{x(t)} = t + 2\sqrt{x_0} \quad \Rightarrow \quad x(t) = (t/2 + \sqrt{x_0})^2$$

Maximales Existenzintervall: $[-2\sqrt{x_0}, \infty)$.
Auch hier erhält man im Endresultat mit $x_0 = 0$ eine Lösung (für $t \geq 0$). Allerdings besitzt dieses Anfangswertproblem auch die zweite Lösung $x(t) = 0$. Es wird sich später herausstellen, dass dieser Mangel an Eindeutigkeit eine Konsequenz der mangelnden *Lipschitzstetigkeit* von f in der Nähe von x_0 ist. $\longrightarrow$ [T, S. 11]

Challenge 1 *Man löse das Anfangswertproblem*

$$\dot{x} = \frac{\sin t}{x}\,, \qquad x(0) = 1\,.$$

Konservative Kraftfelder: Wir betrachten die Newtonschen Gleichungen (1), wobei wir annehmen, dass das Kraftfeld ein *Gradientenfeld* ist, d.h. $F(x) = -\nabla\Phi(x)$ (wobei die Wahl des Vorzeichens willkürlich ist und unten sinnvoll erscheinen wird):

$$m\ddot{x} = -\nabla\Phi(x)\,.$$

Diese Form des Kraftfelds erlaubt die Anwendung eines Tricks, und zwar bilden wir das Skalarprodukt der Gleichung mit $\dot{x}$:

$$m\dot{x} \cdot \ddot{x} + \nabla\Phi(x) \cdot \dot{x} = 0\,.$$

Nun stellt man fest, dass die linke Seite nach der Zeit integriert werden kann:

$$m\frac{|v(t)|^2}{2} + \Phi(x(t)) = E\,.$$

Der Name E für die Integrationskonstante deutet auf ihre physikalische Bedeutung hin. Wir haben mit der obigen Gleichung die Eigenschaft der *Energieerhaltung* hergeleitet. Die linke Seite der Gleichung ist die Summe aus der *kinetischen Energie* $m|v|^2/2$ und der *potentiellen Energie* $\Phi(x)$. Beide können sich mit der Zeit verändern, aber ihre Summe, die *Gesamtenergie* bleibt konstant. Man kann sie aus dem Anfangszustand berechnen.

Eine vollständige Lösung des Problems ist im eindimensionalen Fall $x, v \in \mathbb{R}$ möglich. Zunächst ist in diesem Fall jedes Kraftfeld ein Gradientenfeld. Weiters kann aus der Energieerhaltung $v = \dot{x}$ als Funktion von x berechnet werden. Das ergibt eine autonome Gleichung der Form (2).

Als eindimensionales Beispiel betrachten wir $F(x) = -kx \in \mathbb{R}$ mit $k > 0$. Die Interpretation ist ein Objekt, das sich auf einer Geraden bewegt und das durch eine Feder mit dem Ursprung verbunden ist. Die Federkraft ist proportional zum Abstand des Objektes zum Ursprung. Die Energieerhaltungsgleichung ist

$$mv(t)^2 + kx(t)^2 = 2E\,,$$

d.h. das Objekt bewegt sich in der Zustandsebene (der (x, v)-Ebene) entlang einer Ellipse, also einer geschlossenen Kurve. Daraus folgt, dass alle Lösungen periodisch sind. Dieses Problem wird *harmonischer Oszillator* genannt.

Lineare Gleichungen:

$$\dot{x} = a(t)x + g(t) \tag{3}$$

Wir erinnern daran, dass die Gleichung für $g \equiv 0$ *homogen* genannt wird, sonst *inhomogen*. Seien $x_1(t), x_2(t)$ zwei Lösungen der inhomogenen Gleichung. Dann gilt

$$\frac{d}{dt}(x_1 - x_2) = \dot{x}_1 - \dot{x}_2 = a(x_1 - x_2)\,,$$

d.h. die Differenz $x_1 - x_2$ ist eine Lösung der homogenen Gleichung. Daraus folgt die folgende Vorgangsweise:
Die allgemeine Lösung der inhomogenen Gleichung erhält man als Summe der allgemeinen Lösung der homogenen Gleichung und einer speziellen Lösung (genannt Partikulärlösung*) der inhomogenen Gleichung.*

Es sind also zwei Teilaufgaben zu lösen. Wir beginnen mit der Bestimmung der allgemeinen Lösung der homogenen Gleichung

$$\dot{x} = ax\,. \tag{4}$$

Diese ist separabel. Daher

$$\frac{\dot{x}}{x} = a \quad \Rightarrow \quad \log|x(t)| = \int_{t_0}^{t} a(s)ds + c \quad \Rightarrow \quad x_h(t) = \exp\left(\int_{t_0}^{t} a(s)ds\right) x_0$$

mit $c = \log|x_0|$. Hier haben wir die allgemeine Lösung x_h der homogenen Gleichung durch ihren Wert x_0 an $t = t_0$ parametrisiert.

Die Bestimmung einer Partikulärlösung der inhomogenen Gleichung basiert auf einem Trick, der *Variation der Konstanten* heißt. Er besteht darin, einen Ansatz zu machen, der so aussieht wie die allgemeine Lösung der homogenen Gleichung, allerdings mit einer zeitabhängigen 'Konstanten':

$$x_p(t) = \exp\left(\int_{t_0}^{t} a(s)ds\right) c(t)\,.$$

Differenzieren ergibt

$$\dot{x}_p = a x_p + \exp\left(\int_{t_0}^{t} a(s)ds\right) \dot{c}\,.$$

Daher ist x_p eine Lösung der inhomogenen Gleichung, wenn

$$\exp\left(\int_{t_0}^{t} a(s)ds\right) \dot{c} = g$$

gilt. Das ist eine Differentialgleichung für c, die durch eine Integration gelöst werden kann:

$$c(t) = \int_0^t \exp\left(-\int_{t_0}^{s} a(\tau)d\tau\right) g(s)ds\,.$$

Auf eine Integrationskonstante können wir verzichten, weil wir nur eine Lösung suchen. Für die Partikulärlösung ergibt sich

$$x_p(t) = \int_{t_0}^{t} \exp\left(\int_{s}^{t} a(\tau)d\tau\right) g(s)ds$$

Die allgemeine Lösung der inhomogenen Gleichung (3) ist damit gegeben durch

$$x(t) = x_h(t) + x_p(t) = \exp\left(\int_{t_0}^{t} a(s)ds\right) x_0 + \int_{t_0}^{t} \exp\left(\int_{s}^{t} a(\tau)d\tau\right) g(s)ds\,.$$

Auch hier ist der Parameter x_0 gleichzeitig der Wert der Lösung an der Stelle $t = t_0$.

Man beachte, dass die allgemeine Lösung der homogenen Gleichung einen eindimensionalen *Vektorraum* bildet. Diesen kann man als *Kern* der linearen Abbildung $x \mapsto Lx := \dot{x} - ax$ sehen. Eine Abbildung wie L, die Funktionen auf Funktionen abbildet, wobei $(Lx)(t)$ von Ableitungen von x an der Stelle t abhängt (hier $(Lx)(t) = \dot{x}(t) - a(t)x(t)$) nennt man einen *Differentialoperator.*

Vereinfachung durch Transformation:

Challenge 2 *Mit Hilfe der Transformation $x(t) = ty(t)$ finde man eine implizite Darstellung der Lösung des Anfangswertproblems*

$$\dot{x} = \frac{x - t}{x + t}\,, \qquad x(0) = 1\,.$$

$\longrightarrow$ [T, 1.4]

Bernoulli-Gleichung:

$$\dot{x} = f(t)x + g(t)x^n\,, \qquad n \in \mathbb{R}\,, \quad n \neq 0, 1\,.$$

Transformation: $y = x^{1-n}$

$$\dot{y} = (1-n)x^{-n}\left(fx + gx^n\right) = (1-n)fy + (1-n)g\,, \qquad \text{(linear)}$$

Euler-Gleichung: Linear, homogen, Ordnung k, singulär an $t = 0$.

$$a_k t^k x^{(k)} + a_{k-1} t^{k-1} x^{(k-1)} + \ldots + a_1 t\dot{x} + a_0 x = 0\,, \qquad t > 0\,.$$

Ansatz: $x(t) = t^\alpha$

$$0 = \sum_{l=0}^{k} a_l t^l x^{(l)} = \sum_{l=0}^{k} a_l t^l (\alpha(\alpha-1)\cdots(\alpha-l+1)) t^{\alpha-l} = t^\alpha \sum_{l=0}^{k} a_l (\alpha(\alpha-1)\cdots(\alpha-l+1))$$

Die Summe auf der rechten Seite ist ein Polynom der Ordnung k in α. Hat es k verschiedene reelle Nullstellen $\alpha_1, \ldots, \alpha_k$, dann hat man eine Basis $\{t^{\alpha_1}, \ldots, t^{\alpha_k}\}$ des Lösungsraumes gefunden.

Das kann auch mit komplexen Nullstellen funktionieren: Beispiel:

$$t^2\ddot{x} + t\dot{x} + x = 0$$

Der Ansatz $x = t^\alpha$ gibt

$$0 = \alpha(\alpha-1) + \alpha + 1 = \alpha^2 + 1\,,$$

also $\alpha = \pm i$, und daher

$$x(t) = t^{\pm i} = e^{\pm i \log t} = \cos(\log t) \pm i \sin(\log t)\,.$$

Man rechnet leicht nach, dass $\{\cos(\log t), \sin(\log t)\}$ eine reelle Basis des Lösungsraumes ist.

Reduktion der Ordnung: Lineare homogene Gleichung der Ordnung k:

$$\sum_{l=0}^{k} a_l(t) x^{(l)} = 0\,.$$

Angenommen, man kennt eine Lösung $x_1(t)$. Ansatz (Variation der Konstanten): $x(t) = c(t)x_1(t)$. Differenzieren mit der Leibniz-Formel:

$$x^{(l)} = \sum_{m=0}^{l} \binom{l}{m} c^{(m)} x_1^{(l-m)}$$

Einsetzen und vertauschen der Summationsreihenfolge:

$$0 = \sum_{l=0}^{k} \sum_{m=0}^{l} a_l \binom{l}{m} c^{(m)} x_1^{(l-m)} = \sum_{m=0}^{k} c^{(m)} \sum_{l=m}^{k} a_l \binom{l}{m} x_1^{(l-m)}$$

Auf der rechten Seite verschwindet der Summand mit $m = 0$, weil x_1 eine Lösung der Differentialgleichung ist. Daher, mit $y := \dot{c}$,

$$0 = \sum_{m=1}^{k} b_m(t) y^{(m-1)} = \sum_{m=0}^{k-1} b_{m+1}(t) y^{(m)}\,, \qquad \text{mit } b_m(t) = \sum_{l=m}^{k} a_l(t) \binom{l}{m} x_1^{(l-m)}(t)\,,$$

d.h. y löst eine Gleichung der Ordnung $k-1$.
Beispiel: Euler-Gleichung zweiter Ordnung:

$$t^2\ddot{x} - t\dot{x} + x = 0\,.$$

Ansatz: $x = t^\alpha$. Das ergibt $0 = \alpha(\alpha-1) - \alpha + 1 = \alpha^2 - 2\alpha + 1$,
also die einzige Lösung mit $\alpha = 1$: $x_1(t) = t$.
Daher Reduktion der Ordnung: $x(t) = c(t)x_1(t)$, dann $y = \dot{c}$:

$$t^2(c\ddot{x}_1 + 2\dot{c}\dot{x}_1 + \ddot{c}x_1) - t(c\dot{x}_1 + \dot{c}x_1) + cx_1 = t^2x_1\ddot{c} + (2t^2\dot{x}_1 - tx_1)\dot{c} = t^3\dot{y} + t^2y = 0\,,$$

also eine Gleichung erster Ordnung für y:

$$\dot{y} = -\frac{1}{t}y \quad \Rightarrow \quad y(t) = \exp(-\log t)y_0 = \frac{y_0}{t}\,.$$

Integration der Lösung $y(t) = 1/t$ ergibt $c(t) = \log t$ und damit eine zweite, von x_1 linear unabhängige Lösung der ursprünglichen Gleichung: $x_2(t) = t\log t$.
$\longrightarrow$ [T, 1.4]: MATHEMATICA

1.5 Qualitative Analyse von Gleichungen erster Ordnung

Autonome Gleichungen erster Ordnung: Qualitative Aussagen sind auch ohne explizites Lösen möglich. Anfangswertproblem:

$$\dot{x} = f(x)\,, \qquad x(0) = x_0 \in \mathbb{R}\,.$$

Sei $f : \mathbb{R} \to \mathbb{R}$ glatt. Dann kann man die x-Achse unterteilen in offene Intervalle, in denen f strikt positiv oder strikt negativ ist. Dazwischen liegen Punkte oder abgeschlossene Intervalle in denen $f = 0$ gilt. Wir werden im folgenden Kapitel beweisen, dass das Anfangswertproblem für jedes $x_0 \in \mathbb{R}$ in einem *maximalen Existenzintervall* eine eindeutige Lösung hat.

1. **Fall:** $f(x_0) = 0 \quad \Rightarrow \quad x(t) = x_0$, $t \geq 0$. Man sagt x_0 ist ein *stationärer Punkt.*

2. **Fall:** $f(x_0) > 0$ und es existiert rechts von x_0 eine Nullstelle, d.h.
   $\exists\, x^* := \min\{x > x_0 :\, f(x) = 0\}$.
   Dann existiert die Lösung $x(t)$ für alle $t \geq 0$, sie ist streng monoton wachsend und
   $\lim_{t\to\infty} x(t) = x^*$.

3. **Fall:** $f(x) > 0$ für alle $x \geq x_0$. Dann gibt es ein $T \in (0, \infty]$, sodass die Lösung $x(t)$ existiert für alle $t \in [0, T)$ und $\lim_{t\to T} x(t) = \infty$.

4. **und 5. Fall:** $f(x_0) < 0$. Analog zum 2. und 3. Fall mit den offensichtlichen Änderungen.

Auch diese Aussagen werden im folgenden Kapitel bewiesen.

Beispiel: Fischteich: *Logistisches Modell* mit konstanter Fangrate:

$$\dot{x} = f(x) := rx\left(1 - \frac{x}{x_0}\right) - h\,,$$

mit der Populationsgröße $x(t)$, dem Geburtenüberschuss bei kleinen Populationen $r > 0$, der Sättigungsgrenze x_0 und der Fangrate h. Ist die Fangrate klein genug, und zwar $h < rx_0^2/4$, dann gibt es zwei stationäre Punkte

$$x_{1,2} = \frac{x_0}{2} \mp \sqrt{\frac{x_0^2}{4} - \frac{h}{r}}\,.$$

Es gilt $f(x) > 0$ für $x_1 < x < x_2$. Daher sind alle Lösungen mit einem Anfangswert zwischen x_1 und x_2 streng monoton wachsend und konvergent gegen x_2. Ist der Anfangswert grösss er als x_2, dann ist die Lösung streng monoton fallend und ebenfalls konvergent gegen x_2. Für Anfangswerte kleiner als x_1 ist die Lösung auch streng monoton fallend und erreicht in endlicher Zeit den Wert Null. Wenn man also dafür sorgt, dass die Anfangspopulation groß genug ist (größer als x_1), dann stellt sich ein nachhaltiges Gleichgewicht ($x(t) \to x_2$) ein. Ist die Anfangspopulation zu klein, dann stirbt sie aus. Diese Resultate kann man leicht aus einem *Phasenporträt* (siehe Fig. 1) ablesen. Ist die Fangrate zu groß, d.h. $h > rx_0^2/4$, dann gilt immer $f(x) < 0$, d.h. es gibt keine stationären Punkte und die Population stirbt auf jeden Fall aus.

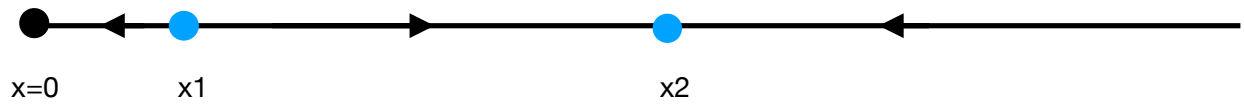

Figure 1: Phasenporträt des logistischen Modells mit nachhaltiger konstanter Fangrate

2 Anfangswertprobleme

2.2 Der Existenz- und Eindeutigkeitssatz

Wir haben gesehen, dass Systeme von gewöhnlichen Differentialgleichungen in expliziter Form immer als explizite Systeme von Gleichungen erster Ordnung geschrieben werden können. Daher beschäftigen wir uns im Folgenden nur mit solchen:

$$\dot{x} = f(t, x)\,, \tag{5}$$

wobei $f : U \to \mathbb{R}^n$ mit einer offenen Menge $U \subset \mathbb{R}^{n+1}$. Wir erwarten, dass durch eine Anfangsbedingung der Form

$$x(t_0) = x_0 \tag{6}$$

eine eindeutige Lösung festgelegt wird. Es ist das Ziel dieses Abschnittes, Existenz und Eindeutigkeit einer Lösung des Problems (5), (6) zu beweisen, wobei auch ein geeignetes Zeitintervall gefunden werden muss (das natürlich $t = t_0$ enthält). Als erste Motivation für die Vorgangsweise betrachten wir einen Schritt des *expliziten Eulerverfahrens* (siehe Abschnitt 1.1), um den Zeitpunkt t zu erreichen:

$$x(t) \approx x_0 + (t - t_0) f(t_0, x_0)\,. \tag{7}$$

Die rechte Seite ist das *Taylorpolynom erster Ordnung,* wenn man die Lösung um $t = t_0$ entwickelt. Das Approximationssymbol $\approx$ ist wahrscheinlich gerechtfertigt, wenn der Zeitschritt $t - t_0$ klein ist.

Die Beobachtung mit dem Taylorpolynom motiviert einen möglichen Zugang zur Lösung des Problems, bei dem man versucht, die *Taylorreihe* der Lösung zu konstruieren. Differenzieren von (5) nach t ergibt

$$\ddot{x} = \partial_t f(t, x) + D_x f(t, x)\dot{x}\,,$$

mit der Jacobimatrix $D_x f$ der Ableitungen der Komponenten von f nach den Komponenten von x. Daher folgt

$$\ddot{x}(t_0) = \partial_t f(t_0, x_0) + D_x f(t_0, x_0) f(t_0, x_0)\,,$$

womit auch das quadratische Taylorpolynom vollständig bestimmt ist. Um diese Idee fortzusetzen, benötigt man alle partiellen Ableitungen von f beliebiger Ordnung. Die Idee stammt von *Augustin-Louis Cauchy* (1798–1857), der folgendes Resultat bewies (Sein Beweis wurde etwas später von *Sofya Kovalevskaya* (1850–1891) verbessert):

Satz 1 *Sei f an der Stelle (t_0, x_0) analytisch. Dann gibt es eine eindeutige Lösung $x(t)$ des Problems (5), (6), die an der Stelle t_0 analytisch ist.*

Challenge 3 *Man verwende die Methode von Cauchy, um die Lösung des Problems $\dot{x} = \lambda x$ ($\lambda \in \mathbb{R}$), $x(0) = 1$, zu berechnen.*

Statt diesen Satz zu beweisen, werden wir uns um einen Zugang bemühen, der mit weniger strengen Glattheitsannahmen auskommt. Im Folgenden werden wir zunächst die Bestandteile eines Beweises motivieren.

Die Approximation (14) kann man so verstehen, dass man die rechte Seite $f(t, x)$ der Differentialgleichung durch $f(t_0, x_0)$ approximiert und dann integriert, um eine verbesserte Approximation zu erhalten. Auch dieses Verfahren kann iteriert werden. Dazu integrieren wir (5) zunächst:

$$x(t) = (Fx)(t) := x_0 + \int_{t_0}^{t} f(s, x(s)) ds\,. \tag{8}$$

Offensichtlich ist diese *Integralgleichung* äquivalent zum Anfangswertproblem (5), (6): Einerseits haben wir sie durch Integration aus dem Anfangswertproblem hergeleitet. Gilt andererseits die Integralgleichung, dann erhält man durch Auswerten an $t = t_0$ die Anfangsbedingung und durch Differenzieren die Differentialgleichung. Es genügt also, wenn wir uns von jetzt an nur mehr mit (8) (dem *Fixpunktproblem* für die Abbildung F) beschäftigen. Das explizite Eulerverfahren erhält man durch Approximation des Integranden durch seinen Wert an der Stelle $t = t_0$. Wir modifizieren das Verfahren leicht, indem wir nur die unbekannte Funktion $x(s)$ durch ihren Anfangswert approximieren. Das ergibt die verbesserte Approximation

$$x_1(t) = x_0 + \int_{t_0}^{t} f(s, x_0) ds\,.$$

Durch Iterieren dieser Vorgangsweise erzeugen wir eine rekursiv definierte Funktionenfolge:

$$x_{k+1} = F(x_k)\,, \qquad k \geq 1\,. \tag{9}$$

Diese *Fixpunktiteration* wird *Picard-Iteration* (*Charles Émile Picard* (1856–1941)) genannt.

Die Idee ist nun eine Vorgangsweise wie bei rekursiv definierten Zahlenfolgen: Zunächst zeigen wir, dass die Folge konvergiert, dann gehen wir in (9) zur Grenze $k \to \infty$ über, um zu zeigen dass der Grenzwert eine Lösung von (8) ist. Was wir zunächst brauchen, ist die Wahl eines *funktionalanalytischen Rahmens,* d.h. die Wahl eines Funktionenraumes samt Konvergenzbegriff. Unter der

Annahme, dass f stetig ist, produziert unsere Rekursion eine Folge $\{x_k\}_{k\in\mathbb{N}}$ von stetigen Funktionen. Als Definitionsbereich wählen wir ein Intervall der Form $[t_0 - T, t_0 + T]$, wobei wir uns die Wahl von T noch offen halten. Der Funktionenraum

$$\mathcal{B}_T := C([t_0 - T, t_0 + T], \mathbb{R}^n)$$

ist ein Vektorraum über $\mathbb{R}$ (bezüglich der punktweisen Addition von Funktionen und Multiplikation mit Skalaren). Als Konvergenzbegriff wählen wir die *gleichmäßige Konvergenz*, die von der *Supremumnorm*

$$\|x\|_\infty := \sup_{|t-t_0|<T} |x(t)|$$

induziert wird. Wir erinnern an einige in der Analysis-Grundvorlesung gezeigte Eigenschaften: Die Supremumnorm erfüllt die Normeigenschaften

1. **Definitheit:** $\|x\|_\infty = 0 \quad \Leftrightarrow \quad x = 0, \qquad \forall x \in B.$
2. **Dreiecksungleichung:** $\|x + \hat{x}\|_\infty \leq \|x\|_\infty + \|\hat{x}\|_\infty, \qquad \forall x, \hat{x} \in B.$
3. **Homogenität:** $\|\lambda x\|_\infty = |\lambda| \, \|x\|_\infty, \qquad \forall x \in B, \, \lambda \in \mathbb{R}.$

Konvergenz bezüglich $\|\cdot\|_\infty$ ist *gleichmäßige Konvergenz*. Sie impliziert *punktweise Konvergenz*. Daraus und aus der Vollständigkeit von $\mathbb{R}$ folgt, dass jede Cauchyfolge in $(\mathcal{B}_T, \|\cdot\|_\infty)$ einen punktweisen Limes besitzt. Aus der Cauchyfolgeneigenschaft folgt, dass die Konvergenz gleichmäßig ist, und daraus, dass der Limes stetig ist. Daher ist $(\mathcal{B}_T, \|\cdot\|_\infty)$ *vollständig* und damit ein *Banachraum*.

Wir werden daher versuchen zu zeigen, dass die durch die Rekursion (9) definierte Folge eine Cauchyfolge ist. Als ersten Schritt betrachten wir den Abschnitt zwischen zwei aufeinanderfolgenden Folgengliedern:

$$\begin{aligned} \|x_{k+1} - x_k\|_\infty &= \sup_{|t-t_0|<T} \left| \int_{t_0}^{t} (f(s, x_k(s)) - f(s, x_{k-1}(s)) ds \right| \\ &\leq \sup_{|t-t_0|<T} \int_{t_0}^{t} |f(s, x_k(s)) - f(s, x_{k-1}(s)| \, ds \, . \end{aligned}$$

An dieser Stelle ist es Zeit, sich Gedanken über Annahmen an f zu machen. Die im Abschnitt 1.4 betrachteten Beispiele autonomer Gleichungen legen nahe, dass für die Eindeutigkeit der Lösung Lipschitzstetigkeit von f als Funktion von x notwendig ist. Das nehmen wir also zusätzlich zur Stetigkeit an:

$$|f(t, x) - f(t, \hat{x})| \leq L|x - \hat{x}| \, ,$$

mit der Lipschitzkonstanten $L > 0$. Damit können wir obige Abschätzung weiterführen:

$$\|x_{k+1} - x_k\|_\infty \leq L \sup_{|t-t_0|<T} \int_{t_0}^{t} |x_k(s) - x_{k-1}(s)| \, ds \, .$$

Jetzt schätzen wir den Integranden mit seinem Supremum ab:

$$\|x_{k+1} - x_k\|_\infty \leq LT \|x_k - x_{k-1}\|_\infty$$

Das Ziel ist natürlich, den Abstand zwischen aufeinanderfolgenden Folgengliedern in jedem Schritt zu verkleinern. Das können wir erreichen, indem wir T klein genug wählen, sodass $LT < 1$:

$$\|x_{k+1} - x_k\|_\infty \leq \cdots \leq (LT)^k \|x_1 - x_0\|_\infty \,.$$

Das ist ein gutes Zeichen, aber noch kein Beweis dafür, dass die Folge eine Cauchyfolge ist. Wählen wir nun Indices $l > k$:

$$\begin{aligned} \|x_l - x_k\|_\infty &= \left\| \sum_{m=k}^{l-1} (x_{m+1} - x_m) \right\|_\infty \leq \sum_{m=k}^{l-1} \|x_{m+1} - x_m\|_\infty \leq \sum_{m=k}^{l-1} (LT)^m \|x_1 - x_0\|_\infty \\ &\leq \sum_{m=k}^{\infty} (LT)^m \|x_1 - x_0\|_\infty = \frac{(LT)^k}{1-LT} \|x_1 - x_0\|_\infty \,, \end{aligned} \tag{10}$$

woraus folgt, das $\{x_l\}_{k\in\mathbb{N}}$ eine Cauchyfolge ist. Das waren die wesentlichen Ideen. Es bleiben noch einige Details zu klären.

Satz 2 *(Picard-Lindelöf) Sei $f \in C(V, \mathbb{R}^n)$ mit $V = [t_0 - \tau, t_0 + \tau] \times \overline{K_\delta(x_0)}$ und $x_0 \in \mathbb{R}^n$, wobei $\tau > 0$, $K_\delta(x_0)$ die offene Kugel im $\mathbb{R}^n$ mit Mittelpunkt x_0 und Radius $\delta > 0$. Sei $f(t,x)$ als Funktion von x lipschitzstetig gleichmäßig in $t \in [t_0 - \tau, t_0 + \tau]$ mit Lipschitzkonstante L. Dann existiert ein $T > 0$, sodass die Gleichung (8), und damit das Anfangswertproblem (5), (6), eine Lösung $x \in \mathcal{B}_T$ besitzt, die eindeutig ist und in $\overline{\mathcal{K}_\delta(x_0)} := \{x \in \mathcal{B}_T : \|x - x_0\|_\infty \leq \delta\}$ liegt.*

Beweis: Da f stetig ist auf der abgeschlossenen Menge V, existiert $M > 0$, sodass

$$|f| \leq M \qquad \text{in } V \,.$$

Daher gilt offensichtlich

$$\|F(x) - x_0\|_\infty \leq MT \,.$$

Um zu garantieren, dass die Iteration im Definitionsbereich von f bleibt, wählen wir T so, dass $MT \leq \delta$, und natürlich $T \leq \tau$. Damit ist die Folge $\{x_k\}_{k\in\mathbb{N}} \subset \overline{\mathcal{K}_\delta(x_0)}$ durch (9) wohldefiniert. Die obigen Abschätzungen sind daher gerechtfertigt, und mit der Wahl

$$T := \min\left\{\tau, \frac{\delta}{M}, \frac{1}{2L}\right\}$$

folgt aus (10), dass $\{x_k\}_{k\in\mathbb{N}}$ eine Cauchyfolge ist und daher konvergiert. Der Limes $x \in \overline{\mathcal{K}_\delta(x_0)}$ löst (8), weil man in (9) den Grenzübergang $k \to \infty$ durchführen kann: Offensichtlich konvergiert die linke Seite gegen x, und die rechte Seite konvergiert gegen $F(x)$ wegen

$$\|F(x_k) - F(x)\|_\infty \leq \frac{1}{2}\|x_k - x\|_\infty \to 0 \quad \text{für } k \to \infty \,.$$

Angenommen, es gäbe zwei Lösungen x und $\hat{x}$. Dann hätten wir

$$\|x - \hat{x}\|_\infty = \|F(x) - F(\hat{x})\|_\infty \leq \frac{1}{2}\|x - \hat{x}\|_\infty \,,$$

woraus $x = \hat{x}$ folgt. ∎

Bemerkung 1 *Eine Abbildung eines Banachraumes in sich selbst, die wie F lipschitzstetig ist mit einer Lipschitzkonstanten kleiner als Eins, nennt man eine* Kontraktion. *Die Tatsache, dass Kontraktionen eindeutige Fixpunkte besitzen, ist Inhalt des* Banachschen Fixpunktsatzes*, der allerdings erst nach dem Satz von Picard-Lindelöf bewiesen wurde.*

Bemerkung 2 *Obwohl der Satz nur eine stetige Lösung garantiert, folgt aus (8) sofort, dass diese auch stetig differenzierbar ist.*

Bemerkung 3 *Man beachte, dass wir im Beweis die Stetigkeit von $f(t,x)$ als Funktion von t nicht wirklich brauchen. Beschränktheit von f und Integrierbarkeit von $f(t,x(t))$ für stetiges $x(t)$ genügen. Allerdings ist x dann nicht mehr unbedingt stetig differenzierbar und nur im Sinne der Integralgleichung eine Lösung (eine sogenannte* Carathéodory-Lösung *der Differentialgleichung).*

Challenge 4 *Man verwende Picard-Iteration, um die Lösung des Problems $\dot{x} = \lambda x$ ($\lambda \in \mathbb{R}$), $x(0) = 1$, zu berechnen.*

Die Lösung ist so glatt, wie es die rechte Seite $f(t,x)$ der Differentialgleichung zulässt:

Lemma 1 *Es gelten die Annahmen von Satz 2 und außerdem $f \in C^k(V, \mathbb{R}^n)$, $k \geq 1$. Dann gilt $x \in C^{k+1}([t_0 - T, t_0 + T], \mathbb{R}^n)$ für die Lösung von (5), (6).*

Beweis: Wie schon in Bemerkung 2 erwähnt, ist x stetig differenzierbar. Die rechte Seite der Differentialgleichung ist daher als Zusammensetzung stetig differenzierbarer Funktionen auch stetig differenzierbar, woraus folgt, dass x zweimal stetig differenzierbar ist. So arbeitet man sich Schritt für Schritt hinauf *(bootstrapping)*, solange die Glattheit von f reicht. ∎

2.3 Erweiterungen: Ungleichungen – Das maximale Existenzintervall

Differentialungleichungen: Wir beginnen mit einem einfachen aber wichtigen Hilfsresultat, dem *Gronwall-Lemma in differentieller Form*:

Lemma 2 *Sei $L \in \mathbb{R}$, $\psi \in C^1([0,T])$ und $\dot{\psi} \leq L\psi$ (bzw. $\dot{\psi} \geq L\psi$) in $[0,T]$. Dann gilt $\psi(t) \leq \psi(0)e^{Lt}$ (bzw. $\psi(t) \geq \psi(0)e^{Lt}$) für $t \in [0,T]$.*

Beweis: (o.B.d.A. nur für $\leq$)

$$\frac{d}{dt}\left(e^{-Lt}\psi\right) = e^{-Lt}(\dot{\psi} - L\psi) \leq 0$$

impliziert

$$e^{-Lt}\psi(t) \leq \psi(0)\,.$$

∎

Die wesentliche Annahme des Lemmas ist eine *Differentialungleichung*. Das Lemma besagt, dass Lösungen der Differentialungleichung durch die Lösung der entsprechenden Differentialgleichung abgeschätzt werden können.

Man beachte, dass die Beweismethode auch verwendet werden kann, um die Differentialgleichung zu lösen. In diesem Zusammenhang nennt man e^{-Lt} einen *integrierenden Faktor*.

Das Hilfsresultat kann verwendet werden, um ein viel allgemeineres *Vergleichsprinzip* zu beweisen.

Satz 3 *Seien $x, y \in C^1([0,T]$, sei $f(t,x)$ stetig und gleichmäßig (in t) lipschitzstetig in x in einer Menge, die die Graphen von x und y enthält. Weiters gelte*

$$x(0) \le y(0) \qquad \textit{und} \qquad \dot{x} - f(t,x(t)) \le \dot{y} - f(t,y(t)) \quad \textit{für } t \in [0,T]\,.$$

Dann gilt $x \le y$ in $[0,T]$. Gilt darüberhinaus $x(0) < y(0)$, dann gilt $x < y$ in $[0,T]$.

Beweis: Sei L die Lipschitzkonstante von f und $\psi(t) := x(t) - y(t)$. Angenommen es existiere $t_1 > 0$ mit $\psi(t_1) > 0$. Dann existiert auch $t_0 \ge 0$ mit $\psi(t_0) = 0$ und $\psi > 0$ in $(t_0, t_1]$. In diesem Intervall gilt dann auch

$$\dot{\psi} \le f(t,x) - f(t,y) \le L|x-y| = L\psi\,.$$

Lemma 2 impliziert daher den Widerspruch $\psi \le 0$ in $[t_0, t_1]$.

Um die letzte Aussage zu beweisen, setzen wir jetzt $\varphi(t) = y(t) - x(t) \ge 0$, für das analog zu oben

$$\dot{\varphi} \ge f(t,y) - f(t,x) \ge -|f(t,y) - f(t,x)| \ge -L|y-x| = -L\varphi$$

gilt und daher, wieder mit Lemma 2, $\varphi(t) \ge \varphi(0)e^{-Lt} > 0$. ∎

Eine Funktion $\underline{x}(t)$ die die Differentialungleichung $\dot{\underline{x}} \le f(t,\underline{x})$ erfüllt, nennt man eine *Unterlösung* der Differentialgleichung $\dot{x} = f(t,x)$. Analog erfüllen *Oberlösungen* $\overline{x}(t)$ die Differentialungleichung $\dot{\overline{x}} \ge f(t,\overline{x})$. Satz 3 impliziert dann die folgende Verallgemeinerung von Lemma 2.

Korollar 1 *Sei $x \in C^1([0,T])$ Lösung des Anfangswertproblems $\dot{x} = f(t,x)$, $x(0) = x_0$, mit gleichmäßig lipschitzstetigem $f(t,\cdot)$. Weiters seien $\underline{x}$ bzw. $\overline{x}$ Unter- bzw. Oberlösung der Differentialgleichung in $[0,T]$ mit $\underline{x}(0) \le x_0 \le \overline{x}(0)$. Dann gilt*

$$\underline{x} \le x \le \overline{x} \qquad \textit{in } [0,T]\,.$$

Daraus folgt auch direkt die Eindeutigkeit der Lösung des Anfangswertproblems.

Integralungleichungen: Wie das Gronwall-Lemma in differentieller Form ist auch das *Gronwall-Lemma in Integralform* ein wichtiges Hilfsresultat:

Lemma 3 *Sei $\psi \in C([0,T])$, $\alpha \in \mathbb{R}$, $\beta \ge 0$ und es gelte die Ungleichung*

$$\psi(t) \le \alpha + \beta \int_0^t \psi(s)ds\,, \qquad 0 \le t \le T\,.$$

Dann gilt $\psi(t) \le \alpha e^{\beta t}$, $0 \le t \le T$.

Beweis: Wir präsentieren einen kurzen Beweis (der wieder den integrierenden Faktor verwendet):

$$\frac{d}{dt}\left(e^{-\beta t}\int_0^t \psi(s)ds\right) = e^{-\beta t}\left(\psi(t) - \beta\int_0^t \psi(s)ds\right) \le e^{-\beta t}\alpha\,.$$

Diese Ungleichung integrieren wir:

$$e^{-\beta t}\int_0^t \psi(s)ds \le \frac{\alpha}{\beta}(1 - e^{-\beta t})\,.$$

Daraus folgt (nach Multiplikation mit $\beta e^{\beta t} \geq 0$)

$$\psi(t) \leq \alpha + \beta \int_0^t \psi(s)ds \leq \alpha + \alpha(e^{\beta t} - 1) = \alpha e^{\beta t} .$$

■

Bemerkung 4 *Man beachte, dass analog zur differentiellen Form Lösungen der Ungleichung durch die Lösung der entsprechenden Gleichung, d.h. des Anfangswertproblems $\dot{\psi} = \beta\psi$, $\psi(0) = \alpha$, abgeschätzt werden können.*

Verallgemeinertes Gronwall-Lemma $\longrightarrow$ [T, 2.4]

Das maximale Existenzintervall: Das Beispiel $\dot{x} = x^2$ aus Abschnitt 1.4 hat gezeigt, dass die Beschränktheit des Existenzintervalles im Satz von Picard-Lindelöf keine Schwäche dieses Resultates ist. In diesem Beispiel hört die Lösung nach endlicher Zeit auf zu existieren, indem sie nach unendlich divergiert. Das folgende Resultat zeigt, dass nichts 'Schlimmeres' passieren kann.

Satz 4 *Sei $f \in C^1(\mathbb{R}^{n+1}, \mathbb{R}^n)$ und $x_0 \in \mathbb{R}^n$. Dann ist das maximale Existenzintervall I der eindeutigen Lösung des Problems (5), (6) offen, d.h. $I = (a, b)$ mit $-\infty \leq a < 0 < b \leq \infty$. In den Fällen $a > -\infty$ bzw. $b < \infty$ gilt*

$$\lim_{t \to a+} |x(t)| = \infty \qquad \textit{bzw.} \qquad \lim_{t \to b-} |x(t)| = \infty .$$

Beweis: Im Fall $I = \mathbb{R}$ ist nichts zu beweisen. Sei also zunächst $b < \infty$. Angenommen, $\lim_{t\to b-} |x(t)| = \infty$ gilt nicht. Dann gibt es eine Folge $t_n \to b-$, sodass die Folge $x(t_n)$ beschränkt ist und daher eine konvergente Teilfolge $x(t_{n_k}) \to \overline{x}$ besitzt (Satz von *Bolzano-Weierstrass*). Aus dem Satz von Picard-Lindelöf folgt, dass für eine Umgebung U des Punktes $(b, \overline{x})$ ein $T > 0$ existiert, sodass für alle $(\tilde{t}, \tilde{x})$ in U die Lösung des Anfangswertproblems (5), $x(\tilde{t}) = \tilde{x}$, im Intervall $(\tilde{t} - T, \tilde{t} + T)$ existiert. Da die Folge $(t_{n_k}, x(t_{n_k}))$ gegen $(b, \overline{x})$ konvergiert, existiert ein n_k, sodass $(t_{n_k}, x(t_{n_k})) \in U$ und $b - t_{n_k} < T$ gilt. Die Lösung kann daher bis zum Zeitpunkt $t_{n_k} + T > b$ fortgesetzt werden, was ein Widerspruch gegen die Tatsache ist, dass b das rechte Ende des Existenzintervalles ist. Die Offenheit des Existenzintervalles bei b ist eine triviale Konsequenz.

Das linke Ende behandelt man analog. ■

Globale Existenz: Mit Hilfe des letzten Resultates kann man in manchen Fällen leicht *globale Existenz* von Lösungen zeigen, d.h. dass die Lösung für alle Zeiten existiert. Der folgende Satz ist ein typisches globales Existenzresultat.

Satz 5 *Seien die Annahmen von Satz 4 erfüllt und sei das Wachstum der rechten Seite höchstens linear in x, d.h. es existieren $0 \leq \lambda, \gamma \in C(\mathbb{R})$ sodass $|f(t, x)| \leq \lambda(t)|x| + \gamma(t)$ für alle $t \in \mathbb{R}$, $x \in \mathbb{R}^n$. Dann existiert die Lösung des Anfangswertproblems (5), (6) global, d.h. für alle $t \in \mathbb{R}$.*

Beweis: Aus der Integralgleichungsdarstellung

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

des Anfangswertproblems folgt

$$|x(t)| \le |x_0| + \int_{t_0}^{t} (\lambda(s)|x(s)| + \gamma(s))ds\,.$$

Mit $\psi(\tau) = |x(t_0+\tau)|$ ergibt sich

$$\psi(\tau) \le |x_0| + \int_{t_0}^{t_0+\tau} (\lambda(s)\psi(s-t_0) + \gamma(s))ds = \alpha(\tau) + \int_0^{\tau} \beta(\sigma)\psi(\sigma)d\sigma$$

mit

$$\alpha(\tau) = |x_0| + \int_{t_0}^{t_0+\tau} \gamma(s)ds\,, \qquad \beta(\tau) = \lambda(t_0+\tau)\,.$$

Da α monoton wachsend ist, folgt aus dem verallgemeinerten Gronwall-Lemma

$$|x(t)| = \psi(t-t_0) \le \left(|x_0| + \int_{t_0}^{t} \gamma(s)ds\right) \exp\left(\int_{t_0}^{t} \lambda(s)ds\right)\,.$$

Daher ist es ausgeschlossen, dass $|x(t)|$ zu einem endlichen Zeitpunkt unbeschränkt wächst, und die Lösung ist daher als Konsequenz aus Satz 4 global. ∎

Challenge 5 *Man gebe eine Bedingung an, unter der das Anfangswertproblem*

$$\dot{x} = \lambda(t)x^a\,, \quad x(0) = x_0\,, \qquad \textit{mit } \lambda(t) \ge 0\,, \quad a > 1\,, \quad x_0 > 0\,,$$

eine globale Lösung besitzt.

2.4 Sachgemäße Gestelltheit des Anfangswertproblems

Sachgemäße Gestelltheit nach *Jacques Hadamard* (1865–1963): Mathematische Modelle physikalischer Phänomene sollten folgende Eigenschaften besitzen:

1. **Existenz** einer Lösung,
2. **Eindeutigkeit** der Lösung,
3. **Stetige Abhängigkeit** der Lösung von den Daten.

Für Anfangswertprobleme der Form (5), (6) ist der Satz von Picard-Lindelöf ein Resultat zu 1. und 2. Die Forderung 3. bedarf einer Präzisierung bzgl. der Fragen, was man als Daten ansieht und wie Fehler in den Daten und in der Lösung gemessen werden. Im folgenden Resultat wird alles berücksichtigt, was das Problem (5), (6) definiert, nämlich die rechte Seite der Differentialgleichung und der Anfangszustand.

Satz 6 *Seien $x, y \in C^1([0,T])$ Lösungen von (5), (6) bzw. von*

$$\dot{y} = g(t,y)\,, \qquad y(0) = y_0\,,$$

wobei $x_0, y_0 \in \mathbb{R}^n$ und $f, g \in C(U, \mathbb{R}^n)$ mit $U \supset \{(t,x(t)) : 0 \le t \le T\} \cup \{(t,y(t)) : 0 \le t \le T\}$. Weiters sei f gleichmäßig in t lipschitzstetig als Funktion von x, d.h. $|f(t,x) - f(t,y)| \le L|x-y|$ für alle $(t,x), (t,y) \in U$. Dann gilt

$$|x(t) - y(t)| \le e^{Lt}|x_0 - y_0| + \frac{e^{Lt}-1}{L} \sup_U |f-g|\,, \qquad 0 \le t \le T\,.$$

Beweis: Aus den integrierten Formen der beiden Anfangswertprobleme folgern wir

$$\begin{aligned} |x(t) - y(t)| &\le |x_0 - y_0| + \int_0^t |f(s,x(s)) - g(s,y(s))|ds \\ &\le |x_0 - y_0| + \int_0^t |f(s,x(s)) - f(s,y(s))|ds + \int_0^t |f(s,y(s)) - g(s,y(s))|ds \\ &\le |x_0 - y_0| + t \sup_U |f-g| + L\int_0^t |x(s-y(s)|ds \end{aligned}$$

Mit $\psi(t) := \frac{1}{L}\sup_U |f-g| + |x(t)-y(t)|$ gilt daher

$$\psi(t) \le \frac{1}{L}\sup_U |f-g| + |x_0 - y_0| + L\int_0^t \psi(s)ds\,.$$

Nun verwenden wir Lemma 3 (das Gronwall-Lemma in Integralform):

$$\psi(t) \le e^{Lt}\left(\frac{1}{L}\sup_U |f-g| + |x_0-y_0|\right)\,,$$

woraus das gesuchte Ergebnis folgt. ∎

Bemerkung 5 *Das Resultat zeigt, dass die Abbildung von den Daten (f, x_0) auf die Lösung x lipschitzstetig ist, wenn man für f und x die Supremum-Normen in $C(U, \mathbb{R}^n)$ bzw. $C([0,T], \mathbb{R}^n)$ verwendet und für x_0 die euklidische Norm im $\mathbb{R}^n$.*

Bemerkung 6 *Man beachte, dass nur eine der beiden verglichenen rechten Seiten lipschitzstetig sein muss.*

Bezüglich der Abhängigkeit vom Anfangszustand gehen wir noch einen Schritt weiter. Wir machen die Abhängigkeit explizit, indem wir die Lösung des Anfangswertproblems (5), (6) schreiben als $\varphi(t, x_0) := x(t)$, d.h.

$$\frac{\partial \varphi}{\partial t} = f(t,\varphi)\,, \qquad \varphi(0,x_0) = x_0\,.$$

Unter der Annahme, dass f glatt ist, werden wir zeigen, dass $x_0 \mapsto \varphi(t,x_0)$ eine glatte Funktion ist. Wir argumentieren zunächst wieder formal. Angenommen φ wäre glatt, dann bezeichnen wir die Jacobimatrix mit $Y(t,x_0) = \frac{\partial \varphi}{\partial x_0}(t,x_0)$. Differenzieren des Anfangswertproblems ergibt dann

$$\frac{\partial Y}{\partial t} = \frac{\partial f}{\partial x}(t,\varphi)Y\,, \qquad Y(0,x_0) = I_n\,,$$

mit der $(n \times n)$-Einheitsmatrix I_n. Um klarer zu machen, was diese Matrix-Differentialgleichung bedeutet, komponentenweise:

$$\frac{\partial Y_{kl}}{\partial t} = \sum_{m=1}^{n} \frac{\partial f_k}{\partial x_m}(t,\varphi) Y_{ml}\,, \qquad Y_{kl}(0,x_0) = \delta_{kl}\,, \qquad 1 \le k,l \le n\,,$$

mit dem Kronecker-Delta δ_{kl}. Das zeigt, dass das Matrixproblem in n unabhängige Vektorprobleme zerlegt werden und Y spaltenweise bestimmt werden kann, wobei der Anfangszustand für die l-te Spalte der l-te kanonische Basisvektor im $\mathbb{R}^n$ ist. Unter der Annahme, dass man φ kennt, kann man das Problem für Y lösen (Sätze 2 und 5). Damit ist aber noch nicht bewiesen, dass die Komponenten von Y die partiellen Ableitungen der Komponenten von φ nach den Komponenten von x_0 sind. Dazu müsste man zeigen, dass

$$|\varphi(t,x_0+h) - \varphi(t,x_0) - Y(t,x_0)h| = o(|h|) \qquad \text{für } h \to 0$$

gilt. Mit der Abkürzung $\varphi_h = \varphi(t,x_0+h)$ und den Integralversionen der Anfangswertprobleme,

$$\varphi_h(t) = x_0 + h + \int_0^t f(s,\varphi_h(s))ds\,, \quad \varphi(t) = x_0 + \int_0^t f(s,\varphi(s))ds\,,$$

$$Y(t)h = h + \int_0^t \frac{\partial f}{\partial x}(s,\varphi(s))Y(s)h\,ds\,,$$

ergibt sich

$$\varphi_h(t) - \varphi(t) - Y(t)h = \int_0^t \left(f(s,\varphi_h(s)) - f(s,\varphi(s)) - \frac{\partial f}{\partial x}(s,\varphi(s))Y(s)h \right) ds\,. \tag{11}$$

Nun berechnen wir

$$\begin{aligned} f(s,\varphi_h(s)) - f(s,\varphi(s)) &= \int_0^1 \frac{\partial f}{\partial x}\Big(s,\varphi(s) + r(\varphi_h(s)-\varphi(s))\Big)(\varphi_h(s)-\varphi(s))dr \\ &= \frac{\partial f}{\partial x}(s,\varphi(s))(\varphi_h(s)-\varphi(s)) + R(s)\,, \end{aligned}$$

mit

$$|R(s)| \le \delta(|\varphi_h(s)-\varphi(s)|)|\varphi_h(s)-\varphi(s)|\,,$$

wobei δ der Stetigkeitsmodul von $\frac{\partial f}{\partial x}$ ist (mit den Eigenschaften stetig, monoton wachsend und $\delta(0)=0$). Aus Satz 6 folgt

$$|\varphi_h(s)-\varphi(s)| \le e^{Ls}|h|\,,$$

mit der Lipschitzkonstanten L von f, und damit $|R(s)| = o(|h|)$. Nun können wir (11) in der Form

$$\varphi_h(t) - \varphi(t) - Y(t)h = \int_0^t \left(\frac{\partial f}{\partial x}(s,\varphi(s))(\varphi_h(s)-\varphi(s)-Y(s)h) + R(s) \right) ds\,,$$

schreiben und erhalten

$$|\varphi_h(t) - \varphi(t) - Y(t)h| \le L\int_0^t |\varphi_h(s)-\varphi(s)-Y(s)h|ds + \int_0^t |R(s)|ds\,,$$

Das verallgemeinerte Gronwall-Lemma impliziert

$$|\varphi_h(t) - \varphi(t) - Y(t)h| \le e^{Lt} \int_0^t |R(s)|ds = o(|h|)\,,$$

woraus folgt dass wirklich $Y = \frac{\partial \varphi}{\partial x_0}$. Eine weitere Anwendung von Satz 6 zeigt, dass Y stetig von x_0 abhängt, weil das für φ gilt. Daher ist φ unter den bisher verwendeten Annahmen (also stetige Differenzierbarkeit von f) stetig differenzierbar nach x_0. Das lässt sich induktiv auf höhere Ableitungen fortsetzen:

Satz 7 *Sei (t_0, x_0) ein innerer Punkt der beschränkten abgeschlossenen Menge $U \subset \mathbb{R}^{n+1}$ und $f \in C^k(U, \mathbb{R}^n)$. Dann ist die Lösung des Anfangswertproblems (5), (6) in ihrem Existenzintervall eine k-mal stetig differenzierbare Funktion von x_0.*

Ergänzung zum Beweis für $k = 1$: Die stetige Differenzierbarkeit von f auf der abgeschlossenen beschränkten Menge V impliziert gleichmäßige Stetigkeit der Ableitungen und damit die Existenz des Stetigkeitsmoduls δ sowie die Endlichkeit der Lipschitzkonstanten L, die gleichzeitig auch Schranke für die Norm der Jacobimatrix $\frac{\partial f}{\partial x}$ ist.

2.5 Reguläre Störungstheorie

In den ersten Übungen wurde für den Abstand eines fallenden Körpers von der Erdoberfläche das skalierte Problem

$$\ddot{x} = -(1 + \varepsilon x)^{-2}\,, \qquad x(0) = 1\,, \quad \dot{x}(0) = 0\,,$$

hergeleitet, wobei der dimensionslose Parameter $\varepsilon > 0$ das Verhältnis aus Anfangshöhe und Erdradius bezeichnet. Für kleine Werte von ε wurde eine formale asymptotische Entwicklung in Potenzen von ε gemäß dem Ansatz

$$x(t, \varepsilon) = x_0(t) + \varepsilon x_1(t) + O(\varepsilon^2)\,, \quad \text{für } \varepsilon \to 0\,,$$

berechnet. Wir werden zeigen, dass die Methode gerechtfertigt ist in dem Sinn, dass eine Konstante $c > 0$ existiert, sodass für $T > 0$ und ε klein genug Folgendes gilt:

$$\sup_{[0,T]} |x - x_0 - \varepsilon x_1| \le c\varepsilon^2\,.$$

Das wird eine Anwendung eines allgemeinen Resultates sein, das wir im Folgenden herleiten.

Challenge 6 *Man betrachte das Problem eines fallenden Körpers mit konstanter Gravitationsbeschleunigung und veräl tnismäßig kleiner Luftreibung in dimensionsloser Form:*

$$\dot{v} = -\varepsilon v - 1\,, \qquad v(0) = 0\,.$$

Man berechne $v_0(t)$ und $v_1(t)$ in $v = v_0 + \varepsilon v_1 + O(\varepsilon^2)$ für $\varepsilon \to 0$.

Wir betrachten Probleme der Form

$$\dot{x} = f(t, x, \varepsilon)\,, \qquad x(0, \varepsilon) = \overline{x}\,, \tag{12}$$

wobei f eine in einer Umgebung von $(0, \overline{x}, 0) \in \mathbb{R}^{n+2}$ definierte glatte Funktion sei. Wir wollen zeigen, dass die Lösung $x(t, \varepsilon)$ eine glatte Funktion von ε ist, woraus folgt, dass sie durch Taylorpolynome approximiert werden kann. Wir nehmen zunächst an, dass $y(t, \varepsilon) = \frac{\partial x}{\partial \varepsilon}(t, \varepsilon)$ existiert und leiten ein Anfangswertproblem dafür her, indem wir (12) nach ε differenzieren:

$$\dot{y} = \frac{\partial f}{\partial x}(t, x, \varepsilon) y + \frac{\partial f}{\partial \varepsilon}(t, x, \varepsilon)\,, \qquad y(0, \varepsilon) = 0\,.$$

Insbesondere werden wir interessiert sein an $x_1(t) = y(t, 0)$, das das Problem

$$\dot{x}_1 = \frac{\partial f}{\partial x}(t, x_0, 0) x_1 + \frac{\partial f}{\partial \varepsilon}(t, x_0, 0)\,, \qquad x_1(0) = 0\,,$$

löst, wobei x_0 Lösung des Näherungsproblems

$$\dot{x}_0 = f(t, x_0, 0)\,, \qquad x_0(0) = \overline{x}\,,$$

ist. Ähnlich zum vorigen Abschnitt definieren wir x_1 als Lösung dieses Problems und zeigen dann, dass $x = x_0 + \varepsilon x_1 + O(\varepsilon^2)$ gilt.

Zunächst folgt aus dem Satz von Picard-Lindelöf, dass die Probleme für x_0 und x_1 in einem Intervall $[-T, T]$ eindeutige Lösungen besitzen. Als nächsten Schritt betrachten wir das Problem für den Fehler $r = x - x_0 - \varepsilon x_1$, wobei wir die Integralversionen der Probleme für x, x_0 und x_1 verwenden:

$$\begin{aligned} r(t) &= \int_0^t \left(f(s, x_0(s) + \varepsilon x_1(s) + r(s), \varepsilon) - f(s, x_0(s), 0) - \varepsilon \frac{\partial f}{\partial x}(s, x_0(s), 0) x_1(s) - \varepsilon \frac{\partial f}{\partial \varepsilon}(s, x_0(s), 0) \right) ds \\ &= \int_0^t \left(\frac{\partial f}{\partial x}\Big(s, x_0(s) + \varepsilon x_1(s), \varepsilon\Big) r(s) + R_1(s, r(s), \varepsilon) + R_2(s, \varepsilon) \right) ds\,, \end{aligned}$$

mit

$$\begin{aligned} R_1(s, r, \varepsilon) &= f(s, x_0(s) + \varepsilon x_1(s) + r, \varepsilon) - f(s, x_0(s) + \varepsilon x_1(s), \varepsilon) - \frac{\partial f}{\partial x}\Big(s, x_0(s) + \varepsilon x_1(s), \varepsilon\Big) r\,, \\ R_2(s, \varepsilon) &= f(s, x_0(s) + \varepsilon x_1(s), \varepsilon) - f(s, x_0(s), 0) - \varepsilon \frac{\partial f}{\partial x}(s, x_0(s), 0) x_1(s) - \varepsilon \frac{\partial f}{\partial \varepsilon}(s, x_0(s), 0)\,. \end{aligned}$$

Offensichtlich sind R_1 und R_2 quadratische Restglieder in Taylorentwicklungen um $r = 0$ bzw. $\varepsilon = 0$. Daher gibt es Konstanten $c_1, c_2 > 0$, sodass

$$|R_1(s, r, \varepsilon)| \le c_1 |r|^2\,, \qquad |R_2(s, \varepsilon)| \le c_2 \varepsilon^2\,.$$

Die nächste Idee ist, das Problem für r durch eine Fixpunktiteration zu lösen, wobei in R_1 statt $r(s)$ die Approximation $r_k(s)$ eingesetzt wird, und sonst r_{k+1}. In jedem Iterationsschritt wird also ein lineares Problem gelöst. Es folgt

$$|r_{k+1}(t)| \le L \int_0^t |r_{k+1}(s)| ds + \int_0^t c_1 |r_k(s)|^2 ds + \varepsilon^2 c_2 t\,,$$

und daher aus dem verallgemeinerten Gronwall-Lemma

$$|r_{k+1}(t)| \le e^{Lt}\left(\int_0^t c_1|r_k(s)|^2 ds + \varepsilon^2 c_2 t\right)$$

Daraus folgt, mit der Supremum-Norm $\|\cdot\|_\infty$ auf dem Intervall $[0,T]$,

$$\|r_{k+1}\|_\infty \le e^{LT} T\left(c_1\|r_k\|_\infty^2 + \varepsilon^2 c_2\right) .$$

Wir behaupten nun, dass für ε klein genug die Fixpunktiteration eine Kugel mit Radius $\varepsilon^2 c$ mit geeignetem $c > 0$ in sich selbst abbildet. Eine Möglichkeit ist die Wahl

$$c = 2e^{LT} T c_2 \,, \qquad \varepsilon^2 c_1 c^2 \le c_2 \,,$$

wie man leicht sieht. Konvergiert die Fixpunktiteration gegen eine Lösung, dann haben wir das erwartete Resultat. Dazu benötigen wir die Kontraktionseigenschaft der Fixpunktabbildung. Diese folgt aber daraus, dass quadratische Restglieder auf kleinen Kugeln kleine Lipschitzkonstanten haben (In der folgenden Rechnung setzen wir zur Übersichtlichkeit $g(r) = f(s, x_0(s)+\varepsilon x_1(s)+r, \varepsilon)$):

$$\begin{aligned} |R_1(s,r,\varepsilon) - R_1(s,\hat r,\varepsilon)| &= \left|g(r) - g(\hat r) - \frac{\partial g}{\partial r}(0)(r-\hat r)\right| = \left|\left(\frac{\partial g}{\partial r}(\tilde r) - \frac{\partial g}{\partial r}(0)\right)(r-\hat r)\right| \\ &\le \delta(\varepsilon^2 c)|r-\hat r| \,, \end{aligned}$$

wobei $\|r\|_\infty, \|\hat r\|_\infty \le \varepsilon^2 c$ angenommen wurde, $\tilde r$ zwischen r und $\hat r$ liegt und δ der Stetigkeitsmodul von $\frac{\partial g}{\partial r}$ ist.

$\longrightarrow$ [T, Satz 2.12]

Singulär gestörte Differentialgleichungen:

Challenge 7 *Man löse das Anfangswertproblem*

$$\varepsilon\dot x = -x + t \,, \qquad x(0) = 1 \qquad 0 < \varepsilon \ll 1 \,.$$

Hinweis: Ansatz für eine Partikulärlösung: $x_p(t) = at + b$
Man berechne den punktweisen Limes der Lösung für $\varepsilon \to 0+$. Ist die Konvergenz gleichmäßig?

Dieses Beispiel fällt natürlich nicht in die oben betrachtete Klasse der *regulär gestörten Probleme,* weil die rechte Seite $f(t,x,\varepsilon) = \frac{-x+t}{\varepsilon}$ an $\varepsilon = 0$ nicht ausgewertet werden kann. Man beachte dass der formale Limes der Gleichung,

$$0 = -x_0 + t \,,$$

keine Differentialgleichung mehr ist. Eine gewöhnliche Differentialgleichung, in der ein kleiner Parameter ε vorkommt, und die für $\varepsilon = 0$ von kleinerer Ordnung ist als für $\varepsilon \neq 0$, nennt man *singulär gestört.* Das Problem mit $\varepsilon = 0$ hat keine Lösung, weil $x_0(t) = t$ die Anfangsbedingung nicht erfüllt.

Die exakte Lösung besitzt einen Anteil, der in Abhängigkeit der schnellen Variablen $\tau := \frac{t}{\varepsilon}$ variiert. Führt man die entsprechende Koordinatentransformation durch, d.h. $\hat x(\tau) = x(\varepsilon\tau)$, dann ergibt sich für $\hat x$ das Problem

$$\frac{d\hat x}{d\tau} = -\hat x + \varepsilon\tau \,, \qquad \hat x(0) = 1 \,.$$

Mit $\varepsilon = 0$ erhält man die Näherung $\hat{x}_0(\tau) = e^{-\tau}$. Interessanterweise ist die Summe der beiden Näherungen,

$$x_0(t) + \hat{x}_0(\tau) = t + e^{-t/\varepsilon}$$

eine gleichmäßig in t gültige Approximation für die exakte Lösung. Das hier auftretende Phänomen eines kurzen Zeitraumes, innerhalb dessen die Lösung sich vom Anfangswert 1 bis nahe an die Null wegbewegt, nennt man eine *Grenzschicht.* Dieser Begriff wurde von *Ludwig Prandtl* (1875–1953) im Zusammenhang mit Problemen aus der Aerodynamik geprägt.

2.6 Das explizite Eulerverfahren – das Peano-Theorem

Das Peano-Theorem ist ein Existenzsatz für Anfangswertprobleme, der ohne die Annahme der Lipschitzstetigkeit der rechten Seite auskommt, dafür aber auch keine Eindeutigkeit liefert. Das Theorem kann auf verschiedene Arten bewiesen werden. Hier gehen wir von einer diskreten Approximation für Lösungen aus, die wir mit Hilfe des expliziten Eulerverfahrens (siehe Abschnitt 1.1) berechnen.

Um das Problem

$$\dot{x} = f(t,x)\,, \qquad x(0) = x_0\,, \tag{13}$$

näherungsweise zu lösen, wählen wir einen Zeitschritt $h > 0$, die diskreten Zeiten $t_k = hk$, $k = 0, 1, \ldots$, und betrachten die Rekursion

$$x_h(t_{k+1}) = x_h(t_k) + hf(t_k, x_h(t_k))\,, \quad k \geq 0\,, \qquad x_h(0) = x_0\,. \tag{14}$$

Um zurückzukommen zur stetigen Zeit, verwenden wir stückweise lineare Interpolation, d.h.

$$x_h(t) = x_h(t_k) + \Big(x_h(t_{k+1}) - x_h(t_k)\Big)\frac{t-t_k}{h} = x_h(t_k) + f(t_k, x_h(t_k))(t-t_k) \qquad \text{für } t_k \leq t \leq t_{k+1}\,. \tag{15}$$

Zur Wohldefiniertheit der Näherungslösung müssen wir garantieren, dass die berechneten Zustände im Definitionsbereich der rechten Seite f bleiben:

Lemma 4 *Sei $x_0 \in \mathbb{R}^n$, $f \in C(V, \mathbb{R}^n)$ mit $V = [0,\tau] \times \overline{B_\delta(x_0)}$, $\delta > 0$, $M := \max_V |f|$ und $h > 0$. Dann gilt für die durch (14) definierte diskrete Lösung:*

$$(t_k, x_h(t_k)) \in V \qquad \textit{für } t_k \leq T := \min\{\tau, \delta/M\}\,.$$

Beweis: Im Rahmen einer vollständigen Induktion nehmen wir an, dass $(t_l, x_h(t_l)) \in V$ für $0 \leq l \leq k-1$. Dann gilt

$$x_h(t_k) = x_0 + h\sum_{l=0}^{k-1} f(t_l, x_h(t_l))\,,$$

und daher

$$|x_h(t_k) - x_0| \leq hkM = t_k M \leq \delta \quad \text{und} \quad t_k \leq \tau\,,$$

solange die Annahmen des Satzes gelten. ∎

Ist $t_k < T \leq t_{k+1}$, dann kann x_h für $t_k \leq t \leq T$ durch (15) definiert werden, und wir haben daher eine Familie von Funktionen $\{x_h : h > 0\} \subset C([0,T])$. Diese liegt in der abgeschlossenen Kugel

$\overline{\mathcal{K}_\delta(x_0)}$. Wäre $C([0,T])$ ein endlichdimensionaler Raum, dann könnte man den Satz von Bolzano-Weierstrass anwenden, um die Existenz einer konvergenten Folge $\{x_{h_n} : n \in \mathbb{N}\}$ mit $h_n \to 0$ zu garantieren.

Genauer gesagt: In endlichdimensionalen Räumen sind beschränkte abgeschlossene Mengen *folgenkompakt,* d.h. dass jede in der Menge liegende Folge einen in der Menge liegenden Häufungspunkt, also eine konvergente Teilfolge, besitzt. In unendlichdimensionalen Räumen sind Beschränktheit und Abgeschlossenheit notwendig aber nicht hinreichend für Folgenkompaktheit.

Der folgende *Satz von Arzelà-Ascoli* zeigt, dass zusätzlich *gleichgradige Stetigkeit* gefordert werden muss. Eine Menge $X \subset C[0,T]$ heißt gleichgradig stetig, wenn sie einen gemeinsamen Stetigkeitsmodul besitzt, d.h. für alle $\varepsilon > 0$ existiert $\delta(\varepsilon) > 0$, sodass $|t-s| < \delta(\varepsilon)$ impliziert, dass

$$|x(t) - x(s)| < \varepsilon \qquad \text{für alle } x \in X\,.$$

Satz 8 (Arzelà-Ascoli) *Sei $X \subset C([0,T])$ beschränkt, abgeschlossen und gleichgradig stetig. Dann ist X folgenkompakt.*

Beweis: Sei (x_n) eine Folge in X und (t_k) eine Nummerierung der rationalen Zahlen in $[0,T]$. Da die Folge $(x_n(t_1))$ beschränkt ist, gibt es eine Teilfolge $(x_{1,n})$ von (x_n), sodass $(x_{1,n}(t_1))$ konvergiert. Induktiv konstruiert man für jedes $k \in \mathbb{N}$ eine Teilfolge $(x_{k,n})$ von $(x_{k-1,n})$, sodass $(x_{k,n}(t_1)), \ldots, (x_{k,n}(t_k))$ konvergieren. Die Diagonalfolge $y_n := x_{n,n}$ konvergiert daher an allen rationalen Punkten in $[0,T]$. Wir werden zeigen, dass $(y_n) \subset (x_n)$ eine Cauchyfolge und daher konvergent in $C([0,T])$ ist.
Sei $\delta(\varepsilon)$ der gemeinsame Stetigkeitsmodul der Elemente von X und es sei $\varepsilon > 0$. Dann gibt es Intervalle $I_1, \ldots, I_J$, deren Länge jeweils höchstens $\delta(\varepsilon)$ ist, sodass

$$[0,T] = \bigcup_{j=1}^{J} I_j\,.$$

Wir wählen rationale Zahlen $t_j \in I_j$, $j = 1, \ldots, J$. Dann gilt

$$|y_m(t_j) - y_n(t_j)| \le \varepsilon \qquad \text{für } m,n \ge N_j(\varepsilon)\,, \quad j = 1, \ldots, J\,.$$

Sei nun $t \in [0,T]$ beliebig und daher $t \in I_j$ für ein $j \in \{1, \ldots, J\}$. Dann gilt

$$|y_m(t) - y_n(t)| \le |y_m(t) - y_m(t_j)| + |y_m(t_j) - y_n(t_j)| + |y_n(t_j) - y_n(t)| \le 3\varepsilon$$

für $m,n \ge N(\varepsilon) := \max\{N_1(\varepsilon), \ldots, N_J(\varepsilon)\}$. ∎

Definieren wir die stückweise konstante Funktion

$$f_h(t) = f(t_k, x_h(t_k)) \qquad \text{für } t_k \le t < t_{k+1}\,,$$

dann gilt für $0 \le s < t \le T$

$$x_h(t) = x_h(s) + \int_s^t f_h(u)du\,, \tag{16}$$

und daher

$$|x_h(t) - x_h(s)| \le M|t-s|\,.$$

Es sind also alle Funktionen x_h lipschitzstetig mit der gemeinsamen Lipschitzkonstanten M. Daraus folgt, dass die Menge $\{x_h : h > 0\}$ gleichgradig stetig ist mit dem Stetigkeitsmodul $\delta(\varepsilon) = \varepsilon/M$. Daher ist nach Arzelà-Ascoli ihr Abschluss folgenkompakt, woraus folgt, dass es eine Folge $h_m \to 0$ gibt, für die die Folge (x_{h_m}) gleichmäßig gegen eine Grenzfunktion x konvergiert.

Setzt man in (16) $s = 0$ und $h = h_m$, so ergibt sich

$$x_{h_m}(t) = x_0 + \int_0^t f_{h_m}(s)ds\,. \tag{17}$$

Wir wollen in dieser Gleichung zum Limes übergehen. Für $t_k \le s \le t_{k+1}$ gilt

$$|f_{h_m}(s) - f(s, x(s))| = |f(t_k, x_{h_m}(t_k)) - f(s, x(s))|\,.$$

Für den Abstand zwischen den Argumenten von f gilt

$$\begin{aligned}&|(t_k, x_{h_m}(t_k)) - (s, x(s))| \le |t_k - s| + |x_{h_m}(t_k) - x(s)|\\ &\le h_m + |x_{h_m}(t_k) - x_{h_m}(s)| + |x_{h_m}(s) - x(s)| \le (1 + M)h_m + \|x_{h_m} - x\|_\infty\,.\end{aligned}$$

Da die rechte Seite für $m \to \infty$ gegen Null konvergiert und f gleichmäßig stetig ist, konvergiert f_{h_m} gleichmäßig gegen $f(\cdot, x(\cdot))$. Das zeigt, dass der Limes $m \to \infty$ in (17) die Integralformulierung des Anfangswertproblems (13) ergibt. Damit haben wir das Peano-Theorem bewiesen:

Satz 9 *Sei $x_0 \in \mathbb{R}^n$, $f \in C(V, \mathbb{R}^n)$ mit $V = [0, \tau] \times \overline{B_\delta(x_0)}$, $\delta > 0$, $M := \max_V |f|$ und $T := \min\{\tau, \delta/M\}$. Dann besitzt das Anfangswertproblem (13) eine Lösung $x \in C^1([0, T])$.*

Das explizite Eulerverfahren ist eine praktisch verwendbare Methode zur numerischen Approximation von Lösungen. In diesem Zusammenhang ist unser Resultat allerdings nicht sehr brauchbar. Man würde Konvergenz der ganzen Familie $\{x_h\}$ für $h \to 0$ und Abschätzungen für den Fehler erwarten. Das lässt sich aber nur unter Bedingungen zeigen, für die die Lösung eindeutig ist. Nehmen wir im Folgenden an, dass f, zusätzlich zu den Annahmen im Peano-Theorem, lipschitzstetig in beiden Argumenten ist, d.h.

$$|f(t, x) - f(s, y)| \le L(|t - s| + |x - y|) \qquad \text{für } (t, x), (s, y) \in V\,.$$

Die Integralformulierung des Problems für den Fehler $r = x - x_h$ ist dann

$$r(t) = \int_0^t (f(s, x(s)) - f_h(s))ds\,.$$

Für $t_k \le s \le t_{k+1}$ schätzen wir den Integranden ab durch

$$\begin{aligned}|f(s, x(s)) - f_h(s)| &\le |f(s, x(s)) - f(s, x_h(s))| + |f(s, x_h(s)) - f(t_k, x_h(t_k)|\\ &\le L|r(s)| + L(|s - t_k| + |x_h(s) - x_h(t_k)|) \le L|r(s)| + L(1 + M)h\end{aligned}$$

Das ergibt, mit $\psi(t) := |r(t)| + (1 + M)h$,

$$\psi(t) \le (1 + M)h + L\int_0^t \psi(s)ds\,,$$

und daher, mit dem Gronwall-Lemma,

$$|x(t) - x_h(t)| = \psi(t) - (1 + M)h \le (e^{Lt} - 1)(1 + M)h\,,$$

also Konvergenz von x_h gegen x mit einem Fehler der Größenordnung h^1. Man sagt daher, dass das explizite Eulerverfahren ein Verfahren mit der *Konvergenzordnung* 1 ist.

Das implizite Eulerverfahren – Stabilität: Das explizite Eulerverfahren heißt *explizit,* weil es auch ein *implizites Eulerverfahren* gibt. Der Unterschied ist, dass die rechte Seite der Gleichung am neuen Zeitpunkt ausgewertet wird:

$$x_h(t_{k+1}) = x_h(t_k) + hf(t_{k+1}, x_h(t_{k+1}))\,.$$

Um $x_h(t_{k+1})$ zu berechnen, muss man also ein im Allgemeinen nichtlineares Gleichungssystem lösen. Man kann zeigen, dass auch dieses Verfahren die Konvergenzordnung 1 hat. Warum sollte man den zusätzlichen Aufwand in jedem Zeitschritt auf sich nehmen?

Challenge 8 *Für die Gleichung $\dot{x} = -\lambda x$, $\lambda > 0$, vergleiche man das explizite und das implizite Eulerverfahren mit der exakten Lösung.*

3 Lineare Systeme

3.1 Homogene Systeme mit konstanten Koeffizienten

⟶ [T, 3.1]

Koordinatentransformation: Um die Gleichung

$$\dot{x} = Ax$$

zu lösen, kann es vorteilhaft sein, eine Koordinatentransformation der Form $x = Uy$ mit einer invertierbaren Matrix U durchzuführen. Für die neue Unbekannte $y(t)$ gilt dann

$$\dot{y} = U^{-1}\dot{x} = U^{-1}Ax = By\,, \qquad \text{mit } B = U^{-1}AU\,.$$

Die neue Koeffizientenmatrix erhält man aus der alten durch eine *Ähnlichkeitstransformation.* Das kann helfen, wenn e^{Bt} leichter zu berechnen ist als e^{At}.

Spezielle Lösungen: Die skalare Version der Gleichung hat als Lösung eine Exponentialfunktion. Es ist also naheliegend zu untersuchen, ob das System Lösungen der Form $x(t) = e^{\lambda t}r$ mit $\lambda \in \mathbb{C}$, $0 \neq r \in \mathbb{C}^n$ besitzt. Einsetzen ergibt

$$\lambda e^{\lambda t}r = Ae^{\lambda t}r \quad \Longrightarrow \quad Ar = \lambda r\,.$$

Es gibt also solche Lösungen, wenn λ ein *Eigenwert* und r ein dazugehöriger *Eigenvektor* von A ist. Die Menge aller Eigenvektoren zum Eigenwert λ bildet einen Unterraum des $\mathbb{R}^n$, den *Eigenraum.*

Eigenwerte sind Nullstellen des charakteristischen Polynoms $p_A(\lambda) = \det(A - \lambda\mathbb{I})$, eines Polynoms n-ten Grades. Betrachten wir zunächst den einfachen Fall, dass es n verschiedene Eigenwerte gibt, d.h. dass das charakteristische Polynom n verschiedene Nullstellen $\lambda_1, \ldots, \lambda_n$ besitzt. Die in diesem Fall eindimensionalen Eigenräume seien aufgespannt von den Eigenvektoren $r_1, \ldots, r_n$, die in diesem Fall eine Basis des $\mathbb{R}^n$ bilden. Stellen wir die Lösung in dieser Basis dar, d.h.

$$x(t) = \sum_{j=1}^{n} y_j(t) r_j\,,$$

dann entspricht das einer Koordinatentransformation wie oben mit der Matrix $U = (r_1, \ldots, r_n)$, deren Spalten die Eigenvektoren sind. Differenzieren ergibt

$$\sum_{j=1}^{n} \dot{y}_j r_j = \dot{x} = Ax = A \sum_{j=1}^{n} y_j r_j = \sum_{j=1}^{n} \lambda_j y_j r_j \,,$$

und daher

$$\dot{y}_j = \lambda_j y_j \,, \qquad j = 1, \ldots, n \,.$$

Das System wurde also durch die Koordinatentransformation entkoppelt. Die transformierte Matrix ist die Diagonalmatrix $U^{-1}AU = J := \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Da $J^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)$ gilt, folgt

$$e^{Jt} = \operatorname{diag}\left(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}\right) .$$

Um das Anfangswertproblem $\dot{x} = Ax$, $x(0) = x_0$ zu lösen, muss auch der Anfangszustand in der neuen Basis dargestellt werden, d.h.

$$x_0 = \sum_{j=1}^{n} y_{0,j} r_j \qquad \text{bzw. } y_0 = U^{-1} x_0 \,.$$

Die Lösung ist dann

$$x(t) = U e^{Jt} y_0 = \sum_{j=1}^{n} e^{\lambda_j t} y_{0,j} r_j \,.$$

Das funktioniert auch noch, wenn für alle Eigenwerte gilt, dass ihre *algebraischen und geometrischen Vielfachheiten* gleich sind, d.h. dass Vielfachheit der Nullstellen von p_A (die algebraische Vielfachheit) gleich der Dimension der entsprechenden Eigenräume (der geometrischen Vielfachheit) ist. Oder: Zu jedem Eigenwert gibt es so viele linear unabhängige Eigenvektoren, wie es seiner (algebraischen) Vielfachheit entspricht. In diesem Fall bleibt in obiger Rechnung alles gleich, außer dass die Eigenwerte nicht alle verschieden sind.

Sonst: Jordansche Normalform $\longrightarrow$ [T, Theorem 3.2]

Sei

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix}$$

die jordansche Normalform von A, d.h. $A = UJU^{-1}$. Dann gilt $A^l = UJ^lU^{-1}$ und daher

$$e^{At} = U e^{Jt} U^{-1} \qquad \text{mit} \quad e^{Jt} = \begin{pmatrix} \exp(J_1 t) & & \\ & \ddots & \\ & & \exp(J_m t) \end{pmatrix} .$$

Sei J_i ein Jordanblock der Größe k mit Eigenwert λ_i. Dann gilt $J_i = \lambda_i \mathbb{I} + N$. Da die beiden Summanden kommutieren (Vielfache der Identität kommutieren mit jeder anderen Matrix), gilt [T, Lemma 3.1]

$$e^{J_i t} = e^{\lambda_i \mathbb{I} t} e^{Nt} = e^{\lambda_i t} e^{Nt} \qquad \left(\text{weil } e^{a\mathbb{I}} = e^a \mathbb{I}\right) .$$

Da die Matrix N *nilpotent vom Grad k* ist, d.h. $N^k = 0$, $N^{k-1} \neq 0$ (siehe auch [T, S. 63]), gilt

$$e^{Nt} = \sum_{l=0}^{k-1} \frac{N^l t^l}{l!}\,,$$

und daher

$$e^{J_i t} = e^{\lambda_i t} \begin{pmatrix} 1 & t & \cdots & \frac{t^{k-1}}{(k-1)!} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Bemerkung: Ist A reell, dann muss auch $e^{At} = Ue^{Jt}U^{-1}$ reell sein, was aus dieser Formel nicht so offensichtlich ist, wenn man bedenkt, dass im Allgemeinen U und die Eigenwerte komplex sein können. Allerdings kann man sagen, dass mit jedem Eigenwert-Eigenvektor-Paar (λ, r) auch $(\overline{\lambda}, \overline{r})$ ein Eigenwert-Eigenvektor-Paar ist (man konjugiere $Ar = \lambda r$), d.h. dass komplexe Eigenwerte nur in konjugiert komplexen Paaren auftreten. Weiters gilt: Ist $\lambda \in \mathbb{C} \setminus \mathbb{R}$ und $r = u + iv$, $u, v \in \mathbb{R}^n$, dann sind u und v linear unabhängig. Wären sie linear abhängig, dann wäre auch u oder v ein Eigenvektor zum Eigenwert λ, das dann reell sein müsste. Für eine Weiterführung dieser Überlegungen siehe [T, S. 65] und den folgenden Abschnitt für $n = 2$.

Nun beschäftigen wir uns noch mit dem Anfangswertproblem für das inhomogene System:

$$\dot{x}(t) = Ax(t) + g(t)\,, \qquad x(0) = x_0\,.$$

Zunächst noch kleine Beobachtungen zum Matrixexponential: Ist $\mathbf{0}$ die Nullmatrix, dann gilt $e^{\mathbf{0}} = \mathbb{I}$. Außerdem gilt $(e^A)^{-1} = e^{-A}$ (Da A und $-A$ kommutieren, gilt $e^A e^{-A} = e^{A-A} = e^{\mathbf{0}} = \mathbb{I}$) und (durch gliedweises Differenzieren der Taylorreihe)

$$\frac{d}{dt} e^{At} = Ae^{At} = e^{At}A\,.$$

Daher gilt

$$\frac{d}{dt}(e^{-At}x) = e^{-At}(\dot{x} - Ax) = e^{-At}g\,.$$

Integration von 0 bis t ergibt

$$e^{-At}x(t) - x_0 = \int_0^t e^{-As}g(s)ds\,,$$

und daher die Variation-der-Konstanten-Formel

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}g(s)ds\,.$$

3.2 Zweidimensionale Systeme – Stabilität

Reelle Berechnung des Matrixexponentials: Für $A \in \mathbb{R}^{2\times 2}$ betrachten wir 3 Fälle:

1. Ein oder zwei reelle Eigenwerte, algebraische Vielfachheit = geometrische Vielfachheit. In diesem Fall ist A durch $U = (r_1, r_2) \in \mathbb{R}^{2\times 2}$ diagonalisierbar, d.h. $J = U^{-1}AU = \mathrm{diag}(\lambda_1, \lambda_2)$ und $e^{Jt} = \mathrm{diag}(e^{\lambda_1 t}, e^{\lambda_2 t})$.

2. Ein reeller Eigenwert λ mit algebraischer Vielfachheit 2 und geometrischer Vielfachheit 1. Die Jordan-Normalform besteht aus einem Jordan-Block:

$$J = U^{-1}AU = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix},$$

wobei $U = (r_1, r_2)$ mit einem Eigenvektor r_1 und einem *Hauptvektor* r_2, d.h. $Ar_2 = \lambda r_2 + r_1$. Es gilt

$$e^{Jt} = e^{\lambda t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

3. Zwei konjugiert komplexe Eigenwerte $a \pm ib$ mit Eigenvektoren $u \pm iv$. In diesem Fall ist A nur durch eine komplexe Transformation diagonalisierbar. Es gilt

$$A(u+iv) = (a+ib)(u+iv) \quad \Longrightarrow \quad Au = au - bv\,,\ Av = bu + av \quad \Longrightarrow \quad AU = UJ\,,$$

mit $U = (u, v)$ und

$$J = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = a\mathbb{I} + bR\,, \qquad R = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

was man als *reelle Normalform* der Matrix A bezeichnen könnte. Man berechnet leicht $R^2 = -\mathbb{I}$ und daher

$$R^{2l} = (-1)^l\mathbb{I}\,, \qquad R^{2l+1} = (-1)^l R\,.$$

Daraus folgt

$$\begin{aligned} e^{Jt} &= e^{at}e^{Rbt} = e^{at}\sum_{k=0}^{\infty}\frac{(bt)^k}{k!}R^k = e^{at}\left(\sum_{l=0}^{\infty}\frac{(bt)^{2l}}{(2l)!}(-1)^l\mathbb{I} + \sum_{l=0}^{\infty}\frac{(bt)^{2l+1}}{(2l+1)!}(-1)^l R\right) \\ &= e^{at}\left(\cos(bt)\mathbb{I} + \sin(bt)R\right) = e^{at}\begin{pmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{pmatrix}. \end{aligned}$$

Alle Komponenten sind entweder Realteil oder Imaginärteil von

$$e^{(a+ib)t} = e^{at}(\cos(bt) + i\sin(bt))\,.$$

Das Phasenporträt: Das Phasenporträt ist eine geometrische Darstellung der Menge aller Lösungen eines autonomen Systems von Differentialgleichungen $\dot{x} = f(x)$. Dabei werden Lösungen $x(t)$, $t \in \mathbb{R}$ (oder t in einem Intervall), als Kurven im *Phasenraum* (das ist der Raum aller Zustände, d.h. der $\mathbb{R}^n$) aufgefasst. Die Gesamtheit aller dieser Kurven *(Trajektorien)* ist das Phasenporträt. Der Begriff wurde schon in Abschnitt 1.5 für $n = 1$ verwendet.

Was die graphische Darstellung betrifft, sind Phasenporträts fast nur im Fall $n = 2$ von Interesse, weil für $n = 1$ nicht viel Interessantes passieren kann und für $n \geq 3$ die graphische Darstellung nur schwer möglich ist.

Trotzdem zunächst einige dimensionsunabhängige Überlegungen unter der Annahme, dass $f : \mathbb{R}^n \to \mathbb{R}^n$ glatt ist. In diesem Fall kann das Anfangswertproblem mit beliebigem $x(0) = x_0 \in \mathbb{R}^n$ zumindest lokal gelöst werden. Ist x_0 kein stationärer Punkte, d.h. $f(x_0) \neq 0$, dann gibt es auch in einer geeigneten Umgebung U von x_0 keine stationären Punkte. In diesem Fall bilden die in U liegenden Teile der Trajektorien eine *schlichte Überdeckung* von U mit *glatten Kurven.*

Zur Erinnerung: Eine Kurve ist glatt, wenn sie eine glatte Parameterdarstellung mit nichtverschwindendem Tangentialvektor besitzt. Eine Menge wird von einer Kurvenschar schlicht überdeckt, wenn jeder Punkt der Menge auf genau einer Kurve liegt.

Insbesondere kann es in einer Menge U, die keine stationären Punkte enthält, keine Schnittpunkte verschiedener Trajektorien geben (wegen der Eindeutigkeit der Lösung des Anfangswertproblems). Besonders interessant sind also Umgebungen stationärer Punkte, aber auch globale Eigenschaften, wie zum Beispiel die Frage, ob es geschlossene Trajektorien gibt, die periodischen Lösungen entsprechen.

Zu homogenen linearen Systemen $\dot{x} = Ax$: Sind die Realteile aller Eigenwerte von A negativ, dann konvergieren alle Lösungen für $t \to \infty$ gegen Null. Der Ursprung wird in diesem Fall eine *Senke* genannt, oder auch ein *asymptotisch stabiler stationärer Punkt*. Sind umgekehrt alle Realteile positiv, dann konvergieren alle Lösungen für $t \to -\infty$ gegen Null. Der Ursprung wird dann eine *Quelle* genannt. Man beachte, dass diese Aussagen auch im Fall von Jordanblöcken gilt, weil die polynomialen Faktoren nichts am qualitativen Verhalten der Exponentialfunktionen verändern.

Insbesondere für $n = 2$: Gilt in obigem Fall 1., dass $\lambda_1 < 0 < \lambda_2$, dann besteht der Eigenraum von λ_1 aus 2 Trajektorien, die für $t \to \infty$ gegen Null konvergieren. Ihre Vereinigung nennt man *stabile Mannigfaltigkeit*. Der Eigenraum von λ_2 besteht aus 2 Trajektorien, die für $t \to -\infty$ gegen Null konvergieren. Ihre Vereinigung nennt man *instabile Mannigfaltigkeit*. Den Ursprung nennt man in diesem Fall einen *Sattelpunkt*.

Schließlich noch der obige Fall 3. mit rein imaginären Eigenwerten, d.h. $a = 0$. In diesem Fall sind alle Lösungen Linearkombinationen von $\cos(bt)$ und $\sin(bt)$ und daher periodisch. Die Trajektorien sind Ellipsen mit dem Ursprung als Mittelpunkt, der in diesem Fall ein *Zentrum* genannt wird.

$\longrightarrow$ [T, S. 68–70]

3.3 Lineare Gleichungen höherer Ordnung mit konstanten Koeffizienten

Wir beginnen mit homogenen Gleichungen

$$(\mathcal{L}x)(t) := x^{(n)} + c_{n-1}x^{(n-1)} + \cdots + c_1\dot{x} + c_0x = 0\,,$$

mit reellen Koeffizienten $c_k \in \mathbb{R}$, $k = 1, \ldots, n-1$. Es ist effizienter, direkt mit dieser Gleichung zu arbeiten, als sie in ein System erster Ordnung zu verwandeln. So führt der Ansatz $x(t) = e^{\lambda t}$ direkt auf das Nullstellenproblem für das *charakteristische Polynom*

$$p_{\mathcal{L}}(\lambda) := \sum_{k=0}^{n} c_k\lambda^k\,, \qquad \text{mit } c_n = 1\,,$$

dessen Faktorisierung wir schreiben als

$$p_{\mathcal{L}}(\lambda) = \prod_{j=1}^{m}(\lambda - \lambda_j)^{\alpha_j}\,,$$

mit $\lambda_j \in \mathbb{C}$, $j = 1, \ldots, m \le n$, $\alpha_j \in \mathbb{N}$, $\sum_{j=1}^{m} \alpha_j = n$.

Sei λ_j ein mehrfacher Eigenwert, d.h. $\alpha_j > 1$. Wir behaupten, dass dann nicht nur $x_{j,0}(t) := e^{\lambda_j t}$, sondern auch $x_{j,1}(t) := te^{\lambda_j t}$ eine Lösung der Differentialgleichung ist. Mit vollständiger Induktion zeigt man

$$x_{j,1}^{(k)} = (k\lambda_j^{k-1} + \lambda_j^k t)e^{\lambda_j t}\,,$$

und daher

$$(\mathcal{L}x_{j,1})(t) = e^{\lambda_j t}\sum_{k=0}^{n} c_k(k\lambda_j^{k-1} + \lambda_j^k t) = e^{\lambda_j t}(p'_{\mathcal{L}}(\lambda_j) + tp_{\mathcal{L}}(\lambda_j)) = 0\,,$$

weil mehrfache Nulllstellen von $p_{\mathcal{L}}$ auch Nullstellen von $p'_{\mathcal{L}}$ sind. Eine ähnliche, etwas kompliziertere Rechnung zeigt, dass $x_{j,k}(t) = t^k e^{\lambda_j t}$, $0 \le k \le \alpha_j - 1$, Lösungen sind. Es gibt also α_j linear unabhängige Lösungen zum Eigenwert λ_j. Damit haben wir eine vollständige Basis

$$\{x_{j,k} : 1 \le j \le m\,,\, 0 \le k \le \alpha_j - 1\}$$

des n-dimensionalen Lösungsraumes gefunden.

Komplexe Eigenwerte treten nur als konjugiert komplexe Paare $\lambda = a \pm ib$ mit denselben Vielfachheiten auf. Eine reelle Basis erhält man in diesem Fall, indem man jedes Paar $(t^k e^{\lambda t}, t^k e^{\overline{\lambda} t})$ durch $(t^k e^{at}\cos(bt), t^k e^{at}\sin(bt))$ ersetzt.

Eine Variation-der-Konstanten-Formel für inhomogene Gleichungen:

Lemma 5 *Die Funktion*

$$x_p(t) := \int_0^t u(t-s)g(s)ds$$

mit $\mathcal{L}u = 0$, $u(0) = \cdots = u^{(n-2)}(0) = 0$, $u^{(n-1)}(0) = 1$, *ist eine Partikulärlösung von* $\mathcal{L}x = g$.

Beweis: Es gilt

$$\dot{x}_p(t) = u(0)g(t) + \int_0^t \dot{u}(t-s)g(s)ds = \int_0^t \dot{u}(t-s)g(s)ds\,,$$

und durch weiteres Differenzieren,

$$x_p^{(k)}(t) = \int_0^t u^{(k)}(t-s)g(s)ds\,, \qquad k = 0, \ldots, n-1\,.$$

Noch einmal Differenzieren:

$$x_p^{(n)}(t) = u^{(n-1)}(0)g(t) + \int_0^t u^{(n)}(t-s)g(s)ds = g(t) + \int_0^t u^{(n)}(t-s)g(s)ds\,.$$

Daher

$$(\mathcal{L}x_p)(t) = g(t) + \int_0^t (\mathcal{L}u)(t-s)g(s)ds = g(t)\,.$$

■

Für Inhomogenitäten, die in endlichdimensionalen, bezüglich Differentiation abgeschlossenen Vektorräumen liegen, ist die Suche nach partikulären Lösungen durch Ansatz sinnvoll. Ein Beispiel ist das folgende Resultat:

Lemma 6 *Sei P ein Polynom und $\lambda \in \mathbb{C}$. Dann besitzt die Gleichung $(\mathcal{L}x)(t) = P(t)e^{\lambda t}$ eine Lösung der Form $x_p(t) = Q(t)e^{\lambda t}$, wobei Q ein Polynom ist. Im Fall $p_{\mathcal{L}}(\lambda) \neq 0$ gilt* $\mathrm{Grad}(Q) = \mathrm{Grad}(P)$. *Ist λ eine α-fache Nullstelle von $p_{\mathcal{L}}$, dann gilt* $\mathrm{Grad}(Q) = \mathrm{Grad}(P) + \alpha$.

Beweis: Wir setzen

$$Q(t) = \sum_{j=0}^{m} Q_j t^j \,,$$

und verwenden die Leibnizsche Produktregel:

$$x_p^{(k)}(t) = e^{\lambda t} \sum_{l=0}^{k} \binom{k}{l} Q^{(l)}(t) \lambda^{k-l} \,.$$

Daher

$$\begin{aligned}
(\mathcal{L}x_p)(t) &= e^{\lambda t} \sum_{k=0}^{n} \sum_{l=0}^{k} c_k \binom{k}{l} Q^{(l)}(t) \lambda^{k-l} = e^{\lambda t} \sum_{l=0}^{m} \frac{Q^{(l)}(t)}{l!} \sum_{k=l}^{n} c_k k(k-1)\cdots(k-l+1)\lambda^{k-l} \\
&= e^{\lambda t} \sum_{l=0}^{m} \frac{Q^{(l)}(t)}{l!} p_{\mathcal{L}}^{(l)}(\lambda) = e^{\lambda t} \sum_{l=0}^{m} \frac{p_{\mathcal{L}}^{(l)}(\lambda)}{l!} \sum_{i=l}^{m} Q_i i(i-1)\cdots(i-l+1) t^{i-l} \\
&= e^{\lambda t} \sum_{l=0}^{m} \frac{p_{\mathcal{L}}^{(l)}(\lambda)}{l!} \sum_{j=0}^{m-l} Q_{j+l}(j+l)(j+l-1)\cdots(j+1) t^j \\
&= e^{\lambda t} \sum_{j=0}^{m} t^j \sum_{l=0}^{m-j} p_{\mathcal{L}}^{(l)}(\lambda) \binom{j+l}{j} Q_{j+l}
\end{aligned}$$

Mit $P(t) = \sum_{j=0}^{m} P_j t^j$ ergibt Koeffizientenvergleich bei den Potenzen von t das Gleichungssystem

$$\sum_{i=j}^{m} p_{\mathcal{L}}^{(i-j)}(\lambda) \binom{i}{j} Q_i = P_j \,, \quad j = 0, \ldots, m \,.$$

Im Fall $p_{\mathcal{L}}(\lambda) \neq 0$ ist die Koeffizientenmatrix eine obere Dreiecksmatrix, in der alle Diagonalelemente gleich $p_{\mathcal{L}}(\lambda)$ sind. Das Gleichungssystem für die Koeffizienten von Q besitzt daher eine eindeutige Lösung. Insbesondere gilt $Q_m = P_m / p_{\mathcal{L}}(\lambda)$, was zeigt, dass die Polynome denselben Grad besitzen.

Ist λ eine α-fache Nullstelle von $p_{\mathcal{L}}$ dann gilt $p_{\mathcal{L}}(\lambda) = \cdots = p_{\mathcal{L}}^{(\alpha-1)}(\lambda) = 0$, $p_{\mathcal{L}}^{(\alpha)}(\lambda) \neq 0$. Daraus folgt, dass in $\mathcal{L}x_p$ die Koeffizienten von t^j für $m - j < \alpha$, d.h. für $j > m - \alpha$, verschwinden. Das Polynom in $\mathcal{L}x_p$ hat den Grad $m-\alpha$, was daher der maximale Grad von P sein kann. Die Koeffizienten $Q_0, \ldots, Q_{\alpha-1}$ kommen in $\mathcal{L}x_p$ nicht vor, was nicht überraschend ist, weil die entsprechenden Lösungsanteile die homogene Gleichung lösen. Die restlichen Koeffizienten $Q_\alpha, \ldots, Q_m$ lösen wieder ein Gleichungssystem mit einer oberen Dreiecksmatrix (und den Koeffizienten $P_0, \ldots, P_{m-\alpha}$ auf der rechten Seite), wobei die Diagonalelemente jetzt $p_{\mathcal{L}}^{(\alpha)}(\lambda)\binom{j+\alpha}{j} \neq 0$ sind. ∎

Kleine Schwingungen eines Fadenpendels: Wir beschreiben die Schwingungen einer Punktmasse, die an einem masselosen Faden der Länge ℓ hängt. Bezeichnet man den Winkel des Fadens zum Lot zum Zeitpunkt t mit $\varphi(t)$, dann ist der Betrag der Tangentialbeschleunigung gegeben durch $|F_{tan}| = g \sin\varphi$, wobei $g = |F_G|$ die als konstant und vertikal angenommene Gravitationsbeschleunigung ist (siehe Fig. 2). Andererseits ist die Tangentialbeschleunigung gegeben durch $\ell\ddot{\varphi}$. Das führt auf die Differentialgleichung

$$\ell\ddot{\varphi} + g \sin\varphi = 0 \,.$$

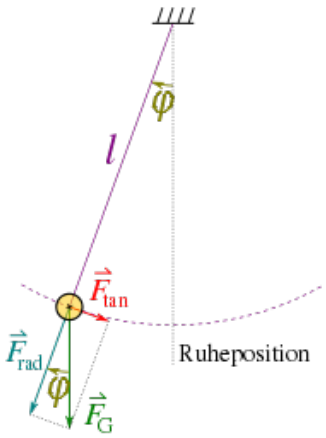

Figure 2: Das Fadenpendel

Kleine Schwingungen können mit Hilfe der Approximation $\sin\varphi \approx \varphi$ durch die lineare Gleichung

$$\ddot{\varphi} + \omega^2\varphi = 0\,,$$

mit der *Frequenz* $\omega := \sqrt{g/\ell}$ beschrieben werden. Diese Gleichung wird auch *harmonischer Oszillator* genannt.

Das charakteristische Polynom $\lambda^2 + \omega^2$ besitzt die rein imaginären Nullstellen $\lambda = \pm i\omega$. Daher ist $\{\cos(\omega t), \sin(\omega t)\}$ eine reelle Lösungsbasis. Alle Lösungen sind periodisch mit der Periode $2\pi/\omega$. Die Periodizität kann man auch mit Hilfe des schon in Abschnitt 1.4 vorgestellten Tricks erhalten, die Gleichung mit $\dot{\varphi}$ zu multiplizieren und dann zu integrieren:

$$\frac{\dot{\varphi}^2}{2} + \omega^2\frac{\varphi^2}{2} = E\,,$$

wobei wir daran erinnern, dass die Integrationskonstante E physikalisch als Gesamtenergie (Summe der kinetischen und der potentiellen Energie) interpretiert wird. Die Trajektorien im $(\varphi, \dot{\varphi})$-Phasenraum sind Ellipsen um den stationären Punkt $(0, 0)$, der daher im Sinne von Abschnitt 3.2 ein *Zentrum* ist.

Die Funktion u aus Lemma 5 löst hier das Problem

$$\ddot{u} + \omega^2 u = 0\,, \qquad u(0) = 0\,, \quad \dot{u}(0) = 1\,,$$

und daher $u(t) = \sin(\omega t)/\omega$. Die inhomogene Gleichung

$$\ddot{\varphi} + \omega^2\varphi = g$$

beschreibt ein Pendel, das durch eine zeitabhängige Kraft $g(t)$ angeregt wird. Eine Partikulärlösung der Gleichung ist laut Lemma 5 gegeben durch

$$\varphi_p(t) = \frac{1}{\omega}\int_0^t \sin(\omega(t-s))g(s)ds\,.$$

Als Anwendung von Lemma 6 betrachten wir eine periodische Anregung der Form $g(t) = \sin(\nu t)$. Ist die Frequenz ν der Anregung verschieden von der *Eigenfrequenz* ω des Oszillators, dann suchen wir nach einer Partikulärlösung der Form $\varphi_p(t) = A\sin(\nu t)$. Den cosinus lassen wir weg, weil in der Gleichung nur Ableitungen gerader Ordnung vorkommen. Einsetzen in die Differentialgleichung ergibt

$$-A\nu^2 + A\omega^2 = 1\,, \qquad \text{also } A = \frac{1}{\omega^2 - \nu^2}\,,$$

und damit die allgemeine Lösung

$$\varphi(t) = a\cos(\omega t) + b\sin(\omega t) + \frac{1}{\omega^2 - \nu^2}\sin(\nu t)\,, \qquad a, b \in \mathbb{R}\,.$$

Das ist eine Überlagerung von Schwingungen mit der Eigenfrequenz und mit der Anregungsfrequenz.

Anregung mit der Eigenfrequenz, d.h. $g(t) = \sin(\omega t)$ entspricht dem Fall $p_L(\lambda) = 0$ in Lemma 6. Da $i\omega$ eine einfache Nullstelle des charakteristischen Polynoms ist, erwarten wir ein Partikulärlösung der (reellen) Form

$$\varphi_p(t) = tA\sin(\omega t) + tB\cos(\omega t)\,.$$

Differenzieren ergibt $\dot\varphi_p(t) = (A - Bt\omega)\sin(\omega t) + (B + At\omega)\cos(\omega t)$, und Einsetzen in die Differentialgleichung daher

$$(-2B\omega - At\omega^2)\sin(\omega t) + (2A\omega - Bt\omega^2)\cos(\omega t) + At\omega^2\sin(\omega t) + Bt\omega^2\cos(\omega t) = \sin(\omega t)\,,$$

und damit

$$A = 0\,, \quad B = -\frac{1}{2\omega}\,.$$

Die Partikulärlösung

$$\varphi_p(t) = -\frac{t}{2\omega}\cos(\omega t)$$

ist eine Schwingung, deren Amplitude immer größer wird. Dieses Phänomen wird *Resonanz* genannt und hat in der Mechanik und in den Ingenieurwissenschaften große Bedeutung.

In der Praxis gibt es bei einem Pendel immer dämpfende Einflüsse, z.B. durch den Luftwiderstand. Dieser produziert eine *Reibungskraft.* Das einfachste Modell dafür ist eine der Bewegungsrichtung entgegengesetzte Kraft, deren Größe proportional zur Geschwindigkeit ist. Für das Pendel ohne Anregung ergibt sich die homogene Gleichung

$$\ddot\varphi + \kappa\dot\varphi + \omega^2\varphi = 0\,,$$

mit der Reibungskraft $-\kappa\dot\varphi$ und dem Reibungskoeffizienten $\kappa > 0$. Für die Nullstellen des charakteristischen Polynoms $\lambda^2 + \kappa\lambda + \omega^2$ gibt es 2 typische Fälle:

1. *Kleine Dämpfung,* $\kappa < 2\omega$: In diesem Fall gibt es die konjugiert komplexen Nullstellen

$$\lambda = -\frac{\kappa}{2} \pm i\sqrt{\omega^2 - \frac{\kappa^2}{4}}\,,$$

mit negativem Realteil. Die Lösungen sind Schwingungen mit der Frequenz $\sqrt{\omega^2 - \kappa^2/4}$, deren Amplitude so wie $e^{-\kappa t/2}$ abklingt.

2. Überdämpfung, $\kappa > 2\omega$: In diesem Fall gibt es 2 reelle negative Nullstellen

$$\lambda = -\frac{\kappa}{2} \pm \sqrt{\frac{\kappa^2}{4} - \omega^2}\,.$$

Das Pendel schwingt nicht, sondern konvergiert monoton (für große t) gegen den Equilibriumszustand $\varphi = 0$ (Pendel in Honig).

Zum Abschluss noch ein gedämpftes Pendel mit periodischer Anregung mit der Eigenfrequenz des ungedämpften Pendels:

$$\ddot{\varphi} + \kappa\dot{\varphi} + \omega^2\varphi = \sin(\omega t)\,.$$

In diesem Fall gibt es die Partikulärlösung

$$\varphi_p(t) = -\frac{1}{\kappa\omega}\cos(\omega t)\,,$$

deren Amplitude, wie zu erwarten ist, für $\kappa \to 0$ gegen unendlich geht. Man beachte, dass die Lösung auch für große Werte von κ funktioniert, d.h. auch ein überdämpftes Pendel kann man durch periodische Anregung zum Schwingen bringen.

3.4 Allgemeine lineare Systeme

Zunächst homogene Systeme:

$$\dot{x} = A(t)x\,, \tag{18}$$

mit $A \in C(I, \mathbb{R}^{n\times n})$, wobei I ein Intervall ist. Aus der Linearität folgt, dass die Menge aller auf I definierten Lösungen einen Vektorraum bildet.

Mit $t_0 \in I$ und $x_0 \in \mathbb{R}^n$ gilt, dass das Anfangswertproblem mit der Anfangsbedingung $x(t_0) = x_0$ eine eindeutige Lösung besitzt, die auf I definiert ist. Das folgt aus Satz 5. Andererseits kann jede auf I definierte Lösung an t_0 ausgewertet werden mit $x_0 := x(t_0)$ und ist daher gleich der eindeutigen Lösung mit gegebenem x_0. Daraus folgt, dass der Lösungsraum aufgespannt wird von den Lösungen $\varphi_j(t; t_0)$, mit den Anfangsbedingungen $\varphi_j(t_0; t_0) = e_j$, der j-te kanonische Basisvektor. Eine *Hauptmatrixlösung* ist dann gegeben durch die Matrix mit den Spalten $\varphi_1, \ldots, \varphi_n$, d.h.

$$\Pi(t; t_0) := (\varphi_1(t; t_0), \ldots, \varphi_n(t, t_0))\,.$$

Diese löst das Matrixproblem

$$\dot{\Pi} = A\Pi\,, \qquad \Pi(t_0; t_0) = \mathbb{I}\,.$$

Die Lösung des Anfangswertproblems mit $x(t_0) = x_0$ ist dann gegeben durch $x(t) = \Pi(t; t_0)x_0$. Für $t_1 \in I$ gilt auch $x(t) = \Pi(t; t_1)x(t_1) = \Pi(t; t_1)\Pi(t_1; t_0)x_0$ und daher, für $t, r, s \in I$,

$$\Pi(t; r) = \Pi(t; s)\Pi(s; r)\,.$$

Insbesondere, mit $r = t$,

$$\Pi(t; s)^{-1} = \Pi(s; t)\,.$$

Also ist $\Pi(t; s)$ für beliebige $s, t \in I$ invertierbar, und die oben eingeführten Lösungen $\varphi_1, \ldots, \varphi_n$ sind nicht nur an $t = t_0$, sondern an jedem $t \in I$ linear unabhängig.

Das wesentliche Resultat ist, dass die Menge der Lösungen des Systems (18) einen n-dimensionalen Vektorraum bildet.

Betrachtet man n beliebige Lösungen $u_1, \ldots, u_n$ von(18) und die entsprechende Matrixlösung $U := (u_1, \ldots, u_n)$, dann nennt man $W := \det(U)$ die *Wronski-Determinante* von $u_1, \ldots, u_n$. Um das nachfolgende Resultat für die Wronski-Determinante beweisen zu können, brauchen wir einige Fakten aus der linearen Algebra. Ohne Beweis werden wir das Multiplikationstheorem für die Determinante verwenden:

$$\det(AB) = \det(A)\det(B)\,, \qquad \forall\, A, B \in \mathbb{R}^{n\times n}\,.$$

Daraus folgt für invertierbares A, dass

$$1 = \det(\mathbb{I}) = \det(A)\det(A^{-1})\,,$$

und weiter die Invarianz der Determinante bezüglich Ähnlichkeitstransformationen:

$$\det(R^{-1}AR) = \det(R^{-1})\det(A)\det(R) = \det(A) = \prod_{i=1}^{n}\lambda_i\,,$$

wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A sind, wobei mehrfache Eigenwerte entsprechend ihrer algebraischen Vielfachheit mehrfach vorkommen. Die letzte Darstellung ist die Determinante der Jordanschen Normalform.

Weiters brauchen wir die *Spur* (engl. *trace*)

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}\,, \qquad A = (A_{ij})_{1\le i,j\le n}\,,$$

mit der Eigenschaft

$$\mathrm{tr}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} B_{ji}A_{ij} = \mathrm{tr}(BA)\,,$$

die daher ebenfalls ähnlichkeitsinvariant ist:

$$\mathrm{tr}(R^{-1}AR) = \mathrm{tr}(ARR^{-1}) = \mathrm{tr}(A) = \sum_{i=1}^{n}\lambda_i\,,$$

wieder aus der Jordanschen Normalform berechnet.

Lemma 7 (Liouvillesche Formel) *Für die Wronski-Determinante von Lösungen des Systems (18) gilt*

$$W(t) = W(s)\exp\left(\int_s^t \mathrm{tr}(A(\tau))d\tau\right)\,.$$

Beweis: Es gilt $U(t+h) = \Pi(t+h;t)U(t)$ und daher $W(t+h) = \det(\Pi(t+h;t))W(t)$. Wegen

$$\frac{d}{dt}\Pi(t;s) = A(t)\Pi(t;s)\,, \qquad \Pi(s,s) = \mathbb{I}\,,$$

folgt mit Taylorentwicklung $\Pi(t+h;t) = \mathbb{I} + hA(t) + O(h^2)$ für $h \to 0$. Ist $J(t) = R(t)^{-1}A(t)R(t)$ die Jordansche Normalform von $A(t)$, dann gilt

$$\begin{aligned}\det(\Pi(t+h;t)) &= \det(\mathbb{I}+hA) + O(h^2) = \det(R^{-1}(\mathbb{I}+hA)R) + O(h^2) = \det(\mathbb{I}+hJ) \\ &= \prod_{i=1}^{n}(1+h\lambda_i) + O(h^2) = 1 + h\sum_{i=1}^{n}\lambda_i + O(h^2) = 1 + h\mathrm{tr}(A) + O(h^2)\,,\end{aligned}$$

und daher

$$W(t+h) = W(t) + hA(t)W(t) + O(h^2)\,,$$

woraus folgt, dass die Wronski-Determinante die Differentialgleichung

$$\dot{W} = \mathrm{tr}(A)W$$

löst. ∎

Die Wronski-Determinante von $u_1, \ldots, u_n$ ist daher entweder identisch Null oder überall verschieden von Null. Im zweiten Fall nennt man $U = (u_1, \ldots, u_n)$ eine Fundamentalmatrixlösung von (18). Aus einer solchen erhält man eine Hauptmatrixlösung durch

$$\Pi(t;s) = U(t)U(s)^{-1}\,.$$

Schließlich noch inhomogene Systeme:

$$\dot{x} = A(t)x + g(t)\,. \tag{19}$$

Angenommen, wir kenne eine Hauptmatrixlösung $\Pi(t;t_0)$. Dann setzen wir den Variation-der-Konstanten-Ansatz $x(t) = \Pi(t;t_0)c(t)$, $c(t) \in \mathbb{R}^n$ ein:

$$A\Pi c + \Pi\dot{c} = A\Pi c + g\,,$$

und daher

$$\dot{c} = \Pi^{-1}g\,.$$

Die Unbekannte c kann also durch Integration bestimmt werden.

$$c(t) = c(t_0) + \int_{t_0}^{t}\Pi(s;t_0)^{-1}g(s)ds = c(t_0) + \int_{t_0}^{t}\Pi(t_0;s)g(s)ds\,.$$

Das beweist:

Lemma 8 *Die Lösung des Anfangswertproblems $x(t_0) = x_0$ für die Gleichung (19) ist gegeben durch*

$$x(t) = \Pi(t;t_0)x_0 + \int_{t_0}^{t}\Pi(t;s)g(s)ds\,.$$

Man vergleiche mit Abschnitt 3.1. Für konstante Koeffizientenmatrizen A gilt $\Pi(t;t_0) = e^{A(t-t_0)}$.

Im Allgemeinen kann man keine expliziten Formeln für $\Pi(t;s)$ erwarten. Kennt man allerdings durch eine glückliche Fügung eine Lösung $X_1(t) \neq 0$ des homogenen Systems $\dot{x} = A(t)x$, dann

kann das Problem durch *Reduktion der Ordnung* etwas leichter gemacht werden. Für skalare Gleichungen höherer Ordnung wurde diese Methode schon in Abschnitt 1.4 vorgestellt. Da X_1 nicht verschwindet, kann es zu jedem Zeitpunkt t zu einer Basis $\{X_1(t), \ldots, X_n(t)\}$ des $\mathbb{R}^n$ ergänzt werden, wobei wir außer der linearen Unabhängigkeit auch stetige Differenzierbarkeit von $X_2, \ldots, X_n$ annehmen, sonst allerdings keinen Zusammenhang mit dem Differentialgleichungssystem. In vielen Fällen ist es möglich, $X_2, \ldots, X_n$ zeitunabhängig zu wählen, z.B. wenn eine bestimmte Komponente von X_1 nie Null wird (siehe [T, S. 84]).

Nun machen wir die Koordinatentransformation $x(t) = X(t)y(t)$ mit $X = (X_1, \ldots, X_n)$. Für die folgende Rechnung benötigen wir die Formel

$$\frac{d}{dt}X^{-1} = -X^{-1}\dot{X}X^{-1}\,,$$

die aus der Ableitung

$$\dot{X}X^{-1} + X\frac{d}{dt}X^{-1} = 0$$

der Gleichung $XX^{-1} = \mathbb{I}$ folgt. Damit haben wir

$$\begin{aligned}\dot{y} &= X^{-1}\dot{x} + \left(\frac{d}{dt}X^{-1}\right)x = X^{-1}\left(Ax - \dot{X}X^{-1}x\right) = X^{-1}\left(AX - \dot{X}\right)y \\ &= X^{-1}\left(0, AX_2 - \dot{X}_2, \ldots, AX_n - \dot{X}_n\right)y = \begin{pmatrix} b^{tr} \\ \hat{A}\end{pmatrix}\hat{y}\,,\end{aligned}$$

mit $\hat{y} = (y_2, \ldots, y_n)$, wobei $b \in \mathbb{R}^{n-1}$ und $\hat{A} \in \mathbb{R}^{(n-1)\times(n-1)}$. Es kann also zunächst das System $\dot{\hat{y}} = \hat{A}\hat{y}$ der Größe $(n-1)$ gelöst und dann y_1 durch Integration von $\dot{y}_1 = b \cdot \hat{y}$ ermittelt werden.

Für den Fall, dass man $k < n$ linear unabhängige Lösungen $X_1, \ldots, X_k$ kennt, kann die Dimension auf $n - k$ reduziert werden, wie die obige Rechnung zeigt.

3.6 Lineare Systeme mit periodischen Koeffizienten

Der Matrix-Logarithmus: (siehe auch [T, Abschnitt 3.8])
Wann gibt es für eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$ eine andere quadratische Matrix B, sodass $A = e^B$? Fangen wir langsam mit dem skalaren Fall $n = 1$ an. Wenn B reell sein soll, dann ist die Antwort natürlich, dass $A > 0$ gelten muss und $B = \log(A)$. Darf B auch komplex sein, dann muss nur $A \neq 0$ gelten, und

$$B_k = \begin{cases} \log(A) + 2k\pi i\,, & A > 0\,, \\ \log(-A) + (2k+1)\pi i\,, & A < 0\,, \end{cases}$$

ist für jedes $k \in \mathbb{Z}$ eine Lösung, weil die Exponentialfunktion die imaginäre Periode $2\pi i$ besitzt. Die Lösung mit $k = 0$ heißt *Hauptzweig des Logarithmus* und wird geschrieben als $B_0 = \log(A)$. Dieser ist auch für komplexe $A \in \mathbb{C}$ definiert durch

$$\log(A) = \log|A| + i\arg(A)\,, \qquad \text{mit } -\pi < \arg(A) \leq \pi\,.$$

Für $|A - 1| < 1$ kann er auch durch die Taylorreihe

$$\log(A) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j}(A-1)^j$$

berechnet werden. Diese kann man bestimmen, indem man in der Gleichung $e^{\log(A)} = A$ einen Ansatz für die Taylorreihe des Logarithmus in die Taylorreihe für die Exponentialfunktion einsetzt und dann einen Koeffizientenvergleich durchführt. Das bedeutet aber, dass man auch für komplexe Matrizen $A \in \mathbb{C}^{n\times n}$ mit $\|A - \mathbb{I}\| < 1$ durch

$$\log(A) := \sum_{j=1}^{\infty} \frac{(-1)^{j+1}}{j} (A - \mathbb{I})^j$$

einen Matrix-Logarithmus definieren kann, der $e^{\log(A)} = A$ erfüllt, und den wir wieder den Hauptzweig nennen.

Für allgemeinere Matrizen können wir uns auf Jordanblöcke berschränken, weil mit $R^{-1}AR = J$ auch

$$J = R^{-1} e^B R = e^{R^{-1}BR}$$

gelten muss. Finden wir als eine Lösung $\hat{B}$ von $J = e^{\hat{B}}$, dann haben wir mit $B = R\hat{B}R^{-1}$ einen Logarithmus von A gefunden. Sei also $J_i = \lambda_i \mathbb{I} + N$ ein Jordanblock der Größe k zum Eigenwert $\lambda_i \neq 0$. Wir machen den Ansatz

$$\hat{B}_i = \log(\lambda_i)\mathbb{I} + \tilde{B}\,.$$

Das ergibt $J_i = \lambda_i e^{\tilde{B}}$ und daher

$$\mathbb{I} + \frac{1}{\lambda_i} N = e^{\tilde{B}}\,.$$

Wegen der Nilpotenz von N kann diese Gleichung mit Hilfe der Taylorreihe gelöst werden:

$$\hat{B}_i = \log(\lambda_i)\mathbb{I} + \sum_{j=1}^{k-1} \frac{(-1)^{j+1}}{j\lambda_i^j} N^j$$

Lemma 9 *Sei $A \in \mathbb{C}^{n\times n}$ und* $\det(A) \neq 0$. *Dann gilt mit obiger Notation für den* Hauptzweig des Matrix-Logarithmus,

$$\log(A) := R\,\mathrm{diag}(\hat{B}_1, \ldots, \hat{B}_m) R^{-1}\,,$$

die Gleichung $e^{\log(A)} = A$.

Natürlich ist $\log(A)$ im Allgemeinen nicht die eindeutige Lösung der Gleichung $A = e^B$, die im Allgemeinen keine reelle Lösung hat, selbst wenn A reell ist.

Wann besitzt eine reelle Matrix einen reellen Logarithmus? $\longrightarrow$ [T, S. 108]

Systeme mit periodischen Koeffizienten: $\longrightarrow$ [T, S. 91]

5 Randwertprobleme

5.1 Einleitung – Wärmeleitung

Wir betrachten das Problem, wie sich die Temperaturverteilung in einem dünnen Stab mit der Zeit verändert. Dabei sei der Stab der Länge L repräsentiert durch das x-Intervall $[0, L]$, und die Temperatur an der Stelle x zum Zeitpunkt t sei gegeben durch $u(x, t)$. Wir erlauben die Möglichkeit, dass der Stab inhomogen ist, das heißt, dass er x-abhängige Materialeigenschaften

besitzt wie zum Beispiel seine *spezifische Wärme* $r(x) > 0$, $0 \le x \le L$. Die zum Zeitpunkt t im Abschnitt $[a, b] \subset [0, L]$ enthaltene *Wärmeenergie* ist dann gegeben durch

$$\int_a^b r(x)u(x,t)dx\,.$$

Wir nehmen an, dass sich diese durch zwei Effekte verändern kann:

- *Wärmeleitung* im Stab, wobei Wärmeenergie durch die Querschnitte bei $x = a$ und $x = b$ transportiert wird, und

- *Wärmezufuhr oder -abfuhr*, z.B. durch den Mantel des Stabes oder durch eine im Stab stattfindende chemische Reaktion.

Diese Annahmen motivieren eine Gleichung der Form

$$\frac{d}{dt}\int_a^b r(x)u(x,t)dx = J(a,t) - J(b,T) + \int_a^b f(x,t)dx\,,$$

mit dem *Wärmefluss* $J(x,t)$ zum Zeitpunkt t durch den Querschnitt x, der positiv (negativ) ist, wenn Wärme in die positive (negative) x-Richtung fließt. Die Wärmezu-/abfuhr wird durch die Dichtefunktion f beschrieben. Mit dem Hauptsatz der Differential- und Integralrechnung kann die obige Gleichung in der Form

$$\int_a^b \left(r\partial_t u + \partial_x J - f\right) dx = 0$$

geschrieben werden. Nimmt man an, dass der Integrand stetig ist, und berücksichtigt, dass a und b beliebig sind, dann folgt die Gleichung

$$r\partial_t u + \partial_x J = f\,. \tag{20}$$

Das gängigste Modell für den Wärmetransport ist das *Fouriersche Gesetz,* das besagt, dass der Wärmefluss in die Richtung zur kleineren Temperatur gerichtet und proportional zur Ortsableitung der Temperatur ist. Das ergibt

$$J(x,t) = -p(x)\partial_x u(x,t)\,,$$

mit der *thermischen Leitfähigkeit* $p(x) > 0$, $0 \le x \le L$. Schließlich nehmen wir noch für die Wärmezu/abfuhr die Form

$$f(x,t) = -q(x)u(x,t) + h(x)$$

mit einer gegebenen zeitunabhängigen Dichte $h(x)$ von Wärmequellen bzw. -senken, sowie $q(x) \ge 0$, $0 \le x \le L$. Die Idee des Termes $-qu$ ist, dass es einen stabilisierenden Effekt gibt, der Wärme zu-(ab-)führt, wenn die Temperatur unter (über) dem Equilibriumswert $u = 0$ liegt. Mit diesen Annahmen wird aus (20) die *eindimensionale Wärmeleitungsgleichung*

$$r\partial_t u = \partial_x(p\partial_x u) - qu + h\,, \tag{21}$$

eine *partielle Differentialgleichung* für die Temperaturverteilung $u(x,t)$. Oft wird die Bezeichnung 'Wärmeleitungsgleichung' auch für den einfachen Fall $r = p = 1$, $q = h = 0$ reserviert.

Nun müssen wir uns noch Gedanken darüber machen, wie der Stab an seinen Enden mit der Umgebung interagiert, d.h. Wärmeenergie austauscht. Eine Idee wäre, dass die Enden Wärmebädern ausgesetzt sind, die sie auf festen Temperaturen halten. Das führt auf sogenannte *Dirichlet-Randbedingungen* der Form

$$u(0,t) = u_0\,, \qquad u(L,t) = u_L\,. \tag{22}$$

Eine andere Möglichkeit sind isolierte Enden, an denen der Wärmefluss verschwindet. Das ergibt sogenannte homogene *Neumann-Randbedingungen,*

$$\partial_x u(0,t) = \partial_x u(L,t) = 0\,. \tag{23}$$

Realistischer als diese beiden Idealisierungen sind die *Abkühlungsbedingungen* (auch *Robin-Randbedingungen*)

$$-p(0)\partial_x u(0,t) = \alpha(u_0 - u(0,t))\,, \qquad -p(L)\partial_x u(L,t) = \beta(u(L,t) - u_L)\,, \tag{24}$$

also eine Wärmefluss durch die Enden, der proportional (mit den Proportionalitätskonstanten $\alpha, \beta > 0$) zum Unterschied zwischen der Temperatur des Stabes und der Umgebungstemparatur (u_0 bzw. u_L) ist. Man beachte, dass es sich um eine Art Kompromiss zwischen Dirichlet- ($\alpha, \beta \to \infty$) und Neumann-Randbedingungen ($\alpha, \beta \to 0$) handelt. Natürlich könnten an den beiden Enden auch verschiedene Randbedingungen gestellt werden.

Eine naheliegende Fragestellung ist die nach einer stationären Temperaturverteilung, d.h. nach einer zeitunabhängigen Lösung $U(x)$. Diese löst ein *Randwertproblem,* bestehend aus der stationären Gleichung

$$0 = (pU')' - qU + h\,, \tag{25}$$

und einem Paar von Randbedingungen (22), (23) oder (24). In den folgenden Abschnitten werden wir uns zunächst mit diesem Problem, dann aber auch mit dem vollen zeitabhängigen Problem beschäftigen.

5.2 Das Dirichlet-Randwertproblem

Wir betrachten das Randwertproblem

$$0 = (pU')' - qU + h \quad \text{in } [0,L]\,, \qquad U(0) = u_0\,, \quad U(L) = u_L\,. \tag{26}$$

Mit $p \in C^1([0,L])$, $q, h \in C([0,L])$, $p > 0$, $q \geq 0$, $u_0, u_L \in \mathbb{R}$. Man beachte, dass daraus die Existenz von $\underline{p} > 0$ mit $p \geq \underline{p}$ folgt.

Zunächst merken wir an, dass zumindest die Anzahl der Gleichungen stimmt, weil die Differentialgleichung zweiter Ordnung ist, und daher zwei Zusatzbedingungen die Lösung eindeutig machen sollten.

Lemma 10 *Das Problem (26) besitzt höchstens eine Lösung.*

Beweis: Seien U, V zwei Lösungen. Dann löst die Differenz $W := U - V$ das homogene Problem:

$$0 = (pW')' - qW \quad \text{in } [0,L]\,, \qquad W(0) = W(L) = 0\,.$$

Nun multiplizieren wir die Differentialgleichung mit $-W$ und integrieren über $[0, L]$:

$$0 = -\int_0^L W(pW')'dx + \int_0^L qW^2 dx \geq \int_0^L p(W')^2 dx \geq \underline{p} \int_0^L (W')^2 dx\,.$$

Daraus folgt $W' \equiv 0$ und daher, wegen der Randbedingungen, auch $W \equiv 0$. ∎

Im Fall $u_0, u_L, h \geq 0$ würden wir auch $U \geq 0$ erwarten. Kann man das beweisen?

Lemma 11 (Schwaches Maximumprinzip) *Sei $U \in C^2([0, L])$, $\mathcal{L}(U) := (pU')' - qU \leq 0$ und $U(0), U(L) \geq 0$. Dann gilt $U(x) \geq 0$ in $[0, L]$. Das Resultat gilt auch mit umgekehrten Ungleichungen.*

Beweis: Machen wir zunächst die stärkere Annahme $\mathcal{L}(U) < 0$. Angenommen U nimmt negative Werte an. Dann muss es ein negatives Minimum $U(x_0) < 0$ mit $0 < x_0 < L$ geben. Daher $U'(x_0) = 0$, $U''(x_0) \geq 0$. Das ergibt den Widerspruch

$$\mathcal{L}(U)\,\big|_{x=x_0} = p(x_0)U''(x_0) + p'(x_0)U'(x_0) - q(x_0)U(x_0) \geq 0\,.$$

Nun wählen wir $\gamma \in \mathbb{R}$ groß genug, dass

$$\mathcal{L}(e^{\gamma L} - e^{\gamma x}) = -e^{\gamma x}\left(p\gamma^2 + p'\gamma\right) - q\left(e^{\gamma L} - e^{\gamma x}\right) \leq -e^{\gamma x}\left(p\gamma^2 + p'\gamma\right) < 0\,.$$

Dann folgt mit $\varepsilon > 0$ aus $\mathcal{L}(U) \leq 0$, dass $\mathcal{L}(U + \varepsilon(e^{\gamma L} - e^{\gamma x})) < 0$ und $U(x) + \varepsilon(e^{\gamma L} - e^{\gamma x}) \geq 0$ an $x = 0, L$. Daraus folgt $U \geq \varepsilon(e^{\gamma x} - e^{\gamma L})$, und daher mit $\varepsilon \to 0$ auch $U \geq 0$. ∎

Definition 1 *Seien $\overline{U}, \underline{U} \in C^2([0, L])$, und es gelte*

$$\mathcal{L}(\overline{U}) + h \leq 0\,, \qquad \overline{U}(0) \geq u_0\,, \quad \overline{U}(L) \geq u_L\,,$$

$$\mathcal{L}(\underline{U}) + h \geq 0\,, \qquad \underline{U}(0) \leq u_0\,, \quad \underline{U}(L) \leq u_L\,.$$

Dann nennt man $\overline{U}$ eine <u>obere Lösung</u> und $\underline{U}$ eine <u>untere Lösung</u> von (26).

Lemma 12 *Seien U, $\overline{U}$ und $\underline{U}$ eine Lösung, eine obere Lösung bzw. eine untere Lösung von (26). Dann gilt*

$$\underline{U} \leq U \leq \overline{U} \qquad \text{in } [0, L]\,.$$

Beweis: Anwendung des schwachen Maximumsprinzips auf $U - \underline{U}$ sowie $\overline{U} - U$. ∎

Obere und untere Lösungen zu finden, ist gewissermaßen eine mathematische Bastelarbeit. Eine Idee ist, die gewünschte Ungleichung mit Hilfe des Terms $(pU')'$ zu erzwingen. Wir suchen daher zunächst eine Funktion U_1, die

$$(pU_1')' = -1\,, \qquad U_1(0) = U_1(L) = 0\,,$$

erfüllt. Integration ergibt $pU_1'(x) = x_0 - x$ und

$$U_1(x) = \int_0^x \frac{x_0 - y}{p(y)} dy\,,$$

wobei x_0 so gewählt wird, dass die rechte Randbedingung erfüllt ist:

$$x_0 = \int_0^L \frac{y\,dy}{p(y)} \left(\int_0^L \frac{dy}{p(y)} \right)^{-1} \in (0, L)\,.$$

Man sieht leicht, dass $U_1(x) \geq 0$ gilt. Behauptung: $\overline{U}(x) := \overline{a}U_1(x) + u_0 + x(u_L - u_0)/L$ und $\underline{U}(x) := -\underline{a}U_1(x) + u_0 + x(u_L - u_0)/L$ sind für $\overline{a}, \underline{a} \geq 0$ groß genug eine obere bzw. untere Lösung. Es gilt

$$\begin{aligned} \mathcal{L}(\overline{U}) + h &= -\overline{a} - \overline{a}qU_1 + p' \frac{u_L - u_0}{L} - q\left(u_0 + \frac{x}{L}(u_L - u_0)\right) + h \\ &\leq -\overline{a} + \max_{[0,L]} \left(p' \frac{u_L - u_0}{L} - q\left(u_0 + \frac{x}{L}(u_L - u_0)\right) + h \right)\,, \end{aligned}$$

und analog

$$\mathcal{L}(\underline{U}) + h \geq \underline{a} + \min_{[0,L]} \left(p' \frac{u_L - u_0}{L} - q\left(u_0 + \frac{x}{L}(u_L - u_0)\right) + h \right)\,,$$

was die Behauptung beweist.

Um das Randwertproblem (26) zu lösen definieren wir zunächst eine Lösungsbasis $\{c, s\}$ für die homogene Differentialgleichung durch die Anfangswertprobleme

$$\mathcal{L}(c) = 0\,, \qquad c(0) = 1\,, \quad p(0)c'(0) = 0\,,$$

$$\mathcal{L}(s) = 0\,, \qquad s(0) = 0\,, \quad p(0)s'(0) = 1\,.$$

Es gilt

$$s(L) \neq 0\,,$$

weil s sonst eine nichttriviale Lösung der homogenen Version von (26) wäre, was wegen des Eindeutigkeitsresultats Lemma 10 unmöglich ist.

Die Kernfunktion in Lemma 5 ist gegeben durch $u(x) = s(x)/p(0)$, und die allgemeine Lösung der Differentialgleichung in (26) ist

$$U(x) = ac(x) + bs(x) - \frac{1}{p(0)} \int_0^x s(x - y)h(y)dy\,, \qquad a, b \in \mathbb{R}\,.$$

Einsetzen in die Randbedingungen ergibt

$$a = u_0\,, \qquad ac(L) + bs(L) - \frac{1}{p(0)} \int_0^L s(L - y)h(y)dy = u_L\,,$$

wodurch a und b eindeutig bestimmt sind. Damit haben wir folgendes Resultat bewiesen:

Satz 10 *Das stationäre Wärmeleitungsproblem mit Dirichlet-Randbedingungen (26) besitzt eine eindeutige Lösung.*

5.3 Das Dirichlet-Randwertproblem für die Wärmeleitungsgleichung (mit konstanten Koeffizienten)

Wir kehren zurück zur zeitabhängigen Gleichung (21) und betrachten zunächst das homogene Dirichlet-Problem für den einfachsten Fall $r = p = 1$, $q = h = 0$, d.h.

$$\partial_t u = \partial_x^2 u\,, \quad \text{in } (0,L)\times(0,\infty)\,, \qquad u(0,t) = u(L,t) = 0\,, \quad t > 0\,.$$

Um die Dynamik der Temperaturverteilung eindeutig zu bestimmen ist wohl auch die Angabe einer Anfangsverteilung notwendig,

$$u(x,0) = u_A(x)\,, \qquad 0 < x < L\,,$$

womit die Formulierung eines *Anfangs-Randwertproblems* vervollständigt wird. Als SpezialistInnen für gewöhnliche Differentialgleichungen versuchen wir zunächst, spezielle Lösungen der partiellen Differentialgleichung zu finden, wobei wir nur Funktionen von einer Veränderlichen verwenden.

Eine erste Idee besteht darin, eine Invarianz der Gleichung zu benützen: Führt man neue Koordinaten $\tau = \alpha^2 t$, $y = \alpha x$ mit $\alpha > 0$ bzw. $U(y,\tau) := u(y/\alpha, \tau/\alpha^2)$, bleibt die Gleichung unverändert, d.h. $\partial_\tau U = \partial_y^2 U$. Wählt man $\alpha = 1/\sqrt{t_0}$, dann ergibt sich $u(x,t_0) = U(x/\sqrt{t_0}, 1)$. Das legt nahe, dass die Wärmeleitungsgleichung Lösungen der Form $u(x,t) = U(\xi)$, $\xi = x/\sqrt{t}$, besitzt. Setzt man diesen Ansatz in die Gleichung ein, dann ergibt sich

$$-\frac{x}{2t^{3/2}}U' = \partial_t u = \partial_x^2 u = \frac{1}{t}U'' \quad \Longrightarrow \quad U'' + \frac{\xi}{2}U' = 0\,.$$

Es hat also funktioniert: Zu lösen ist eine gewöhnliche Differentialgleichung. Wir erhalten

$$U'(\xi) = e^{-\xi^2/4}c\,, \qquad c \in \mathbb{R}\,.$$

Das zeigt, dass diese Lösungen monoton als Funktionen von x sind, woraus wieder folgt, dass sie nicht in der Lage sind, die Dirichlet-Randbedingungen zu erfüllen.

Deshalb versuchen wir noch etwas anderes, und zwar *Separation der Variablen,* d.h. den Ansatz $u(x,t) = \varphi(x)\psi(t)$. Wir fordern, dass die Randbedingungen für u erfüllt sind, d.h. $\varphi(0) = \varphi(L) = 0$. Einsetzen in die Differentialgleichung und Division durch $\varphi\psi$ ergibt

$$\frac{\varphi''(x)}{\varphi(x)} = \frac{\psi'(t)}{\psi(t)}\,.$$

Da die linke Seite eine Funktion von x und die rechte Seite eine Funktion von t ist, folgt aus der Gleichung, dass beide Seiten gleich derselben Konstanten λ sein müssen. Damit ergibt sich zur Bestimmung von φ das *Eigenwertproblem*

$$\varphi'' = \lambda\varphi\,, \qquad \varphi(0) = \varphi(L) = 0\,.$$

Nichttriviale Lösungen existieren nur für bestimmte Werte von λ, die sogenannten *Eigenwerte.* Entsprechende nichttriviale Lösungen φ heißen *Eigenfunktionen.* Wie bei Eigenwertproblemen für Matrizen muss man auch hier mit komplexen Eigenwerten und Eigenfunktionen rechnen. Allerdings: Wir multiplizieren die Gleichung mit der konjugiert Komplexen $\overline{\varphi}$ und integrieren bzgl. x:

$$\lambda\int_0^L |\varphi(x)|^2 dx = \int_0^L \varphi''\overline{\varphi}\,dx = -\int_0^L |\varphi'(x)|^2 dx\,,$$

wobei bei der partiellen Integration die Randterme wegen der Randbedingungen verschwinden. Das zeigt, dass es nur negative reelle Eigenwerte geben kann. Setzen wir daher $\lambda = -\omega^2$, $\omega > 0$, und verwenden die Randbedingung an $x = 0$, dann folgt

$$\varphi(x) = c \sin(\omega x) .$$

Eine nichttriviale Lösung ($c \neq 0$) erfüllt die Randbedingung an $x = L$, wenn $\omega L = k\pi$, $k \in \mathbb{N}$. Es gibt also abzählbar viele Eigenwerte:

$$\lambda_k = -\left(\frac{k\pi}{L}\right)^2 , \qquad \varphi_k(x) = \sin \frac{k\pi x}{L} , \qquad k \in \mathbb{N} .$$

Da die Eigenwerte alle negativ sind, klingen die entsprechenden zeitabhängigen Anteile (Lösungen von $\psi' = \lambda\psi$)

$$\psi_k(t) = e^{\lambda_k t}$$

für $t \to \infty$ exponentiell ab.

Wir versuchen nun, die Lösung des Anfangs-Randwertproblems als Linearkombination der berechneten Produktlösungen zu bestimmen:

$$u(x,t) = \sum_{k=1}^{\infty} u_{Ak} e^{\lambda_k t} \sin \frac{k\pi x}{L} \tag{27}$$

Einsetzen in die Anfangsbedingung liefert

$$u_A(x) = \sum_{k=1}^{\infty} u_{Ak} \sin \frac{k\pi x}{L} .$$

Die u_{Ak} müssen also als Fourierkoeffizienten in der Sinus-Fourierreihe von u_0 gewählt werden:

$$u_{Ak} = \frac{2}{L} \int_0^L u_0(x) \sin \frac{k\pi x}{L} \, dx .$$

Diese Formel ist so einfach, weil die Eigenfunktionen bezüglich des Skalarproduktes

$$\langle u, v \rangle := \int_0^L u(x) \overline{v(x)} \, dx \tag{28}$$

paarweise orthogonal sind. Diese Eigenschaft und die Tatsache, dass die Eigenwerte reell sind, erinnert an symmetrische Matrizen. Tatsächlich ist der Operator $\mathcal{L} := \partial_x^2$ auch symmetrisch, wenn man ihn auf Funktionen anwendet, die die homogenen Dirichlet-Randbedingungen erfüllen:

$$\langle \mathcal{L}u, v \rangle = \int_0^L u'' v \, dx = -\int_0^L u' v' dx = \langle u, \mathcal{L}v \rangle .$$

5.4 Eigenwertprobleme für symmetrische kompakte Operatoren

Wir wollen einge der Überlegungen aus dem vorigen Abschnitt in einen funktionalanalytischen Rahmen bringen. Zunächst beobachten wir, dass durch (28) ein Skalarprodukt auf dem Raum $C([0,L])$ definiert wird, d.h. $\langle\cdot,\cdot\rangle$ ist bilinear, hermitesch und definit. Dieser Raum hat den Nachteil, dass er bezüglich der induzierten Norm $\|u\| = \sqrt{\langle u,u\rangle}$ (bzw. der induzierten Metrik $d(u,v) = \|u-v\|$) nicht vollständig ist. Vervollständigung (wie bei der Konstruktion von $\mathbb{R}$ aus $\mathbb{Q}$) liefert den *Hilbertraum* $L^2(0,L)$. Wie wir später noch beweisen werden, bilden die berechneten Eigenfunktionen ein *vollständiges Orthogonalsystem*. Das bedeutet, dass jedes Element von $L^2(0,L)$ durch seine Fourierreihe dargestellt werden kann. Daraus folgt, dass wir das Anfangs-Randwertproblem für beliebige Anfangsdaten $u_0 \in L^2(0,L)$ lösen können.

Definition 2 *Sei $\left(\mathcal{H},\langle .,.\rangle\right)$ ein Hilbertraum und $\mathcal{L}:\ D(\mathcal{L}) \subset \mathcal{H} \to \mathcal{H}$ ein linearer Operator. Man bezeichnet $\mathcal{L}$ als symmetrisch, wenn*

$$\langle \mathcal{L}u, v\rangle = \langle u, \mathcal{L}v\rangle \qquad \forall u,v \in D(\mathcal{L})$$

gilt.

Satz 11 *Sei (λ,φ) mit $\varphi \in D(\mathcal{L})$ ein Eigenwert-Eigenvektor-Paar zum symmetrischen Operator $\mathcal{L}$, d.h. $\mathcal{L}\varphi = \lambda\varphi$. Dann gilt $\lambda \in \mathbb{R}$. Die Eigenvektoren zu verschiedenen Eigenwerten von $\mathcal{L}$ sind orthogonal.*

Beweis: Es gilt

$$\lambda\,\langle\varphi,\varphi\rangle = \langle\mathcal{L}\varphi,\varphi\rangle = \langle\varphi,\mathcal{L}\varphi\rangle = \overline{\lambda}\,\langle\varphi,\varphi\rangle\,.$$

Daraus folgt $\lambda = \overline{\lambda}$ und damit $\lambda \in \mathbb{R}$.

Seien (λ,φ) und (μ,ψ) Eigenwert-Eigenvektor-Paare mit $\lambda \neq \mu$. Dann gilt

$$\lambda\,\langle\varphi,\psi\rangle = \langle\mathcal{L}\varphi,\psi\rangle = \langle\varphi,\mathcal{L}\psi\rangle = \mu\,\langle\varphi,\psi\rangle\,.$$

Daraus folgt $(\lambda-\mu)\,\langle\varphi,\psi\rangle = 0$ und damit die Orthogonalität der Eigenvektoren. ∎

Auf die Frage, ob die Eigenvektoren ein vollständiges Orthogonalsystem des Hilbertraumes bilden, kann in dieser Allgemeinheit keine positive Antwort gegeben werden. Dazu sind zusätzliche Voraussetzungen an den Operator notwendig. Wir rufen den Begriff der *Kompaktheit* in Erinnerung. Er kann in metrischen Räumen auf verschiedene Art definiert werden. Wie schon im Abschnitt 2.6 ist auch hier *Folgenkompaktheit* die nützliche Definition. Folgenkompakte Mengen sind notwendigerweise abgeschlossen. Verzichtet man auf diese Eigenschaft, dann kommt man zum folgenden Begriff: Eine Teilmenge eines metrischen Raumes nennt man *präkompakt* (oder *relativ kompakt*), wenn ihr Abschluss (folgen-)kompakt ist.

Definition 3 *Der lineare Operator $\mathcal{K}: \mathcal{H} \to \mathcal{H}$ heißt* kompakt*, wenn er beschränkte Mengen auf präkompakte Mengen abbildet.*

Bemerkung 7 *a) In endlichdimensionalen Räumen ist jeder lineare Operator beschränkt und damit auch kompakt, weil beschränkte Mengen präkompakt sind.*
b) Jeder kompakte Operator ist beschränkt.

Für das Eigenwertproblem für symmetrische, kompakte Operatoren gilt der *Entwicklungssatz*:

Satz 12 *Sei $\mathcal{K}$ ein symmetrischer, kompakter Operator, für den Null kein Eigenwert ist. Dann hat $\mathcal{K}$ höchstens abzählbar viele Eigenwerte μ_k, $k \in I$, die alle endliche Vielfachheit haben. In der Aufzählung kommt jeder Eigenwert seiner Vielfachheit entsprechend mehrfach vor. Gibt es unendlich viele Eigenwerte, dann bilden sie eine gegen 0 konvergente Folge. Weiters gibt es ein in $\mathcal{H}$ vollständiges Orthonormalsystem von Eigenvektoren φ_k, $k \in I$, und es gilt*

$$\mathcal{K}u = \sum_{k\in I} \mu_k \langle u, \varphi_k\rangle \varphi_k \,. \tag{29}$$

Für den Beweis des Satzes benötigen wir einige Hilfsresultate.

Lemma 13 *Die Menge der Eigenwerte eines symmetrischen, kompakten Operators kann nur 0 als Häufungspunkt haben.*

Beweis: Sei $\{\mu_n\}$ eine Folge von verschiedenen Eigenwerten, die gegen einen Wert $\mu_0 \neq 0$ konvergiert. Seien φ_n die zugehörigen normierten Eigenvektoren. Dann ist die Folge $\{\varphi_n/\mu_n\}$ für n groß genug beschränkt. Ihr Bild unter dem Operator ist die Folge $\{\varphi_n\}$. Da der Operator kompakt ist, muss diese Folge eine konvergente Teilfolge besitzen, was aber auf Grund der Orthogonalität der φ_n ein Widerspruch ist:

$$\|\varphi_n - \varphi_m\|_{\mathcal{H}}^2 = \langle \varphi_n, \varphi_n\rangle + \langle \varphi_m, \varphi_m\rangle = 2$$

■

Lemma 14 *Für die Norm*

$$\|\mathcal{K}\| := \sup_{\|u\|_{\mathcal{H}}=1} \|\mathcal{K}u\|_{\mathcal{H}}$$

eines symmetrischen, kompakten Operators $\mathcal{K}$ gilt

$$\|\mathcal{K}\| = \sup_{\|u\|_{\mathcal{H}}=1} |\langle \mathcal{K}u, u\rangle| \,.$$

Beweis: Für einen normierten Vektor $u \in \mathcal{H}$ gilt $|\langle \mathcal{K}u, u\rangle| \leq \|\mathcal{K}u\|_{\mathcal{H}} \leq \|\mathcal{K}\|$, und daher

$$\nu(\mathcal{K}) := \sup_{\|u\|_{\mathcal{H}}=1} |\langle \mathcal{K}u, u\rangle| \leq \|\mathcal{K}\| \,.$$

Für beliebige $x, y \in \mathcal{H}$ gilt

$$\begin{aligned} 4\langle \mathcal{K}x, y\rangle &= 2\langle \mathcal{K}x, y\rangle + 2\langle \mathcal{K}y, x\rangle = \langle \mathcal{K}(x+y), x+y\rangle - \langle \mathcal{K}(x-y), x-y\rangle \\ &\leq \nu(\mathcal{K})\big(\|x+y\|_{\mathcal{H}}^2 + \|x-y\|_{\mathcal{H}}^2\big) = 2\nu(\mathcal{K})\big(\|x\|_{\mathcal{H}}^2 + \|y\|_{\mathcal{H}}^2\big) \,, \end{aligned} \tag{30}$$

wobei die letzte Gleichung aus dem *Parallelogrammgesetz* folgt. Für einen normierten Vektor $u \in \mathcal{H}$ nehmen wir an $\mathcal{K}u \neq 0$ und setzen

$$x = u\sqrt{\|\mathcal{K}u\|_{\mathcal{H}}} \,, \qquad y = \frac{\mathcal{K}u}{\sqrt{\|\mathcal{K}u\|_{\mathcal{H}}}} \,.$$

Einsetzen in (30) liefert $4\|\mathcal{K}u\|_{\mathcal{H}}^2 \leq 4\nu(\mathcal{K})\|\mathcal{K}u\|_{\mathcal{H}}$ und daher $\|\mathcal{K}u\|_{\mathcal{H}} \leq \nu(\mathcal{K})$. Diese Ungleichung gilt natürlich auch für $\mathcal{K}u = 0$. Damit haben wir $\|\mathcal{K}\| \leq \nu(\mathcal{K})$ gezeigt, und der Beweis ist vollständig.

■

Lemma 15 *Sei $\{\varphi_k\}$, $k = 1, 2, \ldots$, ein vollständiges Orthonormalsystem in $(\mathcal{H}, \langle ., . \rangle)$ und $\{\alpha_k\}$, $k = 1, 2, \ldots$ eine Folge von reellen Zahlen. Dann konvergiert die Reihe*

$$\sum_{k=1}^{\infty} \alpha_k^2$$

genau dann, wenn die Reihe

$$\sum_{k=1}^{\infty} \alpha_k \varphi_k$$

gegen ein $u \in \mathcal{H}$ konvergiert. In diesem Fall gilt $\alpha_k = \langle u, \varphi_k \rangle$ und die Parsevalsche Gleichung

$$\|u\|_{\mathcal{H}}^2 = \sum_{k=1}^{\infty} \alpha_k^2 \,.$$

Beweis: Für die Partialsummen der betrachteten Reihen führen wir die Bezeichnungen

$$s_n = \sum_{k=1}^{n} \alpha_k \varphi_k \,, \qquad \sigma_n = \sum_{k=1}^{n} \alpha_k^2$$

ein. Dann gilt für $m > n$

$$\|s_m - s_n\|_{\mathcal{H}}^2 = \sum_{k=n+1}^{m} \alpha_k^2 = \sigma_m - \sigma_n \,,$$

woraus folgt, dass $\{s_n\}$ genau dann eine Cauchyfolge in $\mathcal{H}$ ist, wenn $\{\sigma_n\}$ eine Cauchyfolge in $\mathbb{R}$ ist. Damit ist die erste Aussage des Satzes bewiesen. Die weiteren Aussagen folgen direkt aus der Orthonormalität der φ_k. ∎

Beweis: (des Satzes): Aus Lemma 14 folgt, dass es eine Folge $\{u_n\}$ von normierten Vektoren und eine Zahl μ_1 mit $|\mu_1| = \|\mathcal{K}\|$ gibt, sodass $\langle \mathcal{K}u_n, u_n \rangle$ für $n \to \infty$ gegen μ_1 konvergiert. Aus

$$\|\mathcal{K}u_n - \mu_1 u_n\|_{\mathcal{H}}^2 = \|\mathcal{K}u_n\|_{\mathcal{H}}^2 - 2\mu_1 \langle \mathcal{K}u_n, u_n \rangle + \mu_1^2 \|u_n\|_{\mathcal{H}}^2 \leq 2\mu_1(\mu_1 - \langle \mathcal{K}u_n, u_n \rangle)$$

folgt daher

$$\lim_{n \to \infty} (\mathcal{K}u_n - \mu_1 u_n) = 0 \,. \tag{31}$$

Wegen der Kompaktheit von $\mathcal{K}$ hat die Folge $\{\mathcal{K}u_n\}$ eine konvergente Teilfolge $\{\mathcal{K}u_{n_k}\}$. Wegen (31) konvergiert auch die Folge $\{u_{n_k}\}$ gegen einen normierten Vektor φ_1, und es gilt $\mathcal{K}\varphi_1 = \mu_1 \varphi_1$, d.h. φ_1 ist ein Eigenvektor zum Eigenwert μ_1.

Nun gehen wir rekursiv vor. Angenommen, wir hätten paarweise orthogonale, normierte Eigenvektoren $\varphi_1, \ldots, \varphi_{k-1}$ und die entsprechenden Eigenwerte $\mu_1, \ldots, \mu_{k-1}$ konstruiert. Dann bezeichnen wir das orthogonale Komplement der linearen Hülle von $\{\varphi_1, \ldots, \varphi_{k-1}\}$ mit $\mathcal{H}_k$. Offensichtlich ist $\mathcal{H}_k$ ein Hilbertraum. Die Einschränkung des Operators $\mathcal{K}$ auf $\mathcal{H}_k$ nennen wir $\mathcal{K}_k$. Wegen $\mathcal{H}_k \subset \mathcal{H}_{k-1}$ gilt $\|\mathcal{K}_k\| \leq \|\mathcal{K}_{k-1}\|$. Der Operator $\mathcal{K}_k$ ist symmetrisch und kompakt auf $\mathcal{H}_k$. Wir können daher wie oben ein Eigenwert-Eigenvektor-Paar (μ_k, φ_k) mit $|\mu_k| = \|\mathcal{K}_k\|$ bestimmen. Auf

diese Art konstruieren wir eine Folge $\{\mu_k\}$ von Eigenwerten mit $|\mu_k| \leq |\mu_{k-1}|$ und eine orthonormale Folge $\{\varphi_k\}$ von zugehörigen Eigenvektoren.

Die Konstruktion bricht nach endlich vielen Schritten ab, wenn für ein k $\mathcal{H}_k = \{0\}$ gilt, d.h. wenn $\mathcal{H}$ endlichdimensional ist. In diesem Fall sind die Aussagen des Satzes erfüllt. Sei also im Weiteren die oben konstruierte Folge unendlich. Dann folgt aus Lemma 13, dass die Folge der Eigenwerte gegen 0 konvergiert. Das impliziert auch die endliche Vielfachheit der Eigenwerte.

Betrachten wir nun die Vektoren

$$v_n = u - \sum_{k=1}^{n} \langle u, \varphi_k \rangle \varphi_k \in \mathcal{H}_n \,.$$

Es gilt $\|\mathcal{K} v_n\|_{\mathcal{H}} = \|\mathcal{K}_n v_n\|_{\mathcal{H}} \leq |\mu_n| \|v_n\|_{\mathcal{H}}$ und

$$\|v_n\|_{\mathcal{H}}^2 = \|u\|_{\mathcal{H}}^2 - \sum_{k=1}^{n} \langle u, \varphi_k \rangle^2 \leq \|u\|_{\mathcal{H}}^2 \,. \tag{32}$$

Daraus folgt, dass $\mathcal{K} v_n$ gegen 0 konvergiert und daher

$$\mathcal{K} u = \sum_{k=1}^{\infty} \mu_k \langle u, \varphi_k \rangle \varphi_k \qquad \forall\, u \in \mathcal{H} \,.$$

Kämen in der konstruierten Folge nicht alle Eigenwerte von $\mathcal{K}$ entsprechend ihrer Vielfachheit vor, dann gäbe es eine Eigenfunktion φ, die auf alle φ_k, $k = 1, 2, \ldots$, normal steht. Dann folgt aber aus der obigen Gleichung $\mathcal{K}\varphi = 0$, d.h. 0 ist ein Eigenwert von $\mathcal{K}$, was im Widerspruch zu den Annahmen des Satzes steht.

Es bleibt noch, die Vollständigkeit des aus Eigenvektoren bestehenden Orthonoralsystems zu zeigen. Betrachten wir die oben definierte Folge $\{v_n\}$. Aus der Abschätzung (32) folgt die *Besselsche Ungleichung*

$$\sum_{k=1}^{n} \langle u, \varphi_k \rangle^2 \leq \|u\|_{\mathcal{H}}^2 \,,$$

woraus wegen Lemma 15 die Konvergenz von $\{v_n\}$ folgt. Wegen der Stetigkeit von $\mathcal{K}$ gilt für den Grenzwert v die Gleichung $\mathcal{K} v = 0$. Da 0 kein Eigenwert von $\mathcal{K}$ ist, folgt $v = 0$, was gleichbedeutend ist mit

$$u = \sum_{k=1}^{\infty} \langle u, \varphi_k \rangle \varphi_k \qquad \forall\, u \in \mathcal{H} \,.$$

Die Gleichung (29) folgt aus der Stetigkeit von $\mathcal{K}$. ∎

5.5 Das Dirichlet-Problem für die allgemeine Wärmeleitungsgleichung

Wir rufen uns die allgemeine Wärmeleitungsgleichung (21) in Erinnerung,

$$r \partial_t u = \partial_x (p \partial_x u) - q u + h \,,$$

und betrachten das Anfangs-Randwertproblem mit den Randbedingungen

$$u(0,t) = u_0\,, \qquad u(L,t) = u_L\,, \qquad t > 0\,,$$

und den Anfangsbedingungen

$$u(x,0) = u_A(x)\,, \qquad 0 < x < L\,.$$

An die Daten machen wir die Annahmen $r, q, h \in C([0,L])$, $p \in C^1([0,L])$, $r, p > 0$, $q \geq 0$, $u_0, u_L \in \mathbb{R}$, $u_A \in L^2(0,L)$. Weiters erinnern wir daran, dass das stationäre Problem eine eindeutige Lösung besitzt, die in der Form

$$U(x) = u_0 c(x) + \frac{s(x)}{s(L)}\left(u_L - u_0 c(L) + \frac{1}{p(0)}\int_0^L s(L-y)h(y)dy\right) - \frac{1}{p(0)}\int_0^x s(x-y)h(y)dy\,,$$

geschrieben werden kann, wobei $c, s \in C^2([0,L])$ die Anfangswertprobleme

$$(pc')' - qc = 0\,, \quad c(0) = 1\,,\, c'(0) = 0\,, \qquad (ps')' - qs = 0\,, \quad s(0) = 0\,,\, p(0)s'(0) = 1\,,$$

lösen. Als ersten Schritt zur Lösung des Anfangs-Randwertproblems führen wir die neue Unbekannte $v(x,t) := u(x,t) - U(x)$ ein, für die wir das Problem

$$\partial_t v = \mathcal{L}v := \frac{1}{r}\left(\partial_x(p\partial_x v) - qv\right)\,, \qquad v(0,t) = v(L,t) = 0\,, \tag{33}$$

mit den Anfangsbedingungen

$$v(x,0) = v_A(x) := u_A(x) - U(x)\,, \tag{34}$$

erhalten. Man beachte, dass jetzt Differentialgleichung und Randbedingungen homogen sind. Für den Operator $\mathcal{L}$ wählen wir einen Definitionsbereich, der die Randbedingungen enthält, d.h. $v \in D(\mathcal{L}) \quad \Rightarrow \quad v(0) = v(L) = 0$. Wir definieren ein gewichtetes Skalarprodukt und die induzierte Norm durch

$$\langle u, v\rangle_r := \int_0^L u(x)v(x)r(x)dx\,, \qquad \|u\|_r^2 = \langle u, u\rangle_r\,.$$

Diese Norm ist äquivalent zu der in (28) definierten L^2-Norm:

$$\min_{[0,L]} r\,\|v\| \leq \|v\|_r \leq \max_{[0,L]} r\,\|v\|\,,$$

woraus folgt, dass $(L^2(0,L), \langle\cdot,\cdot\rangle_r)$ ein Hilbertraum ist. Es gilt

$$\langle \mathcal{L}u, v\rangle_r = \int_0^L \left((pu')' - qu\right)v\,dx = -\int_0^L \left(pu'v' + quv\right)dx = \langle u, \mathcal{L}v\rangle_r\,, \tag{35}$$

für alle $u, v \in D(\mathcal{L})$, d.h. $\mathcal{L}$ ist symmetrisch. In der partiellen Integration wurden die Randbedingungen verwendet. Die Gleichung $\mathcal{L}v = f$ ist eine Version des stationären Problems mit $h = -rf$, $u_0 = u_L = 0$. Daraus folgt, dass $\mathcal{L}$ invertierbar ist, und mit $\mathcal{K} := \mathcal{L}^{-1}$ gilt

$$(\mathcal{K}f)(x) = -\frac{s(x)}{s(L)p(0)}\int_0^L s(L-y)r(y)f(y)dy + \frac{1}{p(0)}\int_0^x s(x-y)r(y)f(y)dy\,. \tag{36}$$

Mit Hilfe der Cauchy-Schwarzschen Ungleichung schätzen wir ab:

$$\sup_{(0,L)} |\mathcal{K}f| \leq \frac{\sup|s|/|s(L)|+1}{p(0)} \|s\|_r \|f\|_r \,. \tag{37}$$

Da die Supremum-Norm stärker ist als die L^2-Norm $\Big(\|v\| \leq \sqrt{L}\sup|v|\Big)$, zeigt diese Abschätzung, dass $\mathcal{K}$ als beschränkte Abbildung $\mathcal{K}: L^2(0,L) \to L^2(0,L)$ fortgesetzt werden kann. Für $f, g \in L^2(0,L)$ sei $u = \mathcal{K}f$, $v = \mathcal{K}g$. Dann gilt

$$\langle \mathcal{K}f, g\rangle_r = \langle u, \mathcal{L}v\rangle_r = \langle \mathcal{L}u, v\rangle_r = \langle f, \mathcal{K}g\rangle_r \,,$$

d.h. auch $\mathcal{K}$ ist symmetrisch. Differenzieren von (36) gibt

$$(\mathcal{K}f)'(x) = -\frac{s'(x)}{s(L)p(0)} \int_0^L s(L-y)r(y)f(y)dy + \frac{1}{p(0)} \int_0^x s'(x-y)r(y)f(y)dy \,.$$

Eine Abschätzung analog zu (37) zeigt die Existenz einer Konstanten $C > 0$, sodass

$$\sup_{(0,L)} |(\mathcal{K}f)'| \leq C\|f\|_r \,.$$

Das beweist, dass für eine beschränkte Menge $B \subset L^2(0,L)$ die Elemente von $\mathcal{K}(B)$ eine gemeinsamen Schranke für den Betrag ihrer Ableitung besitzen. Daher ist $\mathcal{K}(B)$ gleichgradig stetig und nach Satz 8 (Arzelà-Ascoli) präkompakt in $C([0,L])$. Da die Supremum-Norm immer noch stärker ist als die L^2-Norm, ist $\mathcal{K}(B)$ auch präkompakt in $L^2(0,L)$. Das zeigt, dass $\mathcal{K}$ eine kompakte und (wegen seiner Konstruktion als Inverse) injektive Abbildung ist. Das gestattet die Anwendung des Entwicklungssatzes (Satz 12). Dieser garantiert die Existenz eines in $L^2(0,L)$ vollständigen Orthonormalsystems von Eigenfunktionen $\{\varphi_1, \varphi_2, \ldots\}$ und zugehörigen reellen Eigenwerten $\mu_1, \mu_2 \ldots$ mit $\lim_{k\to\infty} \mu_k = 0$. Wendet man $\mathcal{L}$ auf die Gleichung $\mathcal{K}\varphi_k = \mu_k\varphi_k$ an, dann zeigt das, dass (λ_k, φ_k) mit $\lambda_k = \mu_k^{-1}$ ein Eigenwert-Eigenfunktionspaar für $\mathcal{L}$ ist. Dasselbe Argument funktioniert auch umgekehrt, was zeigt, dass $\mathcal{K}$ und $\mathcal{L}$ dieselben Eigenfunktionen besitzen und die Eigenwerte Kehrwerte von einander sind. Insbesondere gilt auch

$$\mathcal{L}v = \sum_{k=1}^{\infty} \lambda_k \langle v, \varphi_k\rangle_r \varphi_k \,, \qquad v \in D(\mathcal{L}) \,.$$

An dieser Stelle ist es Zeit zuzugeben, dass wir $D(\mathcal{L})$ bisher gar nicht vollständig definiert haben. Dazu bietet die obige Formel eine gute Gelegenheit (siehe die Parsevalsche Gleichung, Lemma 15):

$$v \in D(\mathcal{L}) \quad :\Longleftrightarrow \quad \sum_{k=1}^{\infty} \lambda_k^2 \langle v, \varphi_k\rangle_r^2 < \infty \,.$$

Schließlich noch eine Eigenschaft der Eigenwerte (siehe (35)):

$$\lambda_k = \lambda_k \langle \varphi_k, \varphi_k\rangle_r = \langle \mathcal{L}\varphi_k, \varphi_k\rangle_r = -\int_0^L (p(\varphi_k')^2 + q\varphi_k^2)dx < 0 \,,$$

woraus folgt, dass die Folge der Eigenwerte monoton fällt und nach $-\infty$ divergiert. Daher ist $D(\mathcal{L})$ auch eine echte Teilmenge von $L^2(0,L)$.

Jetzt sind wir in der Lage, das Anfangs-Randwertproblem (33), (34) zu lösen. Mit dem Fourierreihenansatz

$$v(x,t) = \sum_{k=1}^{\infty} v_k(t)\varphi_k(x)\,, \qquad v_k(t) = \langle v(\cdot,t), \varphi_k\rangle_r\,,$$

gilt

$$\dot{v}_k = \langle \partial_t v, \varphi_k\rangle_r = \langle \mathcal{L}v, \varphi_k\rangle_r = \langle v, \mathcal{L}\varphi_k\rangle_r = \lambda_k v_k\,, \qquad v_k(0) = \langle v_A, \varphi_k\rangle_r\,,$$

und daher

$$v(x,t) = \sum_{k=1}^{\infty} \langle v_A, \varphi_k\rangle_r\, e^{\lambda_k t}\varphi_k(x)\,.$$

Mit Hilfe der Parsevalschen Gleichung zeigt man

$$\|v(\cdot,t)\|_r = \left(\sum_{k=1}^{\infty} \langle v_A, \varphi_k\rangle_r^2\, e^{2\lambda_k t}\right)^{1/2} \le e^{\lambda_1 t}\|v_A\|_r\,.$$

Das bedeutet, dass die Lösung u des ursprünglichen Wärmeleitungsproblems im Sinne von $L^2(0,L)$ für $t \to \infty$ exponentiell gegen die stationäre Lösung U konvergiert.

5.6 Oszillationstheorie für Sturm-Liouville-Probleme

Eigenwerteprobleme für den Operator $\mathcal{L}u = \frac{1}{r}((pu')' - qu)$ nennt man *Sturm-Liouville-Probleme.* Wir bleiben bei Dirichlet-Randbedingungen und betrachten daher

$$(p\varphi')' - q\varphi = \lambda r\varphi\,, \qquad \varphi(0) = \varphi(L) = 0\,, \tag{38}$$

mit $p, r > 0$, $q \ge 0$. Wir schreiben die Differentialgleichung als System erster Ordnung mit

$$\varphi' = \frac{y}{p}\,, \qquad y' = (q + \lambda r)\varphi\,,$$

und transformieren (*Prüfer-Transformation)* auf Polarkoordinaten in der (y,φ)-Ebene:

$$\varphi(x) = \varrho(x)\sin\vartheta(x)\,, \qquad y(x) = \varrho(x)\cos\vartheta(x)\,.$$

Einsetzen ergibt

$$\varrho'\sin\vartheta + \varrho\vartheta'\cos\vartheta = \frac{\varrho}{p}\cos\vartheta\,, \qquad \varrho'\cos\vartheta - \varrho\vartheta'\sin\vartheta = (q+\lambda r)\varrho\sin\vartheta\,.$$

Rechnet man daraus die Ableitungen aus, so ergibt sich

$$\varrho' = \varrho\left(\frac{1}{p} + q + \lambda r\right)\sin\vartheta\cos\vartheta\,, \qquad \vartheta' = \frac{1}{p}\cos^2\vartheta - (q+\lambda r)\sin^2\vartheta\,.$$

Dieses System ist nur in eine Richtung gekoppelt. Man kann zunächst die Gleichung für ϑ lösen und dann das Resultat in der ϱ-Gleichung verwenden. Letztere ist dann linear, und aus dem Eindeutigkeitsresultat folgt, dass ϱ entweder immer Null ist (was uns nicht interessiert) oder immer verschieden von Null (was wir im Weiteren annehmen). Dadurch ergibt sich auch in den Randbedingungen eine Entkopplung. Diese reduzieren sich auf $\sin\vartheta(0) = \sin\vartheta(L) = 0$. Die Eigenwerte

können so unabhängig von der Berechnung von ϱ bestimmt werden: Man löst zunächst das Anfangswertproblem

$$\vartheta'_\lambda = \frac{1}{p}\cos^2\vartheta_\lambda - (q+\lambda r)\sin^2\vartheta_\lambda\,, \qquad \vartheta_\lambda(0) = 0\,, \tag{39}$$

und bestimmt dann λ als Lösung der Gleichung $\sin\vartheta_\lambda(L) = 0$. Dazu beobachten wir zunächst, dass, immer wenn $\sin\vartheta_\lambda = 0$ gilt, $\vartheta'_\lambda = 1/p > 0$ folgt. Das bedeutet, dass ϑ_λ zunächst wächst, und dass $\sin\vartheta_\lambda(L) = 0$ nur erfüllt werden kann, indem es ein $k \in \mathbb{N}$ gibt, sodass $\vartheta_\lambda(L) = k\pi$. Wir sammeln zunächst Informationen über ϑ_λ.

Lemma 16 *Die Lösung von (39) ist eine stetige und, für jedes feste $x \in (0, L]$, streng monoton fallende Funktion von λ.*

Beweis: Die Stetigkeit folgt aus Satz 6 (oder auch [T, Satz 2.12]). Für $\lambda_1 < \lambda_2$ gilt

$$\vartheta'_{\lambda_2} - \frac{1}{p}\cos^2\vartheta_{\lambda_2} + (q+\lambda_1 r)\sin^2\vartheta_{\lambda_2} = (\lambda_1-\lambda_2)r\sin^2\vartheta_{\lambda_2} < 0\,,$$

immer außer für $\sin\vartheta_{\lambda_2} = 0$. Daher ist ϑ_{λ_2} eine (strikte) Unterlösung der Gleichung für ϑ_{λ_1}. Ein kleine Verschärfung von Korollar 1 (Abschnitt 2.3) gibt $\vartheta_{\lambda_2}(x) < \vartheta_{\lambda_1}(x)$, $x > 0$. ∎

Lemma 17 *Für die Lösung von (39) gilt*

$$\vartheta_0(L) \le \frac{\pi}{2}$$

Beweis: Die Funktion $\overline{\vartheta}(x) = \pi/2$ erfüllt

$$\overline{\vartheta}' - \frac{1}{p}\cos^2\overline{\vartheta} + (q+\lambda r)\sin^2\overline{\vartheta} = q+\lambda r\,, \qquad \overline{\vartheta}(0) > 0\,.$$

Sie ist daher für $\lambda = 0$ (sogar für alle $\lambda \ge 0$) eine Oberlösung. Das beweist die Aussage. ∎

Das letzte Resultat zeigt noch einmal, dass alle Eigenwerte negativ sind. Weiters wissen wir schon, dass die Eigenwerte eine nach $-\infty$ divergierende Folge bilden. Lemma 16 impliziert auch, dass die Eigenwerte einfach sind. Aus Lemma 16 und 17 folgt

$$\vartheta_k(L) := \vartheta_{\lambda_k}(L) = k\pi\,, \qquad k \in \mathbb{N}\,.$$

Daher gilt auch

Satz 13 *Die Eigenwerte λ_k, $k \in \mathbb{N}$, des Sturm-Liouville-Problems (38) sind einfach. Die normierte Eigenfunktion φ_k zum Eigenwert λ_k hat $k-1$ Nullstellen $x_{k,l}$, $l = 1, \dots, k-1$, in $(0, L)$. Für $k \ge 2$ liegen die Nullstellen von φ_k zwischen den Nullstellen von φ_{k+1}, d.h.*

$$0 < x_{k+1,1} < x_{k,1} < x_{k+1,2} < \cdots < x_{k+1,k-1} < x_{k,k-1} < x_{k+1,k} < L\,,$$

Beweis: Es muss nur mehr die letzte Aussage bewiesen werden. Statt eines allgemeinen Beweises beschränken wir uns auf den Fall $k = 2$, d.h. wir zeigen $x_{3,1} < x_{2,1} < x_{3,2}$. Es gilt $\vartheta_2(x_{2,1}) = \vartheta_3(x_{3,1}) = \pi$ und $\vartheta_3(x_{3,2}) = 2\pi$. Wegen der Monotonie und wegen $\lambda_3 < \lambda_2$ gilt $\vartheta_3 > \vartheta_2$, insbesondere $\vartheta_3(x_{2,1}) > \pi$, und daher $x_{3,1} < x_{2,1}$.

Andererseits vergleichen wir ϑ_3 und $\vartheta_2^* := \vartheta_2 + \pi$. Letzteres erfüllt dieselbe Differentialgleichung wie ϑ_2, und wegen $\vartheta'_3 - \frac{1}{p}\cos^2\vartheta_3 + (q+\lambda_2 r)\sin^2\vartheta_3 = (\lambda_2-\lambda_3)r\sin^2\vartheta_3 > 0$ ist ϑ_3 eine obere Lösung für diese Gleichung. Aus $\vartheta_3(L) = \vartheta_2^*(L) = 3\pi$ folgt $\vartheta_3 < \vartheta_2^*$ und insbesondere $\vartheta_3(x_{2,1}) < 2\pi$, und daher $x_{2,1} < x_{3,2}$. ∎

6 Variationsrechnung

6.1 'Der gerade Weg ist der kürzeste'

Kann man das beweisen? Zunächst ein bisschen exakter formuliert: Unter allen Kurven, die zwei gegebene Punkte verbinden, ist die Verbindungsstrecke die kürzeste. Um das Problem mathematisch zu formulieren, müssen wir uns in Erinnerung rufen, wie man die Länge einer Kurve berechnet. Wir beschränken uns auf Kurven in der Ebene. Seien $x_a, x_b \in \mathbb{R}^2$ die gegebenen Endpunkte. Dann interessieren wir uns für Kurven mit einer *Parameterdarstellung* der Form $K = \{x(t) : \ t \in [0,1]\} \subset \mathbb{R}^2$ mit $x \in C([0,1])$, wobei wir fordern, dass $x(0) = x_a$, $x(1) = x_b$. Approximieren wir die Kurve durch einen Polygonzug K_N mit den Eckpunkten $x(t_k) : 0 \le k \le N$ mit $t_k = k\Delta t = k/N$, dann ist die Länge des Polygonzugs gegeben durch

$$\ell(K_N) = \sum_{k=1}^{N} |x(t_k) - x(t_{k-1})| \approx \sum_{k=1}^{N} |x'(t_k)|\Delta t\,.$$

Der Ausdruck auf der rechten Seite ist eine Riemannsumme. Das motiviert die Definition der *Bogenlänge* der Kurve:

$$\ell(K) := \int_0^1 |x'(t)|dt\,,$$

wenn das Integral auf der rechten Seite existiert. Der Integrand ist die Länge des Tangentialvektors. Jetzt können wir noch exakter formulieren: Wir suchen eine Kurve K mit den Endpunkten x_0, x_1 mit minimaler Länge $\ell(K)$. Wir schränken das Problem etwas ein: Wir wählen $x_0 = (0,a)$, $x_1 = (1,b)$, und lassen nur Kurven zu, die sich als Graph einer Funktion darstellen lassen, d.h. $x(t) = (t, u(t))$ mit $u(0) = a$, $u(1) = b$. Das ergibt $|x'(t)| = \sqrt{1 + u'(t)^2}$ und daher

$$\ell(K) = I[u] := \int_0^1 \sqrt{1 + u'(t)^2}\,dt\,.$$

Wir haben es also mit einer Extremwertaufgabe in einem unendlichdimensionalen Funktionenraum zu tun. Um damit umzugehen, verwenden wir das Konzept der *Richtungsableitung.* Angenommen, u wäre die minimierende Funktion. Dann sind auch Funktionen $u(t) + \varepsilon v(t)$, $\varepsilon \in \mathbb{R}$ zugelassene Funktionen, wenn die *Richtung* v die Randbedingungen $v(0) = v(1) = 0$ erfüllt. Dann nimmt aber die reelle Funktion

$$i(\varepsilon) := I[u + \varepsilon v] = \int_0^1 \sqrt{1 + (u'(t) + \varepsilon v'(t))^2}\,dt$$

ihr Minimum an der Stelle $\varepsilon = 0$ an. Daraus folgt aber $i'(0) = 0$. Berechnen wir zunächst die Ableitung von i:

$$i'(\varepsilon) = \int_0^1 \frac{(u' + \varepsilon v')v'}{\sqrt{1 + (u' + \varepsilon v')^2}}dt$$

Daher

$$0 = i'(0) = \int_0^1 \frac{u'v'}{\sqrt{1 + (u')^2}}dt$$

Was tun wir damit? Wir suchen die optimierende Funktion u. Die obige Gleichung gilt für alle Funktionen v, die die Randbedingungen $v(0) = v(1) = 0$ erfüllen. Zunächst integrieren wir partiell:

$$0 = \left.\frac{u'v}{\sqrt{1 + (u')^2}}\right|_0^1 - \int_0^1 \left(\frac{u'}{\sqrt{1 + (u')^2}}\right)' v\,dt = -\int_0^1 \frac{u''}{(1 + (u')^2)^{3/2}}\,v\,dt \tag{40}$$

Bis jetzt haben wir nichts über die Glattheit der beteiligten Funktionen gesagt. Nehmen wir jetzt an, dass alle im Integral auf der rechten Seite vorkommenden Funktionen stetig sind, dann können wir folgendes Resultat verwenden.

Lemma 18 *Sei $w \in C([0,1])$ und*

$$\int_0^1 w(t)v(t)dt = 0 \qquad \textit{für alle } v \in C([0,1]) \textit{ mit } v(0) = v(1) = 0\,.$$

Dann gilt $w = 0$ in $[0,1]$.

Beweis: Sei $0 < t_0 < 1$. Dann ist

$$v_\delta(t) := \frac{1}{2\delta}\left(1 - \frac{|t - t_0|}{\delta}\right)_+$$

für $\delta > 0$ klein genug ein zugelassenes v. Der Mittelwertsatz der Integralrechnung ergibt

$$0 = \int_0^1 w v_\delta dt = \int_{t_0-\delta}^{t_0+\delta} w v_\delta dt = w(t_\delta)\int_{t_0-\delta}^{t_0+\delta} v_\delta dt = w(t_\delta)\,,$$

mit $|t_\delta - t_0| \le \delta$. Der Limes $\delta \to 0$ ergibt $w(t_0) = 0$. Da t_0 ein beliebiger innerer Punkt ist, ist der Beweis abgeschlossen. ∎

Unter der Annahme $u \in C^2([0,1])$ folgt daher aus (40) die Differentialgleichung

$$u'' = 0\,,$$

und daher $u(t) = a + t(b-a)$, d.h. der gerade Weg ist wirklich der kürzeste. Dazu sollte man aber noch überprüfen, ob wirklich ein Minimum vorliegt. Ein Indiz ergibt sich aus

$$i''(\varepsilon) = \int_0^1 \frac{(v')^2}{(1 + (u' + \varepsilon v')^2)^{3/2}} dt \ge 0\,.$$

Dieses Resultat ist eine Konsequenz aus der Tatsache, dass die Funktion $z \mapsto \sqrt{1+z^2} =: f(z)$, die im Integranden von $I[u]$ auftritt, *konvex* ist, d.h.

$$f(\alpha z_1 + (1-\alpha)z_2) \le \alpha f(z_1) + (1-\alpha) f(z_2)$$

für beliebige z_1, z_2 und $0 \le \alpha \le 1$ (geometrisch: Der Graph liegt unter jeder Sekante). Für glatte Fuktionen genügt auch $f'' \ge 0$ (in diesem Fall $f''(z) = (1+z^2)^{-3/2}$). Wir haben sogar *strikte Konvexität*, d.h.

$$f(\alpha z_1 + (1-\alpha)z_2) < \alpha f(z_1) + (1-\alpha) f(z_2) \quad \text{für } z_1 \ne z_2\,,\ 0 < \alpha < 1\,, \qquad \text{bzw. } f'' > 0\,.$$

Lemma 19 *Sei $f \in C(\mathbb{R})$ strikt konvex und sei $K \subset C^1([a,b])$ ein konvexe Menge. Dann gibt es höchstens eine Funktion $u \in K$, sodass*

$$I[u] = \min_{v \in K} I[v]\,, \qquad \textit{wobei} \quad I[v] := \int_a^b f(v'(t))dt\,.$$

Beweis: Seien $u_1 \neq u_2 \in K$ minimierende Funktionen, d.h.

$$I[u_1] = I[u_2] = I_{min} := \min_{v \in K} I[v]\,.$$

Dann ist wegen der Konvexität von K auch $v := \frac{u_1+u_2}{2} \in K$, und es gilt $f(v') \leq (f(u_1')+f(u_2'))/2$ in $[a,b]$ sowie $f(v') < (f(u_1')+f(u_2'))/2$ überall dort, wo $u_1 \neq u_2$. Wegen der Vorzeichenbeständigkeit stetiger Funktionen folgt der Widerspruch

$$I[v] < \frac{I[u_1] + I[u_2]}{2} = I_{min}\,.$$

■

Wir betrachten noch ein Beispiel, um zu zeigen, dass die strikte Konvexität notwendig ist für dieses Resultat: Anfangs- bzw. Endpunkt einer *Wanderung* liegen a bzw. b Meter über dem Meeresspiegel. Der skalierte Abstand in horizontale Richtung zwischen den beiden Punkten ist 1. Wie muss der Weg beschaffen sein, damit ich mich möglichst wenig anstrenge. Dabei nehmen wir an, dass Bergaufgehen und Bergabgehen für mich gleich anstrengend sind (was in Wahrheit nicht stimmt) und dass die Anstrengung jeweils proportional zur überwundenen Höhe ist. Der Weg wird dann parametrisiert durch $(t, u(t))$, $t \in [0,1]$, mit $u(0) = a$, $u(1) = b$. Angenommen, die Extrema des Weges wären an $a = t_0 < t_1 < \cdots < t_N = b$. Dann ist die Anstrengung proportional zu

$$\sum_{k=1}^{N} |u(t_k) - u(t_{k-1})| = \sum_{k=1}^{N} \left| \int_{t_{k-1}}^{t_k} u'(t)dt \right| = \sum_{k=1}^{N} \int_{t_{k-1}}^{t_k} |u'(t)|dt = \int_a^b |u'(t)|dt\,. \tag{41}$$

Die zweite Gleichung gilt, weil sich das Vorzeichen von u' in den Intervallen (t_{k-1}, t_k) nicht ändert. Wir haben es also mit einem Problem wie oben zu tun mit der konvexen, aber nicht strikt konvexen Funktion $f(z) = |z|$. Der Verdacht liegt nahe, dass ein optimaler Weg monoton verlaufen muss. Für jedes monotone u ergibt sich $\int_a^b |u'|dt = |a-b|$. Mit Hilfe der Dreiecksungleichung zeigt man leicht, dass das eine untere Schranke für die linke Seite von (41) mit beliebigem u ist. Also bei weitem keine Eindeutigkeit der minimierenden Funktion! (Für die ExpertInnen: Die hier zu minimierende Größe (41) ist die *Totalvariation* von u.)

6.2 Variationsprobleme – Die Brachistochrone

Ein *Funktional* ist eine Abbildung von einem Vektorraum in den Skalarkörper. Hier betrachten wir reelle Vektorräume von Funktionen von einer Veränderlichen. Sei $I : K \to \mathbb{R}$ ein Funktional, das auf der konvexen Teilmenge K eines Funktionenraumes definiert ist. Das Problem der Bestimmung von $u \in K$, sodass

$$I[u] = \min_{v \in K} I[v]\,,$$

nennt man ein *Variationsproblem.* Eines der ersten korrekt formulierten und gelösten Variationsprobleme ist die Bestimmung der *Brachistochrone.* Die Formulierung kam von Johann Bernoulli (1696), der es als Herausforderung für andere Mathematiker formuliert hat. Lösungen kamen von ihm selbst, von seinem Bruder Jakob, aber auch von Newton, Leibniz, de l'Hôpital und anderen. Später wurde von Euler und Lagrange eine systematische Theorie für derartige Probleme entwickelt, die von Euler *'Variationsrechnung'* genannt wurde.

Wir werden hier Probleme mit

$$I[u] = \int_{t_0}^{t_1} f(t, u(t), u'(t))dt\,, \qquad K = \{u \in C^1([t_0, t_1]) : \, u(t_0) = u_0\,,\, u(t_1) = u_1\}\,,$$

betrachten. Wie im vorigen Abschnitte definieren wir

$$i(\varepsilon) := I[u + \varepsilon v] = \int_{t_0}^{t_1} f(t, u + \varepsilon v, u' + \varepsilon v')dt\,,$$

wobei u die gesuchte minimierende Funktion ist, $\varepsilon \in \mathbb{R}$ und v so, dass $u + \varepsilon v \in K$, d.h. $v \in C^1([t_0, t_1])$, $v(t_0) = v(t_1) = 0$. Wieder wie im vorigen Abschnitt muss die *Variation von I an der Stelle u in Richtung v*,

$$\delta I[u](v) := i'(0)\,,$$

für alle zulässigen v verschwinden, also

$$\int_{t_0}^{t_1} \left(\partial_u f(t, u, u')v + \partial_{u'} f(t, u, u')v'\right) dt = 0\,.$$

Partielle Integration des zweiten Termes ergibt

$$\int_{t_0}^{t_1} \left(\partial_u f(t, u, u') - \left(\partial_{u'} f(t, u, u')\right)'\right) v\, dt = 0\,,$$

und damit folgt aus Lemma 18 die *Euler-Lagrange-Gleichung*

$$\partial_u f(t, u, u') - \left(\partial_{u'} f(t, u, u')\right)' = 0\,.$$

Das ist im Allgemeinen eine gewöhnliche Differentialgleichung zweiter Ordnung für u, die zusammen mit den Randbedingungen $u(t_0) = u_0$, $u(t_1) = u_1$ zu lösen ist.

Im Fall, dass f nicht explizit von t abhängt, gibt es ein erstes Integral, und zwar

$$H = f(u, u') - \partial_{u'} f(u, u')u'\,.$$

Es gilt nämlich

$$\begin{aligned} H' &= \partial_u f(u, u')u' + \partial_{u'} f(u, u')u'' - (\partial_{u'} f(u, u'))'u' - \partial_{u'} f(u, u')u'' \\ &= \left(\partial_u f(u, u') - (\partial_{u'} f(u, u'))'\right) u' = 0\,. \end{aligned} \tag{42}$$

Das Problem der Brachistochrone besteht darin, zwischen gegebenen Anfangs- und Endpunkten eine Kugelbahn zu bauen, sodass die Kugel, getrieben von der Gravitation, den Weg in möglichst kurzer Zeit zurücklegt. Dabei nimmt man Anfangsgeschwindigkeit Null und konstante vertikale Gravitationsbeschleunigung an, und vernachlässigt Reibungseffekte. Wir formulieren das Problem in einer vertikalen Ebene, wobei die x-Richtung die horizontale und die y-Richtung die vertikale Richtung ist. Anfangs- und Endpunkt legen wir in (x_0, y_0) und (x_1, y_1) mit $x_0 < x_1$ und $y_0 > y_1$ und nehmen an, dass die Kugelbahn als Graph $\{(x, y(x)) : \, x_0 \le x \le x_1\}$ einer Funktion y mit $y(x_0) = y_0$, $y(x_1) = y_1$ geschrieben werden kann (siehe Fig. 3). Im Folgenden werden wir den normierten Tangentialvektor

$$\tau(x) = \frac{(1, y'(x))}{\sqrt{1 + y'(x)^2}}$$

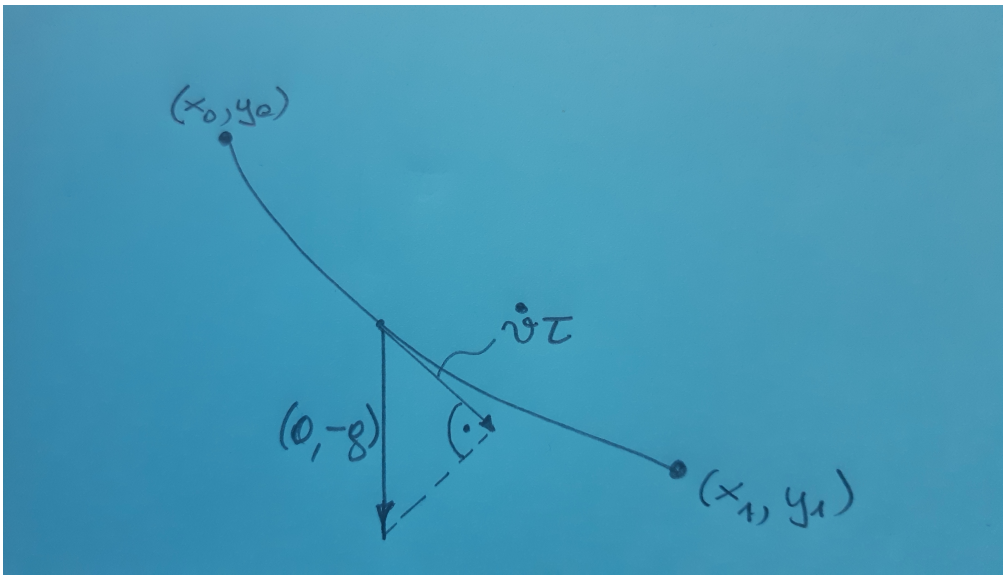

Figure 3: Notation für die Modellierung der Brachistochrone

benötigen. Wenn die Kugel den Weg im Zeitintervall $[0, T]$ zurücklegt, dann ist ihre Bahn in Abhängigkeit von der Zeit t gegeben durch $(x(t), y(x(t)))$, $0 \leq t \leq T$. Schreiben wir den Geschwindigkeitsvektor zum Zeitpunkt t in der Form $v(t)\tau(x(t))$, dann ergibt sich in x-Richtung

$$\dot{x} = \frac{v}{\sqrt{1 + y'(x)^2}}\,. \tag{43}$$

Die Beschleunigung $\dot{v}$ ist gegeben durch den Anteil der Gravitationsbeschleunigung $(0, -g)$ in Richtung τ, d.h.

$$\dot{v} = \frac{-gy'(x)}{\sqrt{1 + y'(x)^2}}\,.$$

Aus diesen beiden Gleichungen folgt

$$\frac{dv}{dx} = \frac{-gy'(x)}{v}\,.$$

Trennung der Variablen und Integration ergibt

$$\frac{v^2}{2} + gy(x) = gy_0\,,$$

wobei die Integrationskonstante aus der Auswertung der linken Seite an $t = 0$ folgt. Dieses Resultat ist nicht überraschend. Multiplikation mit der Masse der Kugel ergäbe das Gesetz der Energieerhaltung. Die linke Seite wird dann zur Summe aus der kinetischen und der potentiellen Energie,

die am Anfang nur aus potentieller Energie besteht. Ausrechnen von v und Einsetzen in (43) ergibt

$$\sqrt{\frac{1+y'(x)^2}{y_0-y(x)}}\,\dot{x} = \sqrt{2g}\,.$$

Diese Gleichung integrieren wir nach der Zeit und verwenden auf der linken Seite die Substitutionsregel:

$$T\sqrt{2g} = \int_{x_0}^{x_1} \sqrt{\frac{1+y'(x)^2}{y_0-y(x)}}\,dx =: I[y]\,.$$

Damit haben wir ein Variationsproblem formuliert. Das Integral $I[y]$ ist zu minimieren über alle Funktionen y, die die Randbedingungen $y(x_0) = y_0$, $y(x_1) = y_1$ erfüllen.

Da der Integrand nicht explizit von x abhängt, können wir (42) verwenden, und es gilt

$$\sqrt{\frac{1+y'(x)^2}{y_0-y(x)}} - \frac{y'(x)^2}{\sqrt{(1+y'(x)^2)(y_0-y(x))}} = H\,,$$

mit einer Konstanten H. Quadrieren ergibt

$$1 + y'(x)^2 + \frac{y'(x)^4}{1+y'(x)^2} - 2y'(x)^2 = H^2(y_0-y(x))\,,$$

und daher

$$\frac{1}{1+y'(x)^2} = H^2(y_0-y(x))\,,$$

und weiter,

$$y'(x)^2 = \frac{1}{H^2(y_0-y(x))} - 1\,.$$

Führen wir statt H einen Parameter $y_{min} = y_0 - H^{-2}$ ein, dann ergibt sich

$$y'(x)^2 = \frac{y(x)-y_{min}}{y_0-y(x)}\,,$$

was den Namen des neuen Parameters erklären sollte. Suchen wir zunächst nach monotonen (d.h. fallenden) Lösungen, so folgt

$$\sqrt{\frac{y_0-y(x)}{y(x)-y_{min}}}\,y'(x) = -1\,,$$

und daher, nach Integration und Verwendung der Randbedingung an $x = x_0$,

$$\int_{y(x)}^{y_0} \sqrt{\frac{y_0-z}{z-y_{min}}}\,dz = x - x_0\,. \tag{44}$$

Das Integral kann mit Hilfe elementarer Funktionen ausgedrückt werden. Diese Darstellung ist allerdings nicht besonders erhellend. Der Parameter y_{min} muss so bestimmt werden, dass auch die Randbedingung an $x = x_1$ erfüllt ist, d.h.

$$\psi_1(y_{min}) := \int_{y_1}^{y_0} \sqrt{\frac{y_0-z}{z-y_{min}}}\,dz = x_1 - x_0\,.$$

Die Funktion $\psi_1 : (-\infty, y_1] \to (0, \psi_1(y_1)]$ ist streng monoton wachsend. Das bedeutet, dass es für $x_1 - x_0 \leq \psi_1(y_1)$ eine eindeutige, monoton fallende Lösung $y(x)$ gibt, die durch die Gleichung (44) implizit gegeben ist. Man beachte, dass die Kugelbahn am Anfangspunkt vertikal ist: $\lim_{x\to x_0+} y'(x) = -\infty$.

Für $x_1 - x_0 > \psi_1(y_1)$ ist die Bahn nicht monoton. Sie ist zunächst durch (44) gegeben, erreicht an der Stelle $x = x_{min}$, gegeben durch

$$\int_{y_{min}}^{y_0} \sqrt{\frac{y_0 - z}{z - y_{min}}}\, dz = x_{min} - x_0\,, \tag{45}$$

den Minimalwert $y_{min} < y_1$ und wächst dann wieder, gemäß

$$\sqrt{\frac{y_0 - y(x)}{y(x) - y_{min}}}\, y'(x) = 1\,,$$

und daher

$$\int_{y_{min}}^{y(x)} \sqrt{\frac{y_0 - z}{z - y_{min}}}\, dz = x - x_{min}\,, \qquad x > x_{min}\,.$$

Die Randbedingung an $x = x_1$ ergibt die Forderung

$$\int_{y_{min}}^{y_1} \sqrt{\frac{y_0 - z}{z - y_{min}}}\, dz = x_1 - x_{min}\,. \tag{46}$$

Nun addieren wir (45) und (46):

$$\psi_2(y_{min}) := \int_{y_{min}}^{y_0} \sqrt{\frac{y_0 - z}{z - y_{min}}}\, dz + \int_{y_{min}}^{y_1} \sqrt{\frac{y_0 - z}{z - y_{min}}}\, dz = x_1 - x_0\,.$$

Die Funktion ψ_2 lässt sich auch schreiben als

$$\psi_2(y_{min}) := \int_0^{y_0 - y_{min}} \sqrt{\frac{y_0 - y_{min} - w}{w}}\, dw + \int_0^{y_1 - y_{min}} \sqrt{\frac{y_0 - y_{min} - w}{w}}\, dw\,,$$

was zeigt, dass sie streng monoton fallend ist und bijektiv als $\psi_2 : (-\infty, y_1] \to [\psi_1(y_1), \infty)$. Daher ist auch für $x_1 - x_0 > \psi_1(y_1)$ der Wert von y_{min} eindeutig bestimmt. In diesem Fall wird er auch angenommen. Es kann auch ein Punkt mit $y_1 = y_0$ erreicht werden.

Ohne weiter auf Details einzugehen, merken wir an, dass die berechneten Kurven *Zykloiden* sind ($\to$ `wikipedia`).

6.3 Isoperimetrische Probleme – Nebenbedingungen

Isoperimetrische Probleme sind weitere klassische Variationsprobleme, die teilweise auf die Antike zurückgehen. Wir beginnen mit dem Problem des Hühnerhofs. Mit einem Zaun der Länge L_1 soll angrenzend an die gerade Hofmauer ein Bereich für die Hühner abgegrenzt werden, der von dem Zaun und einem Stück Mauer der Länge L_2 begrenzt wird. Damit das überhaupt geht, muss $L_1 > L_2$ gelten. Der Bereich für die Hühner soll möglichst großen Flächeninhalt F haben (siehe Fig. 4).

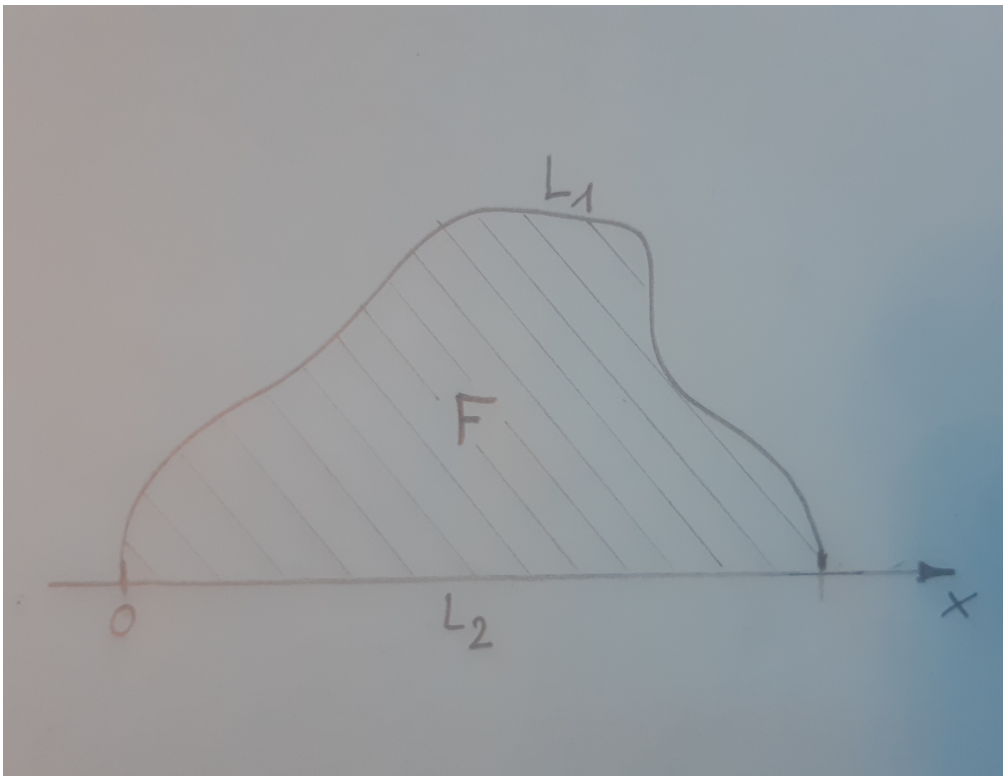

Figure 4: Der Hühnerhof

In der (x, y)-Ebene sei die x-Achse die Hofmauer mit dem Zaun von $x = 0$ bis $x = L_2$ und die Umzäunung beschrieben als Graph einer Funktion $y : [0, L_2] \to [0, \infty)$. Der Flächeninhalt ist daher gegeben durch

$$F[y] = \int_0^{L_2} y(x)dx\,,$$

und die Forderung an die Länge des Zauns durch

$$L[y] = \int_0^{L_2} \sqrt{1 + y'(x)^2}\,dx - L_1 = 0\,. \tag{47}$$

Das Ziel ist es, y unter allen Funktionen, die $y(0) = y(L_2) = 0$ und die *Nebenbedingung* (47) erfüllen, so zu wählen, dass $F[y]$ maximal wird. Wie üblich bei Extremwertaufgaben mit Nebenbedingungen verwenden wir die Methode der *Lagrange-Multiplikatoren*, d.h. wir definieren das *Lagrange-Funktional*

$$I[y, \lambda] := F[y] + \lambda L[y] = \int_0^{L_2} \left(y + \lambda\sqrt{(1 + (y')^2}\right) dx - \lambda L_1\,,$$

mit dem Lagrange-Multiplikator $\lambda \in \mathbb{R}$. Die Methode besteht darin, stationäre Punkte dieses Funktionals zu finden. Da der Integrand nicht explizit von x abhängt, können wir (42) verwenden und erhalten das erste Integral der Euler-Lagrange-Gleichungen

$$H = y + \lambda\sqrt{(1 + (y')^2} - \lambda\frac{(y')^2}{\sqrt{(1 + (y')^2}} = y + \frac{\lambda}{\sqrt{(1 + (y')^2}}$$

Umformen dieser Gleichung ergibt

$$(y')^2 = \frac{\lambda^2 - (y-H)^2}{(y-H)^2}$$

Die Form der rechten Seite motiviert die Einführung der neuen Unbekannten $\vartheta(x)$ durch $y = H + \lambda\cos\vartheta$. Daraus folgt

$$\lambda^2(\vartheta')^2\sin^2\vartheta = \frac{\sin^2\vartheta}{\cos^2\vartheta}\,,$$

und daher

$$\lambda\vartheta'\cos\vartheta = \pm 1\,.$$

Integration mit der Integrationskonstanten $\pm x_0$ ergibt

$$\lambda^2\sin^2\vartheta = (x-x_0)^2\,,$$

und damit

$$\lambda^2 - (y-H)^2 = (x-x_0)^2\,.$$

Die optimale Form des Zauns ist daher die eines *Kreisbogens.* Aus Symmetriegründen muss $x_0 = L_2/2$ gelten. Die Konstanten H und λ müssen so bestimmt werden, dass die Randbedingungen und die Nebenbedingung erfüllt sind.

Das klassische isoperimetrische Problem wird das *Problem der Dido* genannt und hat mit der Gründungslegende der phönizischen Stadt Karthago zu tun: *Welche geschlossene Kurve gegebener Länge umschließt den größten Flächeninhalt?* Für dieses Problem wählen wir zunächst ein beliebiges Parameterintervall, z.B. $t \in [0,1]$. Die Randkurve des Gebietes Ω habe die Parameterdarstellung $\partial\Omega = \{x(t) : 0 \le t \le 1\} \subset \mathbb{R}^2$ wobei x glatt und injektiv sein soll (bis auf die Geschlossenheitsbedingung $x(0) = x(1)$). Wir nehmen an, die Randkurve sei positiv orientiert, d.h. für wachsende t wird sie gegen den Uhrzeigersinn durchlaufen. Um den Flächeninhalt aus der Parameterdarstellung zu berechnen, benötigen wir den *Divergenzsatz:*

$$F[x] := \int_\Omega dx = \frac{1}{2}\int_\Omega \nabla\cdot x\,dx = \frac{1}{2}\int_{\partial\Omega} x\cdot\nu(x)\,ds\,,$$

wobei $\nu(x)$ der nach Außen orientierte Einheitsnormalvektor ist und $ds = |x'|dt$ das Längenelement. Mit $x' = (x_1', x_2')$ ergibt sich $\nu = |x'|^{-1}(x_2', -x_1')$, und daher

$$F[x] = \frac{1}{2}\int_0^1 (x_2'x_1 - x_1'x_2)dt = \int_0^1 x_2'x_1\,dt = -\int_0^1 x_1'x_2\,dt\,.$$

Ist der vorgegebene Umfang gleich L, so ergibt sich die Nebenbedingung

$$U[x] := \int_0^1 |x'|dt = \int_0^1 \sqrt{(x_1')^2 + (x_2')^2}\,dt = L\,.$$

Wir haben es also mit zwei unbekannten Funktionen $x_1(t)$ und $x_2(t)$ zu tun. Wieder bilden wir das Lagrange-Funktional

$$I[x_1, x_2, \lambda] := F[x_1, x_2] + \lambda(U[x_1, x_2] - L)$$

und variieren dieses bezüglich beider unbekannter Funktionen. Das ergibt ein System von zwei Euler-Lagrange-Gleichungen:

$$x_2' = \lambda \left(\frac{x_1'}{\sqrt{(x_1')^2 + (x_2')^2}} \right)' , \qquad x_1' = \lambda \left(\frac{x_2'}{\sqrt{(x_1')^2 + (x_2')^2}} \right)' .$$

Integration liefert

$$x_2 - m_2 = \lambda \frac{x_1'}{\sqrt{(x_1')^2 + (x_2')^2}} , \qquad x_1 - m_1 = \lambda \frac{x_2'}{\sqrt{(x_1')^2 + (x_2')^2}} ,$$

mit den Integrationskonstanten m_1, m_2. Quadrieren und Addieren ergibt das erwartete Resultat einer Kreisgleichung:

$$(x_1 - m_1)^2 + (x_2 - m_2)^2 = \lambda^2 .$$

Die Lage des Mittelpunktes (m_1, m_2) ist natürlich beliebig, und der Lagrange-Multiplikator $\lambda = \frac{L}{2\pi}$ erhält die Bedeutung des Radius.

6.4 Das Sturm-Liouville-Problem als Variationsproblem

Wir kehren zunächst zurück zu Abschnitt 5.4 und zum abstrakten Problem der Eigenwertanalyse für einen kompakten, symmetrischen, injektiven Operator $\mathcal{K} : \mathcal{H} \to \mathcal{H}$. Wir haben gezeigt, dass für den betragsgrößten Eigenwert μ_1 der Zusammenhang

$$|\mu_1| = \|\mathcal{K}\| = \sup_{\|u\|_{\mathcal{H}}=1} |\langle \mathcal{K}u, u \rangle|$$

gilt. Für einen normierten Eigenvektor φ_1 dazu gilt $\mathcal{K}\varphi_1 = \mu_1\varphi_1$, und daher $\langle \mathcal{K}\varphi_1, \varphi_1 \rangle = \mu_1$, woraus folgt, dass das obige Supremum angenommen wird, und daher

$$|\mu_1| = \max_{\|u\|_{\mathcal{H}}=1} |\langle \mathcal{K}u, u \rangle| .$$

Ist der Hilbertraum ein Funktionenraum, dann löst φ_1 also ein Variationsproblem mit Nebenbedingung ($\|u\|_{\mathcal{H}} = 1$). Die Nebenbedingung kann man loswerden, indem man für einen allgemeinen Vektor $0 \neq u \in \mathcal{H}$ die normierte Version einsetzt. Als Konsequenz maximiert man dann

$$\left| \left\langle \mathcal{K} \frac{u}{\|u\|_{\mathcal{H}}}, \frac{u}{\|u\|_{\mathcal{H}}} \right\rangle \right| = \frac{|\langle \mathcal{K}u, u \rangle|}{\|u\|_{\mathcal{H}}^2} , \tag{48}$$

jetzt ohne Nebenbedingung. Jetzt ist jeder (auch nicht normierte) Eigenvektor zu μ_1 ein Maximierer.

Wir erinnern weiters daran, dass es abzählbar viele Eigenwerte $\mu_1, \mu_2, \ldots$ gibt und ein vollständiges Orthonormalsystem $\varphi_1, \varphi_2, \ldots$ zugehöriger Eigenvektoren. Jeder Vektor $u \in \mathcal{H}$ kann daher durch eine Fourierreihe

$$u = \sum_{k=1}^{\infty} u_k \varphi_k$$

dargestellt werden, und es gilt

$$\langle \mathcal{K}u, u \rangle = \sum_{k=1}^{\infty} \mu_k u_k^2 .$$

Wenn wir die Allgemeinheit durch die Annahme $\mu_1 \leq \mu_2 \leq \cdots < 0$ einschränken, die bei Sturm-Liouville-Problemen gilt, Dann gilt weiter

$$|\langle \mathcal{K}u, u\rangle| = \sum_{k=1}^{\infty} (-\mu_k) u_k^2 \,.$$

Für einen neuen Vektor v mit den Fourierkoeffizienten $v_k := \sqrt{-\mu_k}\, u_k$, $k \in \mathbb{N}$, gilt daher

$$|\langle \mathcal{K}u, u\rangle| = \|v\|_{\mathcal{H}}^2 \,.$$

Andererseits

$$\|u\|_{\mathcal{H}}^2 = \sum_{k=1}^{\infty} (-\lambda_k) v_k^2 = |\langle \mathcal{L}v, v\rangle| \,,$$

mit der Notation von Abschnitt 5.5, d.h. $\lambda_k = 1/\mu_k$ und $\mathcal{L} = \mathcal{K}^{-1}$. Das Problem der Maximierung von (48) ist daher äquivalent zum Problem der Minimierung des *Raileigh-Quotienten*

$$-\frac{\langle \mathcal{L}v, v\rangle}{\|v\|_{\mathcal{H}}^2} \,, \tag{49}$$

mit der Lösung

$$-\lambda_1 = \min_{v \in D(\mathcal{L})} \left(-\frac{\langle \mathcal{L}v, v\rangle}{\|v\|_{\mathcal{H}}^2} \right) = \min_{\|v\|_{\mathcal{H}}=1} \left(-\langle \mathcal{L}v, v\rangle \right) = -\langle \mathcal{L}\varphi_1, \varphi_1\rangle \,.$$

Für das in Abschnitt 5.5 behandelte Sturm-Liouville-Problem ist der Raileigh-Quotient gegeben durch

$$-\frac{\langle \mathcal{L}v, v\rangle_r}{\|v\|_r^2} = \frac{\int_0^L \left(p(v')^2 + qv^2 \right) dx}{\int_0^L rv^2 dx} \,.$$

Die Minimierung betrachten wir als Problem mit Nebenbedingung und definieren das Lagrange-Funktional

$$I[v, \lambda] := \int_0^L \left(p(v')^2 + qv^2 \right) dx + \lambda \left(\int_0^L rv^2 dx - 1 \right) = \int_0^L \left(p(v')^2 + (q + \lambda r)v^2 \right) dx - \lambda$$

Die Euler-Lagrange-Gleichung ist

$$(pv')' - qv = \lambda rv \qquad \text{bzw.} \qquad \mathcal{L}v = \lambda v \,,$$

was zu erwarten war. Die weiteren Eigenwerte sind auch durch Minimierung zu bestimmen, indem man den Raum entsprechend einschränkt. Als Beispiel:

$$-\lambda_2 = \min_{v \in \{\varphi_1\}^{\perp}} \left(-\frac{\langle \mathcal{L}v, v\rangle}{\|v\|_{\mathcal{H}}^2} \right) .$$