Mathematik für Physiker III

Michael Dreher
Fachbereich für Mathematik und Statistik
Universität Konstanz

Studienjahr 2012/13

Acknowledgements:

These are the lecture notes to a third semester course on Mathematics for Physicists, and the author is indebted to Nicola Wurz, Maria Lindauer, Florian Franz, Philip Lindner, Simon Schüz, Samuel Greiner, Christian Schoder, Pascal Gumbsheimer for remarks which helped to improve the presentation, and to the audience for appropriating this huge amount of knowledge.

I do not know what I may appear to the world,
but to myself I seem to have been only like a boy playing on the sea–shore,
and diverting myself in now and then finding a smoother pebble
or a prettier shell than ordinary,
whilst the great ocean of truth lay all undiscovered before me.

Sir Isaac Newton (1642 – 1727) [1]

[1] As quoted in [3].

Contents

Part I

Ordinary Differential Equations

Chapter 1

Existence and Uniqueness Results

1.1 An Introductory Example

We consider a thermodynamical system[1] — think of a closed balloon with a certain type of gas in it — and (some of) the thermodynamical quantities are the

pressure p,

temperature T,

specific volume τ, which is the volume per unit mass,

density ϱ with $\varrho\tau = 1$ per definition,

specific entropy S, which is the entropy per unit mass,

specific interior energy e,

specific enthalpy i, defined as $i = e + p\tau$.

We assume that these quantities do not depend on the space variable x in the balloon, and they do not depend on the time variable t.

It turns out that of these 7 quantities (for any fixed system), only two are independent, for instance S and τ. All the other five quantities can be expressed as functions of (S, τ), and these functions depend on the medium under consideration. We now concentrate on the function $e = e(S, \tau)$, and one of the key relations of thermodynamics is the formula

$$\mathrm{d}e = T\,\mathrm{d}S - p\,\mathrm{d}\tau,$$

or expressed in more mathematical style,

$$\frac{\partial e}{\partial S}(S,\tau) = T, \qquad \frac{\partial e}{\partial \tau}(S,\tau) = -p. \tag{1.1}$$

A *caloric equation of state*[2] then is

$$p = f(\varrho, S), \qquad p = g(S, \tau),$$

where f and g depend on the properties of the medium, and f, g are related to each other because of $\varrho = \tau^{-1}$. A physically reasonable assumption is

$$\frac{\partial f}{\partial \varrho} > 0,$$

[1] This exposition follows [5].
[2] kalorische Zustandsgleichung

which immediately implies

$$\frac{\partial g}{\partial \tau} < 0.$$

Then we may define $c := \sqrt{\partial p/\partial \varrho}$ as *sound speed* of the medium. Be careful: this formula holds only if p is written as a function of ϱ and S.

Another reasonable assumption from physics is g to be convex in τ, hence

$$\frac{\partial^2 g}{\partial \tau^2}(S,\tau) > 0,$$

and also

$$\frac{\partial g}{\partial S}(S,\tau) > 0.$$

Definition 1.1. *A medium is called* ideal gas *if*

$$p\tau = RT,$$

with R being a constant depending only on the medium[3].

Proposition 1.2. *Under additional assumptions to appear during the course of the proof, the interior energy of an ideal gas depends only on the temperature T.*

Quasi–Proof. We start with $e = e(S,\tau)$ and (1.1). Then we find

$$R\frac{\partial e}{\partial S} + \tau\frac{\partial e}{\partial \tau} = RT + \tau \cdot (-p) = 0,$$

or more in detail

$$R\frac{\partial e}{\partial S}(S,\tau) + \tau\frac{\partial e}{\partial \tau}(S,\tau) = 0. \tag{1.2}$$

This is a differential equation, and in particular, it is a

partial differential equation (PDE), because partial derivatives appear,

first order differential equation, because higher order derivatives are absent,

linear differential equation, because the operator $R\frac{\partial}{\partial S} + \tau\frac{\partial}{\partial \tau}$ is a linear operator.

PDEs are hard to investigate, which is the reason why this course concentrates on easier equations, so-called *ordinary differential equations (ODE)*, and now we rely on a flash of inspiration which recommends the ansatz

$$e(S,\tau) = h(\tau \cdot H(S)),$$

with unknown functions h and H. However, following this way we will never know if all solutions $e = e(S,\tau)$ to (1.2) can be expressed by this ansatz. Anyway, plugging the ansatz into (1.2) yields

$$Rh' \cdot \tau H'(S) + \tau h' \cdot H(S) = 0,$$

and the physical assumptions

$$\tau > 0, \qquad h' \neq 0$$

make division possible:

$$RH'(S) + H(S) = 0. \tag{1.3}$$

This is an *ordinary differential equation*, because no partial derivatives exist. Moreover, it is

[3] more precisely: R is the universal gas constant divided by the effective molecular weight of the gas under consideration

of first order,

linear,

homogeneous,

with constant coefficients. [4]

One more physical assumption is $H \neq 0$. Then we can divide once again, and

$$\begin{aligned}
R\frac{H'(S)}{H(S)} + 1 &= 0,\\
\frac{\mathrm{d}}{\mathrm{d}S}\ln|H(S)| &= -\frac{1}{R},\\
\int_{S=1}^{S_0}\frac{\mathrm{d}}{\mathrm{d}S}\ln|H(S)|\,\mathrm{d}S &= -\int_{S=1}^{S_0}\frac{1}{R}\,\mathrm{d}S,\\
\ln|H(S_0)| - \ln|H(1)| &= -\frac{1}{R}(S_0-1),\\
\ln\left|\frac{H(S_0)}{H(1)}\right| &= -\frac{1}{R}(S_0-1),\\
|H(S_0)| &= |H(1)|\exp\left(\frac{1}{R}\right)\exp\left(-\frac{S_0}{R}\right).
\end{aligned}$$

If we take the freedom to introduce a constant $C_0 \in \mathbb{R}$, then we can write

$$H(S_0) = C_0\exp(-S_0/R).$$

Incorporating this constant C_0 into an updated version of the function h, we then find

$$e(S,\tau) = h(\tau\exp(-S/R)),$$

and the consequences are then

$$p = -\frac{\partial e}{\partial \tau} = -h'(\tau\exp(-S/R))\exp(-S/R) = -h'(\varrho^{-1}\exp(-S/R))\exp(-S/R).$$

Because of $p > 0$ everywhere, this gives the necessary condition $h' < 0$ on the function h. We also wanted to have $\frac{\partial p(\varrho,S)}{\partial \varrho} > 0$, from which we deduce that $h'' > 0$.
Additionally,

$$T = \frac{\partial e}{\partial S}(S,\tau) = -\frac{1}{R}h'(\tau\exp(-S/R))\cdot\tau\exp(-S/R).$$

We put $y = \tau\exp(-S/R)$, and it follows that

$$T = -\frac{1}{R}h'(y)y,$$

hence T depends on y only. An information from physics is that this dependence is typically monotonically decreasing, and therefore an inverse function exists of the form $y = y(T)$. We can not express this function as a formula, but we know its existence. Then it follows that

$$e = h(\tau\exp(-S/R)) = h(y) = h(y(T)),$$

and indeed e depends on T only, but no other second thermodynamic quantity. □

For completeness, we list the assumptions made:

- e has the form $e = h(\tau H(S))$,
- $h' \neq 0$ and $H \neq 0$ (which means that these functions take nowhere the value zero),

[4] In comparison: (1.2) is also linear and homogeneous, but has variable coefficients.

- the function $T = T(y)$ is strictly monotone.

If the last condition seems to restrictive, we could it replace it by the condition that T has only a finite number of intervals of monotonicity, and consider only such systems where T stays in the same interval of monotonicity.

Corollary 1.3. *Also the sound speed depends only on the temperature, because of*

$$\begin{aligned} c^2 &= \frac{\partial p}{\partial \varrho}(\varrho, S) = -\frac{\partial}{\partial \varrho} h'(\varrho^{-1}) \exp(-S/R) \exp(-S/R) \\ &= h''(\varrho^{-1} \exp(-S/R))(\exp(-S/R))^2 \varrho^{-2} = h''(y) y^2 = h''(y(T)) \cdot (y(T))^2. \end{aligned}$$

Let us take a step back, go to the meta-level, and have a look what we have done so far. It is not the purpose of the math course to teach you thermodynamics. The topic of the first part of this semester are differential equations instead, and we wish to understand their solutions. This can be achieved two ways:

- Sometimes we have an explicit formular of the solution, and from this formula we can read off how the solution behaves. An example is the formula for H, which was found under the additional condition $H \neq 0$.
- In many cases, there is a solution, but we have no formula for it. Then we still have the desire to characterise the solution as thorough as possible. For instance, we have never found the function h, or the function $T = T(y)$, but we were still able to prove that e is a function of one thermodynamical quantity (namely T) alone.

1.2 General Considerations

Definition 1.4. *An* ordinary differential equation of order k[5] *is an equation of the form*

$$f(t, y(t), y'(t), \ldots, y^{(k)}(t)) = 0, \qquad t \in [t_0, t_1]$$

with a given function f and an unknown function $y = y(t)$.

Common variations are the following:

- y and f are vector valued,
- t and y are $\mathbb{C}$–valued, and the condition $t \in [t_0, t_1]$ is to be replaced by the condition of t coming from a closed set of $\mathbb{C}$.

We say that a function $y = y(t)$ is a solution if y is k times continuously differentiable, the vector $(t, y(t), y'(t), \ldots, y^{(k)}(t))$ belongs to the domain of f for all $t \in [t_0, t_1]$, and the differential equation holds for all such t.

Typical questions are:

- are there any solutions at all ?
- how many solutions do exist ? Does the solution set have a special structure ?
- can we give an explicit formula for the solution ?
- can we prove that all solutions have been found ?
- does the solution explode in finite time ? What is its life span ?
- how to find reasonable numerical methods to compute with acceptable effort approximate solutions ?
- can we make qualitative statements about the solutions without computing them ?

[5] gewöhnliche Differentialgleichung k-ter Ordnung

Examples of the above are:

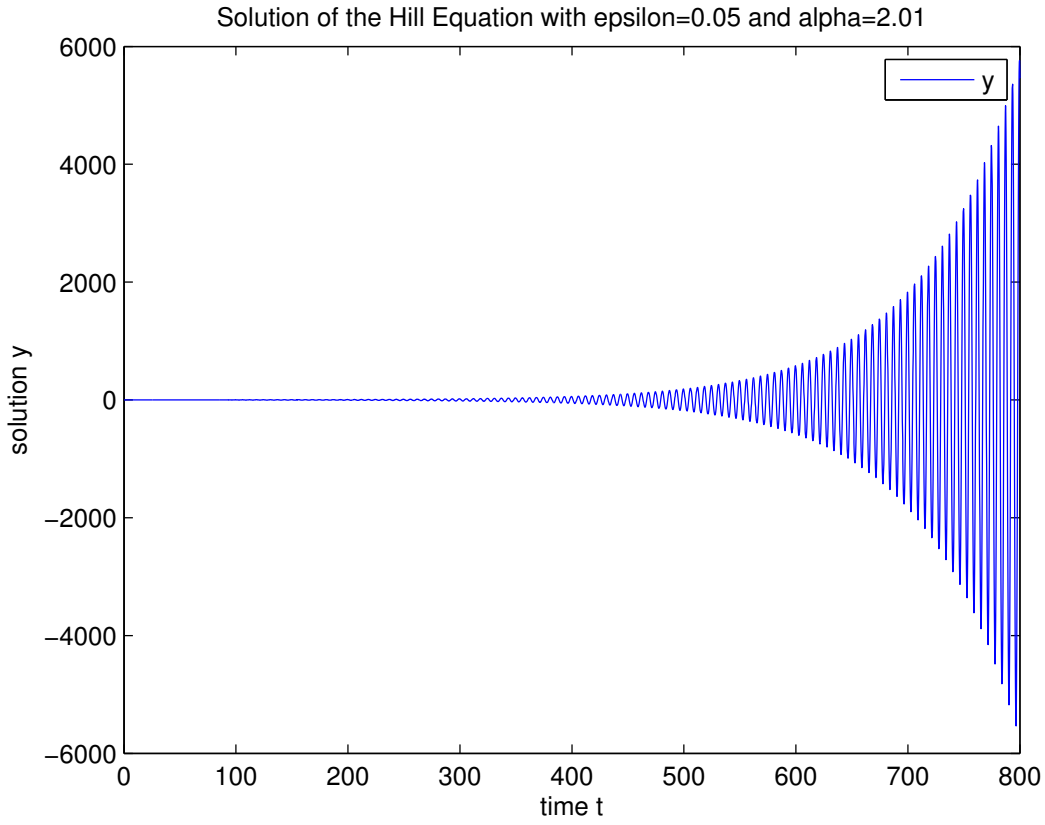
- the ODE $y' - \alpha y = 0$ has solutions $y(t) = Ce^{\alpha t}$ with $C \in \mathbb{R}$, and these are all solutions. If we prescribe an initial condition $y(0) = 29$, then the function $y(t) = 29e^{\alpha t}$ remains as only solution.
- the ODE $y'' + y = 0$ has solutions $y(t) = C_1 \cos t + C_2 \sin t$, for C_1, $C_2 \in \mathbb{R}$, hence the set of all solutions forms a two-dimensional vector space. The solution will be unique if we pose the initial condition $y(0) = 56$, $y'(0) = -27$. There will be *no* solution if we pose the boundary conditions $y(0) = 0$, $y(\pi) = 1$.
- the *initial value problem*[6]

  $$y'(t) = t^2 + y^2(t), \qquad y(0) = 1$$

  has exactly one solution, but a solution formula does not exist. Notwithstanding this obstacle, we are able to prove that the solution y has a pole, and this pole is between $\pi/4$ and 1.
- the following initial value problem (also known as HILL's equation)

  $$y''(t) + (1 + \varepsilon \cos(\alpha t))y(t) = 0, \qquad y(0) = 1, \quad y'(0) = 0$$

  with $0 < \varepsilon \ll 1$ and $\alpha \in \mathbb{R}$ occurs in the investigation of a pendulum whose length is periodically changing with time. There is no solution formula, but one can prove anyway that (assuming that ε and α are suitably chosen) the solution y suffers from an exponential resonance effect (in contrast to the linear resonance effect as it is known from oscillation equations like $y'' + ay' + by = \cos(\omega t + \delta)$). These resonance effects depend in a delicate manner on the values of the parameters, see the figures which have been obtained numerically using the `ode45` method of MATLAB.

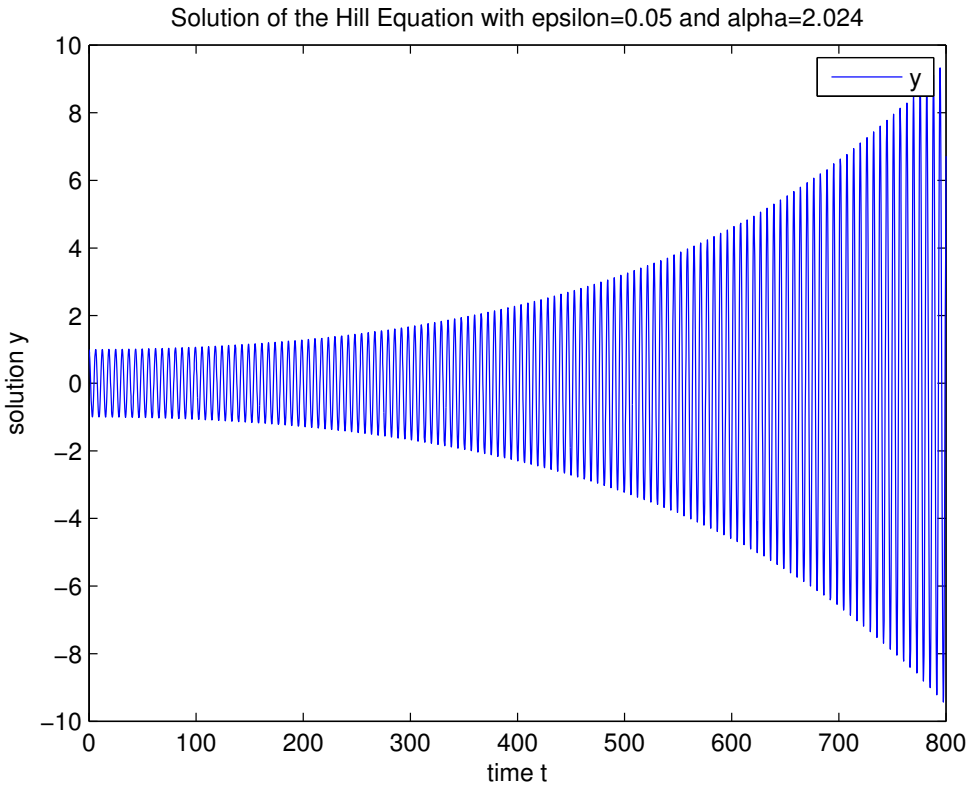


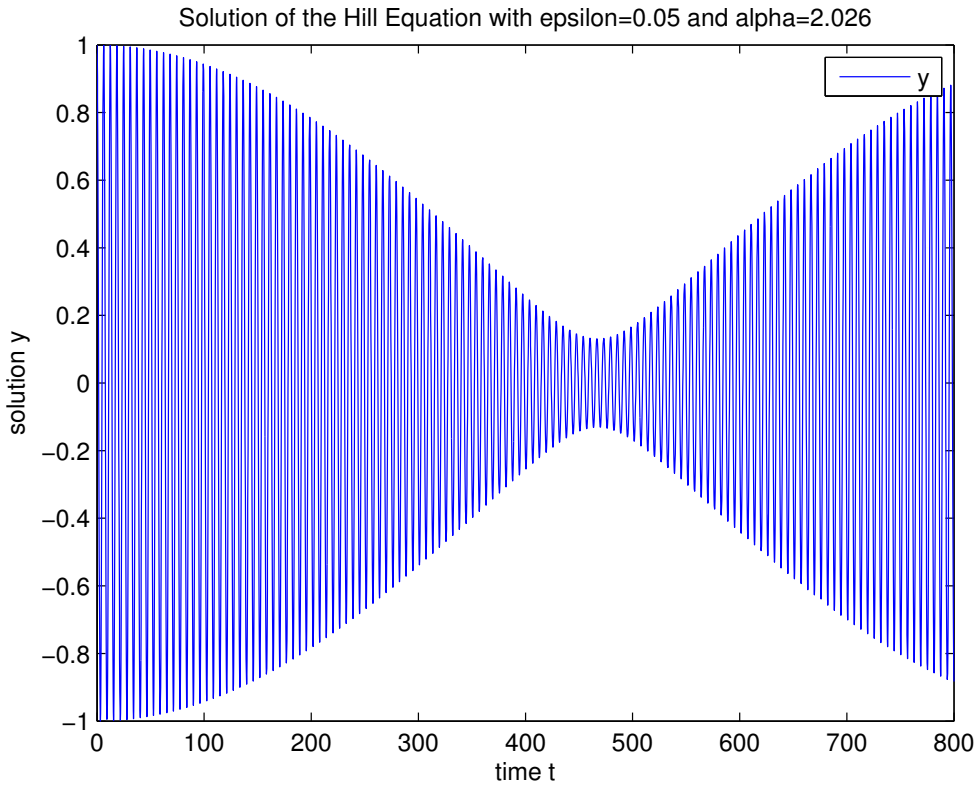
- the initial value problem

  $$y'(t) = \sqrt{|y|}, \qquad y(0) = 0$$

  has an infinite number of solutions.

[6] Anfangswertproblem


Solution of the Hill Equation with epsilon=0.05 and alpha=2.024
y
solution y
time t
10
8
6
4
2
0
−2
−4
−6
−8
−10
0
100
200
300
400
500
600
700
800


Solution of the Hill Equation with epsilon=0.05 and alpha=2.026
y
solution y
time t
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
0
100
200
300
400
500
600
700
800

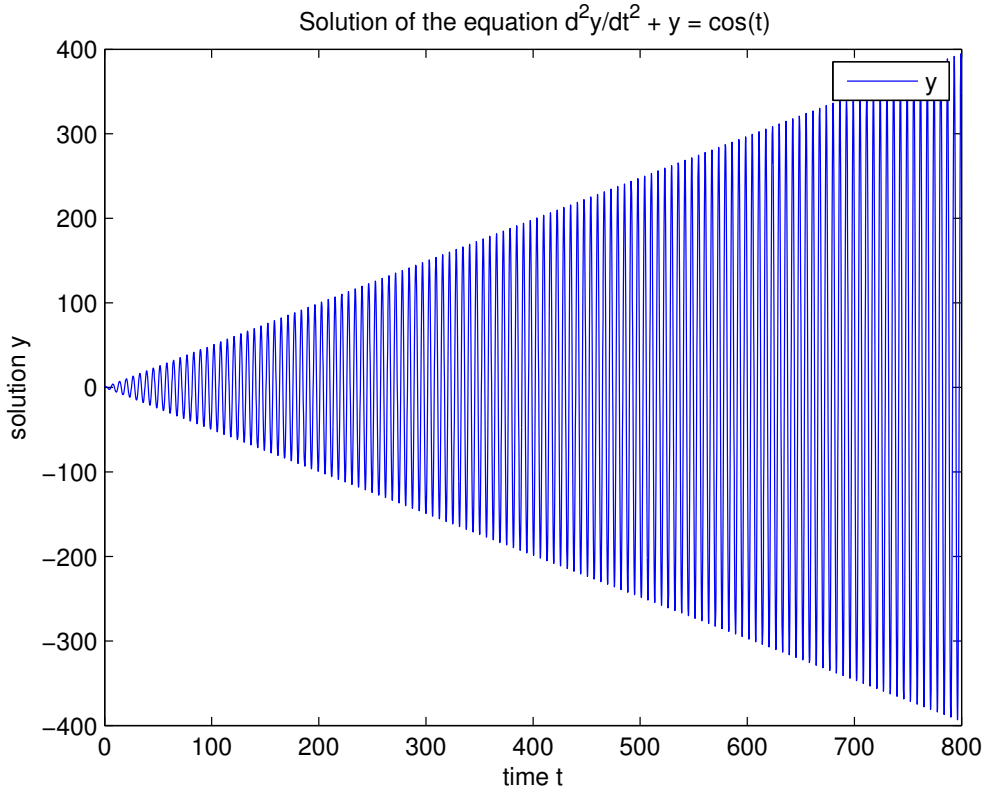


Figure 1.1: Linear resonance

- the Lorenz system is

$$\begin{aligned}\dot{x} &= \sigma(y-x),\\ \dot{y} &= \varrho x - y - xz,\\ \dot{z} &= -\beta z + xy,\end{aligned}$$

  with typical values $\sigma = 10$, $\beta = 8/3$, $\varrho = 28$. This model describes a heavily simplified system from meteorology. It is of first order, a system, nonlinear, and there is definitely no solution formula. But anyway it is possible to give at least a rough description of the long time asymptotics, leading to the celebrated *Lorenz attractor*.

- the logistic growth model is

$$y'(t) = \alpha y(t) - \beta(y(t))^2,$$

  describing the growth of an idealised population of micro-organisms. Here $y(t)$ denotes the mass of the population at time t. A solution formula exists, but independent of that we are able to prove rigorously that the values $y(t)$ can never become negative if $y(0)$ is positive. Clearly, negative values of $y(t)$ would be biological nonsense. We are also able to show that $y(t)$ exists globally in time, whatever the non-negative starting value $y(0)$ is.

Definition 1.5 (**Explicit DE**). *An ODE of order k is called* explicit *if it has the form*

$$y^{(k)}(t) = g(t, y(t), y'(t), \dots, y^{(k-1)}(t)).$$

The implicit function theorem from the second semester gives us a criterion when an implicit ODE can be transformed into an explicit ODE.

Proposition 1.6. *Each explicit ODE of order k can be equivalently transformed into a first order system.*

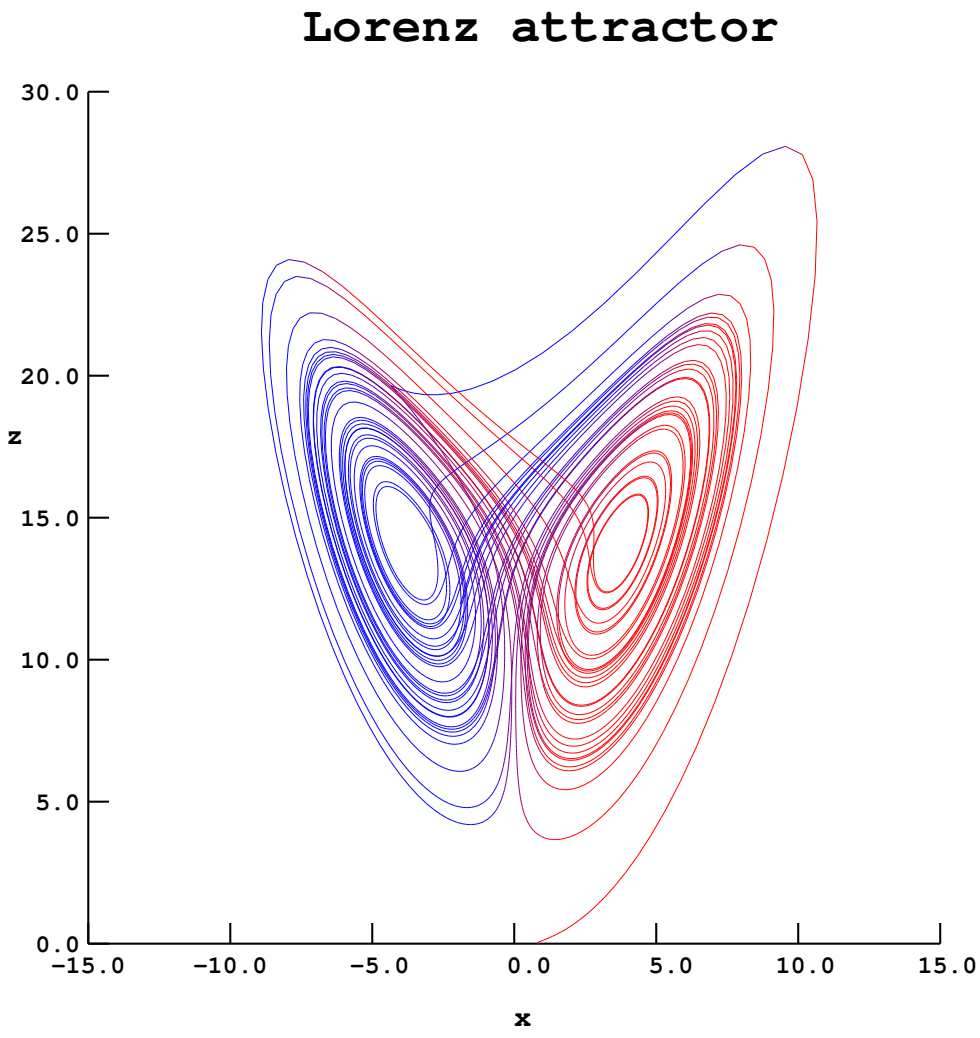


Figure 1.2: Lorenz attractor, generated with `vplot`. The component $y(t)$ is indicated by the colour, with red positive and blue negative.

Proof. Let us be given the explicit ordinary differential equation

$$y^{(k)}(t) = g(t, y, y', \ldots, y^{(k-1)}), \tag{1.4}$$

and define

$$u_1(t) := y(t), \qquad u_2(t) := y'(t), \quad \ldots, \quad u_k(t) := y^{(k-1)}(t),$$
$$U := (u_1, \ldots, u_k)^\top.$$

Then we find the system

$$U'(t) = \begin{pmatrix} u_2 \\ u_3 \\ \vdots \\ u_k(t) \\ g(t, u_1, u_2, \ldots, u_k) \end{pmatrix}. \tag{1.5}$$

And now it is obvious: if y solves (1.4), then the vector U constructed above solves the first order system (1.5). And conversely: if U solves (1.5), then the first component u_1 of U is a solution to (1.4). □

The key advantage of this result is that it enables us to focus our further studies on explicit first order systems.

For a single equation $y' = f(t, y)$ without solution formula, it might be helpful to draw so-called *slope fields*[7] to obtain a rough idea how the solutions look like.

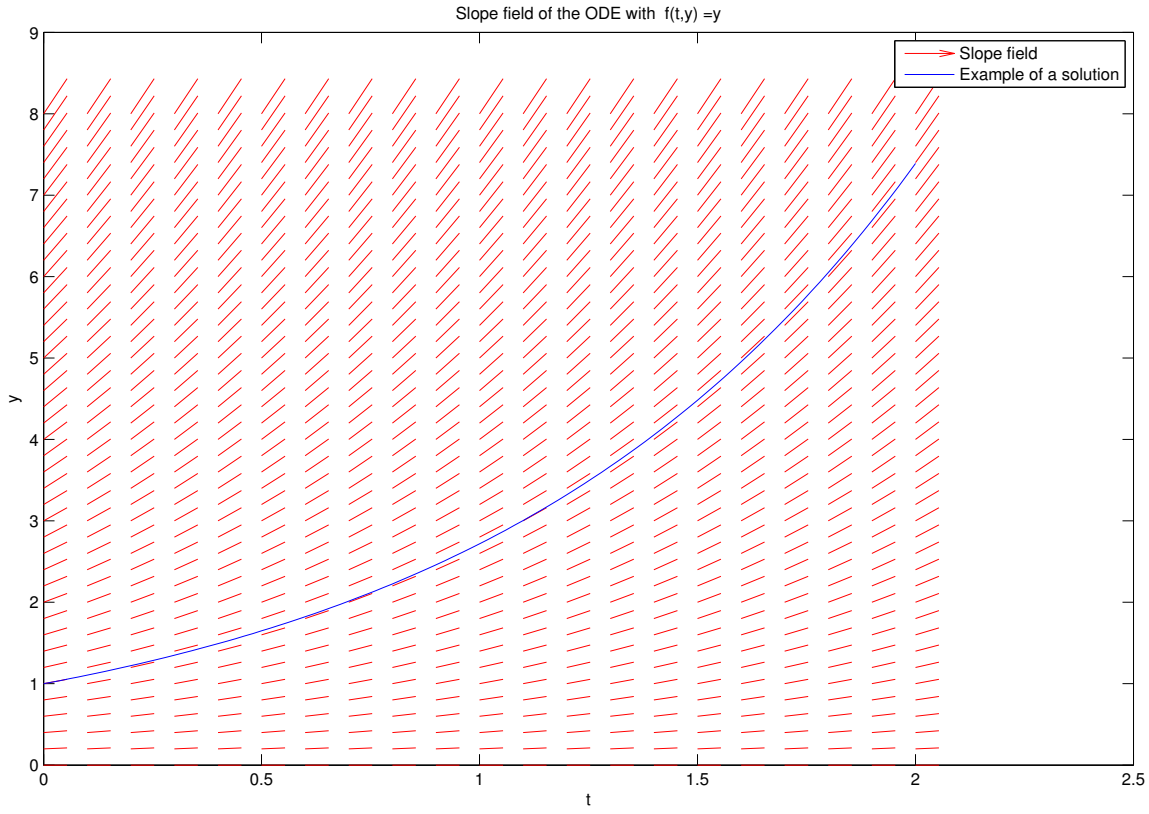


Figure 1.3: A slope field for the differential equation $y' = f(t, y) = y$

[7] Richtungsfelder

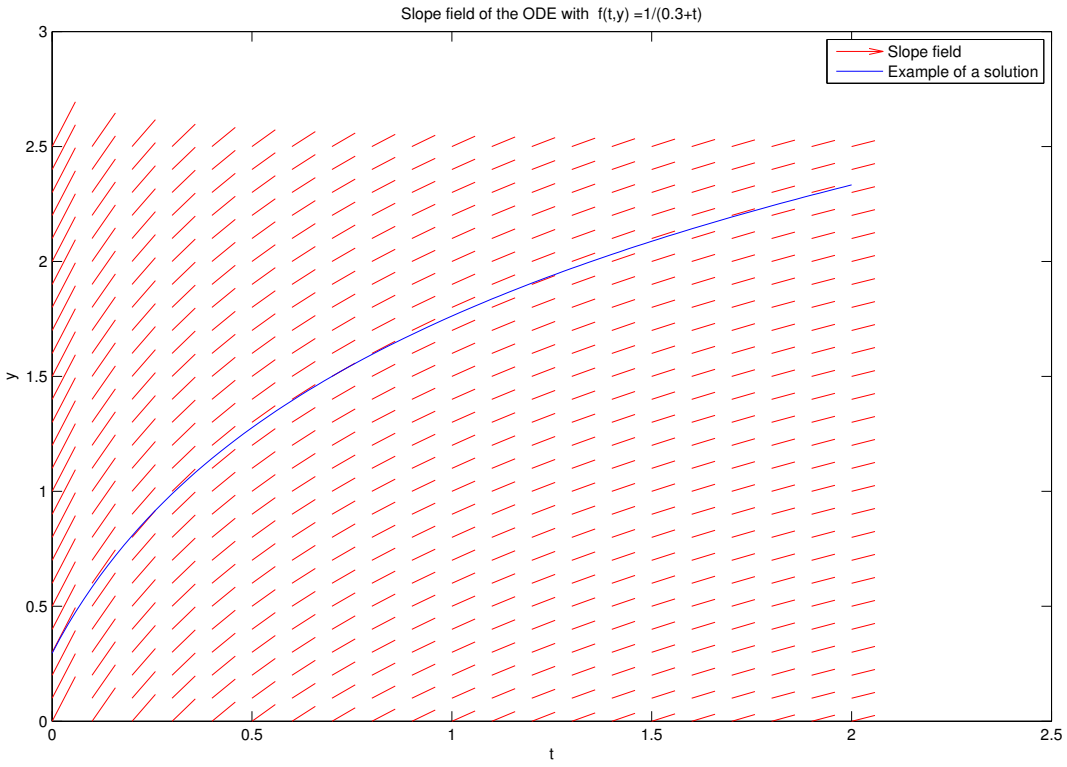


Figure 1.4: A slope field for the differential equation $y' = f(t, y) = 1/(0.3 + t)$

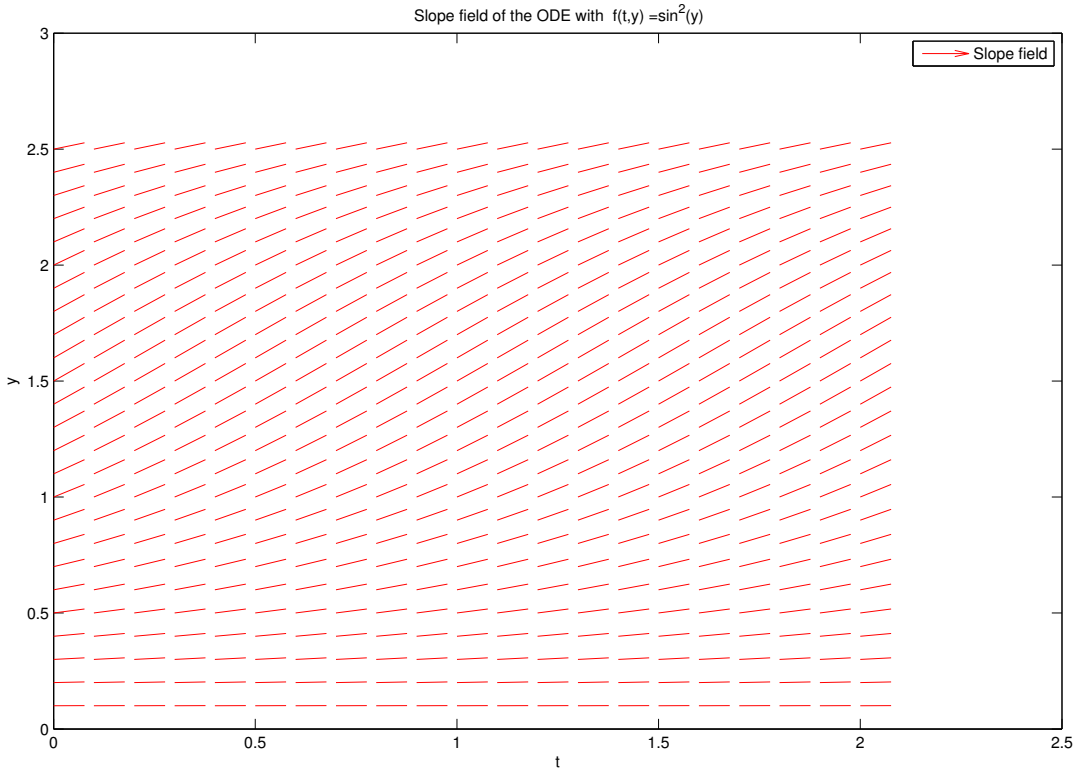


Figure 1.5: A slope field for the differential equation $y' = f(t, y) = (\sin y)^2$

1.3 The Theorem of Picard and Lindelöf

8 9

We recall the *Banach fixed point theorem* from the first semester:

Theorem 1.7. *Let U be a Banach space, $M \subset U$ a closed subset, and Φ a mapping with the properties*

- *Φ maps M into itself,*
- *Φ is contractive, which means that there is a number α less than one such that $\|\Phi(u) - \Phi(\tilde{u})\|_U \leq \alpha \|u - \tilde{u}\|_U$ for all u, $\tilde{u} \in M$.*

Then there is a unique fixed point u^* of Φ, in the sense of $\Phi(u^*) = u^*$, and any sequence $(u_n)_{n\in\mathbb{N}}$ defined by selecting $u_0 \in M$ freely and setting $u_{n+1} := \Phi(u_n)$ converges to this fixed point u^*, and we have the error estimate

$$\|u_n - u^*\|_U \leq \frac{\alpha^n}{1-\alpha} \|u_1 - u_0\|_U .$$

Theorem 1.8 (Global Version of the Picard-Lindelöf Theorem). *Let $I := [a, b] \subset \mathbb{R}$, $t_0 \in I$ and $y_0 \in \mathbb{R}^n$, and $f\colon I \times \mathbb{R}^n \to \mathbb{R}^n$ be a continuous function which satisfies a* global ***Lipschitz***[10] *condition with respect to y as follows:*

$$\exists L > 0\colon \quad \forall t \in I, \quad \forall y_1, y_2 \in \mathbb{R}^n\colon \ \|f(t, y_1) - f(t, y_2)\|_{\mathbb{R}^n} \leq L \|y_1 - y_2\|_{\mathbb{R}^n} .$$

Then the initial value problem

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0 \tag{1.6}$$

possesses a unique solution $y \in C^1(I \to \mathbb{R}^n)$.

Proof.

Step 1: transformation to an integral equation

If $y \in C^1(I \to \mathbb{R}^n)$ is a solution to (1.6), then the Fundamental Theorem of Calculus asserts

$$y(t) = y_0 + \int_{s=t_0}^{t} f(s, y(s))\, \mathrm{d}s =: (\Phi(y))(t), \qquad t \in I. \tag{1.7}$$

Conversely: if $y \in C(I \to \mathbb{R}^n)$ is a solution to (1.7), then (again by the Fundamental Theorem of Calculus), the integral in the right–hand side of (1.7) is a differentiable function of t, hence $y \in C^1(I \to \mathbb{R}^n)$, and we are able to differentiate (1.7), obtaining (1.6).

Step 2: applying the Banach Fixed Point Theorem

We choose the Banach space $U = C(I \to \mathbb{R}^n)$ with the special norm

$$\|u\|_U := \sup_{t\in I} e^{-(L+1)|t-t_0|} \|u(t)\|_{\mathbb{R}^n} ,$$

with L being the constant appearing in the Lipschitz condition. The closed set M shall be $M = U$, and the map Φ is as in (1.7). Obviously, Φ maps M into itself, and we only have to check whether Φ contracts.

First we note that

$$\Big(\Phi(u) - \Phi(\tilde{u})\Big)(t) = \int_{s=t_0}^{t} f(s, u(s)) - f(s, \tilde{u}(s))\, \mathrm{d}s,$$

[8] Charles Émile Picard, 1856 – 1941
[9] Ernst Leonard Lindelöf, 1870 – 1946
[10] Rudolf Otto Sigismund Lipschitz, 1832 – 1903

and therefore we find

$$
\begin{aligned}
e^{-(L+1)|t-t_0|} \left\| \Big(\Phi(u) - \Phi(\tilde{u})\Big)(t) \right\|_{\mathbb{R}^n} &= e^{-(L+1)|t-t_0|} \left\| \int_{s=t_0}^{t} f(s,u(s)) - f(s,\tilde{u}(s))\,\mathrm{d}s \right\|_{\mathbb{R}^n} \\
&\le e^{-(L+1)|t-t_0|} \int_{s=\min(t,t_0)}^{\max(t,t_0)} \|f(s,u(s)) - f(s,\tilde{u}(s))\|_{\mathbb{R}^n}\,\mathrm{d}s \\
&\le e^{-(L+1)|t-t_0|} \int_{s=\min(t,t_0)}^{\max(t,t_0)} L\,\|u(s) - \tilde{u}(s)\|_{\mathbb{R}^n}\,\mathrm{d}s \\
&\le e^{-(L+1)|t-t_0|} L \int_{s=\min(t,t_0)}^{\max(t,t_0)} e^{(L+1)|s-t_0|} \underbrace{e^{-(L+1)|s-t_0|}\,\|u(s) - \tilde{u}(s)\|_{\mathbb{R}^n}}_{\le \|u-\tilde{u}\|_U}\,\mathrm{d}s \\
&\le \|u-\tilde{u}\|_U\, e^{-(L+1)|t-t_0|} L \int_{s=\min(t,t_0)}^{\max(t,t_0)} e^{(L+1)|s-t_0|}\,\mathrm{d}s \\
&= \|u-\tilde{u}\|_U\, e^{-(L+1)|t-t_0|} L \cdot \frac{1}{L+1}\left(e^{(L+1)|t-t_0|} - e^0\right) \\
&\le \frac{L}{L+1}\,\|u-\tilde{u}\|_U\,.
\end{aligned}
$$

This shows that Φ is contractive with constant $\alpha = L/(L+1) < 1$. Therefore the Banach fixed point theorem assures us that Φ has exactly one fixed point y^* in $M = U$.

This fixed point is then the desired solution of (1.7), hence also of (1.6). □

The proof remains the same if we consider a function $f\colon I \times \mathbb{C}^n \to \mathbb{C}^n$ and look for a solution $y \in C^1(I \to \mathbb{C}^n)$.

The global Lipschitz condition can be written as

$$\frac{\|f(t,y_1) - f(t,y_2)\|_{\mathbb{R}^n}}{\|y_1 - y_2\|_{\mathbb{R}^n}} \le L < \infty$$

for all $(t, y_1, y_2) \in I \times \mathbb{R}^n \times \mathbb{R}^n$, which can be understood as a limit on the slope of the secant lines of f. Unfortunately, the function $f(t,y) = t^2 + y^2$ does not satisfy this global Lipschitz condition because of

$$\frac{|f(t,y_1) - f(t,y_2)|}{|y_1 - y_2|} = \frac{|y_1^2 - y_2^2|}{|y_1 - y_2|} = |y_1 + y_2|,$$

which can become arbitrarily large if y_1 and y_2 are allowed to move freely in the whole $\mathbb{R}^1$. Therefore the global version of the Picard–Lindelöf Theorem is not applicable to this function f. We can overcome this trouble if we refine the Picard–Lindelöf Theorem, at the price of a slightly more technical proof.

Theorem 1.9 (Local Version of the Picard-Lindelöf Theorem). *Let $I := [a,b] \subset \mathbb{R}$, t_0 be an interior point of I, and $y_0 \in \mathbb{R}^n$, and define B_R as a closed ball about y_0 with radius R:*

$$B_R := \{y \in \mathbb{R}^n\colon\ \|y - y_0\|_{\mathbb{R}^n} \le R\}.$$

Let a function $f\colon I \times B_R \to \mathbb{R}^n$ be continuous and assume that it satisfies a local *Lipschitz condition with respect to y as follows:*

$$\exists L > 0\colon \quad \forall t \in I, \quad \forall y_1, y_2 \in B_R\colon\ \|f(t,y_1) - f(t,y_2)\|_{\mathbb{R}^n} \le L\,\|y_1 - y_2\|_{\mathbb{R}^n}\,.$$

Then there is a positive ε such that the initial value problem

$$y'(t) = f(t,y(t)), \qquad y(t_0) = y_0$$

possesses a unique solution $y \in C^1([t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n)$.

Sketch of proof. Choose $U = C([t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n)$ with an ε not yet specified, and with the same norm as in the old proof. The closed set $M \subset U$ shall consist of all those $y \in U$ whose values $y(t)$ stay inside B_R for $t_0 - \varepsilon \le t \le t_0 + \varepsilon$.

If you choose ε sufficiently small, you then can show that Φ maps M into itself, and the proof of contractivity is the same as before. □

Remark 1.10. *In case of the initial value problem*

$$y'(t) = \sqrt{|y(t)|}, \qquad y(0) = 0, \tag{1.8}$$

we have $f(t,y) = \sqrt{|y|}$ which violates the (global and local) Lipschitz condition. The Picard–Lindelöf Theorem can not be applied, and it is a wonderful homework to find at least two different solution $y = y(t)$ to (1.8).

Lemma 1.11. *If a solution $y = y(t)$ to a differential equation $y' = f(t,y)$ ceases to exist at a time t_{end}, then one of the following two events occurs:*

- *the function y has a pole at t_{end},*
- *the point $(t_{\text{end}}, y(t_{\text{end}}))$ is at the boundary of the domain of definition of f.*

Sketch of proof. Otherwise $y(t_{\text{end}})$ is finite and the point $(t_{\text{end}}, y(t_{\text{end}}))$ is in the interior of the domain of definition of f. Then we can apply the local version of the Picard–Lindelöf theorem once again, but now with starting time t_{end}, giving us an extension of the function y beyond the time t_{end}, which is a contradiction. □

In theory, one could use the Picard–Lindelöf method numerically for ODEs where an explicit solution formula can not be found; however, later we will learn numerical methods which are much stronger at similar computational effort.

We come to a crucial consequence.

Proposition 1.12. *Let $y_1 = y_1(t)$ and $y_2 = y_2(t)$ be two solutions to*

$$y'(t) = f(t, y(t)),$$

where f is continuous, with continuous derivative $\partial f/\partial y$.

Then the following holds: if y_1 and y_2 coincide at a certain time t^*, then they coincide always.

Proof. The initial value problem with starting time t^* has exactly one solution, not two. □

Example 1.13. *The model of logistic growth*

$$u'(t) = \alpha u(t) - \beta (u(t))^2, \qquad \alpha, \beta > 0,$$

has two stationary solutions: $u \equiv 0$ and $u \equiv \alpha/\beta$. If $0 < u(0) < \alpha/\beta$, then also $u(t)$ for all $t \in \mathbb{R}$ must remain between 0 and α/β, which is what we wanted to know. And by Lemma 1.11, this solution u exists for eternity.

Then $u'(t)$ is positive because of

$$u'(t) = \alpha u(t) \cdot \left(1 - \frac{\beta}{\alpha} u(t)\right).$$

Therefore u is growing and bounded from above, hence $\lim_{t\to\infty} u(t)$ exists, and it must be equal to α/β, because otherwise u would not stop growing.

Next we wish to understand how errors in the initial data propagate. Consider the initial value problems

$$\begin{aligned} u'(t) &= f(t,u), & u(0) &= u_0, \\ v'(t) &= f(t,v), & v(0) &= v_0, \end{aligned}$$

with $|u_0 - v_0| \ll 1$. Our goal is to find an estimate for $|u(t) - v(t)|$, where we assume the usual Lipschitz condition on f.

Lemma 1.14. *The value $u(t)$ depends (for each fixed t) continuously on u_0.*

Proof. We start with

$$u(t) = u_0 + \int_{s=0}^{t} f(s, u(s))\,\mathrm{d}s,$$
$$v(t) = v_0 + \int_{s=0}^{t} f(s, v(s))\,\mathrm{d}s,$$

hence also (assuming $t > 0$ for simplicity)

$$\begin{aligned} |u(t) - v(t)| &\le |u_0 - v_0| + \int_{s=0}^{t} |f(s, u(s)) - f(s, v(s))|\,\mathrm{d}s \\ &\le |u_0 - v_0| + L \int_{s=0}^{t} |u(s) - v(s)|\,\mathrm{d}s, \\ e^{-(L+1)t}|u(t) - v(t)| &\le e^{-(L+1)t}|u_0 - v_0| + Le^{-(L+1)t} \int_{s=0}^{t} e^{(L+1)s} e^{-(L+1)s} |u(s) - v(s)|\,\mathrm{d}s. \end{aligned}$$

We are interested in t running in a time interval $[0, t_{\max}]$, hence we have $0 \le s \le t \le t_{\max}$.
We put $w(t) = e^{-(L+1)t}|u(t) - v(t)|$ for brevity of notation, and then it follows that

$$\begin{aligned} w(t) &\le w(0) + Le^{-(L+1)t} \int_{s=0}^{t} e^{(L+1)s} w(s)\,\mathrm{d}s \\ &\le w(0) + Le^{-(L+1)t} \int_{s=0}^{t} e^{(L+1)s} \left(\sup_{0 \le z \le t_{\max}} w(z) \right) \mathrm{d}s \\ &= w(0) + Le^{-(L+1)t} \left(\sup_{0 \le z \le t_{\max}} w(z) \right) \int_{s=0}^{t} e^{(L+1)s}\,\mathrm{d}s \\ &\le w(0) + Le^{-(L+1)t} \left(\sup_{0 \le z \le t_{\max}} w(z) \right) \cdot \frac{1}{L+1} e^{(L+1)t} \\ &= w(0) + \frac{L}{L+1} \left(\sup_{0 \le z \le t_{\max}} w(z) \right), \end{aligned}$$

from which we then obtain that

$$\left(\sup_{0 \le t \le t_{\max}} w(t) \right) \le w(0) + \frac{L}{L+1} \left(\sup_{0 \le z \le t_{\max}} w(z) \right),$$

hence also

$$\frac{1}{L+1} \left(\sup_{0 \le t \le t_{\max}} w(t) \right) \le w(0),$$

which can be re-arranged to

$$|u(t) - v(t)| \le (L+1)e^{(L+1)t}|u_0 - v_0|, \qquad \forall\, t \in [0, t_{\max}].$$

Therefore $|u(t) - v(t)|$ will be small if $|u_0 - v_0|$ is small. Note however that this estimate will be only of limited use for large values of the product $(L+1)t$. □

This estimate of $|u(t) - v(t)|$ against $|u_0 - v_0|$ will in general not hold if the function f does not satisfy a Lipschitz condition. As an example, we consider $y' = \sqrt{|y|}$ with the initial values $u_0 = 0$ and $v_0 = 10^{-12}$. The solutions are computed numerically with the `ode45` method of MATLAB, and it can be seen how even a tiny error in the initial values grows extremely fast.

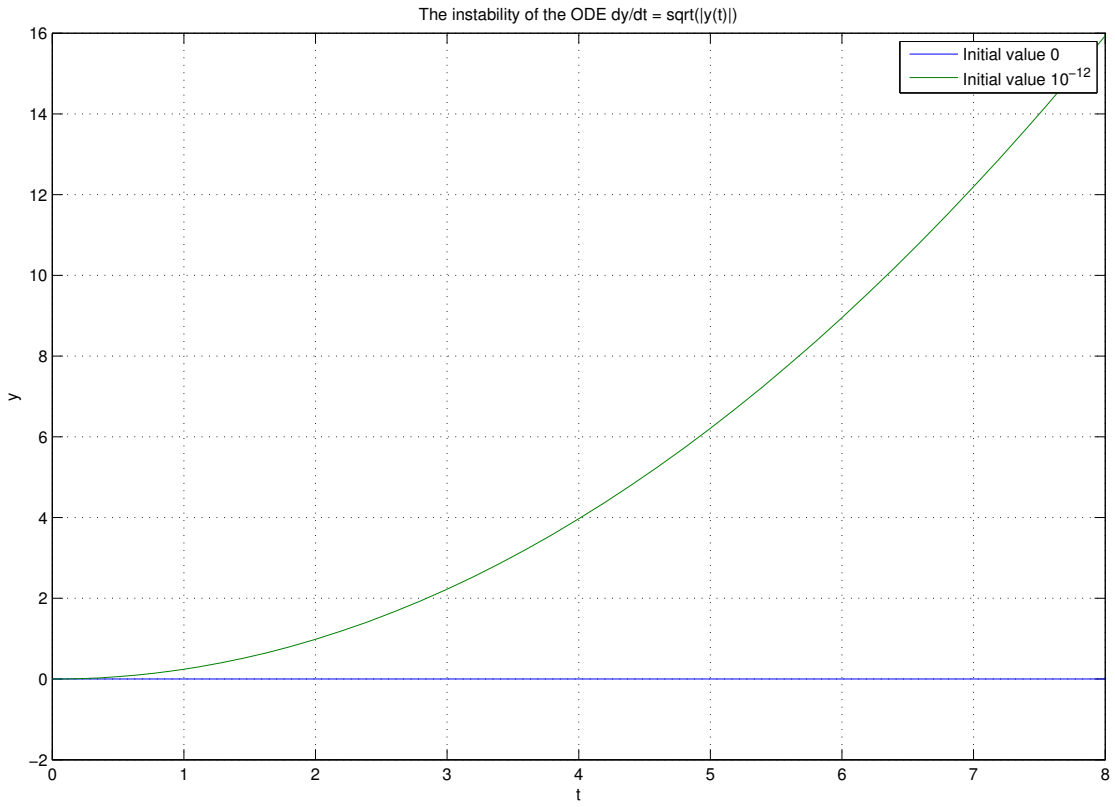
Just for completeness, we mention one final existence result:

Theorem 1.15 (PEANO[11]). *Let $I \subset \mathbb{R}$, t_0 be an interior point of I, and let $f = f(t, y)$ be a continuous function mapping from $I \times B_R$ into $\mathbb{R}^n$, with B_R being a ball in $\mathbb{R}^n$ with radius R and centre y_0. Then the system*

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0$$

has at least one local solution $u \in C^1([t_0 - \varepsilon, t_0 + \varepsilon] \to \mathbb{R}^n)$, for some small positive ε.

[11] GIUSEPPE PEANO, 1858 – 1932, mathematician and linguist, inventor of the symbols $\in$, $\cap$, $\cup$, $\exists$. He was the first to give axiomatic definitions of the natural numbers and of abstract vector spaces. Also famous for the crazy Peano curve.



The proof is quite long, and it would require more knowledge than we have, therefore we skip it. Note that nothing is said about the uniqueness of the solution, and indeed the example $y' = \sqrt{|y|}$ with $y_0 = 0$ shows that uniqueness can not be expected.

1.4 Comparison Principles

Now we want to compare different solutions, which is of course possible only for y and f taking values in $\mathbb{R}^1$ (not $\mathbb{R}^n$ or $\mathbb{C}^n$).

We have already one comparison principle: if u and v are solutions to

$$\begin{cases} u'(t) = f(t, u(t)), \\ u(0) = u_0, \end{cases} \qquad \begin{cases} v'(t) = f(t, v(t)), \\ v(0) = v_0, \end{cases}$$

with $u_0 < v_0$, then $u(t) < v(t)$ for all times t for which both solutions exist, by Proposition 1.12.

Now we generalise this to initial value problems with differing right–hand sides.

Proposition 1.16. *Let u and v be solutions to*

$$\begin{cases} u'(t) = f(t, u(t)), \\ u(0) = u_0, \end{cases} \qquad \begin{cases} v'(t) = g(t, v(t)), \\ v(0) = v_0, \end{cases}$$

which exist on the time interval $[0, T]$. Suppose that

$$\begin{aligned} f(t, y) \le g(t, y), \qquad & \forall\ (t, y) \in [0, T] \times \mathbb{R}, \\ u_0 < v_0, & \end{aligned}$$

and that f, g satisfy the usual Lipschitz condition with constant L. Then $u(t) < v(t)$ for $0 \le t \le T$.

Proof. The inequality $u(t) < v(t)$ holds for $t = 0$ by assumption, and both u, v are continuous functions. Therefore $u(t) < v(t)$ at least on a short time interval $[0, \varepsilon)$. Consequently, we have to bring the following situation

$$\begin{aligned} u(t) &< v(t), && 0 \le t < T_1 \le T, \\ u(T_1) &= v(T_1) \end{aligned}$$

to a contradiction (draw a picture !). Take $t \in [0, T_1)$. Then

$$\begin{aligned} v'(t) - u'(t) &= g(t, v(t)) - f(t, u(t)) \\ &\ge f(t, v(t)) - f(t, u(t)) \\ &\ge -L|v(t) - u(t)| \\ &= -L(v(t) - u(t)). \end{aligned}$$

We can divide by a positive number, since $t < T_1$:

$$\begin{aligned} \frac{v'(t) - u'(t)}{v(t) - u(t)} &\ge -L, \\ \frac{\mathrm{d}}{\mathrm{d}t} \ln\left(v(t) - u(t)\right) &\ge -L, \\ \int_{t=0}^{T_1 - \delta} \frac{\mathrm{d}}{\mathrm{d}t} \ln(v(t) - u(t))\, \mathrm{d}t &\ge -L(T_1 - \delta), \qquad (0 < \delta \ll 1), \\ \ln(v(T_1 - \delta) - u(T_1 - \delta)) - \ln(v(0) - u(0)) &\ge -L(T_1 - \delta), \\ \ln \frac{v(T_1 - \delta) - u(T_1 - \delta)}{v_0 - u_0} &\ge -L(T_1 - \delta), \\ v(T_1 - \delta) - u(T_1 - \delta) &\ge e^{-L(T_1 - \delta)}(v_0 - u_0). \end{aligned}$$

Now we send δ to zero and find

$$v(T_1) - u(T_1) \ge e^{-LT_1}(v_0 - u_0) > 0.$$

But the assumption of our situation was $u(T_1) = v(T_1)$, giving us $0 > 0$, which is nonsense. □

Question: Why did we first integrate only till $T_1 - \delta$, and have then sent δ to zero ?

This result can be made a bit stronger:

Lemma 1.17. *Let the assumptions of the previous Proposition hold, but now with $u_0 \le v_0$ instead of $u_0 < v_0$. Then $u(t) \le v(t)$ for $0 \le t \le T$.*

Proof. For $0 < \varepsilon \ll 1$, put $u_{0,\varepsilon} := u_0 - \varepsilon$ and let u_ε be the solution to

$$u_\varepsilon'(t) = f(t, u_\varepsilon(t)), \qquad u_\varepsilon(0) = u_{\varepsilon,0}.$$

Then u_ε exists up to $t = T$ (if ε is small enough), and $u_\varepsilon(t) < v(t)$ by Proposition 1.16. However, Lemma 1.14 gives us $u(t) = \lim_{\varepsilon \to 0} u_\varepsilon(t)$, which implies $u(t) \le v(t)$. □

Example 1.18. *Let $y = y(t)$ be the solution to*

$$y'(t) = t^2 + y^2(t), \qquad y(0) = 1.$$

There is no solution formula for this initial value problem. However, if $x = x(t)$ and $z = z(t)$ are the solutions to

$$\begin{aligned} x'(t) &= x^2(t), && x(0) = 1, \\ z'(t) &= 1 + z^2(t), && z(0) = 1, \end{aligned}$$

then $x(t) \le y(t) \le z(t)$ as long as all three solutions exist and $t \le 1$. The advantage is that x and z are easy to guess, and then it follows that y must have a pole between $\pi/4$ and 1.

Figuring out the details is an excellent homework.

Chapter 2

Special Solution Methods

In the previous chapter, we have learned under which conditions solutions *exist*, and now we want to find *solution formulae* if possible. All equations here are scalar equations (not systems).

2.1 Equations with Separable Variables

Equations with separable variables[1] are equations of the form

$$y'(t) = g(t) \cdot h(y(t)),$$

with continuous functions g and h, and we start our investigations with the example

$$y'(t) = -2ty^2(t), \qquad y(t_0) = y_0.$$

Here $g(t) = -2t$ and $h(y) = y^2$. Assuming $y(t) \neq 0$ we can divide:

$$\frac{y'}{y^2} = -2t.$$

Note that the two variables y and t have been separated: on the left-hand side only terms with y appear, and on the right-hand side only terms with t appear.

We suppose that $y^2(s) \neq 0$ for $t_0 \leq s \leq t$, and integrate:

$$\begin{aligned}
&\int_{s=t_0}^{t} \frac{y'(s)}{y^2(s)}\,\mathrm{d}s = -\int_{s=t_0}^{t} 2s\,\mathrm{d}s, \\
\implies \quad &-\left(\frac{1}{y(t)} - \frac{1}{y(t_0)}\right) = -(t^2 - t_0^2), \\
\implies \quad &\frac{1}{y(t)} = \frac{1}{y_0} + t^2 - t_0^2, \\
\implies \quad &y(t) = \frac{1}{t^2 - t_0^2 + \frac{1}{y_0}}.
\end{aligned}$$

A further example is $y' = \cos^2(y)$, and if we apply the method mindlessly, we might get a calculation like this:

$$y' = \cos^2(y) \quad \overset{?}{\iff} \quad \frac{\mathrm{d}y}{\mathrm{d}t} = \cos^2(y) \quad \overset{?}{\iff} \quad \frac{\mathrm{d}y}{\cos^2(y)} = \mathrm{d}t$$
$$\overset{?}{\iff} \quad \int \frac{\mathrm{d}y}{\cos^2(y)} = \int \mathrm{d}t \quad \overset{?}{\iff} \quad \tan(y) = t + C \quad \overset{?}{\iff} \quad y(t) = \arctan(t + C),$$

for arbitrary $C \in \mathbb{R}$. After finishing this calculation we may test this candidate for a solution, and it turns out that these functions are indeed solutions, for each $C \in \mathbb{R}$.

[1] Gleichungen mit trennbaren Variablen

However, now we have lost the solutions $y \equiv \pi/2$ and $y \equiv -\pi/2$, so we should try to exercise more care.

An important example are linear ODEs

$$y'(t) = a(t)y(t), \qquad y(t_0) = y_0,$$

with a continuous function $a = a(t)$. Before embarking on the calculations, let us think about the shape of the solutions. The right-hand side $f = f(t, y) = a(t)y$ is continuous, and its derivative $\partial f/\partial y = a(t)$ is also continuous, making the local Picard–Lindelöf theorem available, and in particular we know that different solution trajectories do not cross, see Proposition 1.12. Now it is obvious that $y \equiv 0$ is a (quite boring) solution, and we are interested in the other solutions. Then we can conclude that such an other solution can never change its sign, and if it is zero at some time, it must be zero always.

Assuming $y_0 \neq 0$ is enough to deduce that $y(t) \neq 0$ for all $t \in \mathbb{R}$, and we can divide:

$$\begin{aligned}
&\frac{y'(t)}{y(t)} = a(t) \qquad \Longrightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t} \ln|y(t)| = a(t) \qquad \Longrightarrow \qquad \int_{s=t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}s} \ln|y(s)| \,\mathrm{d}s = \int_{s=t_0}^{t} a(s) \,\mathrm{d}s \\
&\Longrightarrow \qquad \ln|y(s)|\Big|_{s=t_0}^{s=t} = \int_{s=t_0}^{t} a(s) \,\mathrm{d}s \qquad \Longrightarrow \qquad \ln\left|\frac{y(t)}{y(t_0)}\right| = \int_{s=t_0}^{t} a(s) \,\mathrm{d}s \\
&\Longrightarrow \qquad |y(t)| = |y_0| \exp\left(\int_{s=t_0}^{t} a(s) \,\mathrm{d}s\right),
\end{aligned}$$

and here the modulus bars can be omitted because $y(t)$ and y_0 have the same sign.

Lemma 2.1. *If $a = a(t)$ is a continuous function, then*

$$y(t) = y_0 \exp\left(\int_{s=t_0}^{t} a(s) \,\mathrm{d}s\right)$$

is the only solution to the initial value problem

$$y'(t) = a(t)y(t), \qquad y(t_0) = y_0.$$

The life span of the solution y is infinite.

The general solution formula is the following:

Proposition 2.2. *Let I, $J \subset \mathbb{R}$ be open intervals, and $g \in C(I \to \mathbb{R})$, $h \in C(J \to \mathbb{R})$, with $h(y) \neq 0$ for all $y \in J$. Suppose $(t_0, y_0) \in I \times J$. Define*

$$G(t) := \int_{s=t_0}^{t} g(s) \,\mathrm{d}s, \qquad H(y) := \int_{z=y_0}^{y} \frac{1}{h(z)} \,\mathrm{d}z,$$

and assume $G(I) \subset H(J)$.

Then there is exactly one solution $y \in C^1(I \to \mathbb{R})$ to the initial value problem

$$y'(t) = g(t)h(y(t)), \qquad y(t_0) = y_0, \tag{2.1}$$

and this solution satisfies

$$H(y(t)) = G(t), \qquad t \in I. \tag{2.2}$$

Proof. The proof consists of three parts: existence of the solution, uniqueness of the solution, and demonstrating (2.2).

The existence of the solution is secured by the Peano theorem. Now we show (2.2). If y is any solution to (2.1), then (2.2) holds for $t = t_0$ because of $0 = 0$. On the other hand,

$$\frac{\mathrm{d}}{\mathrm{d}t} H(y(t)) = \frac{1}{h(y(t))} \cdot y'(t) = \frac{1}{h(y(t))} \cdot g(t) \cdot h(y(t)) = g(t) = \frac{\mathrm{d}}{\mathrm{d}t} G(t).$$

Therefore, both sides of (2.2) always have the same time derivative, and both sides coincide at one time, hence they coincide always.

Now we come to the uniqueness of the solution y. The function $H = H(y)$ has derivative $1/h(y)$ which never changes its sign, making H strictly monotone, hence invertible with inverse function H^{-1}. This gives us

$$y(t) = H^{-1}(G(t))$$

for each solution y to (2.1). Because of $G(I) \subset H(J)$, the expression $H^{-1}(G(t))$ is meaningful. □

Note that there are three obstacles to overcome:

- finding a primitive function G of g,
- finding a primitive function H of $1/h$,
- finding an inverse function H^{-1} to H,

and each obstacle could be insurmountable.

2.2 Substitution and Homogeneous Differential Equations

Consider the ODE

$$y'(t) = f(ay + bt + c)$$

with fixed real parameters a, b, c. Setting

$$z(t) := ay(t) + bt + c$$

gives

$$z'(t) = ay'(t) + b = af(z) + b,$$

which is an ODE whose variables (t, z) are separated.

Homogeneous ODE

Definition 2.3. *A term $P(u, v)$ is* positively homogeneous of order α *if*

$$P(\lambda u, \lambda v) = \lambda^\alpha P(u, v)$$

for all u, v and all $\lambda \in \mathbb{R}_+$.

In this sense, the left-hand side as well as the right-hand side of a linear homogeneous equation $Ax = 0$ are both homogeneous of order one, which explains the expression *homogeneous linear system*.

Consider the ODE

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

with a right-hand side homogeneous of order zero in the variables (t, y). A substitution $z(t) = y(t)/t$ gives

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

and again the variables (t, z) can be separated. Another example for a homogeneous ODE is

$$y'(t) = g\left(\frac{ay + bt}{cy + dt}\right), \qquad a, b, c, d \in \mathbb{R},$$

because of $(ay + bt)/(cy + dt) = (ay/t + b)/(cy/t + d)$.

Bernoulli DE

The BERNOULLI[2] equation has the form

$$y'(t) = a(t)y(t) + b(t)(y(t))^\alpha$$

for some $\alpha \in \mathbb{R}$. If α is a fractional number, $y(t)$ should be nonnegative. In case of $\alpha = 0$, this is a linear inhomogeneous ODE (to be studied in the next chapter), and the equation is separable for $\alpha = 1$. Hence we may assume $\alpha \neq 0, 1$, and then the substitution $z(t) := y(t)^{1-\alpha}$ gives

$$z'(t) = (1-\alpha)y^{-\alpha}y'(t) = (1-\alpha)y^{-\alpha}(ay + by^\alpha) = (1-\alpha)a(t)z(t) + (1-\alpha)b(t),$$

which is a linear inhomogeneous ODE.

Riccati DE

The RICCATI[3] equation can be understood as an "inhomogeneous" version of the Bernoulli equation,

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

with continuous functions k, g, h.

We present two substitutions which bring us to other differential equations.

First we suppose that y_* is a known solution, and we wish to find all the other solutions y. Then we can write $y = y_* + u$ with unknown u, and it turns out that

$$\begin{aligned} u' &= y' - y'_* = (ky^2 + gy + h) - (ky_*^2 + gy_* + h) \\ &= k(y_* + u)^2 - ky_*^2 + g(y - y_*) = k(2y_*u + u^2) + gu \\ &= k(t)u^2 + (2k(t)y_*(t) + g(t))u, \end{aligned}$$

which is a Bernoulli equation.

Second we consider the substitution

$$w(t) = \exp\left(-\int_{t_0}^{t} k(s)y(s)\,\mathrm{d}s\right),$$

for some fixed value t_0. Then $w' = w \cdot (-ky)$, and therefore

$$\begin{aligned} w'' &= w' \cdot (-ky) - wk'y - wky' = w \cdot (-ky)^2 - wk'y - wk\left(ky^2 + gy + h\right) \\ &= -wk'y - wkgy - wkh = -wk'\frac{w'}{-wk} - wkg\frac{w'}{-wk} - wkh \\ &= \frac{k'}{k}w' + gw' - wkh, \end{aligned}$$

and now the advantage is that this differential equation is *linear* (although still hard to solve by a formula).

It should be noted that Riccati equations (in particular in matrix form) appear in the theory of controlling vibrations in mechanical systems.

2.3 Power Series Expansions (Or How to Determine the Sound of a Drum)

We are interested in the eigenfrequencies of a drum, which is geometrically described by

$$\Omega = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 < R^2\},$$

[2] named after JACOB BERNOULLI, 1654 – 1705, not to be confused with his doctoral students NICOLAUS I BERNOULLI or JOHANN BERNOULLI (who found the B.-l'Hospital rule and was doctoral adviser of Euler) or the other family members: NICOLAUS II BERNOULLI, DANIEL BERNOULLI (well-known for the B. principle in aerodynamics), JOHANN II BERNOULLI and his sons JOHANN III BERNOULLI and JAKOB II BERNOULLI, all of them mathematicians / physicists.

[3] JACOPO FRANCESCO RICCATI, 1676 – 1754

and $u = u(t, x, y)$ denotes the elongation at time t and position (x, y) of the membrane. From physics we get the PDE

$$mu_{tt} = \operatorname{div}(K(x, y) \operatorname{grad} u)$$

with $m = m(x, y)$ as mass density and $K = K(x, y)$ characterising the elasticity properties of the membrane. The operators div and grad only act on (x, y), not t.

The membrane is clamped at the boundary $\partial\Omega$, hence

$$u(t, x, y) = 0 \qquad \text{if } x^2 + y^2 = R^2.$$

Eigenfrequencies can be found by the ansatz

$$u(t, x, y) = \cos(\lambda t)v(x, y),$$

giving us

$$-m\lambda^2 v = \operatorname{div}(K \operatorname{grad} v).$$

The function v should not be zero everywhere (this would correspond to silence), and this problem is also called an *eigenvalue problem*.

Experience tells us that a drum can not produce every frequency of sound, and our goal is to find the possible ones. This is hard for general functions m and K, which is why we assume $m = 1$ and $K \equiv \text{const.}$.

Then our problem is to find non-zero solutions v to

$$\begin{cases} -\lambda^2 v(x, y) = K \triangle v(x, y), & (x, y) \in \Omega, \\ v(x, y) = 0, & (x, y) \in \partial\Omega. \end{cases}$$

Remark 2.4. *Consider the vector space $U = L^2(\Omega)$ with the scalar product $\langle f, g\rangle_U := \int_\Omega f\overline{g}\,\mathrm{d}x$, and the operator $\triangle$ with domain of definition*

$$D(\triangle) := \{f \in L^2(\Omega) \colon \triangle f \in L^2(\Omega), \qquad f = 0 \text{ on } \partial\Omega\}.$$

Then $\triangle$ is a symmetric operator:

$$\langle \triangle f, g\rangle_U = \langle f, \triangle g\rangle_U, \qquad f, g \in D(\triangle),$$

by Green*'s formula from the second semester. We could consider $\triangle$ as a self-adjoint operator, and the theory of self-adjoint matrices nourishes the hope that the eigenvalues of $\triangle$ are real, and that the eigenfunctions of $\triangle$ give rise to an orthonormal basis of $U = L^2(\Omega)$. They do indeed, but we can neither prove nor explain this.*

We introduce polar coordinates,

$$x = r\cos\varphi, \qquad y = r\sin\varphi,$$

and then we have the Jacobi matrix

$$\frac{\partial(x, y)}{\partial(r, \varphi)} = \begin{pmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix},$$

and the inverse map $(x, y) \mapsto (r, \varphi)$ has as Jacobi matrix the inverse matrix:

$$\frac{\partial(r, \varphi)}{\partial(x, y)} = \begin{pmatrix} r_x & r_y \\ \varphi_x & \varphi_y \end{pmatrix} = \begin{pmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{pmatrix}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\varphi & r\sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}.$$

Put $w(r, \varphi) = v(x, y)$. Then

$$v_x = w_r r_x + w_\varphi \varphi_x = \cos\varphi \cdot w_r - \frac{\sin\varphi}{r} \cdot w_\varphi,$$
$$v_y = w_r r_y + w_\varphi \varphi_y = \sin\varphi \cdot w_r + \frac{\cos\varphi}{r} \cdot w_\varphi,$$

or, written as operator identity,

$$\frac{\partial}{\partial x} = \cos\varphi \frac{\partial}{\partial r} - \frac{\sin\varphi}{r}\frac{\partial}{\partial \varphi},$$
$$\frac{\partial}{\partial y} = \sin\varphi \frac{\partial}{\partial r} + \frac{\cos\varphi}{r}\frac{\partial}{\partial \varphi}.$$

Then we quickly find that

$$\triangle = \left(\frac{\partial}{\partial r}\right)^2 + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2},$$

and our eigenvalue problem turns into

$$\begin{cases} -\lambda^2 w(r,\varphi) = K\left(\dfrac{\partial^2}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2}{\partial \varphi^2}\right) w(r,\varphi), & (r,\varphi) \in (0,R)\times[0,2\pi], \\ w(R,\varphi) = 0, & \varphi \in [0,2\pi], \\ \lim\limits_{r\to 0} |w(r,\varphi)| < \infty, & \varphi \in [0,2\pi]. \end{cases} \tag{2.3}$$

Question: What is the purpose of the last condition in (2.3) ?

Our permanent assumption is that a sufficiently large number of derivatives of w exists. From $w(r,\varphi) = w(r,\varphi+2\pi)$ for all (r,φ) we then learn that a Fourier series expansion is possible:

$$w(r,\varphi) = \frac{a_0(r)}{2} + \sum_{n=1}^{\infty} \big(a_n(r)\cos(n\varphi) + b_n(r)\sin(n\varphi)\big),$$

and the convergence of the series is fast because of the smoothness assumption on w.

Question: Prove the following: if a function $f = f(\varphi)$ is 2π–periodic and L times continuously differentiable, then the Fourier coefficients a_n, b_n behave like $\mathfrak{O}(n^{-L})$ for $n \to \infty$.

Next we plug the Fourier series expansion into (2.3):

$$-\frac{\lambda^2}{K}\left(\frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos(n\varphi) + b_n\sin(n\varphi)\right)\right)$$
$$= \frac{1}{2}\left(a_0'' + \frac{1}{r}a_0'\right) + \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right)\sum_{n=1}^{\infty}\left(a_n\cos(n\varphi) + b_n\sin(n\varphi)\right).$$

We commute the Laplacian and $\sum_{n=1}^{\infty}(\dots)$ (this step needs a justification !):

$$-\frac{\lambda^2}{K}\left(\frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos(n\varphi) + b_n\sin(n\varphi)\right)\right)$$
$$= \frac{1}{2}\left(a_0'' + \frac{1}{r}a_0'\right) + \sum_{n=1}^{\infty}\left(\cos(n\varphi)\left(a_n'' + \frac{1}{r}a_n' - \frac{n^2}{r^2}a_n\right) + \sin(n\varphi)\left(b_n'' + \frac{1}{r}b_n' - \frac{n^2}{r^2}b_n\right)\right),$$

which can be re-ordered to

$$0 = \frac{1}{2}\left(a_0'' + \frac{1}{r}a_0' + \frac{\lambda^2}{K}a_0\right) + \sum_{n=1}^{\infty}\cos(n\varphi)\left(a_n'' + \frac{1}{r}a_n' + \left(\frac{\lambda^2}{K} - \frac{n^2}{r^2}\right)a_n\right)$$
$$+ \sum_{n=1}^{\infty}\sin(n\varphi)\left(b_n'' + \frac{1}{r}b_n' + \left(\frac{\lambda^2}{K} - \frac{n^2}{r^2}\right)b_n\right).$$

This is a Fourier series expansion of the zero function (on the left–hand side), but all Fourier coefficients of the zero function are zero, which means

$$0 = a_0''(r) + \frac{1}{r}a_0'(r) + \frac{\lambda^2}{K}a_0(r), \qquad \forall\, r \in (0,R),$$
$$0 = a_n''(r) + \frac{1}{r}a_n'(r) + \left(\frac{\lambda^2}{K} - \frac{n^2}{r^2}\right)a_n(r), \qquad \forall\, r \in (0,R),$$
$$0 = b_n''(r) + \frac{1}{r}b_n'(r) + \left(\frac{\lambda^2}{K} - \frac{n^2}{r^2}\right)b_n(r), \qquad \forall\, r \in (0,R).$$

Moreover, we have the boundary conditions

$$a_0(R) = a_n(R) = b_n(R) = 0,$$
$$\lim_{r\to 0} |a_0(r)| < \infty, \quad \lim_{r\to 0} |a_n(r)| < \infty, \quad \lim_{r\to 0} |b_n(r)| < \infty.$$

These differential equations are all very similar, which is the reason why we now only consider a_n. For $n \geq 0$, we set

$$a_n(r) =: c_n\left(\frac{\lambda}{\sqrt{K}} r\right), \qquad s := \frac{\lambda}{\sqrt{K}} r,$$

and then we get

$$a_n'(r) = \frac{\lambda}{\sqrt{K}} c_n'(s), \qquad a_n''(r) = \frac{\lambda^2}{K} c_n''(s),$$
$$0 = \frac{\lambda^2}{K} c_n''(s) + \frac{\lambda}{\sqrt{K}} \cdot \frac{1}{s} \cdot \frac{\lambda}{\sqrt{K}} c_n'(s) + \left(\frac{\lambda^2}{K} - \frac{\lambda^2}{K} \cdot \frac{n^2}{s^2}\right) c_n(s),$$

which simplifies to

$$\begin{cases} c_n''(s) + \frac{1}{s} c_n'(s) + \left(1 - \frac{n^2}{s^2}\right) c_n(s) = 0, & 0 < s < \frac{\lambda}{\sqrt{K}} R, \\ c_n\left(\frac{\lambda}{\sqrt{K}} R\right) = 0, \\ \lim_{s\to 0} |c_n(s)| < \infty. \end{cases}$$

This ODE is called Bessel[4] differential equation, and we want to solve it.

And here comes the method: a **power series expansion**. We make the ansatz

$$c_n(s) = \sum_{k=0}^{\infty} \gamma_k s^{\alpha+k}, \qquad \gamma_0 \neq 0,$$

with some unknown coefficients γ_k (which also depend on n) and some parameter $\alpha \in \mathbb{R}$ (which is maybe not an integer). By linearity, we may assume $\gamma_0 = 1$. A converging power series can be differentiated term-wise, as we have learned in the second semester. Then

$$c_n'(s) = \sum_{k=0}^{\infty} \gamma_k (\alpha + k) s^{\alpha+k-1},$$
$$c_n''(s) = \sum_{k=0}^{\infty} \gamma_k (\alpha + k)(\alpha + k - 1) s^{\alpha+k-2},$$

and plugging this into the Bessel differential equation gives

$$\begin{aligned} 0 &= \sum_{k=0}^{\infty} \gamma_k (\alpha + k)(\alpha + k - 1) s^{\alpha+k-2} + \sum_{k=0}^{\infty} \gamma_k (\alpha + k) s^{\alpha+k-2} + \sum_{k=0}^{\infty} (-\gamma_k) n^2 s^{\alpha+k-2} + \sum_{k=0}^{\infty} \gamma_k s^{\alpha+k} \\ &= \sum_{k=0}^{\infty} \gamma_k \Big((\alpha + k)^2 - n^2\Big) s^{\alpha+k-2} + \sum_{k=0}^{\infty} \gamma_k s^{\alpha+k}. \end{aligned}$$

The smallest power $s^{\alpha-2}$ occurs for $k = 0$ in the left sum, but $\gamma_0 = 1 \neq 0$, hence we find

$$0 \stackrel{!}{=} (\alpha^2 - n^2)$$

with the two solutions $\alpha = n$ and $\alpha = -n$. The last one violates $\lim_{s\to 0} |c_n(s)| < \infty$, hence $\alpha = n$. Then we have to solve

$$0 \stackrel{!}{=} \sum_{k=1}^{\infty} \gamma_k \Big((n + k)^2 - n^2\Big) s^{n+k-2} + \sum_{k=0}^{\infty} \gamma_k s^{n+k} = \sum_{k=1}^{\infty} \gamma_k (2nk + k^2) s^{n+k-2} + \sum_{k=0}^{\infty} \gamma_k s^{n+k}.$$

[4] Friedrich Wilhelm Bessel, 1784 – 1846. He generalised the Bessel functions which were defined by Daniel Bernoulli.

Comparing powers of s^{n-1} gives $\gamma_1 = 0$, and in general we have

$$\gamma_k(k^2 + 2kn) + \gamma_{k-2} = 0,$$

consequently $\gamma_3 = \gamma_5 = \cdots = 0$. And for $k = 2m + 2$ as an even number, it follows that

$$\gamma_{2m+2} = -\frac{\gamma_{2m}}{(2m+2)^2 + 2 \cdot 2n \cdot (2m+2)} = -\frac{\gamma_{2m}}{(2m+2)(2m+2n+2)}.$$

From $\gamma_0 = 1$ we then find that

$$\begin{aligned}\gamma_{2m} &= \frac{(-1)^m}{2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2m) \cdot (2+2n) \cdot (4+2n) \cdot \ldots \cdot (2m+2n)} \\ &= \frac{(-1)^m}{2^m \cdot m! \cdot 2^m (1+n)(2+n)(3+n) \cdot \ldots \cdot (m+n)}.\end{aligned}$$

Now the function c_n has been found, and we only have to make sure that the power series for c_n converges (if this series had a convergence radius of zero, we would not have gained anything). However, the estimate

$$|\gamma_{2m}| \leq \frac{1}{m!}$$

is quite easy to see, and then the convergence radius is $+\infty$. (In the first year, we had learned how to compute the convergence radius of a power series using a root criterion or a quotient criterion.)

It is common practice to define γ_0 slightly different from our choice, leading to

$$J_n(s) := \frac{1}{2^n n!} c_n(s) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+n)!} \left(\frac{s}{2}\right)^{2m+n}, \qquad s \in \mathbb{R} \subset \mathbb{C}. \tag{2.4}$$

These are the *Bessel functions of first kind.*

We have almost forgotten the boundary condition

$$c_n\left(\frac{\lambda}{\sqrt{K}} R\right) = 0,$$

which will determine λ.

Lemma 2.5. *The eigenfrequencies λ of a drum are given by*

$$\lambda = j_{n,i} \frac{\sqrt{K}}{R}, \qquad n \in \mathbb{N}_0, \quad i \in \mathbb{N}_+,$$

with $j_{n,i}$ as the ith positive zero of the Bessel function J_n, and R the radius of the drum, K the elasticity coefficient.

Hence we need a deeper understanding of where the zeroes of the Bessel functions are located. Absolutely everything (including a list of zeroes) about these functions can be found in the 800 pages of [23], but also [1] is an excellent reference. From there, the following formulas can be extracted:

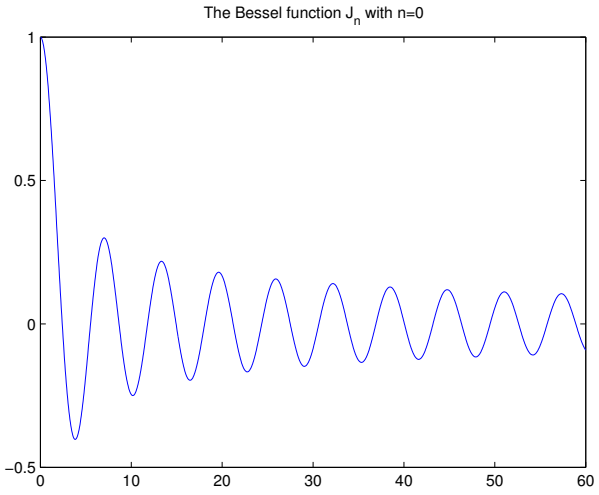
$$J_n(x) = \left(\frac{x}{2}\right)^n \frac{1}{n!} + \mathfrak{O}(x^{n+2}) \quad \text{for } x \to 0, \quad \text{usable for } 0 < x \lesssim \frac{n}{2}, \tag{2.5}$$

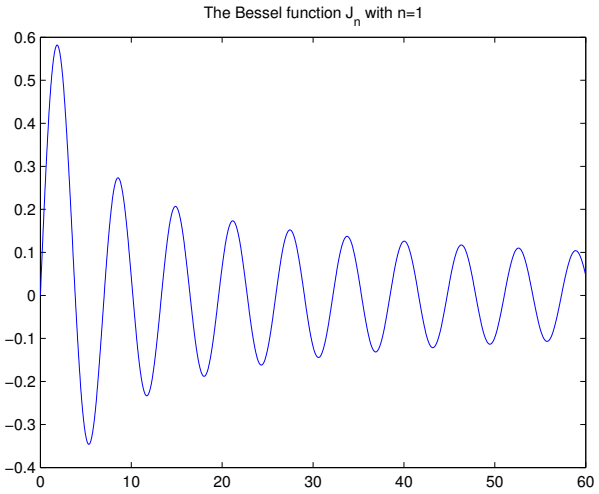
$$J_n(x) = \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{n\pi}{2} - \frac{\pi}{4}\right) + \mathfrak{O}(x^{-3/2}) \quad \text{for } x \to +\infty, \quad \text{usable for } x > 2n, \tag{2.6}$$

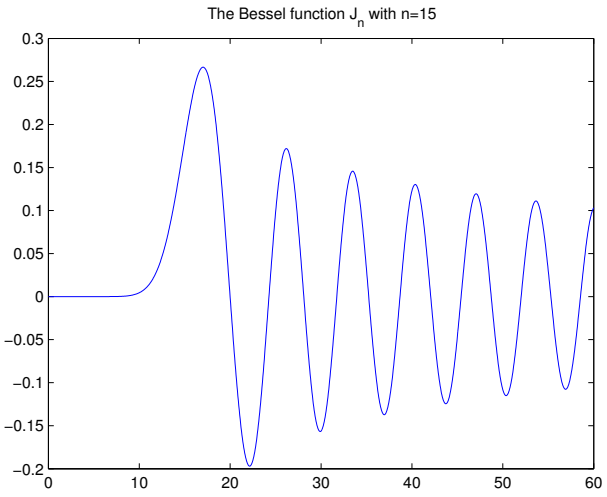
$$j_{n,1} = n + 1.8557571 n^{1/3} + \mathfrak{O}(n^{-1/3}) \quad \text{for } n \to \infty.$$

In particular, the smallest zero of J_0 is $j_{0,1} = 2.4048256\ldots$ which gives us the ground frequency of our drum. For growing R, the ground frequency gets lower, which matches our experience. And for higher tension of the membrane, K gets bigger, and also the pitch gets higher, as expected.

One more interesting property of the Bessel function is that the zeroes of J_{n+1} are always between the zeroes of J_n, in the sense of $j_{n,1} < j_{n+1,1} < j_{n,2} < j_{n+1,2} < \ldots$, and this can be proved analytically.


The Bessel function J_n with n=0
1
0.5
0
−0.5
0
10
20
30
40
50
60


The Bessel function J_n with n=1
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
0
10
20
30
40
50
60


The Bessel function J_n with n=15
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
0
10
20
30
40
50
60

2.4 Exact Differential Equations

In explaining what exact differential equations are, it is perhaps easier to start with the strategy of how to solve them, and then it will naturally follow how exact differential equations look like.

The key idea is that solving an equation

$$E(x,y) = \text{const.}$$

for y is considered easy, but solving an ODE of the form $y'(x) = f(x, y(x))$ is considered hard. More theoretically: if we assume that grad E is never $(0,0)$, then the implicit function theorem from the second semester tells us that near each point (x_0, y_0), the equation $E(x,y) = \text{const.}$ can be transformed into $x = x(y)$ or $y = y(x)$, and the solution set of all the points $(x,y) \in \mathbb{R}^2$ for which $E(x,y) = \text{const.}$ is a curve in the plane. For instance, $E(x,y) = x^2 + y^2 \stackrel{!}{=} R^2$ describes a circle of radius R, and this circle can also be expressed as

$$\begin{aligned}
&y = y(x) = +\sqrt{R^2 - x^2} \quad \text{or} \quad y = y(x) = -\sqrt{R^2 - x^2},\\
&x = x(y) = +\sqrt{R^2 - y^2} \quad \text{or} \quad x = x(y) = -\sqrt{R^2 - y^2},\\
&x = x(t), \qquad y = y(t),
\end{aligned}$$

where the representations in the first line do *not* hold in neighbourhoods of $(+R, 0)$, $(-R, 0)$; the representations of the second line are not valid in neighbourhoods of $(0, +R)$, $(0, -R)$; and the representations of the third line are not uniquely determined (we know already from the second semester that various parametrisations can describe the same curve).

In general, $E(x,y) = \text{const.}$ implies after differentiating with respect to x that

$$E_x + E_y y'(x) = 0,$$

and also

$$E(x,y) = \text{const.} \qquad \Longrightarrow \qquad E_x x'(y) + E_y = 0.$$

And finally

$$E(x,y) = \text{const.} \qquad \Longrightarrow \qquad E_x x'(t) + E_y y'(t) = 0.$$

We have $y'(x) = \frac{\mathrm{d}y}{\mathrm{d}x}$, $x'(y) = \frac{\mathrm{d}x}{\mathrm{d}y}$, and then the three ODEs can be written in a *formally* unified form as

$$E_x(x,y)\,\mathrm{d}x + E_y(x,y)\,\mathrm{d}y = 0.$$

In the example of the circle, we have

$$\begin{aligned}
&x + y(x)y'(x) = 0,\\
&x(y)x'(y) + y = 0,\\
&x(t)x'(t) + y(t)y'(t) = 0.
\end{aligned}$$

Definition 2.6. *A differential equation*

$$f(x,y) + g(x,y)y'(x) = 0 \qquad (\textit{or } f(x,y)\,\mathrm{d}x + g(x,y)\,\mathrm{d}y = 0)$$

is called exact *if a scalar function $E = E(x,y)$ (called* potential*) exists with*

$$E_x(x,y) = f(x,y), \qquad E_y(x,y) = g(x,y),$$

valid for all (x,y) under consideration.

Lemma 2.7. *Let the ODE $f(x,y) + g(x,y)y'(x) = 0$ be exact, with the initial condition $y(x_0) = y_0$. Then $f_y(x,y) = g_x(x,y)$ for all (x,y), and if E is the potential of the vector field (f,g), then the solution curve $y = y(x)$ (if it exists) satisfies*

$$E(x, y(x)) = E_0 \qquad \textit{for all } x,$$

where E_0 is determined by $E_0 := E(x_0, y_0)$.

Proof. By the Schwarz theorem,

$$f_y = E_{xy} = E_{yx} = g_x,$$

giving the first claim. The second claim follows from

$$\frac{\mathrm{d}}{\mathrm{d}x} E(x, y(x)) = E_x + E_y y'(x) = f + g y' = 0,$$

hence E is constant along a solution curve. □

In many cases, the function E is the "total energy" of a closed system.

Consider the equation for a pendulum,

$$\varphi''(t) + \sin \varphi(t) = 0,$$

which can not be solved by a solution formula (involving only functions known from school). But we can set

$$x(t) := \varphi(t), \qquad y(t) := \varphi'(t),$$

with the conclusion

$$x'(t) = y(t) =: g(x, y), \qquad y'(t) = -\sin x(t) =: -f(x, y),$$

and now we trivially have

$$f(x, y) \cdot g(x, y) + g(x, y) \cdot (-f(x, y)) = 0,$$
$$\sin x \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + y(t) \cdot \frac{\mathrm{d}y}{\mathrm{d}t} = 0.$$

The integrability condition $f_y = g_x$ holds, and $\mathbb{R}^2$ is simply connected, hence a potential E exists, which turns out to be

$$E(x, y) = -\cos x + \frac{1}{2} y^2 + C,$$

with C as a constant of integration which can be set to zero. Therefore, the solution $\varphi = \varphi(t)$ to the pendulum equation satisfies

$$-\cos \varphi(t) + \frac{1}{2} \Big(\varphi'(t)\Big)^2 = -\cos \varphi_0 + \frac{1}{2} \Big(\varphi'(0)\Big)^2$$

for all $t \in \mathbb{R}$, which is of course known as conservation of mechanical energy.

We still have not found $\varphi = \varphi(t)$, but we know $E(\varphi, \varphi') \equiv E_0$, giving the possibility of expressing φ' in terms of φ (or φ in terms of φ').

As a second practical example, we consider the famous model of Lotka[5] and Volterra[6] about a prey population of size $x(t)$ and a predator population of size $y(t)$ which solve the system

$$x'(t) = x(t) \cdot (\alpha - \beta y(t)), \qquad y'(t) = y(t) \cdot (-\gamma + \delta x(t)), \qquad \alpha, \beta, \gamma, \delta > 0.$$

Because of the nonlinearities, we can not expect to find a solution formula. As before, we set

$$g(x, y) = x \cdot (\alpha - \beta y) = x', \qquad -f(x, y) = y \cdot (-\gamma + \delta x) = y'.$$

Then trivially

$$f \cdot g + g \cdot (-f) = 0,$$
$$y(\gamma - \delta x) \cdot \frac{\mathrm{d}x}{\mathrm{d}t} + x(\alpha - \beta y) \cdot \frac{\mathrm{d}y}{\mathrm{d}t} = 0.$$

[5] Alfred James Lotka, 1880 – 1949, american mathematician, statistician, biophysicist
[6] Vito Volterra, 1860 – 1940, italian mathematician and physicist

This is *not* an exact differential equation, because $f_y \neq g_x$. But if we succeed in finding a multiplier $M = M(x, y)$ with the property that $(Mf)_y = (Mg)_x$ then the ODE is transformed into an exact DE. Geometrically spoken: the solution curve in the $x - y-$plane remains the same, we are simply choosing another parametrisation of the curve (M should never be zero). In the 19th century, solving an ODE was called *integrating this ODE*, and from that epoch such a function M is called *integrating factor*[7].

In general, finding such a factor M is highly non-trivial, but for the Lotka–Volterra model, $M(x, y) = 1/(xy)$ does the trick, giving us

$$\left(\frac{\gamma}{x} - \delta\right)\frac{\mathrm{d}x}{\mathrm{d}t} + \left(\frac{\alpha}{y} - \beta\right)\frac{\mathrm{d}y}{\mathrm{d}t} = 0,$$

and now a potential is

$$E(x, y) = \gamma \ln x - \delta x + \alpha \ln y - \beta y, \qquad x, y > 0.$$

This function is concave (meaning that the graph is always under the tangent plane), and for (x, y) approaching the boundary of $(0, \infty) \times (0, \infty)$, E goes to $-\infty$. Then the points (x, y) solving $E(x, y) = E_0$ form a loop in the $x - y-$plane.

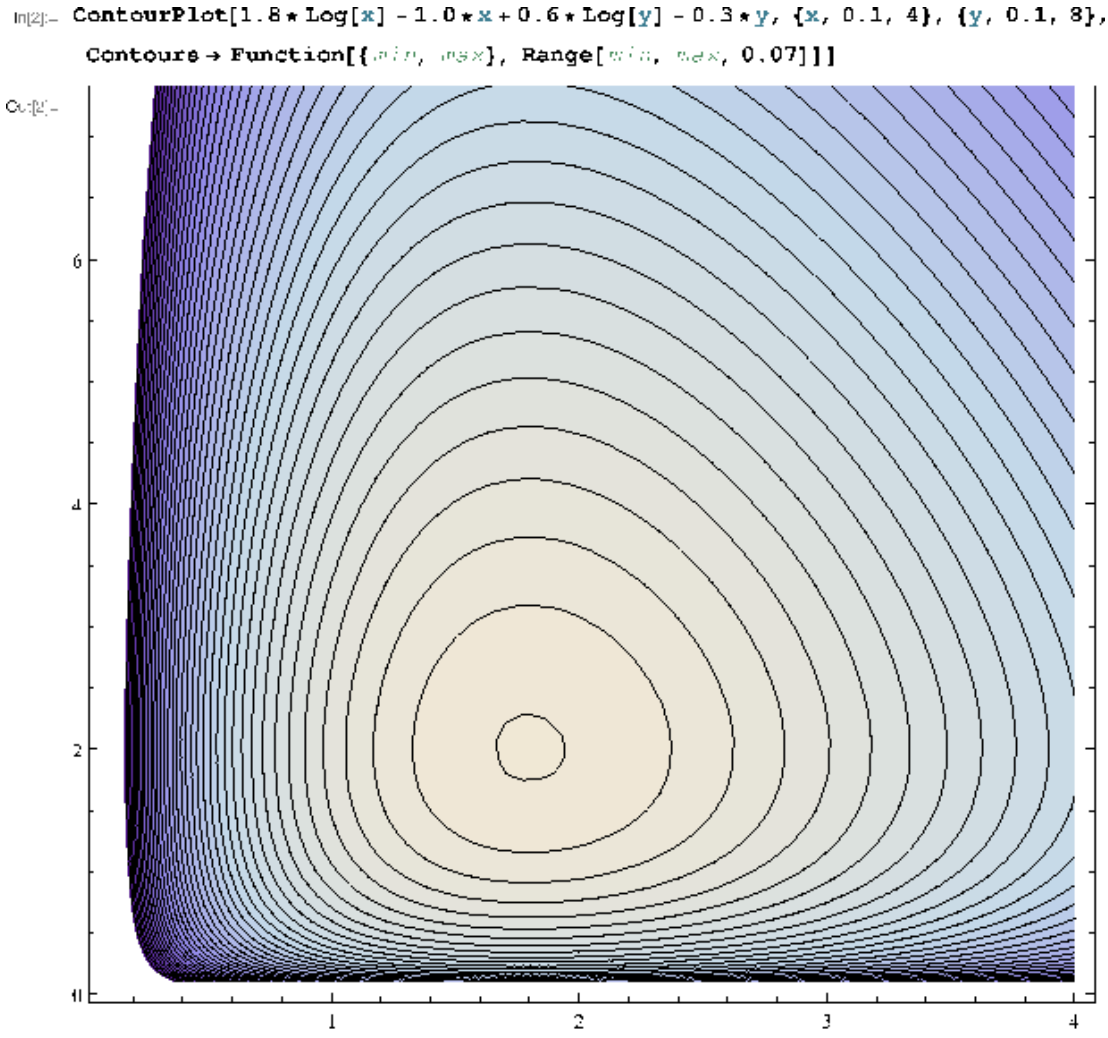


Figure 2.1: The potential E of the Lotka–Volterra model. In this diagram, E is constant along the egg-shaped curves.

[7] integrierender Faktor

Chapter 3

Linear Differential Equations and Systems

3.1 Linear Differential Equations

A *linear differential equation* is

$$y'(t) = a(t)y(t) + f(t), \tag{3.1}$$

with y, a, f as continuous functions from $\mathbb{R}$ to $\mathbb{R}$, and an *initial value problem* is obtained when (3.1) is complemented with the initial condition

$$y(t_0) = y_0. \tag{3.2}$$

By the Picard–Lindelöf theory, we know that (3.1), (3.2) is uniquely solvable, and the solution y exists on all of $\mathbb{R}^1$.

Our strategy is the following: first we forget about (3.2) and find all solutions to (3.1), and among them we then select that one which also fulfils (3.2).

We simplify even more and neglect f (only for a moment). Then all solutions to $y'(t) = a(t)y(t)$ are given by

$$y(t) = C \exp\left(\int_{s=t_0}^{t} a(s)\,\mathrm{d}s\right),$$

with C running through $\mathbb{R}$. For brevity of notation, we put

$$a_+(t) := \int_{s=t_0}^{t} a(s)\,\mathrm{d}s.$$

Now we bring back f into the differential equation, and our hope is to find a solution to (3.1) via the ansatz

$$y(t) = C(t)\exp(a_+(t)),$$

with C as a function. *Every* solution y can be expressed like this, because if y is a solution, then we can always pull out a factor $\exp(a_+(t))$ since exp takes never the value zero. Then it should hold that

$$\begin{aligned} y'(t) &= C'(t)\exp(a_+(t)) + C(t)\exp(a_+(t))a(t) \\ &\overset{!}{=} a(t)y(t) + f(t) \\ &= a(t)C(t)\exp(a_+(t)) + f(t), \end{aligned}$$

which turns into

$$C'(t)\exp(a_+(t)) \overset{!}{=} f(t),$$

$$\implies \quad C'(t) = \exp(-a_+(t))f(t) \qquad \implies \quad C(t) = \int_{s=t_0}^{t} \exp(-a_+(s))f(s)\,\mathrm{d}s + C_0,$$

with $C_0 \in \mathbb{R}$ as an additional integration constant. This brings us to

$$\begin{aligned} y(t) &= C(t)\exp(a_+(t)) \\ &= \exp(a_+(t)) \int_{s=t_0}^{t} \exp(-a_+(s)) f(s)\,\mathrm{d}s + C_0 \exp(a_+(t)), \end{aligned}$$

and C_0 is still available for choice. To satisfy the initial condition (3.2), we observe that $a_+(t_0) = 0$, giving us $C_0 = y_0$.

Lemma 3.1. *Let $a = a(t)$ and $f = f(t)$ be continuous functions. Then the unique solution to the initial value problem (3.1), (3.2) is given by*

$$y(t) = \exp\left(\int_{s=t_0}^{t} a(s)\,\mathrm{d}s\right) y_0 + \int_{s=t_0}^{t} \exp\left(\int_{r=s}^{t} a(r)\,\mathrm{d}r\right) f(s)\,\mathrm{d}s. \tag{3.3}$$

This formula is known as Duhamel*'s formula*[1], or as *variation of constants formula.*

We conclude the consideration of linear differential equations with the remark that

- the solutions to the linear homogeneous DE $y'(t) = a(t)y(t)$ form a linear space (vector space) of dimension one. Recall that $L := \frac{\mathrm{d}}{\mathrm{d}t} - a(t)$ is a linear operator, and the set of solutions y to $y' = ay$ is exactly $\ker L$, and the kernel of a linear map is always a vector space, as we have shown it in the first semester.
- the solutions to the linear inhomogeneous DE $y'(t) = a(t)y(t)+f(t)$ form an *affine* space of dimension one.

In a first attempt at generalising the above results to systems, we consider homogeneous systems

$$y'(t) = A(t)y(t), \qquad y\colon \mathbb{R} \to \mathbb{R}^n, \qquad A\colon \mathbb{R} \to \mathbb{R}^{n\times n},$$

and a first guess of the solution formula is

$$y(t) \stackrel{?}{=} \exp\left(A_+(t)\right) C, \qquad A_+(t) := \int_{s=t_0}^{t} A(s)\,\mathrm{d}s,$$

with a fixed vector $C \in \mathbb{R}^n$. Here, exp of a matrix is defined via the power series:

$$\exp B := \sum_{k=0}^{\infty} \frac{1}{k!} B^k, \qquad B \in \mathbb{R}^{n\times n}.$$

But the solution formula is wrong, because

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp(A_+(t))C \neq A(t)\exp(A_+(t))C, \tag{3.4}$$

since

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp(A_+(t)) = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{k=0}^{\infty} \frac{1}{k!} (A_+(t))^k = \sum_{k=0}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{k!}(A_+(t))^k\right),$$

and plugging this into (3.4) gives

$$0 \cdot C + \frac{1}{1!}AC + \frac{1}{2!}(A_+A + AA_+)C + \frac{1}{3!}(A_+A_+A + A_+AA_+ + AA_+A_+)C + \dots$$

on the left-hand side, but the right-hand side of (3.4) is

$$AIC + \frac{1}{1!}AA_+C + \frac{1}{2!}AA_+A_+C + \dots,$$

and now we are stuck because (in general) $AA_+ \neq A_+A$.

Unfortunately, the solution formula to $y'(t) = A(t)y(t) + F(t)$ will be quite a bit more complicated !

[1] Jean–Marie Constant Duhamel, 1797 – 1872, french mathematician and physicist

3.2 Exp of a Matrix, and $(\det A)'$

For our studies on systems, we need some tools. The first shall be the exp of a matrix, defined as

$$\exp A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k, \qquad A \in \mathbb{C}^{n\times n}.$$

To answer the question of convergence of this series, we need appropriate norms for a matrix.

Norms have the purpose of making our work easier. To this end, some conditions have to be met.

For the space $\mathbb{C}^n$, the norm $\|x\|_2 := \sqrt{|x_1|^2 + \cdots + |x_n|^2}$ is natural. We wish a matrix norm $\|\cdot\|_2$ to behave like this:

$$\begin{aligned} &\|Ax\|_2 \le \|A\|_2 \cdot \|x\|_2\,, \qquad \forall\, A \in \mathbb{C}^{n\times n}, \quad \forall\, x \in \mathbb{C}^n, \\ &\|AB\|_2 \le \|A\|_2 \cdot \|B\|_2\,, \qquad \forall\, A, B \in \mathbb{C}^{n\times n}. \end{aligned} \tag{3.5}$$

Note how these inequalities make it possible to "pull the norm bars onto each factor".

Moreover, we wish these inequalities to be *sharp*: for each matrix A, a vector $x_* \neq \vec{0}$ shall exist such that $\|Ax_*\|_2 = \|A\|_2 \cdot \|x_*\|_2$ with equality sign (otherwise we always waste something when we pull the norm bars onto each factor).

The first stab at the definition of $\|A\|$ is the Frobenius[2] norm $\|A\|_F := \sqrt{\sum_{j,k} |a_{jk}|^2}$, which indeed satisfies (3.5), because

$$(Ax)_1 = \sum_{j=1}^{n} a_{1j}x_j = (a_{11}, a_{12}, \ldots, a_{1n}) \cdot (x_1, \ldots, x_n)^\top,$$

and now the Cauchy–Schwarz inequality gives

$$|(Ax)_1| \le \sqrt{|a_{11}|^2 + \cdots + |a_{1n}|^2} \cdot \sqrt{|x_1|^2 + \cdots + |x_n|^2},$$

similarly for the other components $(Ax)_k$.

However, now the unit matrix has norm $\sqrt{n}$, which is quite a waste for large n.

The correct choice is

$$\|A\|_2 := \sup\{\|Ax\|_2 :\ x \in \mathbb{C}^n \text{ with } \|x\|_2 \le 1\}.$$

The unit ball $\{x \in \mathbb{C}^n :\ \|x\|_2 \le 1\}$ is a compact subset of $\mathbb{C}^n$, and from the first semester we know that a continuous function on a compact set attains its supremum, which in our situation means that there is an x_* with $\|x_*\|_2 = 1$ such that $\|A\|_2 = \|Ax_*\|_2$.

By the very definition of this norm via the supremum, we have (3.5), and this is sharp. Moreover, we have

$$\begin{aligned} \|AB\|_2 &= \max\{\|ABx\|_2 :\ \|x\|_2 \le 1\} \\ &\le \max\{\|A\|_2 \cdot \|Bx\|_2 :\ \|x\|_2 \le 1\} \\ &= \|A\|_2 \cdot \max\{\|Bx\|_2 :\ \|x\|_2 \le 1\} = \|A\|_2 \cdot \|B\|_2\,. \end{aligned}$$

Question: Show that $\|A\|_2$ can be computed via $\|A\|_2 = \sqrt{\lambda_{\max}(A^*A)}$, with $\lambda_{\max}$ as the largest eigenvalue of the self-adjoint positive semi-definite matrix A^*A.

Now we have constructed a norm on the vector space $\mathbb{C}^{n\times n}$, which is compatible to all the operations (multiplication with a scalar, with a vector, with another matrix, addition of matrices) in a most beautiful manner, and in particular, we have

$$\left\| \frac{1}{k!} A^k \right\|_2 = \frac{1}{k!} \|A \cdot \ldots \cdot A\|_2 \le \frac{1}{k!} \|A\|_2^k\,.$$

[2] Ferdinand Georg Frobenius, 1849 – 1917

Then the convergence of the series defining $\exp A$ follows by the general majorisation principle from the first semester, since $\sum_{k=0}^{\infty} \frac{1}{k!} M^k$ is finite for all $M \in \mathbb{R}$, in particular for $M = \|A\|_2$.

As expected from a similar relation in $\mathbb{C}$, we have

$$\exp A = \lim_{m\to\infty} \left(I + \frac{A}{m}\right)^m, \tag{3.6}$$

with I as identity matrix, and a proof can be found in § 15 of [2]. Further rules in the matrix calculus are

- if $AB = BA$ then $(A+B)^m = \sum_{k=0}^{m} \binom{m}{k} A^k B^{m-k}$,
- if $AB = BA$ then $\exp(A+B) = \exp(A)\exp(B)$,
- if $A = A^*$ then $\exp(\mathrm{i}A)$ is unitary,
- if $A \in \mathbb{R}^{n\times n}$ is skew-symmetric (that means $A^\top = -A$) then $\exp(A)$ is orthogonal.

Try to prove them all !

Next we need to differentiate a determinant with respect to a parameter.

Lemma 3.2. *Let P, $Q(\varepsilon) \in \mathbb{C}^{n\times n}$ with $\|Q(\varepsilon)\|_2 = \mathfrak{O}(\varepsilon^2)$ for $\varepsilon \to 0$. Then, as $\varepsilon \to 0$,*

$$\det(I + \varepsilon P + Q(\varepsilon)) = 1 + \varepsilon \operatorname{trace} P + \mathfrak{O}(\varepsilon^2),$$

with $\operatorname{trace} P = \sum_{j=1}^{n} p_{jj}$ as usual.

Proof. Put $A = I + \varepsilon P + Q(\varepsilon)$ with entries a_{jk}. Expanding $\det A$ gives us $n!$ products to be summed up. And we distinguish these $n!$ products as follows:

all n factors of this product are on the diagonal of A: there is only one product of that form, namely $a_{11}a_{22}\dots a_{nn}$.

exactly $n-1$ factors of this product are on the diagonal of A: this never happens (you can try it for $n = 3, 4$).

at least two factors of this product are off-diagonal: then this product is $\mathfrak{O}(\varepsilon^2)$.

Therefore, we have

$$\begin{aligned}\det A &= a_{11}a_{22}\dots a_{nn} + \mathfrak{O}(\varepsilon^2)\\ &= (1+\varepsilon p_{11})(1+\varepsilon p_{22})\dots(1+\varepsilon p_{nn}) + \mathfrak{O}(\varepsilon^2)\\ &= 1 + \varepsilon(p_{11}+p_{22}+\dots+p_{nn}) + \mathfrak{O}(\varepsilon^2),\end{aligned}$$

which finishes the proof. □

Example 3.3. *Put $P = B$, $Q = 0$, $\varepsilon = 1/m$ for $m \gg 1$. Then*

$$\begin{aligned}&\det\left(I + \frac{B}{m}\right) = 1 + \frac{\operatorname{trace} B}{m} + \mathfrak{O}(m^{-2}),\\ \Longrightarrow\quad &\det\left(\left(I + \frac{B}{m}\right)^m\right) = \left(\det\left(I + \frac{B}{m}\right)\right)^m = \left(1 + \frac{\operatorname{trace} B}{m} + \mathfrak{O}(m^{-2})\right)^m.\end{aligned}$$

Sending m to infinity and utilising (3.6) give

$$\det\exp(B) = e^{\operatorname{trace} B}. \tag{3.7}$$

Lemma 3.4. *Let $B = B(t)$ be twice continuously differentiable, and $B(t_0)$ invertible. Then*

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\det B\right)(t_0) = \det B(t_0)\operatorname{trace}\left(B^{-1}(t_0)B'(t_0)\right) = \det B(t_0)\operatorname{trace}\left(B'(t_0)B^{-1}(t_0)\right).$$

Proof. We have, for $t \to t_0$,

$$\begin{aligned} B(t) &= B(t_0) + B'(t_0) \cdot (t - t_0) + \mathfrak{O}((t-t_0)^2) \\ &= B(t_0) \left(I + B^{-1}(t_0) B'(t_0)(t-t_0) + \mathfrak{O}((t-t_0)^2) \right) \\ &= \left(I + B'(t_0) B^{-1}(t_0)(t-t_0) + \mathfrak{O}((t-t_0)^2) \right) B(t_0), \end{aligned}$$

and consequently

$$\begin{aligned} \det B(t) &= \det B(t_0) \det \left(I + B^{-1}(t_0) B'(t_0)(t-t_0) + \mathfrak{O}((t-t_0)^2) \right) \\ &= \det B(t_0) \left(1 + \operatorname{trace}(B^{-1}(t_0) B'(t_0))(t-t_0) + \mathfrak{O}((t-t_0)^2) \right) \\ &= \det B(t_0) + \det B(t_0) \operatorname{trace}(B^{-1}(t_0) B'(t_0)) \cdot (t-t_0) + \mathfrak{O}((t-t_0)^2), \\ \frac{\det B(t) - \det B(t_0)}{t - t_0} &= \det B(t_0) \operatorname{trace}(B^{-1}(t_0) B'(t_0)) + \mathfrak{O}(t - t_0), \end{aligned}$$

which gives the first formula, and the second is shown similarly. □

It is not necessary that B be twice differentiable; once is enough (but the proof less easy).

3.3 Linear Systems with General Coefficients

We start with homogeneous systems

$$y'(t) = A(t) y(t), \qquad y \colon \mathbb{R} \to \mathbb{C}^n, \quad A \colon \mathbb{R} \to \mathbb{C}^{n \times n}. \tag{3.8}$$

Proposition 3.5. *Let $y_{(1)}, \dots, y_{(n)}$ be solutions to (3.8). If the vectors $y_{(1)}(t_0), \dots, y_{(n)}(t_0)$ are linearly independent vectors in $\mathbb{C}^n$ for some $t_0 \in \mathbb{R}$, then the vectors $y_{(1)}(t), \dots, y_{(n)}(t)$ are linearly independent vectors in $\mathbb{C}^n$ for all times t.*

Proof. We arrange the vector functions $y_{(1)}, \dots, y_{(n)}$ to a matrix Y:

$$Y(t) := \begin{pmatrix} | & & | \\ y_{(1)}(t) & \dots & y_{(n)}(t) \\ | & & | \end{pmatrix}.$$

Then the systems (3.8) turn into $Y'(t) = A(t) Y(t)$.

We know that $Y(t_0)$ has full rank, hence $\det Y(t_0) \neq 0$, and by continuity then also $\det Y(t) \neq 0$ for t near t_0. For such t, we then have

$$\frac{\mathrm{d}}{\mathrm{d}t} \det Y(t) = \det Y(t) \cdot \operatorname{trace}(Y'(t) Y^{-1}(t)) = \det Y(t) \cdot \operatorname{trace} A(t),$$

which is a scalar linear homogeneous ODE, which can be solved by the solution formula from Lemma 2.1,

$$\det Y(t) = \det Y(t_0) \cdot \exp\left(\int_{s=t_0}^{t} \operatorname{trace} A(s) \, \mathrm{d}s \right), \qquad \text{for } t \text{ near } t_0.$$

Clearly, $\exp(\dots)$ never vanishes, and then it follows that $\det Y(t)$ can never be zero. □

Definition 3.6 (Wronski determinant). *For solutions $y_{(1)}, \dots, y_{(n)}$ to (3.8), the determinant of $Y(t)$ is called* Wronski[3] determinant, *and $Y(t)$ is called* Wronski matrix.

This determinant has the following use: showing directly that solutions $y_{(1)}, \dots, y_{(n)}$ are linearly independent is hard, because they live in the vector space $C^1(\mathbb{R} \to \mathbb{C}^n)$ which is of infinite dimension. However, now we have learned that it suffices to pick a time t_0 and check the linear independence of the vectors $y_{(1)}(t_0), \dots, y_{(n)}(t_0)$ as vectors in $\mathbb{C}^n$.

[3] Josef–Maria Hoëné de Wronski, 1778 – 1853, czech / polish / french mathematician

Example 3.7. *We want to find all solutions $u = u(t)$ to the scalar ODE*

$$u'''(t) + a_2(t)u''(t) + a_1(t)u'(t) + a_0(t)u(t) = 0. \tag{3.9}$$

To this end, we define a vector function

$$y(t) = \begin{pmatrix} u(t) \\ u'(t) \\ u''(t) \end{pmatrix},$$

and then we obtain the system

$$y'(t) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0(t) & -a_1(t) & -a_2(t) \end{pmatrix} y(t) = A(t)y(t),$$

which is an equivalent transformation, by Proposition 1.6. We define $y_{(1)}$, $y_{(2)}$, $y_{(3)}$ as solutions to $y'(t) = A(t)y(t)$ with the initial conditions

$$y_{(1)}(t_0) = (1,0,0)^\top, \qquad y_{(2)}(t_0) = (0,1,0)^\top, \qquad y_{(3)}(t_0) = (0,0,1)^\top.$$

Then $\det Y(t_0) = 1$, hence these functions $y_{(1)}$, $y_{(2)}$, $y_{(3)}$ are linearly independent, and then also the first components $u_{(1)}$, $u_{(2)}$, $u_{(3)}$ of the vectors $y_{(1)}$, $y_{(2)}$, $y_{(3)}$ must be linearly independent. But these $u_{(1)}$, $u_{(2)}$, $u_{(3)}$ solve (3.9), with the initial conditions

$$\begin{aligned} u_{(1)}(t_0) &= 1, & u'_{(1)}(t_0) &= 0, & u''_{(1)}(t_0) &= 0, \\ u_{(2)}(t_0) &= 0, & u'_{(2)}(t_0) &= 1, & u''_{(2)}(t_0) &= 0, \\ u_{(3)}(t_0) &= 0, & u'_{(3)}(t_0) &= 0, & u''_{(3)}(t_0) &= 1. \end{aligned}$$

Our remote goal is to find a solution formula like (3.3). Note that there expressions like $\exp(\int_{s=t_0}^{t} a(s)\,\mathrm{d}s)$ played a crucial rôle . The equivalent to these exponentials will now be written as $X(t, t_0)$, to be defined as follows:

Definition 3.8 (Fundamental solution). *A function $X = X(t_1, t_2)$ which maps from $\mathbb{R} \times \mathbb{R}$ into $\mathbb{C}^{n\times n}$ is called* fundamental solution *if*

- *the dependence of X from the first time argument is of regularity*[4] *C^1,*
- *for all t, $s \in \mathbb{R}$, we have $\frac{\partial}{\partial t}X(t,s) = A(t)X(t,s)$,*
- *for all $t \in \mathbb{R}$, we have $X(t,t) = I$, the identity matrix.*

Lemma 3.9. *If $A = A(t)$ is continuous, then exactly one fundamental solution X exists.*

Proof. We believe in a matrix version of the Picard–Lindelöf theorem. □

Lemma 3.10. *If $A = A(t)$ is continuous, then the following holds:*

- *the solution $y = y(t)$ to $y'(t) = A(t)y(t)$ with initial condition $y(t_0) = y_0$ is given by*

$$y(t) = X(t, t_0)y_0, \qquad \forall\, t \in \mathbb{R},$$

- *for arbitrary t_0, t_1, $t_2 \in \mathbb{R}$, we have*

$$X(t_2, t_1)X(t_1, t_0) = X(t_2, t_0).$$

Proof.

[4] regularity means *smoothness*

- Put $z(t) := X(t,t_0)y_0$. Then we find $z'(t) = \partial_t X(t,t_0)y_0 = A(t)X(t,t_0)y_0 = A(t)z(t)$, as well as $z(t_0) = X(t_0,t_0)y_0 = y_0$. Therefore the functions y and z solve the same initial value problem, hence $z(t) = y(t)$ for all times.
- Choose $y_0 \in \mathbb{C}^n$ freely, and let y be the solution to $y' = Ay$ with $y(t_0) = y_0$. From the first •, we get

  $$y(t_1) = X(t_1,t_0)y_0, \qquad y(t_2) = X(t_2,t_0)y_0,$$

  but also $y(t_2) = X(t_2,t_1)y(t_1) = X(t_2,t_1)X(t_1,t_0)y_0$, which brings us to

  $$X(t_2,t_1)X(t_1,t_0)y_0 = X(t_2,t_0)y_0.$$

  But if this equality holds for *all* $y_0 \in \mathbb{C}^n$, then the matrices must be equal.

□

We may set $t_2 = t_0$ with the consequence $X(t_0,t_1)X(t_1,t_0) = I$, or

$$X(t_0,t_1) = \Big(X(t_1,t_0)\Big)^{-1},$$

showing us that the dependence of $X(t_0,t_1)$ from the second variable t_1 must also be of regularity C^1. The invertibility of the matrix X is no surprise, because (the proof of) Proposition 3.5 gives us

$$\det X(t_1,t_0) = \exp\left(\int_{s=t_0}^{t_1} \operatorname{trace} A(s)\,\mathrm{d}s\right) \neq 0.$$

Now that the fundamental solution X has been found (at least in an abstract sense), we have a deep look at (3.3) and guess the following:

Lemma 3.11. *Let $A = A(t)$ and $f = f(t)$ be continuous functions. Then the unique solution to the initial value problem*

$$y'(t) = A(t)y(t) + f(t), \qquad y(t_0) = y_0$$

is given by

$$y(t) = X(t,t_0)y_0 + \int_{s=t_0}^{t} X(t,s)f(s)\,\mathrm{d}s. \tag{3.10}$$

Proof. For $t = t_0$, we find the expression

$$X(t_0,t_0)y_0 + \int_{s=t_0}^{t_0} \ldots\,\mathrm{d}s = y_0,$$

and therefore the right-hand side of (3.10) satisfies the initial condition. And differentiating the right-hand side of (3.10) with respect to t gives

$$\begin{aligned}
&\frac{\partial}{\partial t}X(t,t_0)y_0 + \frac{\partial}{\partial t}\int_{s=t_0}^{t} X(t,s)f(s)\,\mathrm{d}s \\
&\quad = A(t)X(t,t_0)y_0 + X(t,t)f(t) + \int_{s=t_0}^{t} \frac{\partial}{\partial t}X(t,s)f(s)\,\mathrm{d}s \\
&\quad = A(t)X(t,t_0)y_0 + f(t) + \int_{s=t_0}^{t} A(t)X(t,s)f(s)\,\mathrm{d}s \\
&\quad = A(t)X(t,t_0)y_0 + A(t)\int_{s=t_0}^{t} X(t,s)f(s)\,\mathrm{d}s + f(t) \\
&\quad = A(t)\left(X(t,t_0)y_0 + \int_{s=t_0}^{t} X(t,s)f(s)\,\mathrm{d}s\right) + f(t),
\end{aligned}$$

which was our aim. □

Up to now, the fundamental solution X has been found only in an abstract way, but we have no formula for X. This gap will be closed now.

Proposition 3.12. *The fundamental solution $X = X(t, t_0)$ to the matrix $A = A(t)$ is given by*

$$X(t, t_0) = \sum_{k=0}^{\infty} X^{(k)}(t, t_0)$$

with $X^{(0)}(t, t_0) \equiv I$, and the other $X^{(k)}$ are recursively defined by

$$X^{(k)}(t, t_0) = \int_{s=t_0}^{t} A(s) X^{(k-1)}(s, t_0)\, \mathrm{d}s, \qquad k \geq 1.$$

Sketch of proof. Keep t_0 fixed, and we restrict the variable t to a time interval $[t_0 - T, t_0 + T]$, for some T, which can be huge. We choose a number M with $\|A(s)\|_2 \leq M$ for all $s \in [t_0 - T, t_0 + T]$ (maybe M depends on T, but this is no problem). Then we find the estimates

$$\begin{aligned}
\left\| X^{(0)}(t, t_0) \right\|_2 &= 1, \\
\left\| X^{(1)}(t, t_0) \right\|_2 &\leq \int_{s=\min(t,t_0)}^{\max(t,t_0)} \|A(s)\|_2 \cdot \left\| X^{(0)}(s, t_0) \right\|_2 \, \mathrm{d}s \leq M|t - t_0|, \\
\left\| X^{(2)}(t, t_0) \right\|_2 &\leq \int_{s=\min(t,t_0)}^{\max(t,t_0)} \|A(s)\|_2 \cdot M|s - t_0| \, \mathrm{d}s \leq \frac{(M|t - t_0|)^2}{2}, \\
\left\| X^{(3)}(t, t_0) \right\|_2 &\leq \int_{s=\min(t,t_0)}^{\max(t,t_0)} \|A(s)\|_2 \cdot \frac{(M|s - t_0|)^2}{2} \, \mathrm{d}s \leq \frac{(M|t - t_0|)^3}{3!},
\end{aligned}$$

and continuing in this manner gives

$$\left\| X^{(k)}(t, t_0) \right\|_2 \leq \frac{(M|t - t_0|)^k}{k!}, \qquad k \geq 0,$$

and therefore the series $\sum_{k=0}^{\infty} X^{(k)}(t, t_0)$ indeed converges uniformly.

By definition, X solves

$$\partial_t X(t, t_0) = A(t) X(t, t_0), \qquad X(t_0, t_0) = I.$$

As in the proof of the Picard–Lindelöf Theorem, we transform this equivalently into an integral equation,

$$X(t, t_0) = I + \int_{s=t_0}^{t} A(s) X(t, t_0)\, \mathrm{d}s.$$

It remains to verify that

$$\sum_{k=0}^{\infty} X^{(k)}(t, t_0) \stackrel{?}{=} I + \int_{s=t_0}^{t} A(s) \sum_{k=0}^{\infty} X^{(k)}(s, t_0)\, \mathrm{d}s,$$

but this is checked quickly: just note that the uniform convergence of the series $\sum_{k=0}^{\infty} \dots$ allows to interchange the integration and the summation, as we have learned in the second semester. □

This explicit formula for the fundamental solution is mainly of theoretical interest, since computing all the matrices $X^{(k)}$ gets tedious very quickly, and (carefully chosen) numerical methods have a better ratio between precision and effort anyway, as we will see in a later chapter.

And finally, some comments on the geometric structure of solution sets.

- solutions to the linear homogeneous system $y'(t) = A(t)y(t)$ form a vector space of dimension n. A basis is given by that functions $y_{(1)}, \dots, y_{(n)}$ with $y_{(j)}(t_0) = e_j$, where $e_j = (0, \dots, 1, \dots, 0)$ is the canonical j-th basis vector.
- solutions to the inhomogeneous system $y'(t) = A(t)y(t) + f(t)$ form an affine space of dimension n.

3.4 Linear Systems and Equations with Constant Coefficients

Now we consider a constant matrix $A \in \mathbb{C}^{n\times n}$, and the initial value problem is

$$y'(t) = Ay(t) + f(t), \qquad y(t_0) = y_0. \tag{3.11}$$

Lemma 3.13. *The fundamental solution to a constant matrix A is*

$$X(t, t_0) = \exp(A(t - t_0)) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k (t - t_0)^k.$$

Proof. Either we check directly the definition, or we follow Proposition 3.12. □

Then the solution to (3.11) is

$$y(t) = e^{A(t-t_0)} y_0 + \int_{s=t_0}^{t} e^{A(t-s)} f(s)\, \mathrm{d}s.$$

Next we show how to compute $\exp(At)$, and to make the idea more clear, we choose a matrix $A \in \mathbb{C}^{6\times 6}$ with

- an eigenvalue λ_1 of algebraic multiplicity 1 and eigenvector u_1,
- an eigenvalue λ_2 of algebraic multiplicity 5 and geometric multiplicity 2, with chains of eigenvectors and principal vectors

$$u_2 \to p_{2,1}, \qquad u_3 \to p_{3,1} \to p_{3,2}.$$

Then we put

$$S = \begin{pmatrix} | & | & | & | & | & | \\ u_1 & u_2 & p_{2,1} & u_3 & p_{3,1} & p_{3,2} \\ | & | & | & | & | & | \end{pmatrix},$$

and the Jordan normal form then is

$$S^{-1}AS = D + N = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} + \begin{pmatrix} \boxed{0} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{0} & \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{0} & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{0} & \boxed{1} & \boxed{0} \\ 0 & 0 & 0 & \boxed{0} & \boxed{0} & \boxed{1} \\ 0 & 0 & 0 & \boxed{0} & \boxed{0} & \boxed{0} \end{pmatrix},$$

or $A = S(D + N)S^{-1}$.
Now $\exp(At) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k t^k$, and

$$A^k = S(D + N)S^{-1} \cdot S(D + N)S^{-1} \cdot \ldots \cdot S(D + N)S^{-1} = S(D + N)^k S^{-1},$$

giving us

$$\exp(At) = S \exp(Dt + Nt) S^{-1}.$$

We quickly check that $DN = ND$, and therefore

$$\exp(At) = S \exp(Dt) \exp(Nt) S^{-1}.$$

Now $\exp(Dt)$ is easy:

$$\exp(Dt) = \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{\lambda_2 t} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{\lambda_2 t} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{\lambda_2 t} \end{pmatrix},$$

and $\exp(Nt)$ needs a bit more care:

$$\exp(Nt) = I + (Nt) + \frac{1}{2!}N^2 t^2,$$

and the power series stops here since $N^3 = 0$. Such matrices N are called *nilpotent*, which means that a certain power of N is the zero matrix. Observe that

$$N^2 = \begin{pmatrix} \boxed{0} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}} & & 0 & 0 & 0 \\ 0 & & & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{\begin{matrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}} & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \end{pmatrix}, \qquad \exp(Nt) = \begin{pmatrix} \boxed{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{\begin{matrix} 1 & t \\ 0 & 1 \end{matrix}} & & 0 & 0 & 0 \\ 0 & & & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{\begin{matrix} 1 & t & t^2/2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{matrix}} & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \end{pmatrix},$$

and then we combine these results to

$$\begin{aligned} \exp(At) &= S \exp(Dt) \exp(Nt) S^{-1} \\ &= S \begin{pmatrix} \boxed{e^{\lambda_1 t}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{\begin{matrix} e^{\lambda_2 t} & te^{\lambda_2 t} \\ 0 & e^{\lambda_2 t} \end{matrix}} & & 0 & 0 & 0 \\ 0 & & & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{\begin{matrix} e^{\lambda_2 t} & te^{\lambda_2 t} & t^2 e^{\lambda_2 t}/2 \\ 0 & e^{\lambda_2 t} & te^{\lambda_2 t} \\ 0 & 0 & e^{\lambda_2 t} \end{matrix}} & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & & & \end{pmatrix} S^{-1}. \end{aligned}$$

Now we wish to evaluate $\exp(At)y_0$. We expand y_0 into the new basis $(u_1, u_2, p_{2,1}, u_3, p_{3,1}, p_{3,2})$:

$$y_0 = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_{2,1} p_{2,1} + \alpha_3 u_3 + \alpha_{3,1} p_{3,1} + \alpha_{3,2} p_{3,2} = S(\alpha_1, \alpha_2, \ldots, \alpha_{3,2})^\top,$$

$$\begin{aligned} &\exp(At)y_0 = S \exp(Dt) \exp(Nt)(\alpha_1, \ldots, \alpha_{3,2})^\top \\ &= S \left(\alpha_1 \begin{pmatrix} e^{\lambda_1 t} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ e^{\lambda_2 t} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_{2,1} \begin{pmatrix} 0 \\ te^{\lambda_2 t} \\ e^{\lambda_2 t} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ e^{\lambda_2 t} \\ 0 \\ 0 \end{pmatrix} + \alpha_{3,1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ te^{\lambda_2 t} \\ e^{\lambda_2 t} \\ 0 \end{pmatrix} + \alpha_{3,2} \begin{pmatrix} 0 \\ 0 \\ 0 \\ t^2 e^{\lambda_2 t}/2 \\ te^{\lambda_2 t} \\ e^{\lambda_2 t} \end{pmatrix} \right) \\ &= \alpha_1 e^{\lambda_1 t} u_1 + (\alpha_2 + \alpha_{2,1} t) e^{\lambda_2 t} u_2 + \alpha_{2,1} e^{\lambda_2 t} p_{2,1} + (\alpha_3 + \alpha_{3,1} t + \alpha_{3,2} t^2/2) e^{\lambda_2 t} u_3 \\ &\quad + (\alpha_{3,1} + \alpha_{3,2} t) e^{\lambda_2 t} p_{3,1} + \alpha_{3,2} e^{\lambda_2 t} p_{3,2}. \end{aligned}$$

Now the term $\exp(At)y_0$ is understood, and $\int_{s=0}^{t} \exp(A(t-s))f(s)\,\mathrm{d}s$ can be handled in the same style: for each s, expand $f(s)$ into the new basis of eigenvectors and principal vectors, and continue as above.

Now we come to scalar equations of higher order n, with constant coefficients:

$$u^{(n)}(t) + a_{n-1} u^{(n-1)}(t) + \cdots + a_2 u''(t) + a_1 u'(t) + a_0 u(t) = f(t), \quad u^{(k)}(0) = u_{0,k}, \quad 0 \le k \le n-1, \tag{3.12}$$

with the $u_{0,0}$, $u_{0,1}$, …, $u_{0,n-1}$ as initial values for the derivatives of order up to $n-1$.

Theoretically, we could transfer this higher order scalar equation into a first order system, as we have done it in Example 3.7, and then apply the solution formula from (3.10). This approach works, but often we are faster when we directly deal with the higher order equation.

Constructing the solution is done in three steps:

- first, we find *all* solutions to the homogeneous problem, with $f \equiv 0$ and ignoring all $u_{0,k}$. This representation involves n freely selectable constants $\alpha_1, \ldots, \alpha_n$.
- second, we find *one* solution to the inhomogeneous problem with the original f, still ignoring all $u_{0,k}$.
- third, we add the above obtained parts of the solutions and choose the constants $\alpha_1, \ldots, \alpha_n$ in such a way that the initial conditions hold.

We come to the **first step**, finding all solutions to

$$u^{(n)}(t) + a_{n-1}u^{(n-1)}(t) + \cdots + a_2u''(t) + a_1u'(t) + a_0u(t) = 0,$$

where $a_{n-1}, \ldots, a_0 \in \mathbb{C}$ are constants. We can also say that we want to determine the kernel of a differential operator L,

$$L = \frac{\mathrm{d}^n}{\mathrm{d}t^n} + a_{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}} + \cdots + a_2\frac{\mathrm{d}^2}{\mathrm{d}t^2} + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0.$$

Definition 3.14 (**characteristic polynomial**). *The* characteristic polynomial *to the operator L is*

$$\chi(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

By the Fundamental Theorem of Algebra, this polynomial has n roots λ_j in the complex plane (counted according to their multiplicity), and we can factorise:

$$\chi(\lambda) = (\lambda - \lambda_1)^{\mu_1}(\lambda - \lambda_2)^{\mu_2} \cdots (\lambda - \lambda_k)^{\mu_k},$$

with μ_j as multiplicity of the zero λ_j, and $\lambda_i \neq \lambda_j$ for $j \neq j$. We clearly have $\mu_1 + \mu_2 + \cdots + \mu_k = n$.
Note that also the operator L can be factorised:

$$L = \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)^{m_1}\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)^{m_2} \cdots \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_k\right)^{m_k}$$

This is possible because the a_j are constant.
Question: Think about why

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} - 2t\frac{\mathrm{d}}{\mathrm{d}t} + t^2 \neq \left(\frac{\mathrm{d}}{\mathrm{d}t} - t\right)^2.$$

Lemma 3.15. *The kernel of L is of dimension n, and its basis is given by functions $t \mapsto t^l \exp(\lambda_j t)$, for $0 \leq l \leq \mu_j - 1$.*

Sketch of proof. In order to not get drowned in an ocean of indices, we take $n = 5$ and consider the special fifth order operator

$$L = \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^3$$

as an example. A solution to $Lu = 0$ is uniquely determined by the values of $u(0)$, $u'(0)$, $u''(0)$, $u^{(3)}(0)$, $u^{(4)}(0)$, which are five numbers. Therefore the dimension of L can not be more than five, and we only have to guess five linearly independent elements of $\ker L$.
The function $u_1(t) = \exp(\lambda_1 t)$ belongs to $\ker L$, because of

$$Lu_1 = \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^3\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)u_1 = \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^3 0 = 0.$$

Similarly for the functions $u_2(t) = \exp(\lambda_2 t)$ and $u_3(t) = \exp(\lambda_3 t)$. Next we note that

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda\right)\left(t^p \exp(\lambda t)\right) = pt^{p-1}\exp(\lambda t), \qquad p \in \mathbb{N}_0,$$

and consequently also the function $u_4(t) = t\exp(\lambda_3 t)$ belongs to the kernel, since

$$\begin{aligned} Lu_4 &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^3 \Big(t\exp(\lambda_3 t)\Big) \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^2 \Big(\exp(\lambda_3 t)\Big) \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right) 0 = 0. \end{aligned}$$

And finally, we have $u_5(t) = t^2\exp(\lambda_3 t)$:

$$\begin{aligned} Lu_5 &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^3 \Big(t^2\exp(\lambda_3 t)\Big) \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right)^2 \Big(2t\exp(\lambda_3 t)\Big) \\ &= \left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_1\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_2\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_3\right) \Big(2\exp(\lambda_3 t)\Big) = 0. \end{aligned}$$

It only remains to check the linear independence of u_1, …, u_5, which could be done with the Wronski matrix $Y(t)$, and we can choose $t = 0$:

$$\begin{aligned} Y(0) &= \begin{pmatrix} u_1(0) & u_2(0) & u_3(0) & u_4(0) & u_5(0) \\ u_1'(0) & u_2'(0) & u_3'(0) & u_4'(0) & u_5'(0) \\ u_1''(0) & u_2''(0) & u_3''(0) & u_4''(0) & u_5''(0) \\ u_1'''(0) & u_2'''(0) & u_3'''(0) & u_4'''(0) & u_5'''(0) \\ u_1''''(0) & u_2''''(0) & u_3''''(0) & u_4''''(0) & u_5''''(0) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ \lambda_1 & \lambda_2 & \lambda_3 & 1 & 0 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & 2\lambda_2 & 2 \\ \lambda_1^3 & \lambda_2^3 & \lambda_3^3 & 3\lambda_2^2 & 6\lambda_3 \\ \lambda_1^4 & \lambda_2^4 & \lambda_3^4 & 4\lambda_2^3 & 12\lambda_3^2 \end{pmatrix}. \end{aligned}$$

Here we have made fruitful use of the Leibniz formula:

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^m (v(t)w(t)) = \sum_{l=0}^{m} \binom{m}{l} v^{(l)}(t) w^{(m-l)}(t).$$

Question: Prove the Leibniz formula.

Now we can check by direct computation that $\det Y(0) \neq 0$. This is doable, but tedious, and difficult to generalise to arbitrary operators L.

Another approach to the proof of the linear independence of u_1, …, u_5 is more algebraic in nature. Consider the equation

$$\alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3 + \alpha_4 u_4 + \alpha_5 u_5 = 0, \qquad \alpha_1, \ldots, \alpha_5 \in \mathbb{C},$$

with 0 as the zero function, and we want to show that $\alpha_1 = \cdots = \alpha_5 = 0$. If $\alpha_5 \neq 0$, then u_5 can be written as linear combination of the other u_j,

$$u_5 = \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 + \beta_4 u_4, \qquad \beta_1, \ldots, \beta_4 \in \mathbb{C}.$$

Now remove one factor $(\partial_t - \lambda_3)$ from L, giving us the operator $L_1 = (\partial_t - \lambda_1)(\partial_t - \lambda_2)(\partial_t - \lambda_3)^2$. Then u_1, …, u_4 belong to $\ker L_1$, but u_5 does not. Recalling that $\ker L_1$ is a vector space, we find a contradiction, hence $\alpha_5 = 0$. Now we assume $\alpha_4 \neq 0$, and then we get

$$u_4 = \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3, \qquad \beta_1, \ldots, \beta_3 \in \mathbb{C},$$

and considering the operator $L_2 = (\partial_t - \lambda_1)(\partial_t - \lambda_2)(\partial_t - \lambda_3)$ brings us again a contradiction, hence $\alpha_4 = 0$. Therefore, $0 = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3$, with u_1, u_2, $u_3 \in \ker L_2$, and looking at L_2 instead of L gives us the Wronski matrix

$$Y(t) = \begin{pmatrix} u_1(t) & u_2(t) & u_3(t) \\ u_1'(t) & u_2'(t) & u_3'(t) \\ u_1''(t) & u_2''(t) & u_3''(t) \end{pmatrix}, \qquad Y(0) = \begin{pmatrix} 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{pmatrix}.$$

To show that $\det Y(0) \neq 0$, we recall the theory of VANDERMONDE[5] determinants

$$V_n(\lambda_1, \dots, \lambda_n) := \det \begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{pmatrix}.$$

Some cunning ideas allow us to evaluate these determinants quickly:

Keep $\lambda_1, \dots, \lambda_{n-1}$ frozen and let λ_n run through $\mathbb{C}$. Then $V_n(\lambda_1, \dots, \lambda_n)$ is a polynomial in the only variable λ_n of degree $n-1$, and it can be found by expanding the determinant along the last column:

$$V_n(\lambda_1, \dots, \lambda_n) = V_{n-1}(\lambda_1, \dots, \lambda_{n-1}) \cdot \lambda_n^{n-1} + \boxed{?}\, \lambda_n^{n-2} + \dots + \boxed{?}\, \lambda_n + \boxed{?}, \tag{3.13}$$

where we do not care what the missing coefficients are. On the other hand, the determinant vanishes if two columns coincide, and therefore the polynomial (3.13) has the $n-1$ zeroes $\lambda_n = \lambda_1, \dots, \lambda_n = \lambda_{n-1}$, giving us the decomposition of V_n into linear factors:

$$V_n(\lambda_1, \dots, \lambda_n) = \boxed{?}\, (\lambda_n - \lambda_1) \cdot \ldots \cdot (\lambda_n - \lambda_{n-1}), \tag{3.14}$$

with some unknown leading coefficient in the box. Comparing (3.13) and (3.14) gives us

$$V_n(\lambda_1, \dots, \lambda_n) = V_{n-1}(\lambda_1, \dots, \lambda_{n-1}) \cdot (\lambda_n - \lambda_1)(\lambda_n - \lambda_2) \cdot \ldots \cdot (\lambda_n - \lambda_{n-1}),$$

and by induction we then find

$$V_n(\lambda_1, \dots, \lambda_n) = \prod_{j<k} (\lambda_k - \lambda_j),$$

which is not zero if all the λ_i are distinct. □

In the **second step**, we intend to find one solution u_{inh} to the inhomogeneous problem

$$u^{(n)}(t) + a_{n-1} u^{(n-1)}(t) + \dots + a_1 u'(t) + a_0 u(t) = f(t).$$

One approach is the *variation of constants*. We make the ansatz

$$u(t) = C_1(t) u_1(t) + \dots + C_n(t) u_n(t),$$

with the u_j as the solutions found in the first step. Then we find (and set)

$$\begin{aligned}
u'(t) &= \sum_{j=1}^{n} \big(C_j'(t) u_j(t) + C_j(t) u_j'(t) \big) \overset{!}{=} \sum_{j=1}^{n} C_j(t) u_j'(t), \\
u''(t) &= \sum_{j=1}^{n} \big(C_j'(t) u_j'(t) + C_j(t) u_j''(t) \big) \overset{!}{=} \sum_{j=1}^{n} C_j(t) u_j''(t), \\
&\dots \\
u^{(n-1)}(t) &= \sum_{j=1}^{n} \Big(C_j'(t) u_j^{(n-2)}(t) + C_j(t) u_j^{(n-1)}(t) \Big) \overset{!}{=} \sum_{j=1}^{n} C_j(t) u_j^{(n-1)}(t), \\
u^{(n)}(t) &= \sum_{j=1}^{n} \Big(C_j'(t) u_j^{(n-1)}(t) + C_j(t) u_j^{(n)}(t) \Big) \\
&\overset{!}{=} f(t) - \sum_{k=0}^{n-1} a_k u^{(k)}(t) = f(t) - \sum_{j=1}^{n} C_j(t) \sum_{k=0}^{n-1} a_k u_j^{(k)}(t),
\end{aligned}$$

[5] ALEXANDRE–THÉOPHILE VANDERMONDE, 1735–1796, french musician, chemist and mathematician. This determinant appears nowhere in his four mathematical papers, and it is unknown why it is named after him.

which turns into the system

$$\begin{pmatrix} u_1 & u_2 & \dots & u_n \\ u_1' & u_2' & \dots & u_n' \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{(n-2)} & u_2^{(n-2)} & \dots & u_n^{(n-2)} \\ u_1^{(n-1)} & u_2^{(n-1)} & \dots & u_n^{(n-21} \end{pmatrix} \begin{pmatrix} C_1' \\ C_2' \\ \vdots \\ C_{n-1}' \\ C_n' \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ f \end{pmatrix},$$

or $(C_1', \dots, C_n')^\top = Y^{-1}(t)(0, \dots, 0, f)^\top$, from which the C_j can be found by integration.

Another approach is to guess the u_{inh} by a special ansatz:

- if $f = f(t)$ is a polynomial in t, then u_{inh} can be found as a polynomial, typically of the same degree, in exceptional cases of higher degree,
- if $f = f(t)$ is a multiple of $\exp(\kappa t)$, then u_{inh} can be found as a multiple of $\exp(\kappa t)$, in exceptional cases as $\exp(\kappa t)$ multiplied by a polynomial of t,
- if f is a linear combination of the above, then u_{inh} can be found as appropriate linear combination,
- if $f = f(t) = \exp(at)\cos(bt)$, then we can either write $f(t) = (\exp((a + \mathrm{i}b)t) + \exp((a - \mathrm{i}b)t))/2$ and proceed as above, or we try to find u_{inh} as linear combination of $\exp(at)\cos(bt)$ and $\exp(at)\sin(bt)$.

And in the **third step**, we add up:

$$u(t) = \alpha_1 u_1(t) + \dots + \alpha_n u_n(t) + u_{inh}(t),$$

and choose the α_j in such a way that the initial conditions in (3.12) are fulfilled.

Chapter 4

Flows

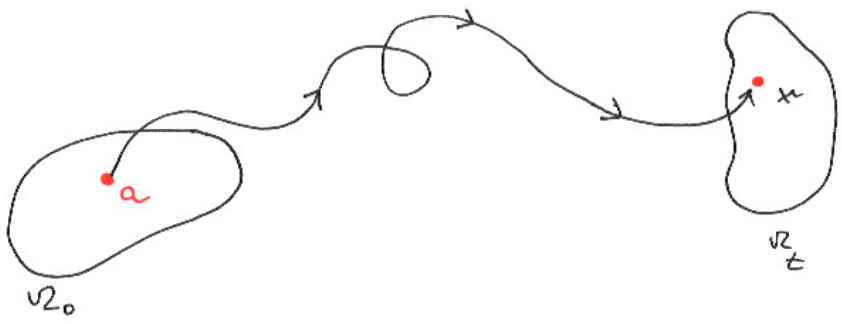
4.1 General Remarks

Imagine a flowing fluid or gas, or a solid which is able to be deformed elastically (or plastically). At time 0, it occupies a domain $\Omega_0 \subset \mathbb{R}^n$, and at time t, a domain Ω_t. Of course, these domains may overlap. A particle, which is at the position $a \in \Omega_0$ at time 0, moves to a position $x \in \Omega_t$ at time t, and we write this as

$$x = \Phi(t, t_0, a).$$

We can write $\Phi(t, t_0)$ for this map $a \mapsto x$, and the following properties are physically plausible:

$$\begin{aligned} \Phi(t,t) &= \mathrm{id}, \qquad \forall\, t \in \mathbb{R}, \\ \Phi(t_2,t_1) \circ \Phi(t_1,t_0) &= \Phi(t_2,t_0), \qquad \forall\, t_0,t_1,t_2 \in \mathbb{R}. \end{aligned}$$



We also suppose that this map $\Phi(t, t_0)$ is a smooth *diffeomorphism*. This means that $\Phi(t, t_0)$ maps Ω_{t_0} bijectively onto Ω_t, the map $a \mapsto x$ is infinitely differentiable, and the inverse map $x \mapsto a$ is also infinitely differentiable. We also assume that the derivatives of $\Phi(t, t_0, a)$ with respect to t and to t_0 exist, up to any order.

Now we look at one special particle which is at $a \in \Omega_0$ at the starting time 0. This particle moves along the curve

$$t \mapsto \Phi(t, 0, a), \qquad t \in [0, \infty),$$

and its velocity is

$$U(t,x) := \frac{\partial}{\partial t}\Phi(t,0,a) \qquad \text{if } x = \Phi(t,0,a).$$

Think of a physical quantity of this particle (and its neighbour particles): velocity, acceleration, temperature, mass density, pressure. A formula of this quantity could refer to this particle via the variables (t, x), with x being the position at time t, which are the *Euler coordinates*.

Or the formula for this quantity at time t could refer to this particle using the variable a, which was the position at time zero. These are the *Lagrange coordinates*.

Example 4.1. *The Eulerian velocity at a particle x at time t is $U(t, x)$.*

The Lagrangian velocity of the same particle is $\partial_t \Phi(t, 0, a)$, where a and x are related via $x = \Phi(t, 0, a)$, or $a = \Phi(0, t, x)$.

The Eulerian description is more popular because it needs no translation step from x back to a.

How to do this translation if only the velocity field $U = U(t, x)$ is known, but Φ is not ?

Given are a time $t^* > 0$ and a point $x^* \in \Omega_{t^*}$. Then the trajectory $x = x(t)$ of that particle solves the initial value problem

$$x'(t) = U(t, x(t)), \qquad x(t^*) = x^*.$$

We just have to solve this initial value problem, and then the starting position is found as $a = x(0)$.

Although the Lagrangian coordinates are less used, they are still needed.

Definition 4.2. *The Lagrangian acceleration is defined as $\partial_t^2 \Phi(t, 0, a)$.*

Lemma 4.3. *Then the Eulerian acceleration is computed as*

$$\gamma(t, x) = \partial_t U(t, x) + ((U \cdot \nabla)U)(t, x),$$

with $U \cdot \nabla := \sum_{j=1}^n U_j \partial_j$, and $\nabla = (\partial_1, \dots, \partial_n)$ contains only the spatial derivatives.

Proof. We have (remember that $x = \Phi(t, 0, a)$, and U, Φ are vector-valued)

$$\begin{aligned}\partial_t^2 \Phi(t, 0, a) &= \frac{\partial}{\partial t} \frac{\partial}{\partial t} \Phi(t, 0, a) = \frac{\mathrm{d}}{\mathrm{d}t} U(t, x(t)) = \frac{\mathrm{d}}{\mathrm{d}t} U(t, \Phi(t, 0, a)) \\ &= \partial_t U(t, x) + \sum_{j=1}^n \frac{\partial U}{\partial x_j}(t, x) \cdot \frac{\partial \Phi_j}{\partial t}(t, 0, a) = \partial_t U(t, x) + \sum_{j=1}^n U_j(t, x) \cdot \frac{\partial U}{\partial x_j}(t, x).\end{aligned}$$

□

More generally, we have: if $H = H(t, x)$ is a physical quantity expressed in Eulerian coordinates, and $G = G(t, a)$ is the same quantity expressed in Lagrangian coordinates, both connected via $x = \Phi(t, 0, a)$, then the derivatives transform like this:

$$\frac{\partial G}{\partial a_j} = \sum_{k=1}^n \frac{\partial H}{\partial x_k} \cdot \frac{\partial \Phi_k}{\partial a_j}, \qquad \frac{\partial G}{\partial t} = \frac{\partial H}{\partial t} + \sum_{k=1}^n \frac{\partial H}{\partial x_k} \cdot U_k(t, x) = \frac{\partial H}{\partial t} + (U \cdot \nabla) H.$$

Now we consider *balance equations*[1]. Consider a scalar function $C = C(t, x)$, and define

$$K(t) = \int_{\Omega_t} C(t, x) \, \mathrm{d}x.$$

For instance, C could be the mass density (mass per volume), and then $K(t)$ would be the mass of the domain Ω_t. Or C could be the energy density (energy per volume), giving K as total energy of Ω_t. We would like to know how $K(t)$ changes with time.

Lemma 4.4. *It holds*

$$\frac{\mathrm{d}}{\mathrm{d}t} K(t) = \int_{\Omega_t} C_t(t, x) \, \mathrm{d}x + \int_{\Omega_t} \operatorname{div}(CU)(t, x) \, \mathrm{d}x.$$

[1] Bilanzgleichungen

Proof. The key problem is that the domain of integration, appearing in the definition of K, is changing with time. We overcome this trouble by first going back to the Lagrangian description, then evaluating the time derivative, and finally transforming to the Euler coordinates.

The map $\Phi(t,0)\colon a \to x$ is a diffeomorphism, which implies that the Jacobi matrix $\frac{\partial\Phi}{\partial a}(t,0,a)$ is always invertible. Then the determinant of this Jacobi matrix is never zero. And because this determinant is one for $t=0$, we find that

$$\det \frac{\partial\Phi}{\partial a}(t,0,a) > 0$$

always. This gives, by substitution of x against a,

$$K(t) = \int_{\Omega_0} C(t,\Phi(t,0,a)) \left|\det \frac{\partial\Phi}{\partial a}(t,0,a)\right| \,\mathrm{d}a,$$

and the modulus bars around the determinant are not needed, as already seen. Write $J(t,a) := \det \frac{\partial\Phi}{\partial a}(t,0,a)$. Then $\mathrm{d}x = J(t,a)\,\mathrm{d}a$ and

$$K'(t) = \int_{\Omega_0} \left(\frac{\mathrm{d}}{\mathrm{d}t} C(t,\Phi(t,0,a))\right) \cdot J(t,a)\,\mathrm{d}a + \int_{\Omega_0} C(t,\Phi(t,0,a)) \cdot (\partial_t J(t,a))\,\mathrm{d}a.$$

We know already that

$$\frac{\mathrm{d}}{\mathrm{d}t} C(t,\Phi(t,0,a)) = C_t(t,x) + (U(t,x)\cdot\nabla)C(t,x),$$

and then the first integral turns into

$$\int_{\Omega_t} C_t(t,x) + (U(t,x)\cdot\nabla)C(t,x)\,\mathrm{d}x.$$

Now we only have to understand the second integral. By our formula for the derivative of a determinant,

$$\partial_t J(t,a) = J(t,a)\cdot \operatorname{trace}\left((\partial_t\partial_a\Phi(t,0,a))\cdot(\partial_a\Phi(t,0,a))^{-1}\right).$$

Here $\partial_a\Phi(t,0,a)$ is the Jacobi matrix of the map $a \mapsto x$.

On the other hand, the Jacobi matrix of the inverse map $x \mapsto a$ is $\partial_x\Phi(0,t,x)$, and this is the inverse matrix of $\partial_a\Phi(t,0,a)$. Therefore we have found

$$\partial_t J(t,a) = J(t,a)\cdot \operatorname{trace}\left((\partial_t\partial_a\Phi(t,0,a))\cdot(\partial_x\Phi(0,t,x))\right).$$

We continue to see a as $a = a(x) = \Phi(0,t,x)$. Then we have

$$\frac{\partial U_i}{\partial x_j}(t,x) = \frac{\partial}{\partial x_j}\frac{\partial\Phi_i}{\partial t}(t,0,a(x)) = \sum_{k=1}^n \frac{\partial^2\Phi_i}{\partial t\partial a_k}\cdot\frac{\partial a_k}{\partial x_j} = \sum_{k=1}^n \frac{\partial^2\Phi_i(t,0,a)}{\partial t\partial a_k}\cdot\frac{\partial\Phi_k(0,t,x)}{\partial x_j},$$

which brings us

$$\begin{aligned}\operatorname{div} U(t,x) &= \sum_{j=1}^n \frac{\partial U_j}{\partial x_j}(t,x) = \sum_{j=1}^n\sum_{k=1}^n \frac{\partial^2\Phi_j(t,0,a)}{\partial t\partial a_k}\cdot\frac{\partial\Phi_k(0,t,x)}{\partial x_j}\\ &= \operatorname{trace}\left((\partial_t\partial_a\Phi(t,0,a))\cdot(\partial_x\Phi(0,t,x))\right).\end{aligned}$$

Therefore, $\partial_t J(t,a) = \operatorname{div} U(t,x)\cdot J(t,a)$, and the second integral becomes

$$\int_{\Omega_0} C(t,\Phi(t,0,a))\cdot\operatorname{div} U(t,\Phi(t,0,a))\cdot J(t,a)\,\mathrm{d}a = \int_{\Omega_t} C(t,x)\operatorname{div} U(t,x)\,\mathrm{d}x,$$

hence we conclude that

$$K'(t) = \int_{\Omega_t} C_t(t,x) + (U(t,x)\cdot\nabla)C(t,x)\,\mathrm{d}x + \int_{\Omega_t} C(t,x)\operatorname{div} U(t,x)\,\mathrm{d}x.$$

The remainder of the proof is left to the reader. □

Example 4.5. *Put $C(t,x) = \varrho(t,x)$ as mass density. Then $K(t)$ is the mass of the moving domain Ω_t, and an axiom from physics is that $K(t) \equiv \text{const.}$. Consequently,*

$$0 = \int_{\Omega_t} \varrho_t(t,x) + \operatorname{div}(\varrho U)(t,x)\,\mathrm{d}x.$$

However, the domain Ω_t can be chosen arbitrarily, because we are allowed to focus our attention to a subset of all the particles. Then we find

$$\partial_t \varrho + \operatorname{div}(\varrho U) = 0, \tag{4.1}$$

which is known as conservation of mass *in Eulerian coordinates.*

Example 4.6. *Put $C(t,x) \equiv 1$. Then $K(t)$ equals the volume of the moving domain Ω_t, and our conclusion then is*

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{vol}(\Omega_t) = \int_{\Omega_t} \operatorname{div} U(t,x)\,\mathrm{d}x.$$

Example 4.7. *Consider the Lorenz system from meteorology, which we can rewrite as $x'(t) = U(x)$, or*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}' = U(x) = \begin{pmatrix} \sigma(x_2 - x_1) \\ \varrho x_1 - x_2 - x_1 x_3 \\ -\beta x_3 + x_1 x_2 \end{pmatrix},$$

with positive parameters σ, β, ϱ. We find that $\operatorname{div} U = -\sigma - 1 - \beta$ which is a negative constant. Hence the volume $\operatorname{vol}(\Omega_t)$ solves

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{vol}(\Omega_t) = -(\sigma + 1 + \beta)\operatorname{vol}(\Omega_t),$$

with the explicit solution $\operatorname{vol}(\Omega_t) = \exp(-(\sigma+1+\beta)t)\operatorname{vol}(\Omega_0)$.

We are still not able to solve the Lorenz system (and we will never be), but at least we can say that, for large times, the solution trajectories must stay inside a domain of $\mathbb{R}^3$ with extremely small volume if the initial values at time zero are chosen from a bounded set of $\mathbb{R}^3$.

Example 4.8. *Consider the system $x'(t) = Bx$, with a constant matrix $B \in \mathbb{R}^{n\times n}$. Then $\Phi(t,0,a) = X(t,0)a = \exp(Bt)a$, and the flow transports the domain Ω_0 to $\Omega_t = \exp(Bt)\Omega_0$, which is obtained via multiplication of the matrix $\exp(Bt)$ to any point $a \in \Omega_0$. Then the Jacobian matrix is $\frac{\partial \Phi}{\partial a}(t,0,a) = X(t,0) = \exp(Bt)$, hence $\mathrm{d}x = J(t,a)\,\mathrm{d}a = \det(\exp(Bt))\,\mathrm{d}a$, and the substitution rule then gives*

$$\operatorname{vol}(\Omega_t) = \int_{\Omega_t} 1\,\mathrm{d}x = \int_{\Omega_0} 1 \cdot J(t,a)\,\mathrm{d}a = \det(\exp(Bt)) \int_{\Omega_0} 1\,\mathrm{d}a = \det(\exp(Bt))\operatorname{vol}(\Omega_0).$$

On the other hand, we have $\operatorname{div} U = \operatorname{trace} B$, hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{vol}(\Omega_t) = \int_{\Omega_t} \operatorname{trace}(B)\,\mathrm{d}x = \operatorname{trace}(B)\operatorname{vol}(\Omega_t),$$

with the explicit solution $\operatorname{vol}(\Omega_t) = e^{\operatorname{trace}(B)t}\operatorname{vol}(\Omega_0)$. Comparing both formulae for $\operatorname{vol}(\Omega_t)$ gives $\det(\exp(Bt)) = e^{\operatorname{trace}(B)t}$.

We have proved (3.7) a second time !

Example 4.9 (Conservation of phase space volume[2]). *Consider a mechanical system with d degrees of freedom, with the generalised coordinates $q_1, \ldots, q_d$, and the generalised momentums $p_1, \ldots, p_d$. We assume that this is a* Hamiltonian system[3]*, which means that there is a function $H = H(q,p)$ with*

$$q'(t) = \frac{\partial H}{\partial p}(q,p),$$
$$p'(t) = -\frac{\partial H}{\partial q}(q,p).$$

[2] Erhaltung des Phasenraumvolumens
[3] Hamiltonsches System

We can put $z(t) := (q(t), p(t))^\top$ *and obtain a system* $z'(t) = U(z)$ *which generates a flow in the* $2d$*-dimensional space. We quickly check that* $\operatorname{div} U \equiv 0$*, and the conclusion then is*

$$\operatorname{vol}(\Omega_t) = \operatorname{vol}(\Omega_0).$$

The domain $\mathbb{R}^d_q \times \mathbb{R}^d_p$ *is called* phase space[4]*, and* Ω_0*,* Ω_t *are subsets in it. Therefore we have shown one of the key results of analytical mechanics and statistical physics: the flow generated by a Hamiltonian system preserves the phase space volume.*

This is also known as Theorem of Liouville[5].

Remark 4.10. *For completeness, we mention that two further balance equations are of high importance in fluid dynamics, complementing the local conservation of mass from (4.1). The first is the* local conservation of momentum*,*

$$\partial_t(\varrho U) + \operatorname{div}(\varrho U \otimes U) = f + \operatorname{div}\sigma.$$

The left-hand side is the total acceleration, and the right-hand side contains the vectorial density f *of external forces acting on each molecule (imagine gravity) and the divergence of the stress tensor* σ*, which is a* 3×3 *matrix introduced in the script of the first semester. We remark that* $U \otimes U$ *is a symmetric* 3×3 *matrix with entries* $U_j U_k$*, and* div *is applied on each row separately.*

And the final balance equation is the local conservation of energy

$$\partial_t\left(\varrho\left(e + \frac{1}{2}|U|^2\right)\right) + \operatorname{div}\left(\varrho U\left(e + \frac{1}{2}|U|^2\right)\right) = \langle f, U\rangle + r + \operatorname{div}(\sigma U - q),$$

with $e = e(t,x)$ *being the interior energy per mass (a certain thermodynamical quantity), and* $\frac{\varrho}{2}|U|^2$ *is the volume density of the kinetic energy. The left-hand side is the total time derivative of both energies together. On the right-hand side, we have* $\langle f, U\rangle$ *as the mechanical power*[6]*,* r *is the density of generated heat (imagine an exothermic chemical reaction), and* q *as the heat flux vector (the heat flows into the colder region, therefore the minus).*

4.2 Dynamical Systems and Stability

Definition 4.11. *A* dynamical system *consists of the phase space* $\mathbb{R}^n$*, the additive group* $(\mathbb{R}, +)$*, and a flow*

$$\Phi\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n,$$
$$\Phi\colon (t,a) \mapsto \Phi(t,a),$$

with the following properties:

$$\Phi(0,a) = a, \qquad \forall\, a \in \mathbb{R}^n,$$
$$\Phi(t,\Phi(s,a)) = \Phi(t+s,a), \qquad \forall\, t,s \in \mathbb{R}, \quad \forall\, a \in \mathbb{R}^n,$$
$$\forall\, t \in \mathbb{R}: \quad \Phi(t,\cdot)\colon a \mapsto x \text{ is a diffeomorphism of } \mathbb{R}^n \text{ onto } \mathbb{R}^n.$$

Here a map of $\mathbb{R}^n$ *onto* $\mathbb{R}^n$ *is called a diffeomorphism if it is bijective, differentiable, and the inverse maps is also differentiable.*

Example 4.12. *Let* $x = x(t)$ *be the solution to*

$$x'(t) = U(x(t)), \qquad x(0) = a \in \mathbb{R}^n,$$

with $U \in C^1(\mathbb{R}^n \to \mathbb{R}^n)$*. Set* $\Phi(t,a) := x(t)$*, assuming that* x *has infinite life span.*

Because U does not depend itself on t, we can restrict Φ to depend only on one time variable. Our new notation shall be to write x_0 instead of a.

[4] Phasenraum
[5] Joseph Liouville, 1809 – 1882, french mathematician
[6] mechanische Leistung

Definition 4.13 (**Orbits**). *We call* $\gamma(x_0) := \{\Phi(t, x_0) \colon t \in \mathbb{R}\}$ *the* orbit *of* $x_0 \in \mathbb{R}^n$*, and*

$$\gamma_+(x_0) := \{\Phi(t, x_0) \colon t \geq 0\},$$
$$\gamma_-(x_0) := \{\Phi(t, x_0) \colon t \leq 0\}$$

are called the forward orbit *and* backward orbit *of* x_0*.*

Definition 4.14. *A point* $x_* \in \mathbb{R}^n$ *is called a* stationary point *or* resting point *of* Φ *if* $\gamma(x_*) = \{x_*\}$*.*

A point $x_* \in \mathbb{R}^n$ *is a* periodic orbit of minimal period p *if* $\Phi(p, x_*) = x_*$*, but* $\Phi(t, x_*) \neq x_*$ *for all* t *with* $0 < t < p$*.*

An orbit $\gamma(x_0)$ *is a* hetero-clinic orbit *if there are resting points* x_- *and* x_+ *such that*

$$\lim_{t \to -\infty} \Phi(t, x_0) = x_-, \qquad \lim_{t \to +\infty} \Phi(t, x_0) = x_+.$$

Example 4.15. *Consider the logistic growth model* $x'(t) = \alpha x - \beta x^2$ *with the resting points* $x_* = 0$ *and* $x^* = \alpha/\beta$*. If* $0 < x_0 < \alpha/\beta$ *then* $\gamma(x_0)$ *is a hetero-clinic orbit.*

Lemma 4.16. *Consider the differential equation*

$$x'(t) = U(x(t))$$

with the associated flow Φ*. Then* x_* *is a resting point of* Φ *if and only if* $U(x_*) = 0$*.*

Proof. This is the uniqueness part of the Picard–Lindelöf theorem. □

Definition 4.17 (**Invariant sets**). *A set* $M \subset \mathbb{R}^n$ *is called* positive invariant *if* $\Phi(t, \cdot)$ *maps* M *into itself for all* $t \geq 0$*.*

A set $M \subset \mathbb{R}^n$ *is called* negative invariant *if* $\Phi(t, \cdot)$ *maps* M *into itself for all* $t \leq 0$*.*

A set $M \subset \mathbb{R}^n$ *is called* invariant *if it is positive invariant and negative invariant.*

Example 4.18. *Consider again the logistic growth model. This has the invariant sets*

$$M_1 = (\alpha/\beta, \infty), \qquad M_2 = \{\alpha/\beta\}, \qquad M_3 = (0, \alpha/\beta), \qquad M_4 = \{0\}, \qquad M_5 = (-\infty, 0),$$

and all unions of these sets.

Definition 4.19 (**Stable points**). *A resting point* $x_* \in \mathbb{R}^n$ *is a* stable point of Φ *if*

$$\forall\, \varepsilon > 0 \quad \exists\, \delta > 0 \colon \ \|x_0 - x_*\| < \delta \implies \|\Phi(t, x_0) - x_*\| < \varepsilon, \quad \forall\, t \geq 0.$$

A resting point $x_* \in \mathbb{R}^n$ *is an* asymptotically stable point of Φ *if it is stable and there is a positive* δ_0 *such that*

$$\|x_0 - x_*\| < \delta_0 \implies \lim_{t \to +\infty} \|\Phi(t, x_0) - x_*\| = 0.$$

A point of $\mathbb{R}^n$ *is called* unstable *if it is not stable.*

Example 4.20. *The logistic growth model has the unstable resting point* $x_* = 0$ *and the asymptotically stable point* $x^* = \alpha/\beta$*.*

Example 4.21. *A wooden pendulum with friction has two resting positions: pendulum down (asymptotically stable) and pendulum up (unstable).*

The above definition of stability does not cover stable limit cycles as sketched in the figure.

Definition 4.22 (**Stable sets**). *A set* $M \subset \mathbb{R}^n$ *is called* stable *if*

$$\forall\, \varepsilon > 0 \quad \exists\, \delta > 0 \colon \ \operatorname{dist}(x_0, M) < \delta \implies \operatorname{dist}(\Phi(t, x_0), M) < \varepsilon, \quad \forall\, t \geq 0.$$

A set $M \subset \mathbb{R}^n$ *is called* asymptotically stable *if it is stable and there is a positive* δ_0 *such that*

$$\operatorname{dist}(x_0, M) < \delta_0 \implies \lim_{t \to +\infty} \operatorname{dist}(\Phi(t, x_0), M) = 0.$$

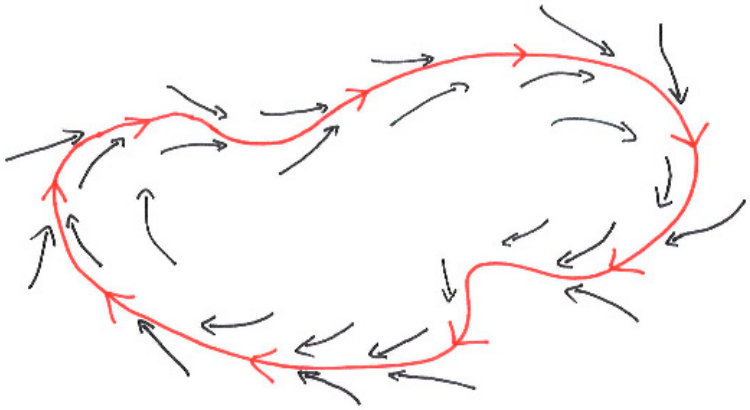


Figure 4.1: A stable limit cycle

Now we look at $x'(t) = U(x(t))$ and ask how we can determine the stability type of a resting point.

As an example, we take the logistic growth model $x'(t) = \alpha x - \beta x^2$, and call y the deviation from the stationary point $x^* = \alpha/\beta$:

$$x(t) = x^* + y(t).$$

Then it follows that

$$\begin{aligned} y'(t) = x'(t) &= \alpha(x^* + y) - \beta(x^* + y)^2 = \alpha x^* + \alpha y - \beta(x^*)^2 - 2\beta x^* y - \beta y^2 \\ &= -\alpha y - \beta y^2, \end{aligned}$$

and for $y \approx 0$ we have $|\beta y^2| \ll |\alpha y|$, giving us the vague hope that the solution y behaves similarly to the solution of the decay equation

$$z'(t) = -\alpha z, \qquad |z(0)| \ll 1.$$

Theorem 4.23. *Consider the system*

$$x'(t) = U(x(t))$$

with $U \in C^2(\mathbb{R}^n \to \mathbb{R}^n)$. Then a point x^* is asymptotically stable if $U(x^*) = 0$, and additionally all eigenvalues of $U'(x^*)$ have negative real part. If one eigenvalue of $U'(x^*)$ has positive real part, then x^* is unstable.

Sketch of proof. We set $y(t) = x(t) - x^*$ and find

$$y'(t) = x'(t) = U(x^* + y(t)) = U(x^*) + U'(x^*)y + \mathfrak{O}(\|y(t)\|^2).$$

Note that $y(t)$ can be written as

$$y(t) = e^{U'(x^*)t}y(0) + \int_{s=0}^{t} e^{U'(x^*)(t-s)}\mathfrak{O}(\|y(s)\|^2)\,\mathrm{d}s,$$

and now we have to show that the integral term is smaller than the first term for small $\|y(0)\|$. □

Now we forget about the remainder term and consider only linear systems for two unknown functions $x_1 = x_1(t)$ and $x_2 = x_2(t)$. For given real parameters a, b, c, d, this system shall have the form

$$x'(t) = Ax(t), \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

and the eigenvalues of A are determined as

$$\lambda_{1,2} = \frac{1}{2}\left(T \pm \sqrt{T^2 - 4\Delta}\right),$$

with $T = \operatorname{trace} A = a + d$ being the trace, and $\Delta = \det A$ the determinant of A.
Depending on these two values, the flow will have one of the following types.

Case A: $T^2 - 4\Delta > 0$**:** Then λ_1 and λ_2 are both real.

Case A1: $\Delta < 0$**:** Because of $\lambda_1\lambda_2 = \Delta$ the eigenvalues λ_1, λ_2 have different sign, the resting point is $(x_1, x_2) = (0, 0)$. By a suitable rotation of the coordinate system (in other words: a diagonalisation of the matrix A), the system has new unknown functions (y_1, y_2), and the system will be $y_1' = \lambda_1 y_1$, $y_2' = \lambda_2 y_2$ with the solutions $y_1(t) = \exp(\lambda_1 t) y_{1,0}$ and $y_2(t) = \exp(\lambda_2 t) y_{2,0}$, respectively. Expressed in the phase space, the solution will be

$$y_2(t) = \text{const.}(y_1(t))^{\frac{\lambda_2}{\lambda_1}}, \qquad \frac{\lambda_2}{\lambda_1} < 0.$$

The stationary point is called a *saddle point*.

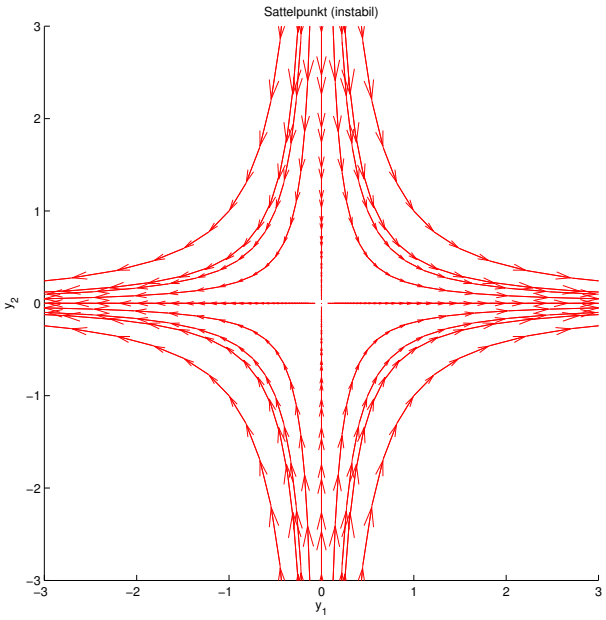


Figure 4.2: saddle point

Case A2: $\Delta > 0$ **and** $T > 0$**:** Then the λ_j are both positive, and we find the representation

$$y_2(t) = \text{const.}(y_1(t))^{\frac{\lambda_2}{\lambda_1}}, \qquad \frac{\lambda_2}{\lambda_1} > 0,$$

and the stationary point $(0, 0)$ is an *unstable node*.

Case A3: $\Delta > 0$ **and** $T < 0$**:** then the λ_j are both negative, and we have a *stable node*.

Case A4: $\Delta = 0$ **and** $T > 0$**:** then $\lambda_1 = 0$ and $\lambda_2 = S > 0$. The differential system turns into $y_1' = 0$ and $y_2' = \lambda_2 y_2$. The resting position $(y_1, y_2) = (0, 0)$ is part of a *line of unstable resting positions*.

Case A5: $\Delta = 0$ **and** $T < 0$**:** then $\lambda_1 = 0$ and $\lambda_2 = S < 0$. The resting position $(y_1, y_2) = (0, 0)$ is part of a *line of stable resting positions*.

Case B: $T^2 - 4\Delta < 0$**:** then the eigenvalues λ_1, λ_2 can not be real numbers, but we have $\lambda_1 = \overline{\lambda_2}$.

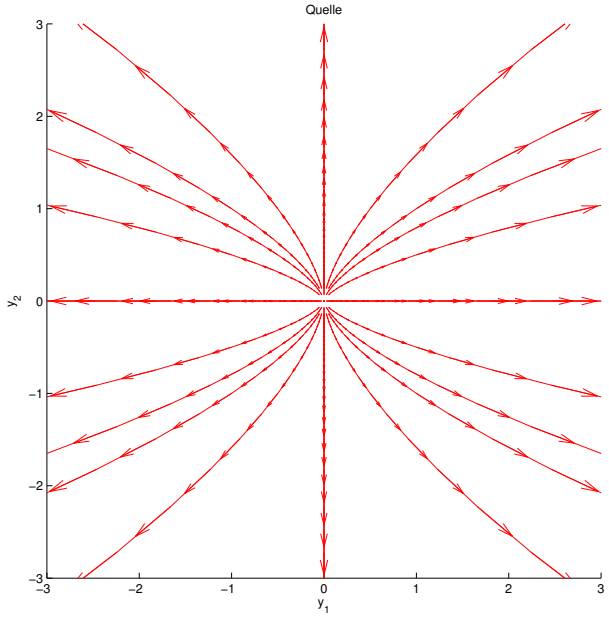


Figure 4.3: unstable node

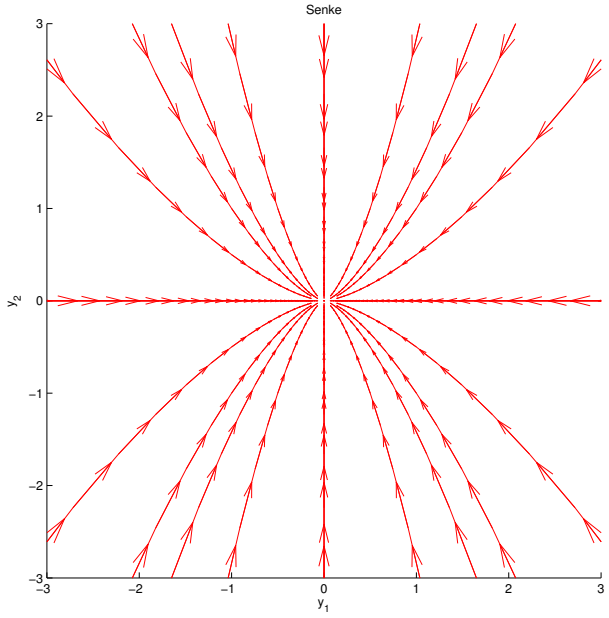


Figure 4.4: stable node

Case B1: $T > 0$**:** Then $\Re\lambda_j > 0$, and we have, after a rotation,

$$y_1(t) = y_{1,0}e^{\alpha t}\cos(\omega t), \qquad y_2(t) = y_{2,0}e^{\alpha t}\sin(\omega t), \qquad \alpha = \Re\lambda_j > 0.$$

The position $(y_1, y_2) = (0, 0)$ is an *unstable vortex*[7].

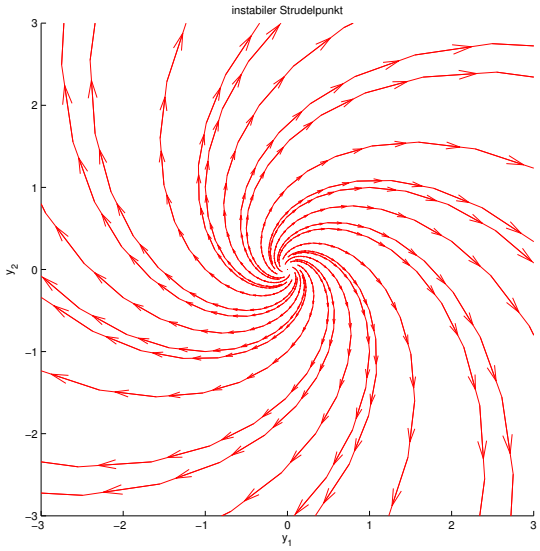


Figure 4.5: unstable vortex

Case B2: $T < 0$**:** Then $\Re\lambda_j < 0$, and we have a *stable vortex*.

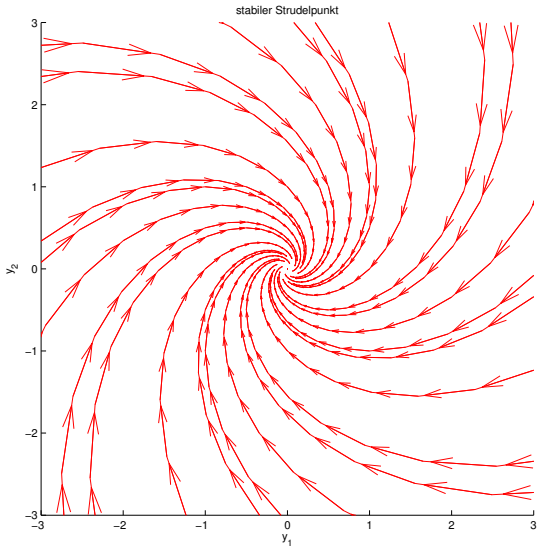


Figure 4.6: stable vortex

Case B3: $T = 0$**:** The eigenvalues are, $\lambda_{1,2} = \pm\mathrm{i}\sqrt{-\Delta}$, and the solution curves in the phase space are ellipses. The resting position is called a *centre*.

Case C: $T^2 - 4\Delta = 0$**:**

[7] instabiler Strudelpunkt

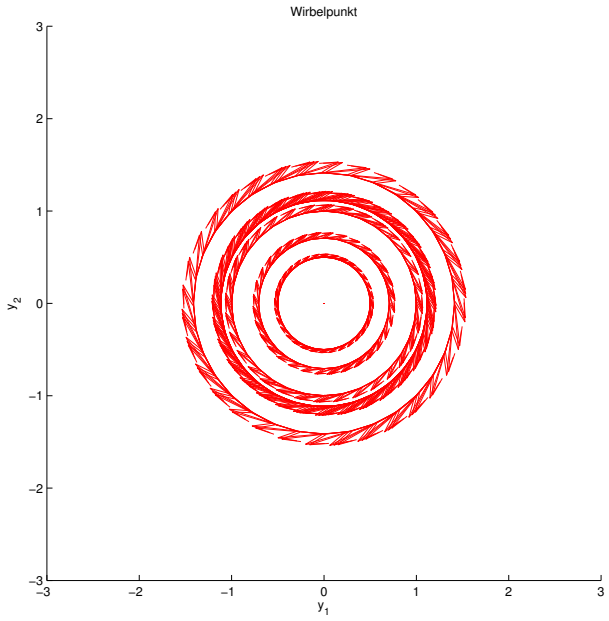


Figure 4.7: center

Case C1: $T > 0$ Then A possesses an eigenvalue $\lambda > 0$ of algebraic multiplicity 2 and geometric multiplicity 1 or 2 (we ignore the latter case). After a transformation into Jordan normal form, we find

$$y_1(t) = e^{\lambda t} y_{1,0} + t e^{\lambda t} y_{2,0}, \qquad y_2(t) = e^{\lambda t} y_{2,0},$$

and the resting position $(y_1, y_2) = (0, 0)$ is known as *degenerate unstable node.*

Case C2: $T < 0$**:** Now we have $\lambda < 0$, and the resting position $(y_1, y_2) = (0, 0)$ is called *degenerate stable node.*

4.3 Outlook: Stability of Periodic Solutions, and The Over Head Pendulum

(Outlook sections are not relevant to examinations)

Consider the initial value problem

$$z'(t) = U(z(t)), \qquad z(0) = z_0,$$

and assume that $z = z(t)$ is a periodic solution. Here z takes values in $\mathbb{R}^n$. We would like to know how the solution chances if z_0 is chosen slightly different. Write $z(t; z_0)$ for the solution with initial value z_0.

Therefore we should discuss $x(t) := \frac{\partial z(t;z_0)}{\partial z_0}$ which is a function from $\mathbb{R}$ to $\mathbb{R}^{n\times n}$. Then $x = x(t)$ solves the matrix differential equation

$$x'(t) = U'(z(t; z_0))x(t), \qquad x(0) = I.$$

Question: Prove this.

The question now is whether $x(t)$ stays bounded for $t \to \infty$, which, by linearity of the equation, is equivalent to the stability of the zero solution to $x'(t) = U'(z(t; z_0))x(t)$.

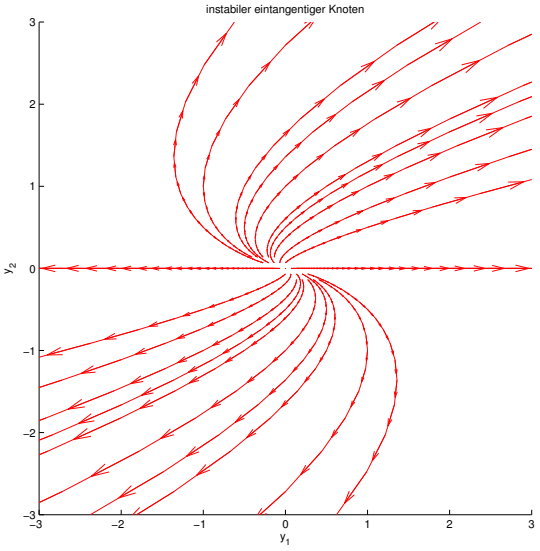


Figure 4.8: degenerate unstable node

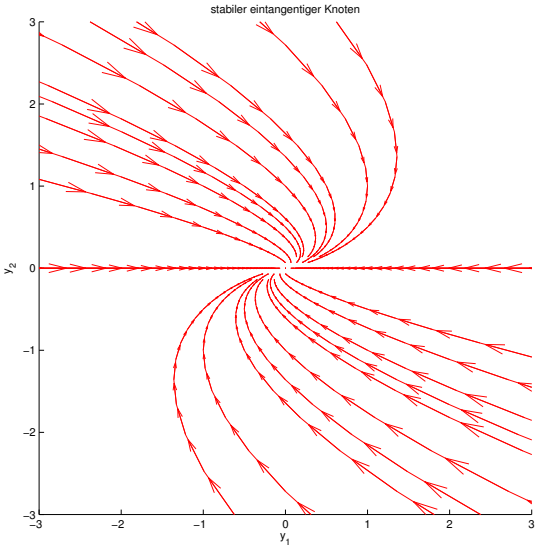


Figure 4.9: degenerate stable node

A more general situation is the following. Consider

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

with a matrix function $A = A(t)$ which is periodic: $A(t+p) = A(t)$ for all t. Then we can write

$$x(t) = X(t, t_0)x_0,$$

with X as the well-known fundamental solution. We ask for the stability of the zero solution $x \equiv 0$. The answer will flow out of the following celebrated result:

Theorem 4.24 (Floquet[8]**'s Theorem**). *Let $A \in C(\mathbb{R} \to \mathbb{C}^{n\times n})$ be periodic with period p. Then there is a p-periodic function $Z = Z(t)$ and a constant matrix $B \in \mathbb{C}^{n\times n}$ with*

$$X(t, 0) = Z(t)\exp(Bt), \qquad \forall\, t \in \mathbb{R}.$$

A consequence then is

$$\begin{aligned} X(t, t_0) &= X(t, 0)X(0, t_0) = X(t, 0)(X(t_0, 0))^{-1} \\ &= Z(t)\exp(Bt)\exp(-Bt_0)(Z(t_0))^{-1} = Z(t)\exp(B(t - t_0))(Z(t_0))^{-1}. \end{aligned}$$

Sketch of proof. First we have

$$I = X(0, 0) = Z(0)\exp(0B) \quad \Longrightarrow \quad Z(0) = I.$$

If the function X has the above representation, then it follows that

$$C := X(p, 0) = Z(p)\exp(Bp) = Z(0)\exp(Bp) = \exp(Bp).$$

Now we wish to determine a matrix $B \in \mathbb{C}^{n\times n}$ with $C = \exp(Bp)$. For simplicity, we assume that C can be diagonalised:

$$C = S^{-1}\Lambda S, \qquad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

Since $X(t, t_0)$ is always invertible, also C is invertible, hence each λ_j is non-zero. Choose a $\gamma_j \in \mathbb{C}$ with $\exp(\gamma_j) = \lambda_j$. Then γ_j are uniquely determined up to multiples of $2\pi\mathrm{i}$. Now put

$$B := \frac{1}{p}S^{-1}\Gamma S, \qquad \Gamma := \operatorname{diag}(\gamma_1, \ldots, \gamma_n).$$

Having chosen B, we next define $Z(t) := X(t, 0)\exp(-Bt)$, and it remains to check whether Z is indeed p-periodic:

$$\begin{aligned} Z(t+p) &= X(t+p, 0)\exp(-B(t+p)) = X(t, -p)\exp(-Bp)\exp(-Bt) = X(t, -p)C^{-1}\exp(-Bt) \\ &= X(t, -p)X(0, p)\exp(-Bt) = X(t, -p)X(-p, 0)\exp(-Bt) = X(t, 0)\exp(-Bt) = Z(t). \end{aligned}$$

Here we have made repeated use of $X(t+p, s+p) = X(t, s)$ for all t, s. This is true because $t \mapsto X(t+p, s+p)$ solves

$$\partial_t\Psi(t) = A(t+p)\Psi(t) = A(t)\Psi(t), \qquad \Psi(s) = I,$$

and the matrix function $t \mapsto X(t, s)$ solves

$$\partial_t\tilde{\Psi}(t) = A(t)\tilde{\Psi}(t), \qquad \tilde{\Psi}(s) = I.$$

By the uniqueness statement in the Picard–Lindelöf theorem, we have $\Psi = \tilde{\Psi}$, hence $X(t+p, s+p) = X(t, s)$ for all $t, s \in \mathbb{R}$. □

Definition 4.25 (Floquet **exponents and** Floquet **multipliers**). *The eigenvalues μ_1 , …, μ_n of B are called* Floquet exponents*, and they are unique up to multiples of $2\pi\mathrm{i}/p$.*

The eigenvalues λ_1, …, λ_n of $X(p, 0)$ are called Floquet multipliers*, and they are unique.*

[8] Achille Marie Gaston Floquet, 1847–1920

Now the structure of X has been determined, and we quickly find the stability behaviour of the zero solution $x \equiv 0$ to $x'(t) = A(t)x(t)$:

the zero state is stable if and only if all Floquet exponents $\mu_1, \ldots, \mu_n$ have real part ≤ 0, and if $\mu_j \in \mathrm{i}\mathbb{R}$, then its algebraic and geometric multiplicities are the same,

the zero state is asymptotically stable if and only if all Floquet exponents have negative real part,

the zero state is asymptotically stable if and only if all Floquet multipliers are in the interior of the unit disk of $\mathbb{C}$.

Warning: *It can indeed happen that the eigenvalues of $A(t)$ are always in the open left half-plane of $\mathbb{C}$, but B has an eigenvalue of positive real part. Therefore the stability behaviour of the zero state can not be determined from A alone. An example is given in [12], Chapter 8:*

$$A(t) = \begin{pmatrix} -1 + (3/2)\cos^2 t & 1 - (3/2)\cos t \sin t \\ -1 - (3/2)\sin t \cos t & -1 + (3/2)\sin^2 t \end{pmatrix}$$

with eigenvalues $-\frac{1}{4} \pm \frac{\mathrm{i}}{4}\sqrt{7}$ (definitely being in the left half-plane for all t), but the vector function $x(t) = (-\cos t, \sin t)^\top e^{t/2}$ solves $x'(t) = A(t)x(t)$.

We come to the example of the **over head pendulum** which solves

$$x''(t) = \omega^2 x, \qquad \omega^2 = \frac{g}{l},$$

where g is the gravitational acceleration on the earth, l the length of the pendulum. The zero state is unstable. Now we show that it can be stabilised by vertical vibrations of the point of suspension. Suppose that the suspension point moves up and down with period 2τ and acceleration $\pm c$. Then we get

$$x''(t) = \big(\omega^2 + h(t)\big)\, x(t),$$
$$h(t) = \begin{cases} -\alpha^2 & : 0 < t < \tau, \\ +\alpha^2 & : \tau < t < 2\tau, \end{cases} \qquad \alpha^2 = \frac{c}{l}.$$

The amplitude of this vibration shall be a. The times of maximal elongation are $t = \tau/2$ and $t = 3\tau/2$, hence

$$a = \frac{c}{2}\left(\frac{\tau}{2}\right)^2 = \frac{c\tau^2}{8}, \qquad \alpha^2 = \frac{c}{l} = \frac{8a}{l\tau^2}.$$

In the sequel, we will select a and α suitably, and then τ will follow.

We know: if $a = 0$, then the resting position $x = 0$ is a saddle point with eigenvalues $\pm\omega$, hence unstable.

To transfer into a first order system, we set $y = x$ and $z = \dot{x}$, with the consequence

$$\begin{pmatrix} \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \omega^2 + h(t) & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} = A(t) \begin{pmatrix} y \\ z \end{pmatrix}.$$

The matrix A has jumps, but we ignore this.

The sum of the eigenvalues of $A(t)$ is zero.

If $\omega^2 + h(t) \geq 0$, then A has two real eigenvalues.

If $\omega^2 + h(t) < 0$, then A has two imaginary eigenvalues.

To obtain stability, it sounds reasonable that (at least for certain times) $A(t)$ should not have eigenvalues in the right half-plane. Therefore we prefer to arrange the constants in such a way that

$$\alpha^2 > \omega^2.$$

The fundamental solution then is

$$X(2\tau, 0) = X(2\tau, \tau)X(\tau, 0) = \begin{pmatrix} \cosh k\tau & \frac{1}{k}\sinh k\tau \\ k\sinh k\tau & \cosh k\tau \end{pmatrix} \begin{pmatrix} \cos \Omega\tau & \frac{1}{\Omega}\sin \Omega\tau \\ -\Omega \sin \Omega\tau & \cos \Omega\tau \end{pmatrix},$$

where we have put $k^2 := \alpha^2 + \omega^2$ and $\Omega^2 = \alpha^2 - \omega^2$. The reason is that $A(t)$ is a constant function on each half-period, and on these intervals, X is given by Lemma 3.13.

Next we need the eigenvalues of $X(2\tau, 0)$. Because $\operatorname{trace} A(t) = 0$ for each t, (3.7) gives $\det X(t, t_0) = 1$ for all t, t_0, in particular $\det X(2\tau, 0) = 1$, and hence $\lambda_+ \cdot \lambda_- = 1$. Then the eigenvalues are given by

$$\lambda_\pm = \frac{1}{2}\left(\operatorname{trace} X(2\tau,0) \pm \sqrt{(\operatorname{trace} X(2\tau,0))^2 - 4}\right).$$

Our goal is $|\lambda_\pm| \leq 1$, and to this end it is necessary to have $|\operatorname{trace} X(2\tau, 0)| \leq 2$, which is equivalent to

$$\left|2\cosh(k\tau)\cos(\Omega\tau) + \left(\frac{k}{\Omega} - \frac{\Omega}{k}\right)\sinh(k\tau)\sinh(\Omega\tau)\right| \leq 2,$$

as can be found quickly if we know that $\operatorname{trace}(PQ) = \sum_{j,k} p_{jk}q_{kj}$ for all matrices P, Q of quadratic shape. Now our assumptions shall be:

$$\frac{a}{l} =: \varepsilon^2 \ll 1, \qquad \frac{g}{c} =: \mu^2 \ll 1,$$

and from this we conclude that

$$\begin{aligned}
k\tau &= \sqrt{\alpha^2+\omega^2}\sqrt{\frac{8a}{c}} = \sqrt{8}\sqrt{\frac{c}{l}+\frac{g}{l}}\sqrt{\frac{a}{c}} = 2\sqrt{2}\sqrt{\frac{a}{l}+\frac{a}{l}\cdot\frac{g}{c}} = 2\sqrt{2}\varepsilon\sqrt{1+\mu^2},\\
\Omega\tau &= \sqrt{\alpha^2-\omega^2}\sqrt{\frac{8a}{c}} = \sqrt{8}\sqrt{\frac{c}{l}-\frac{g}{l}}\sqrt{\frac{a}{c}} = 2\sqrt{2}\sqrt{\frac{a}{l}-\frac{a}{l}\cdot\frac{g}{c}} = 2\sqrt{2}\varepsilon\sqrt{1-\mu^2},\\
\frac{k}{\Omega} &= \frac{\sqrt{1+\mu^2}}{\sqrt{1-\mu^2}} = \frac{1+\frac{1}{2}\mu^2}{1-\frac{1}{2}\mu^2} + \mathfrak{O}(\mu^4) = 1+\mu^2+\mathfrak{O}(\mu^4).
\end{aligned}$$

Then we directly get

$$\begin{aligned}
\frac{k}{\Omega} - \frac{\Omega}{k} &= 2\mu^2 + \mathfrak{O}(\mu^4),\\
\cosh(k\tau) &= 1 + 4\varepsilon^2(1+\mu^2) + \frac{8}{3}\varepsilon^4 + \mathfrak{O}(\mu^6+\varepsilon^6),\\
\cos(\Omega\tau) &= 1 - 4\varepsilon^2(1-\mu^2) + \frac{8}{3}\varepsilon^4 + \mathfrak{O}(\mu^6+\varepsilon^6),\\
\left(\frac{k}{\Omega} - \frac{\Omega}{k}\right)\sinh(k\tau)\sin(\Omega\tau) &= 16\varepsilon^2\mu^2 + \mathfrak{O}(\mu^6+\varepsilon^6),
\end{aligned}$$

hence we wish to arrange that

$$2\left(1 + 8\varepsilon^2\mu^2 + \frac{16}{3}\varepsilon^4 - 16\varepsilon^4\right) + 16\varepsilon^2\mu^2 < 2,$$

or $3\mu^2 < \varepsilon^2$ or $\frac{g}{c} < \frac{a}{3l}$ or $\tau^2 < \frac{8a^2}{3lg}$.

We take a concrete example: if $l = 20$cm and $a = 1$cm then $\tau < 0.01166$, which corresponds to a vibration frequency of at least 43Hz.

A thorough presentation of this topic can be found in § 28 of [2].

4.4 Geometric Investigations of Dynamical Systems

The system $x'(t) = f(x)$ with stationary state x^* leads, after setting $x(t) =: x^* + y(t)$, to

$$y'(t) = f'(x^*)y + R(y), \qquad R(y) = \mathfrak{O}(\|y\|^2),$$

and neglecting the quadratic remainder term then brings us to the system

$$u'(t) = f'(x^*)u,$$

called the *linearisation* of $x'(t) = f(x)$.

We know

- if all the eigenvalues of $f'(x^*)$ are in $\mathbb{C}_- = \{z \in \mathbb{C}\colon \Re z < 0\}$, then x^* is an asymptotically stable stationary state of the nonlinear system,
- if at least one eigenvalue of $f'(x^*)$ is in $\mathbb{C}_+ = \{z \in \mathbb{C}\colon \Re z > 0\}$, then x^* is an instable resting point of the nonlinear system.

And in both cases, the orbits of the nonlinear system and of the linearised systems look very similar (up to minor deformations) near x^* and 0, respectively.

However, the situation changes completely if $f'(x^*)$ has eigenvalues on the imaginary axis — then the stability behaviour of x^* can depend heavily on the nonlinear term R.

To understand the situation better, so-called LYAPUNOV[9] functions may be helpful. These are positive definite functions near x^*.

Definition 4.26. *A function $V \in C^1(\Omega \to \mathbb{R})$ with $\Omega \subset \mathbb{R}^n$ as a neighbourhood of x^* is called* positive definite with respect to x^* *if*

$$V(x^*) = 0, \qquad V(x) > 0 \quad \forall\, x \in \Omega \setminus \{x^*\}.$$

If $V \in C^2(\Omega \to \mathbb{R})$ is positive definite, and if the Hessian $(\nabla \otimes \nabla V)(x^*)$ is a strictly positive definite matrix then the *level sets*[10] $\{x \in \Omega\colon V(x) = \text{const.}\}$ are diffeomorphic to the unit sphere[11], at least near x^*. Compare Figure 2.1.

Now the key idea is to look whether the vectors $f(x)$ cross the level sets everywhere from the outside to the inner side. If yes, then x^* is asymptotically stable. Note that the scalar function $t \mapsto V(x(t))$ has derivative

$$\partial_t V(x(t)) = \nabla V(x) \cdot x'(t) = \langle \nabla V(x), f(x) \rangle\,,$$

and ∇V is perpendicular to the level sets of V, and points outwards.

Proposition 4.27. *Let x^* be a stationary point to the dynamical system governed by $x'(t) = f(x)$, and V be positive definite with respect to x^*. Then*

- *if $\langle \nabla V(x), f(x) \rangle \le 0$ for all $x \in \Omega$, then x^* is stable,*
- *if $\langle \nabla V(x), f(x) \rangle < 0$ for all $x \in \Omega \setminus \{x^*\}$, then x^* is asymptotically stable,*
- *if $\langle \nabla V(x), f(x) \rangle > 0$ for all $x \in \Omega \setminus \{x^*\}$, then x^* is unstable.*

One example is the differential equation $z'' + 2az' + z + z^3 = 0$, with $0 < a < 1$. Setting $x_1 = z$ and $x_2 = z'$, we find the system

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & -2a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -x_1^3 \end{pmatrix},$$

with the only stationary point $x^* = (0,0)^\top$. The Jacobi matrix is

$$f'(x^*) = \begin{pmatrix} 0 & 1 \\ -1 & -2a \end{pmatrix},$$

having the eigenvalues $-a \pm \mathrm{i}\sqrt{1-a^2}$, making x^* asymptotically stable. An interesting question is about the size of its catchment basin[12], and we will find an estimate of this size using a carefully constructed Lyapunov functional.

First we bring the matrix $f'(x^*)$ into a normal form. The substitution $y = Px$ (with $P \in \mathbb{R}^{2\times 2}$ as a matrix not yet chosen) brings us

$$y' = Pf'(x^*)P^{-1} \cdot y + \mathfrak{O}(\|y\|^2),$$

[9] ALEKSANDR MIKHAILOVICH LYAPUNOV, 1857 – 1918
[10] Niveaulinien, Höhenlinien
[11] this means that there is a diffeomorphism which maps them onto the unit sphere
[12] Einzugsgebiet (z.B. hydrogeologisch)

and now P shall be selected in such a way that $Pf'(x^*)P^{-1}$ is "as nice as possible". The typical Jordan normal form will *not* work, because the eigenvalues of $f'(x^*)$ are non-real, and then also the eigenvectors of $f'(x^*)$ will be non-real. Instead, we use the *real Jordan normal form*:

Lemma 4.28 (**Real Jordan Normal Form**). *Let $A \in \mathbb{R}^{2\times 2}$. Then there is a matrix $P \in \mathbb{R}^{2\times 2}$ such that $J = PAP^{-1}$ has one of the following three forms:*

$$J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \qquad J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \qquad J = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix},$$

where $\lambda_1, \lambda_2, \lambda, \alpha, \beta \in \mathbb{R}$ and $\beta \neq 0$.

The proof is a nice exercise.

In our example, we take $\beta = \sqrt{1-a^2}$ and

$$P = \begin{pmatrix} 1 & 0 \\ -a & \beta \end{pmatrix}, \quad P^{-1} = \frac{1}{\beta}\begin{pmatrix} \beta & 0 \\ a & 1 \end{pmatrix},$$

as well as $y = Px$. In particular, this brings $y_1 = x_1$, which is a useful information for the transformation of the nonlinear terms. Then we get

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} -a & +\beta \\ -\beta & -a \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{\beta}y_1^3 \end{pmatrix}.$$

Now the variables y_1 and y_2 play the same rôle in the linear principal part (as can be seen from the coinciding entries on the diagonal), and it seems reasonable to try our luck with

$$V(y_1, y_2) = \frac{1}{2a}(y_1^2 + y_2^2).$$

This gives, after some computation,

$$\langle \nabla_y V, f(P^{-1}x)\rangle = -(y_1^2 + y_2^2) - \frac{1}{a\beta}y_1^3 y_2,$$

and now the question comes up where this is negative. We conjecture that this expression is negative at least in a circle with radius r_0 about the origin. If $y_1^2 + y_2^2 \leq r_0^2$, then $|y_1^3 y_2| = y_1^2|y_1 y_2| \leq r_0^2|y_1 y_2| \leq r_0^2 \frac{1}{2}(y_1^2 + y_2^2)$, by the Binomi[13] formula, and now we should make sure that

$$\frac{1}{a\beta}|y_1^3 y_2| \leq y_1^2 + y_2^2 \quad \Longleftarrow \quad \frac{r_0^2}{2a\beta} \leq 1$$

which holds, for instance, if $r_0 = \sqrt{2a\beta}$.

Hence we have shown: if the starting point $(y_1(0), y_2(0))^\top$ is in a ball about the origin with radius r_0, then the trajectory is attracted to the origin. This is just a lower estimate; in reality, the catchment basin is certainly larger. After transforming back to the x variables, the basin becomes an ellipse.

For describing the long time asymptotics of a dynamical system, we need one more concept.

Definition 4.29. *Let Φ be a flow on $\mathbb{R}^n$, and $x_0 \in \mathbb{R}^n$. Then the sets*

$$\omega(x_0) := \left\{ y \in \mathbb{R}^n : \exists \text{ sequence } (t_k)_{k\in\mathbb{N}} \text{ with } t_k \nearrow +\infty \text{ and } y = \lim_{k\to+\infty} \Phi(t_k, x_0) \right\},$$

$$\alpha(x_0) := \left\{ y \in \mathbb{R}^n : \exists \text{ sequence } (t_k)_{k\in\mathbb{N}} \text{ with } t_k \searrow -\infty \text{ and } y = \lim_{k\to+\infty} \Phi(t_k, x_0) \right\}$$

are called ω limit set and α limit set, respectively.

[13] Ernesto Binomi, 1210–1331, sikinian mathematician

For an understanding of the choice of the letters, look at the greek ABC.

From the definition we quickly deduce that[14]

$$\omega(x_0) = \bigcap_{t\geq 0} \overline{\gamma_+(\Phi(t,x_0))},$$
$$\alpha(x_0) = \bigcap_{t\leq 0} \overline{\gamma_-(\Phi(t,x_0))}.$$

If the trajectory starting in x_0 approaches a limit point, then $\omega(x_0)$ is exactly that one limit point. If the trajectory starting in x_0 approaches a periodic cycle (in the sense of Figure 4.1), then $\omega(x_0)$ is that limit cycle. And if the forward orbit $\gamma_+(x_0)$ is unbounded, then $\omega(x_0)$ might be the empty set.

Proposition 4.30. *Let $\gamma_+(x_0)$ be bounded. Then $\omega(x_0)$ is compact, non-empty, invariant, and connected.*

Proposition 4.31 (Invariance principle). *Let $V \in C(\mathbb{R}^n \to \mathbb{R})$ and $k \in \mathbb{R}$. Put*

$$\Omega = \{x \in \mathbb{R}^n \colon V(x) < k\}.$$

We assume that $V \in C^1(\Omega \to \mathbb{R})$, and that

$$\langle \nabla V(x), f(x)\rangle \leq 0 \qquad \forall x \in \Omega.$$

Define $S = \{x \in \Omega \colon \langle \nabla V(x), f(x)\rangle = 0\}$, and M as the biggest invariant subset of S.

Then the following holds: each forward orbit that starts in Ω and stays bounded possesses an ω limit set which is contained in M.

In many situations, the function V grows to infinity at the "boundary" of $\mathbb{R}^n$:

Proposition 4.32. *Let $V \in C^1(\mathbb{R}^n \to \mathbb{R})$ with $V(x) \to +\infty$ for $\|x\| \to +\infty$. Furthermore, suppose $\langle \nabla V(x), f(x)\rangle \leq 0$ for all $x \in \mathbb{R}^n$.*

Then each forward orbit is bounded, and each forward orbit has its ω limit set in M, the biggest invariant subset of

$$\{x \in \mathbb{R}^n \colon \langle \nabla V(x), f(x)\rangle = 0\}.$$

To apply this knowledge, we return to the differential equation

$$z'' + 2az' + z + z^3 = 0.$$

We set $x_1 = z$ and $x_2 = z'$, hence

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & -2a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 0 \\ x_1^3 \end{pmatrix}.$$

The positive number a can be understood as friction coefficient. If $a = 0$, then we obtain a conservative system with the energy

$$V(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{4}x_1^4.$$

In that case, the orbits are running along the level sets $V(x_1, x_2) = \text{const}$.

Now assume $a > 0$:

$$\partial_t V(x_1(t), x_2(t)) = \langle \nabla V(x), f(x)\rangle = \left\langle \begin{pmatrix} x_1 + x_1^3 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_2 \\ -x_1 - 2ax_2 - x_1^3 \end{pmatrix} \right\rangle = -2ax_2^2 \leq 0.$$

[14] Here the over-bar denotes the *topological closure*, which means that you augment the set $\gamma_\pm(\Phi(t,x_0))$ with all its cluster points.

For the invariance principle, we note that

$$V \in C^1(\mathbb{R}^2 \to \mathbb{R}),$$
$$V(x) \to +\infty \quad \text{if} \quad \|x\| = \sqrt{x_1^2 + x_2^2} \to \infty.$$

Hence each forward orbit is bounded, and its ω limit is contained in the biggest invariant subset of

$$\left\{x \in \mathbb{R}^2 \colon \langle \nabla V(x), f(x)\rangle = 0\right\} = \left\{x \in \mathbb{R}^2 \colon x_2^2 = 0\right\}.$$

Therefore, we should look for the biggest invariant subset M of $\{x \in \mathbb{R}^2 \colon x_2 = 0\}$.

For $x^* \in M$ we have the representation $x^* = (x_1^*, x_2^*) = (x_1^*, 0)$. Next $\Phi(t, x_*)$ remains in M, because M is invariant. Consequently we have

$$\Phi(t, x^*) = (x_1(t), 0).$$

On the other hand, $\partial_t x_1(t) = x_2 = 0$, which gives us $x_1(t) = x_1^*$ for all $t \in \mathbb{R}$. Additionally,

$$0 = \partial_t 0 = \partial_t x_2(t) = -x_1 - 2ax_2 - x_1^3 = -x_1 - x_1^3 = -x_1(1 + x_1^2),$$

and therefore $x_1^* = x_1(t) = 0$. As a consequence, the biggest invariant subset M of $\{x \in \mathbb{R}^2 \colon x_2 = 0\}$ is $\{(0,0)\}$, and each forward orbit is approaching this point.

Therefore, the catchment basin of the origin is the whole $\mathbb{R}^2$.

Concerning a system $x' = f(x)$, we would like to know how $\omega(x_0)$ might look like. A particularly nice answer is possible in the case of $n = 2$.

Theorem 4.33 (Poincare[15] — Bendixson[16]). *Let $f \in C^2(\mathbb{R}^2 \to \mathbb{R}^2)$, and assume the forward orbit $\gamma_+(x_0)$ as bounded. Then exactly one of the following cases occurs:*

1. *$\omega(x_0)$ is a periodic orbit,*
2. *for each $y \in \omega(x_0)$ the following holds: $\alpha(y)$ as well as $\omega(y)$ consist only of resting points.*

Key ideas of the proof are:

- the Jordan curve theorem which states that each closed Jordan curve splits $\mathbb{R}^2$ into two parts: an interior part, and an exterior part,
- different solution trajectories can not cross.

Possible ω limit sets are the red parts in Figure 4.10. Red balls denote the resting points.

Note that an analogous version of the Theorem of Poincare–Bendixson can not hold in $\mathbb{R}^3$, as the example of the Lorenz model shows:

$$x' = \sigma(y - x),$$
$$y' = \varrho x - y - xz,$$
$$z' = -\beta z + xy,$$

where $\sigma = 10$, $\varrho = 28$ and $\beta = 8/3$. Then one can show that the solution curves remain bounded if the initial value is in a small neighbourhood of the origin. The orbit then will stay near a surface in $\mathbb{R}^3$ consisting of two sheets, but one can not predict when the orbit will be on one sheet, or the other sheet, because this depends in a very sensitive way on the initial values. Therefore we have: the

[15] Jules Henri Poincaré, 1854–1912, *The Last Universalist*
[16] Ivar Otto Bendixson, 1861–1935, swedish mathematician

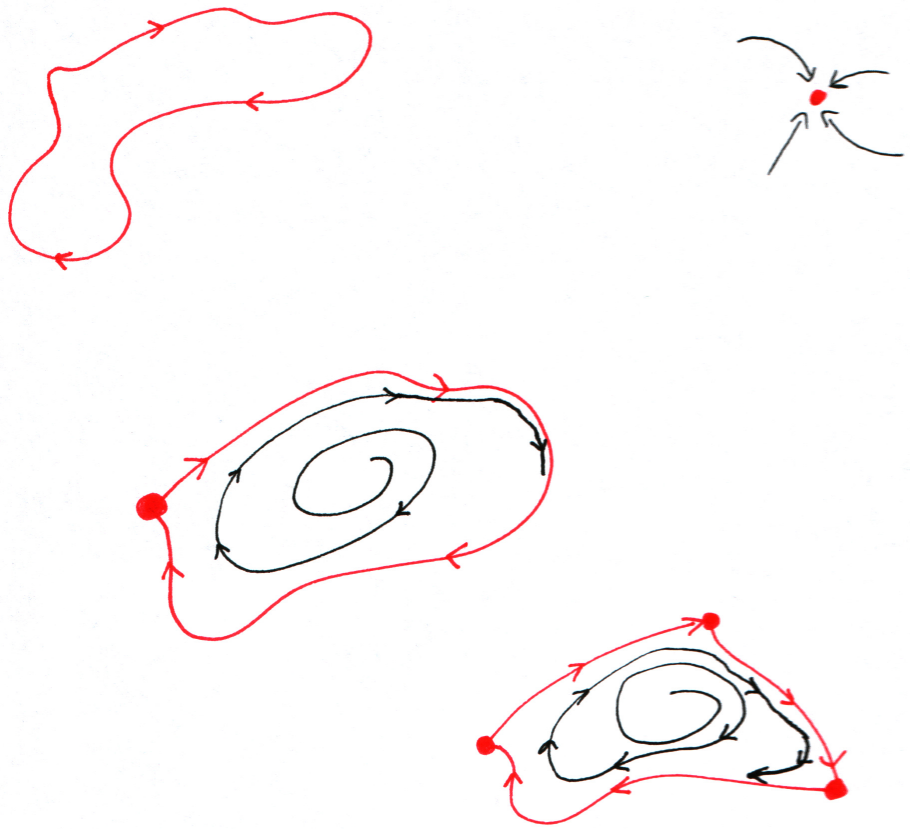


Figure 4.10:

forward orbit $\gamma_+(x_0)$ is bounded, but the ω limit set of x_0 is not a periodic orbit, and also the relation "$y_0 \in \omega(x_0) \Longrightarrow \omega(y_0)$ is a resting point" does not hold.

As an application, we consider the Brusselator, which was proposed by Prigogine and Lefever ([18]) 1968 as a simplified model of the Belouzov Zhabotinskii reaction:

$$x' = a - x - bx + x^2 y,$$
$$y' = bx - x^2 y,$$

with a and b as positive constants. The quantities x und y describe concentrations of chemical substances, and therefore they should never be negative.

To determine resting positions, we add up the equations $x' = 0$ and $y' = 0$, giving us $a - x = 0$, hence $x = a$. Together with $bx - x^2 y = 0$ we then find $y = \frac{b}{a}$. Linearisation then gives

$$f' = \begin{pmatrix} -1 - b + 2xy & x^2 \\ b - 2xy & -x^2 \end{pmatrix}, \qquad f'(a, b/a) = \begin{pmatrix} b - 1 & a^2 \\ -b & -a^2 \end{pmatrix}.$$

The trace is $b - a^2 - 1$, and the determinant is $\det f' = a^2 > 0$. If we suppose

$$b > a^2 + 1,$$

then the matrix f' possesses two eigenvalues in the right half-plane, making the resting point $(a, b/a)$ unstable.

Lemma 4.34. *If $b > a^2 + 1$, then this dynamical system has a periodic orbit in the first quadrant $\{(x, y) \in \mathbb{R}^2 : x > 0,\ y > 0\}$ of $\mathbb{R}^2$.*

Proof. We choose a domain $\Omega \subset \mathbb{R}^2$ by the following inequalities:

$$x > 0, \qquad y > 0, \qquad x + y < c_1, \qquad y - x < c_2.$$

If the parameters c_1 und c_2 are selected sufficiently large and positive, then one can show that, for each of the four boundary lines, the vector field f points *into* Ω. Therefore Ω is positive invariant.

Now choose an arbitrary point $p \in \Omega$. Then $\gamma_+(p)$ stays forever in Ω, hence $\gamma_+(p)$ is bounded. If $\omega(p)$ were not periodic, then, according to the Theorem of Poincare and Bendixson, $\omega(q)$ would consist solely of resting points of f, for each $q \in \omega(p)$. However, f has exactly one resting point, namely $(a, b/a)$, and this one is repulsive. Therefore, $\omega(p)$ must be periodic. □

To be specific, we explain how to choose the numbers c_1 and c_2. We start with the line $x + y = c_1$, which has the normal vector $(1, 1)^\top$, not necessarily normalized. The vector field f points inwards if and only if the scalar product $(x', y') \cdot (1, 1)$ is non-positive, which boils down to

$$0 \overset{!}{\geq} x' + y' = (a - x - bx + x^2 y) + (bx - x^2 y) = a - x,$$

which is true for $x \geq a$, whatever the value c_1 is.

Next we consider the line $y = x + c_2$ with normal vector $(-1, 1)^\top$. Here the condition on the scalar product becomes

$$\begin{aligned} 0 &\overset{!}{>} (-1, 1) \cdot (x', y') = (-a + x + bx - x^2 y) + (bx - x^2 y) = -a + (2b + 1)x - 2x^2 y \\ &= -a + (2b + 1)x - 2x^2 (x + c_2), \end{aligned}$$

and we wish this to be negative, where it is enough to consider x between 0 and a. Using the elementary inequality $|uv| \leq \frac{1}{2}(u^2 + v^2)$, we then find

$$\begin{aligned} -a + (2b + 1)x - 2x^2 (x + c_2) &= -a + \sqrt{a} \cdot \frac{(2b + 1)x}{\sqrt{a}} - 2x^2 (x + c_2) \\ &\leq -a + \frac{1}{2} \left(a + \frac{(2b + 1)^2 x^2}{a} \right) - 2x^2 (x + c_2) \\ &\leq -\frac{1}{2} a + x^2 \left(\frac{(2b + 1)^2}{2a} - 2c_2 \right), \end{aligned}$$

and this will be negative if we choose c_2 large enough. After that we select c_1 in such a way that the lines $y = x + c_2$ and $x + y = c_1$ intersect at a point which has x-coordinate larger than a.

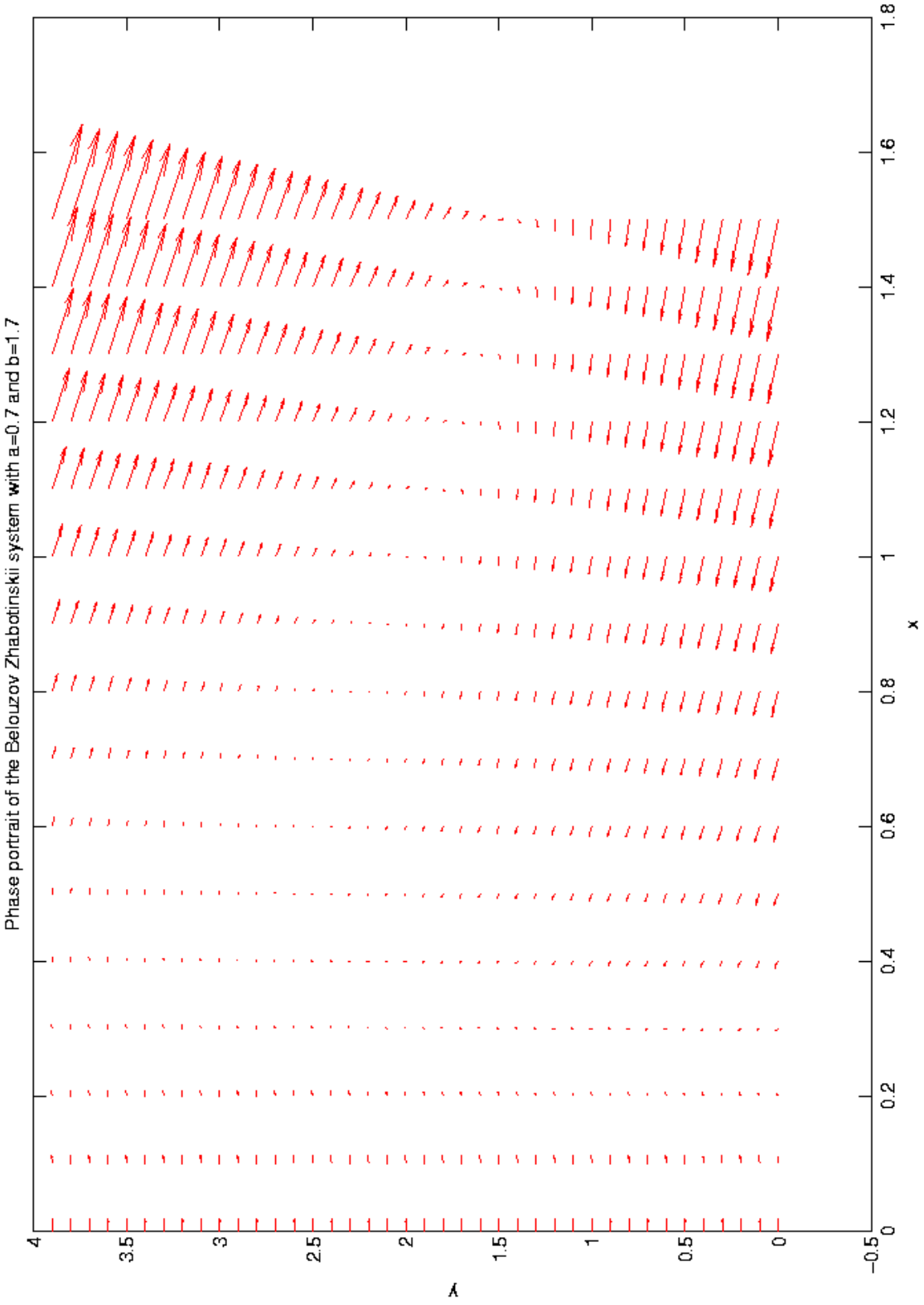
We conclude this chapter with a criterion which can exclude periodic orbits in the plane.

Proposition 4.35 (**Criterion of Bendixson**). *Let $\Omega \subset \mathbb{R}^2$ be a simply connected domain, and $f \in C^2(\Omega \to \mathbb{R}^2)$, with $\operatorname{div} f$ of constant sign and nowhere zero (except isolated points). Then the system $x'(t) = f(x)$ has no periodic orbit in Ω.*

Proof. Assume that Γ were a periodic orbit in Ω. Then the interior S of Γ is simply connected, and the Gauß integral theorem gives us

$$\oint_\Gamma f_1 \, \mathrm{d}x_2 - f_2 \, \mathrm{d}x_1 = \iint_S \operatorname{div} f \, \mathrm{d}x_1 \, \mathrm{d}x_2.$$

The left side is zero, because $(f_1, f_2)^\top$ is a tangential vector on Γ, but the right side is not zero because $\operatorname{div} f$ has constant sign. This is a contradiction. □


Phase portrait of the Belouzov Zhabotinskii system with a=0.7 and b=1.7
y
x
4
3.5
3
2.5
2
1.5
1
0.5
0
−0.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8

Chapter 5

Numerical Methods

This part follows [21].

We wish to solve a differential equation or a system

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0 \tag{5.1}$$

using numerical methods. The key idea is to define points in time

$$t_j = t_0 + jh, \qquad j = 1, 2, 3, \ldots, \qquad 0 < h \ll 1,$$

and to search for values η_j which approximate $y(t_j)$, with a good relation between error and effort. The number h is called *time step size*.

The exact solution will always be $y = y(t)$, and the approximate values are called η_j (but $\eta_j(h)$ would be more precise).

5.1 Explicit Methods

By Taylor expansion, we have

$$\begin{aligned} y(t_{j+1}) = y(t_j + h) &= y(t_j) + y'(t_j)h + \frac{1}{2}y''(t_j + \theta h)h^2 \qquad (0 < \theta < 1) \\ &= y(t_j) + f(t_j, y(t_j))h + \frac{1}{2}y''(t_j + \theta h)h^2, \end{aligned}$$

and neglecting the quadratic term gives us a first algorithm:

- set $\eta_0 := y_0$,
- for $j = 1, 2, \ldots$, set $\eta_{j+1} := \eta_j + f(t_j, \eta_j)h$.

This is known as the *Explicit Euler Method*. As a toy model, we solve $y' = y$ with initial value $y(0) = 1$ and ask for the numerical approximation at time $t = 1$:

h	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
error	0.12	0.013	0.0014	0.00014	$1.359 \cdot 10^{-5}$	$1.359 \cdot 10^{-6}$	$1.359 \cdot 10^{-7}$
runtime (seconds)	0.0005	0.001	0.005	0.037	0.27	2.76	27.1

As a side remark, the toy model $y' = -y$ has smaller errors. For instance, $h = 10^{-7}$ gives the error $1.84 \cdot 10^{-8}$ then.

It seems that we need better methods, and a more systematic treatment.

Definition 5.1. *A* one-step method[1] *to the initial value problem (5.1) has the form*

[1] Einschrittverfahren

- $\eta_0 := y_0$,
- *for* $j = 1, 2, \ldots$, *set* $\eta_{j+1} := \eta_j + \Phi(t_j, \eta_j, h)h$,

where Φ *is a certain function.*

The explicit Euler method is described by $\Phi(t, \eta, h) = f(t, \eta)$, and our goal shall be to no longer neglect the quadratic term $y''(t_j + \theta h)h^2$ as we did above.

Definition 5.2 (**local error, consistent method**). *For a time* t_* *and a point* y_*, *let* $y = y(t)$ *be the exact solution to* $y'(t) = f(t, y(t))$ *with initial condition* $y(t_*) = y_*$. *Put*

$$\Delta(t_*, y_*, h) := \begin{cases} \frac{1}{h}(y(t_* + h) - y(t_*)) & : h > 0, \\ f(t_*, y_*) & : h = 0. \end{cases}$$

Then $\tau(t_*, y_*, h) := \Delta(t_*, y_*, h) - \Phi(t_*, y_*, h)$ *is called the* local discretisation error of the method Φ.

The method Φ *is called* consistent *if* $\lim_{h\to 0} \tau(t_*, y_*, h) = 0$.

The method Φ *is* of order p *if* $\tau(t_*, h_*, y) = \mathfrak{O}(h^p)$ *for* $h \to 0$.

The term τ measures the deviation between the exact difference quotient and the approximate difference quotient when we go from j to $j + 1$.

Example 5.3. *Concerning the explicit Euler method, we note*

$$y(t_* + h) = y(t_*) + y'(t_*)h + \frac{1}{2}y''(t_* + \theta h)h^2 = y(t_*) + f(t_*, y_*)h + \frac{1}{2}y''(t_* + \theta h)h^2,$$

from which we obtain that (for $h > 0$*)*

$$\Delta(t_*, y_*, h) = f(t_*, y_*) + \frac{1}{2}y''(t_* + \theta h)h,$$

which brings us to

$$\tau(t_*, y_*, h) = \frac{h}{2}y''(t_* + \theta h) = \frac{h}{2}\frac{\mathrm{d}}{\mathrm{d}t}f(t, y(t))\Big|_{t=t_*+\theta h}$$

if we are willing to assume that f *is sufficiently smooth (in this case,* $f \in C^1$ *is enough). Therefore the Euler method has consistence order one.*

Another example is the HEUN[2] method with

$$\Phi(t_j, \eta_j, h) := \frac{1}{2}\Big(f(t_j, \eta_j) + f(t_j + h, \eta_j + f(t_j, \eta_j)h)\Big).$$

Writing f_1 and f_2 for the partial derivatives of f, we then get

$$\begin{aligned} \Phi(t_*, y_*, h) &= \frac{1}{2}\left(2f(t_*, y_*) + f_1(t_*, y_*) \cdot h + f_2(t_*, y_*) \cdot f(t_*, y_*)h + \mathfrak{O}(h^2)\right) \\ &= f(t_*, y_*) + \frac{f_1(t_*, y_*) + f_2(t_*, y_*)f(t_*, y_*)}{2}h + \mathfrak{O}(h^2). \end{aligned}$$

The local discretisation error then is

$$\begin{aligned} \tau(t_*, y_*, h) &= \frac{1}{h}(y(t_* + h) - y(t_*)) - \Phi(t_*, y_*, h) \\ &= y'(t_*) + \frac{h}{2}y''(t_*) + \frac{h^2}{3!}y'''(t_*) + \mathfrak{O}(h^3) \\ &\quad - \left(f(t_*, y_*) + \frac{f_1(t_*, y_*) + f_2(t_*, y_*)f(t_*, y_*)}{2}h + \mathfrak{O}(h^2)\right) \\ &= \frac{h}{2}\Big(y''(t_*) - f_1(t_*, y_*) - f_2(t_*, y_*)f(t_*, y_*)\Big) + \mathfrak{O}(h^2). \end{aligned}$$

h	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
error	0.0042	$4.5 \cdot 10^{-5}$	$4.53 \cdot 10^{-7}$	$4.53 \cdot 10^{-9}$	$4.53 \cdot 10^{-11}$	$4.08 \cdot 10^{-13}$	$5.9 \cdot 10^{-13}$
runtime	0.0085	0.0014	0.0084	0.058	0.53	5.47	53.04

Table 5.1: The Heun method applied to $y' = y$ with $y(0) = 1$ and $y(1) =?$

Now y is the exact solution, hence $y' = f(t, y(t))$, and then also $y''(t) = f_1(t, y(t)) + f_2(t, y(t))y'(t)$, which implies $\tau(t_*, y_*, h) = \mathfrak{O}(h^2)$. The consistency order of the Heun method is two.

Again, we play with our toy model $y' = y$, $y(0) = 1$, getting results as in Table 5.1. For a chosen step size h, the effort (and the runtime) has doubled in comparison to the explicit Euler method, but the errors are much smaller.

One more example is the classical method of Runge[3] and Kutta[4]

$$\begin{aligned}
\Phi(t_j, \eta_j, h) &= \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4), \\
K_1 &= f(t_j, \eta_j), \\
K_2 &= f\left(t_j + \frac{h}{2}, \eta_j + \frac{h}{2}K_1\right), \\
K_3 &= f\left(t_j + \frac{h}{2}, \eta_j + \frac{h}{2}K_2\right), \\
K_4 &= f\left(t_j + h, \eta_j + hK_3\right)
\end{aligned}$$

which has the consistency order 4, as can be shown by a lengthy calculation, see also the table for a numerical example.

h	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
error	$2.08 \cdot 10^{-6}$	$2.25 \cdot 10^{-10}$	$2.09 \cdot 10^{-14}$	$1.11 \cdot 10^{-14}$	$5.77 \cdot 10^{-15}$	$5.77 \cdot 10^{-14}$	$5.93 \cdot 10^{-13}$
runtime	0.003	0.0021	0.016	0.11	1.06	11.5	106

Table 5.2: The Runge Kutta method applied to $y' = y$ with $y(0) = 1$ and $y(1) =?$

The three methods presented so far fit into the framework of the *general Runge Kutta methods with s stages*, which are of the form

$$\begin{aligned}
\Phi(t_j, \eta_j, h) &= \sum_{m=1}^{s} b_m K_m, \\
K_m &= f\left(t_j + c_m h, \eta_j + h \sum_{l=1}^{m-1} a_{ml} K_l\right), \qquad m = 1, \ldots, s, \qquad \sum_{l=1}^{0} := 0.
\end{aligned}$$

It is custom to arrange the coefficients in a table:

The discretisation error measures how much we deviate from the exact solution after one time step. This is just a local description, and what is more interesting is a *global* error.

Definition 5.4 (global error, convergent method). *The* global discretisation error *is defined as*

$$e(t_j, h) := \eta_j(h) - y(t_j), \qquad t_0 \le t_j = t_0 + jh \le T.$$

For a fixed t and some $m \in \mathbb{N}$, we define a step size

$$h_m := \frac{t - t_0}{m}.$$

Then the method Φ is called convergent *if*

$$\lim_{m \to \infty} e(t, h_m) = 0$$

for all $t \in [t_0, T]$ and all $f \in C^1([t_0, T] \times \mathbb{R}^n)$ with bounded derivatives f_1, f_2.

[2] Karl Heun, 1859 – 1929
[3] Carl David Tolmé Runge, 1856 – 1927
[4] Martin Wilhelm Kutta, 1867 – 1944

$$\begin{array}{c|c} c_1 & \\ \hline & b_1 \end{array} \quad = \quad \begin{array}{c|c} 0 & \\ \hline & 1 \end{array} \qquad \text{(explicit Euler method)}$$

$$\begin{array}{c|cc} c_1 & & \\ c_2 & a_{21} & \\ \hline & b_1 & b_2 \end{array} \quad = \quad \begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \qquad \text{(Heun's method)}$$

$$\begin{array}{c|ccc} c_1 & & & \\ c_2 & a_{21} & & \\ c_3 & a_{31} & a_{32} & \\ c_4 & a_{41} & a_{42} & a_{43} \\ \hline & b_1 & b_2 & b_3 & b_4 \end{array} \quad = \quad \begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \qquad \text{(Runge Kutta method)}$$

The local and global errors are connected by the following result:

Proposition 5.5. *Let $y = y(t)$ be the exact solution to (5.1), and Φ be continuous on*

$$G := \{(t, z, h) \in [t_0, T] \times \mathbb{R}^n \times \mathbb{R} \colon |z - y(t)| \le \gamma, \quad 0 \le h \le h_0\}$$

for some positive numbers γ and h_0. Suppose that there are numbers M and N such that

$$\left\| \Phi(t, \eta^{(1)}, h) - \Phi(t, \eta^{(2)}, h) \right\| \le M \left\| \eta^{(1)} - \eta^{(2)} \right\|$$

for all $(t, \eta^{(1)}, h)$, $(t, \eta^{(2)}, h) \in G$, and

$$\|\tau(t, y(t), h)\| \le N h^p, \qquad \forall\, t \in [t_0, T], \quad \forall\, h \le h_0.$$

Then there is a number h^* with $0 < h^* \le h_0$ such that the global discretisation error is bounded like this:

$$\|e(t, h_m)\| \le h_m^p \frac{N}{M} \left(e^{M(t - t_0)} - 1 \right)$$

for all $t \in [t_0, T]$ and all $h_m = (t - t_0)/m$ with $h_m \le h^*$.

A proof can be found in [21].

In theory, this estimate of $e(t, h_m)$ could be used to choose that h_m which is optimal for the desired accuracy. In practice however, the constants M and N are almost never known.

Choosing a good step size is not easy. As the table for the Runge Kutta method shows, the error will **increase** if h becomes too small. This seems to contradict the above estimate of $e(t, h_m)$, but that estimate does not take into account the finite computing precision of the microprocessors: typically, they compute with about 16 decimal digits, and in our case the solution is between 10^0 and 10^1, which forecasts rounding errors of size 10^{-15} which will dominate the overall error for small step size h. Moreover: the number of time steps is inversely proportional to the time step size, and each time step introduces rounding errors into the computation. Clearly, these rounding errors could accumulate.

Therefore it is desirable to find a method of automatically selecting a good step size, without human intervention. And indeed, this can be done. A deeper look at the table for the Euler method and the Heun method suggests that the error is not just a random number, but has some structure in it. More precisely, we have:

Proposition 5.6 (**Asymptotic expansion of the global error**). *Suppose that the function $f = f(t, y)$ is $(N + 2)$ times continuously differentiable on $[t_0, T] \times \mathbb{R}^n$, and all the derivatives are bounded. Let $\eta = \eta_j(h)$ be the approximate solution to (5.1), constructed by a one-step method Φ of order p, with $p \le N$. Then the error $e(t, h)$ possesses an asymptotic expansion of the form*

$$e(t, h) = h^p e_p(t) + h^{p+1} e_{p+1}(t) + \cdots + h^N e_N(t) + h^{N+1} E_{N+1}(t, h),$$

with $e_p(t_0) = 0$, and for all $t \in [t_0, T]$, all $h = (t - t_0)/m$, $m \in \mathbb{N}$. The functions $e_k = e_k(t)$ are independent of h, and the remainder term E_{N+1} is bounded in h, for all t.

A very elegant proof is in [21].

Typically, the functions e_k are unknown, but this does not matter, as we see now: Suppose that $h^p e_p(t)$ is the biggest term in the asymptotic expansion, which is certainly true for small h. Perform the computations with step size h, and with step size $h/2$. Then we get known values $\eta(h)$, $\eta(h/2)$ both referring to the same time t, and $y(t)$ is unknown. But we have

$$\begin{aligned}\eta(h) &= y(t) + h^p e_p(t) + \mathfrak{O}(h^{p+1}),\\ \eta(h/2) &= y(t) + 2^{-p} h^p e_p(t) + \mathfrak{O}(h^{p+1}),\end{aligned}$$

which directly gives

$$2^{-p} h^p e_p(t) = \frac{\eta(h) - \eta(h/2)}{2^p - 1} + \mathfrak{O}(h^{p+1}).$$

Our assumption was that the remainder term $\mathfrak{O}(h^{p+1})$ is much smaller than the leading term $h^p e_p$ of the asymptotic expansion, and we find the approximation

$$\eta(h/2) - y(t) \approx \frac{\eta(h) - \eta(h/2)}{2^p - 1}$$

for the global discretisation error. This gives us a tool for the automatic selection of the optimal step size h, adjusted to the desired accuracy of the numerical solution.

Another method of step size control[5] is to couple a Runge Kutta method of order p with a Runge Kutta method of order $p+1$, and to use the difference of the approximations as a means for estimating the global error, and then to adjust the time step size for the next step. If both methods use the same parameters c_m and a_{ml}, but differ only in the parameters b_m, then it is only a negligible additional effort to use both methods in parallel, since the number of evaluations of f remains the same. This is of particular relevance if evaluating the function f requires high numerical cost. For instance, the `ode45` method of matlab couples a fourth order method and a fifth order method, exploiting the famous scheme of Dormand and Prince established in 1980.

Finally, a remark about where the expression *one-step method*[6] comes from. By definition of the method Φ, the new value η_{j+1} is computed (in a sometimes quite complicated way) from the old value η_j alone. However, it might be a good idea to use more information for the computation of η_{j+1}, for instance, the values of η_{j-1}, η_{j-2}, $\dots \eta_{j-d}$ for some fixed d. Such methods are called *multi-step methods*[7], which will not be discussed in this course. They can be quite difficult to handle: for instance, a consistent method need not be convergent, in contrast to Proposition 5.5. And also the step size control is obviously more complicated.

5.2 Implicit Methods

Consider the initial value problem

$$y'(t) = Ay, \qquad y(0) = \begin{pmatrix}1\\0\end{pmatrix}, \qquad A = \begin{pmatrix}998 & 1998\\ -999 & -1999\end{pmatrix}$$

as a toy model. The eigenvalues of A are -1 and -1000, and the exact solution is

$$y(t) = \begin{pmatrix}2e^{-t} - e^{-1000t}\\ -e^{-t} + e^{-1000t}\end{pmatrix}.$$

Now we try to solve this problem with the explicit Euler method of step size h. Then we have the iteration scheme $\eta_{j+1} = (I + hA)\eta_j$ for $j \geq 1$, which has the explicit solution

$$\eta_j = \begin{pmatrix}2(1-h)^j - (1-1000h)^j\\ -(1-h)^j + (1-1000h)^j\end{pmatrix}.$$

[5] Schrittweitensteuerung
[6] Einzelschrittverfahren
[7] Mehrschrittverfahren

The exact solution $y = y(t)$ decays if t approaches $+\infty$, but $\lim_{j\to\infty} \eta_j = 0$ holds only if $|1 - 1000h| < 1$, with the consequence $h < \frac{2}{1000}$. We make two observations: the term e^{-1000t} does not contribute anything to the solution $y(t)$ for $t \gtrsim 1$, but this term forces us to choose h very small, which makes the computational effort very large. All the explicit Runge Kutta methods have the same drawback.

Definition 5.7 (**Stiff differential equations**). *A linear system of differential equations $y' = Ay$ with constant coefficients is called* stiff[8] *if A has at least one eigenvalue with real part $\ll 1$, and if*

$$S(A) := \frac{\max_j |\Re\lambda_j(A)|}{\min_j |\Re\lambda_j(A)|}$$

is large (typical values are $S(A) \sim 10^3 \dots 10^6$).

A system $y'(t) = f(t, y)$ (not necessarily linear) is called stiff *if the value $S(A)$ with $A = f_y(t, y)$ as Jacobi matrix is large.*

Stiff systems always come up when the physical system under consideration has effects which live on very different time scales. Many problems from chemistry or biology are stiff.

We go back to the introductory example. The matrix A from there has eigenvalues λ_1, λ_2, both in the left half-plane. The exact solution is $y(t) = \exp(At)y_0$, and $\exp(At)$ has eigenvalues $\exp(\lambda_1 t)$, $\exp(\lambda_2 t)$, both in the interior of the unit ball of $\mathbb{C}$. Therefore, $y(t)$ decays for $t \to \infty$. On the other hand, the approximate solution has the representation $\eta_j = (I + hA)^j y_0$, and $I + hA$ has eigenvalues $1 + h\lambda_1$ and $1 + h\lambda_2$. These should also be in the interior of the unit ball, otherwise $\lim_{j\to\infty} \eta_j \neq 0$.

Replacing the explicit Euler method by another explicit Runge Kutta method would give us the recursion $\eta_{j+1} = g(hA)\eta_j$ for some function g, and then also $\eta_j = (g(hA))^j y_0$. For instance, the explicit Euler method has $g(z) = 1 + z$. The behaviour of η_j for large j will only be correct if the numbers $g(h\lambda_1)$ and $g(h\lambda_2)$ are in the unit ball. This function g is called *stability* function.

Definition 5.8 (**A–stable**). *A one-step method is called* A–stable *(or* absolutely stable*) if its stability function g satisfies*

$$\Re z < 0 \quad \Longrightarrow \quad |g(z)| < 1.$$

We directly see that the explicit Euler method is not A–stable, and with some effort one can show that explicit methods of Runge Kutta type are *never* A–stable.

This trouble can be resolved when we go to *implicit* methods. As an example, we construct the implicit Euler method:

$$\begin{aligned} y(t_j) = y(t_{j+1} - h) &= y(t_{j+1}) - y'(t_{j+1})h + \frac{1}{2}h^2 y''(t_{j+1} - \theta h) \qquad (0 < \theta < 1) \\ &= y(t_{j+1}) - f(t_{j+1}, y_{j+1})h + \frac{1}{2}h^2 y''(t_{j+1} - \theta h), \end{aligned}$$

and neglecting the remainder term gives us the recursion

$$\eta_{j+1} = \eta_j + hf(t_{j+1}, \eta_{j+1}).$$

The unknown term η_{j+1} appears on both sides, which explains the name of the method.

Going back to the introductory example once again, we have $f(y) = Ay$, hence $\eta_{j+1} = \eta_j + hA\eta_{j+1}$, which brings us

$$\eta_0 := y_0, \qquad \eta_{j+1} = (I - hA)^{-1}\eta_j,$$

and then also

$$\eta_j = (I - hA)^{-j} y_0.$$

[8] steif

Now the eigenvalues of $(1 - hA)$ are $1 - h\lambda_1$ and $1 - h\lambda_2$ which have real part > 1, because λ_1, λ_2 are in the left half-plane. Therefore the eigenvalues of $(I - hA)^{-1}$ are automatically in the unit ball, as desired.

The stability function of the implicit Euler method is $g(z) = 1/(1 - z)$.

The implicit Euler method fits into the framework of *general implicit Runge Kutta methods* which are of the form

$$\Phi(t_j, \eta_j, h) = \sum_{m=1}^{s} b_m K_m,$$
$$K_m = f\left(t_j + c_m h, \eta_j + h \sum_{l=1}^{s} a_{ml} K_l\right), \qquad m = 1, \ldots, s,$$

and they can be symbolised by a scheme like this:

c_1	a_{11}	a_{12}	$\ldots$	a_{1s}
c_2	a_{21}	a_{22}	$\ldots$	a_{2s}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
c_s	a_{s1}	a_{s2}	$\ldots$	a_{ss}
	b_1	b_2	$\ldots$	b_s

In each step $\eta_j \to \eta_{j+1}$, a nonlinear system has to be solved, which can be accomplished by a variant of Newton's method, for instance.

5.3 Symplectic Methods

[9]

We consider the harmonic oscillator equation $z''(t) + z = 0$ with initial values $z(0) = 1$ and $z'(0) = 0$. Setting $y = (y_1, y_2)^\top = (z, z')^\top$ and transferring to a first order system we get

$$y' = Ay, \qquad y(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{5.2}$$

Then $y(t) = \exp(At)y(0)$ as usual. Note that A is skew-adjoint (which means $A^\top = -A$), and we know already that then $\exp(A)$ is a unitary matrix, hence also $\exp(At)$. In the second semester we have learnt that multiplication with a unitary matrix does not change the length of a vector, and therefore

$$\|y(t)\|^2 = \|y(0)\|^2, \qquad \forall\, t \in \mathbb{R}.$$

This is no surprise because $\|y(t)\|^2 = |z(t)|^2 + |z'(t)|^2$ is just the mechanical energy of the system. On the other hand, $\Omega_0 \subset \mathbb{R}^2$ is mapped by the flow to $\Omega_t = \exp(At)\Omega_0$, which is simply a rotated copy of Ω_0. This corresponds to the Theorem of Liouville about the preservation of the phase space volume.

When we try to solve numerically the system (5.2) by the explicit and implicit Euler methods, we get approximate solutions as in the Figures 5.1 and 5.2, and the long time behaviour is completely wrong.

The deeper reason: the explicit Euler scheme is $\eta_{j+1} = (I + hA)\eta_j$, hence $\eta_j = (I + hA)^j \eta_0$, and the matrix $(I + hA)^j$ is not orthogonal, since $\det(I + hA) = 1 + h^2 > 1$. Therefore the volume of the phase space grows by the factor $(1 + h^2)$ at each iteration step. Conversely, the implicit Euler scheme is $\eta_{j+1} = (I - hA)^{-1}\eta_j$, hence $\eta_j = (I - hA)^{-j}\eta_0$, and now $\det(I - hA)^{-1} = (1 + h^2)^1 < 1$, which makes the phase space volume shrink by the factor $(1 + h^2)^{-1}$ at each step.

Choosing a smaller h would not help much, since then the number of time steps would grow. Considered at the end time T, the total factor then is

$$(1 + h^2)^{T/h} = \left(1 + \frac{hT}{T/h}\right)^{T/h} \approx e^{hT} \qquad \text{if } \frac{T}{h} \gg 1.$$

[9] See [20] and [11].

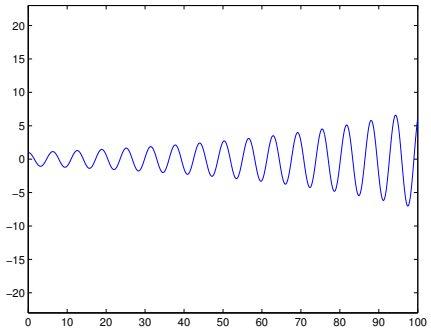


Figure 5.1: Solving numerically the harmonic oscillator with the explicit Euler method

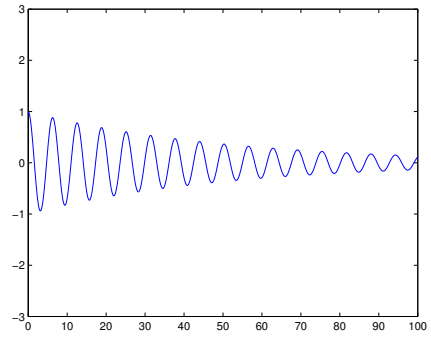


Figure 5.2: Solving numerically the harmonic oscillator with the implicit Euler method

Looking at these examples we get the impression that, for instance, simulating the motion of the planets of our solar system numerically over a period of millions of years would be infeasible using typical explicit or implicit Runge Kutta methods.

The main problem is that the system (5.2) possesses conservation properties which are no longer present in the above numerical schemes.

Definition 5.9. *A system of differential equations is called a* Hamiltonian[10] system *if it has the form*

$$q'(t) = \frac{\partial H(q,p)}{\partial p}, \qquad p'(t) = -\frac{\partial H(q,p)}{\partial q},$$

with $q\colon \mathbb{R}_t \to \mathbb{R}^n$, $p\colon \mathbb{R}_t \to \mathbb{R}^n$, $H\colon \mathbb{R}^{2n} \to \mathbb{R}$. *The function* H *is called* Hamiltonian.

In case of (5.2), we have $n = 1$ and $y_1 = q_1$, $y_1' = y_2 = p_1$, and $H = p_1^2 + q_1^2$.

We know already that, along a solution $(q,p) = (q,p)(t)$,

- the energy $H(q(t), p(t))$ is constant,
- the phase space volume $\int_{\Omega_t} \mathrm{d}q\,\mathrm{d}p$ is conserved (Theorem of Liouville).

We can improve the second • heavily:

Theorem 5.10. *The Hamiltonian flow preserves the symplectic*[11] *form* $\omega = \sum_{j=1}^{n} \mathrm{d}q_j \wedge \mathrm{d}p_j$.

[10] Sir William Rowan Hamilton, 1805–1865, irish physicist, astronomer, mathematician

[11] Where does the adjective come from ? All linear mappings in the $\mathbb{R}^{2n}$ that preserve a nondegenerate, skew-symmetric, bilinear form are elements of the *symplectic group*. And the name *symplectic group* goes back to Hermann Weyl (1885–1955), see [24] for the following quotation:

> The name "complex group" formerly advocated by me in allusion to line complexes, as these are defined by the

The symplectic form is a two-form, and we should explain how to work with such objects:

- a zero-form is a smooth scalar function $\mathbb{R}^N \to \mathbb{R}$,
- a one-form is written as $f_1(y)\,\mathrm{d}y_1 + f_2(y)\,\mathrm{d}y_2 + \cdots + f_N(y)\,\mathrm{d}y_N$, and it can be integrated along a curve in $\mathbb{R}^N$, giving us a curve integral of second kind,
- a two-form is written as

  $$\sum_{j<k} f_{jk}(y)\,\mathrm{d}y_j \wedge \mathrm{d}y_k,$$

  and it can be integrated over a two-dimensional surface patch $S \subset \mathbb{R}^N$, which is done as follows:

  Parametrise S with parameters $(u_1, u_2) \in \mathbb{R}^2$. Then $y_1 = y_1(u_1, u_2)$, ..., $y_N = y_N(u_1, u_2)$ for the points on S. Cut the parameter domain into many small rectangles using a grid whose axis are parallel to the u_1- and u_2- axis, and whose mesh widths are $\Delta u_1 = \mathrm{d}u_1$ and $\Delta u_2 = \mathrm{d}u_2$. This grid in the parameter domain corresponds to a grid on the surface patch S, which cuts S into many small pieces which look like slightly deformed parallelograms. Then the integral

  $$I_{jk} = \int_S f_{jk}(y)\,\mathrm{d}y_j \wedge \mathrm{d}y_k$$

  is evaluated via

  $$\begin{aligned}
  \mathrm{d}y_j &= \frac{\partial y_j}{\partial u_1}\,\mathrm{d}u_1 + \frac{\partial y_j}{\partial u_2}\,\mathrm{d}u_2, \\
  \mathrm{d}y_j \wedge \mathrm{d}y_k &= \left(\frac{\partial y_j}{\partial u_1}\,\mathrm{d}u_1 + \frac{\partial y_j}{\partial u_2}\,\mathrm{d}u_2\right) \wedge \left(\frac{\partial y_k}{\partial u_1}\,\mathrm{d}u_1 + \frac{\partial y_k}{\partial u_2}\,\mathrm{d}u_2\right) \\
  &= \frac{\partial y_j}{\partial u_1}\frac{\partial y_k}{\partial u_1}\,\mathrm{d}u_1 \wedge \mathrm{d}u_1 + \frac{\partial y_j}{\partial u_2}\frac{\partial y_k}{\partial u_2}\,\mathrm{d}u_2 \wedge \mathrm{d}u_2 \\
  &\quad + \frac{\partial y_j}{\partial u_1}\frac{\partial y_k}{\partial u_2}\,\mathrm{d}u_1 \wedge \mathrm{d}u_2 + \frac{\partial y_j}{\partial u_2}\frac{\partial y_k}{\partial u_1}\,\mathrm{d}u_2 \wedge \mathrm{d}u_1.
  \end{aligned}$$

  Now $\mathrm{d}u_l \wedge \mathrm{d}u_m = -\,\mathrm{d}u_m \wedge \mathrm{d}u_l$, in particular $\mathrm{d}u_l \wedge \mathrm{d}u_l = 0$, hence

  $$\mathrm{d}y_j \wedge \mathrm{d}y_k = \det \begin{pmatrix} \partial_1 y_j & \partial_2 y_j \\ \partial_1 y_k & \partial_2 y_k \end{pmatrix} \cdot \mathrm{d}u_1 \wedge \mathrm{d}u_2,$$

  and $\mathrm{d}u_1 \wedge \mathrm{d}u_2$ equals the area of the small grid rectangle in the parameter domain.
- in $\mathbb{R}^N$, an N–form is

  $$f(y)\,\mathrm{d}y_1 \wedge \mathrm{d}y_2 \wedge \ldots \wedge \mathrm{d}y_N = f(y)\,\mathrm{d}y_1 \ldots \mathrm{d}y_N,$$

  which can be integrated over domains of $\mathbb{R}^N$.

The wedge product has the following properties:

- it is linear in each factor,
- it is anti-commutative and associative,
- the wedge product with N factors from $\mathbb{R}^N$ behaves like a determinant function:

  $$v_1 \wedge v_2 \wedge \ldots \wedge v_N = \Delta_N(v_1, \ldots, v_N).$$

vanishing of antisymmetric bilinear forms, has become more and more embarrassing through collision with the word "complex" in the connotation of complex number. I therefore propose to replace it by the corresponding Greek adjective "symplectic". Dickson calls the group the "Abelian linear group" in homage to Abel who first studied it.

The exterior derivative d behaves as follows:

it turns K–forms into $(K+1)$–forms, in a linear manner,

$$\text{if } f = f(y) \text{ is a zero-form (hence a function), then } \mathrm{d}f = \sum_{j=1}^{N} \frac{\partial f}{\partial y_j}\,\mathrm{d}y_j, \tag{5.3}$$

$$\text{if } \varrho = f(y)\,\mathrm{d}y_{i_1} \wedge \mathrm{d}y_{i_2} \wedge \ldots \wedge \mathrm{d}y_{i_K} \text{ is a } K\text{–form, then } \mathrm{d}\varrho = (\mathrm{d}f) \wedge \mathrm{d}y_{i_1} \wedge \mathrm{d}y_{i_2} \wedge \ldots \wedge \mathrm{d}y_{i_K}, \tag{5.4}$$

with $\mathrm{d}f$ evaluated by (5.3),

$$\mathrm{d}\,\mathrm{d} = 0. \tag{5.5}$$

As an example, take $N = 2$ and f as a function. Then

$$\begin{aligned}
\mathrm{d}f &= (\partial_1 f)\,\mathrm{d}y_1 + (\partial_2 f)\,\mathrm{d}y_2,\\
\mathrm{d}\,\mathrm{d}f &= (\,\mathrm{d}\partial_1 f) \wedge \mathrm{d}y_1 + (\,\mathrm{d}\partial_2 f) \wedge \mathrm{d}y_2\\
&= (\partial_1^2 f\,\mathrm{d}y_1 + \partial_1\partial_2 f\,\mathrm{d}y_2) \wedge \mathrm{d}y_1 + (\partial_1\partial_2 f\,\mathrm{d}y_1 + \partial_2^2 f\,\mathrm{d}y_2) \wedge \mathrm{d}y_2\\
&= \partial_1^2 f\,\mathrm{d}y_1 \wedge \mathrm{d}y_1 + \partial_2^2 f\,\mathrm{d}y_2 \wedge \mathrm{d}y_2 + (\partial_1\partial_2 f)(\,\mathrm{d}y_2 \wedge \mathrm{d}y_1 + \mathrm{d}y_1 \wedge \mathrm{d}y_2)\\
&= 0.
\end{aligned}$$

Note that the classical calculus rules $\operatorname{rot}\operatorname{grad} = 0$ and $\operatorname{div}\operatorname{rot} = 0$ have their deeper origin in $\mathrm{d}\,\mathrm{d} = 0$.

Now we come back to the Hamiltonian system with n degrees of freedom. Put $y = (y_1, \ldots, y_{2n})^\top = (q_1, \ldots, q_n, p_1, \ldots, p_n)^\top$ and write the system as

$$y'(t) = J\nabla_y H(y), \qquad J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}.$$

The volume form in the phase space $\mathbb{R}^n_q \times \mathbb{R}^n_p$ is $\mathrm{d}q\,\mathrm{d}p = \mathrm{d}q_1 \ldots \mathrm{d}q_n\,\mathrm{d}p_1 \ldots \mathrm{d}p_n = \mathrm{d}q_1 \wedge \ldots \wedge \mathrm{d}p_n$.

Lemma 5.11. *The volume form is a multiple of ω^n:*

$$\mathrm{d}q\,\mathrm{d}p = \pm\frac{1}{n!}\omega^n, \qquad \omega = \sum_{j=1}^{n} \mathrm{d}q_j \wedge \mathrm{d}p_j.$$

Proof. We consider $n = 2$ only and leave the other n to the reader:

$$\begin{aligned}
\omega^2 &= (\,\mathrm{d}q_1 \wedge \mathrm{d}p_1 + \mathrm{d}q_2 \wedge \mathrm{d}p_2)^2\\
&= \Big(\mathrm{d}q_1 \wedge \mathrm{d}p_1 \wedge \mathrm{d}q_1 \wedge \mathrm{d}p_1\Big) + \Big(\mathrm{d}q_1 \wedge \mathrm{d}p_1 \wedge \mathrm{d}q_2 \wedge \mathrm{d}p_2\Big) + \Big(\mathrm{d}q_2 \wedge \mathrm{d}p_2 \wedge \mathrm{d}q_1 \wedge \mathrm{d}p_1\Big)\\
&\quad + \Big(\mathrm{d}q_2 \wedge \mathrm{d}p_2 \wedge \mathrm{d}q_2 \wedge \mathrm{d}p_2\Big)\\
&= 0 - \Big(\mathrm{d}q_1 \wedge \mathrm{d}q_2 \wedge \mathrm{d}p_1 \wedge \mathrm{d}p_2\Big) + \Big(\mathrm{d}q_1 \wedge \mathrm{d}q_2 \wedge \mathrm{d}p_2 \wedge \mathrm{d}p_1\Big) + 0 = -2\,\mathrm{d}q\,\mathrm{d}p.
\end{aligned}$$

□

The Liouville Theorem as proved in Example 4.9 corresponds to the conservation of the volume form, hence the conservation of ω^n.

But Theorem 5.10, which we prove right now, is much stronger than the Liouville theorem, because the conservation of ω implies the conservation of ω^2, ω^3, …, ω^n.

Proof of Theorem 5.10. Let $y = y(t) = (q(t), p(t))$ be the solution to $y'(t) = J\nabla H(y)$ with the initial values

$$y^0 = (y_1^0, \ldots, y_{2n}^0)^\top = (q_1^0, \ldots, p_n^0)^\top.$$

Then $q(t)$ and $p(t)$ depend on the initial values, and in the sense of a total differential we have

$$\mathrm{d}y_l = \sum_{k=1}^{2n} \frac{\partial y_l}{\partial y_k^0}\,\mathrm{d}y_k^0, \qquad 1 \le l \le 2n. \tag{5.6}$$

We wish to show that

$$0 \overset{!}{=} \frac{\mathrm{d}}{\mathrm{d}t}\omega = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{j=1}^{n} \mathrm{d}q_j \wedge \mathrm{d}p_j = \sum_{j=1}^{n}\left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}q_j\right) \wedge \mathrm{d}p_j + \mathrm{d}q_j \wedge \left(\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}p_j\right)\right).$$

Now we have (using (5.6) backwards in the last step)

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}q_j &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_{k=1}^{2n}\frac{\partial q_j}{\partial y_k^0}\,\mathrm{d}y_k^0 = \sum_{k=1}^{2n}\left(\frac{\partial}{\partial y_k^0}\frac{\partial q_j}{\partial t}\right)\mathrm{d}y_k^0 = \sum_{k=1}^{2n}\left(\frac{\partial}{\partial y_k^0}\frac{\partial H(q,p)}{\partial p_j}\right)\mathrm{d}y_k^0 \\
&= \sum_{k=1}^{2n}\sum_{l=1}^{2n}\frac{\partial^2 H(q,p)}{\partial y_l \partial p_j}\frac{\partial y_l}{\partial y_k^0}\,\mathrm{d}y_k^0 = \sum_{l=1}^{2n}\frac{\partial^2 H(q,p)}{\partial y_l \partial p_j}\,\mathrm{d}y_l, \\
\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}p_j &= -\sum_{l=1}^{2n}\frac{\partial^2 H(q,p)}{\partial y_l \partial q_j}\,\mathrm{d}y_l,
\end{aligned}$$

and collecting the pieces we then find

$$\begin{aligned}
&\sum_{j=1}^{n}\left(\left(\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}q_j\right) \wedge \mathrm{d}p_j + \mathrm{d}q_j \wedge \left(\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{d}p_j\right)\right) \\
&\quad= \sum_{j=1}^{n}\sum_{l=1}^{2n}\left(\frac{\partial^2 H(q,p)}{\partial y_l \partial p_j}\,\mathrm{d}y_l \wedge \mathrm{d}p_j - \frac{\partial^2 H(q,p)}{\partial y_l \partial q_j}\,\mathrm{d}q_j \wedge \mathrm{d}y_l\right) && \big| \ \wedge \text{ anti-commutates} \\
&\quad= \sum_{j=1}^{n}\sum_{l=1}^{2n}\left(\frac{\partial^2 H(q,p)}{\partial p_j \partial y_l}\,\mathrm{d}y_l \wedge \mathrm{d}p_j + \frac{\partial^2 H(q,p)}{\partial q_j \partial y_l}\,\mathrm{d}y_l \wedge \mathrm{d}q_j\right) \\
&\quad= \sum_{k=1}^{2n}\sum_{l=1}^{2n}\frac{\partial^2 H(q,p)}{\partial y_k \partial y_l}\,\mathrm{d}y_l \wedge \mathrm{d}y_k = \sum_{k=1}^{2n}\left(\sum_{l=1}^{2n}\frac{\partial^2 H(q,p)}{\partial y_k \partial y_l}\,\mathrm{d}y_l\right) \wedge \mathrm{d}y_k && \big| \ (5.3) \text{ backwards} \\
&\quad= \sum_{k=1}^{2n}\left(\mathrm{d}\frac{\partial H}{\partial y_k}\right) \wedge \mathrm{d}y_k && \big| \ (5.4) \text{ backwards, with } f = \tfrac{\partial H}{\partial y_k} \text{ and } K = 1, \\
&\quad= \sum_{k=1}^{2n}\mathrm{d}\left(\frac{\partial H}{\partial y_k}\,\mathrm{d}y_k\right) && \big| \ \mathrm{d} \text{ is linear} \\
&\quad= \mathrm{d}\left(\sum_{k=1}^{2n}\frac{\partial H}{\partial y_k}\,\mathrm{d}y_k\right) && \big| \ (5.3) \text{ backwards} \\
&\quad= \mathrm{d}\,\mathrm{d}H && \big| \ (5.5) \\
&\quad= 0,
\end{aligned}$$

which was our goal. □

Definition 5.12. *A numerical scheme $\eta_j \mapsto \eta_{j+1} = \eta_j + h\Phi(t_j, \eta_j, h)$ is called* symplectic *if it preserves the symplectic form, i.e.,*

$$\sum_{k=1}^{n}\mathrm{d}\eta_{j+1,k} \wedge \mathrm{d}\eta_{j+1,k+n} = \sum_{k=1}^{n}\mathrm{d}\eta_{j,k} \wedge \mathrm{d}\eta_{j,k+n}.$$

To construct a symplectic scheme for $y'(t) = J\nabla H(y)$, we make the ansatz of a Runge Kutta scheme of one stage (see Table 5.3), which is a compact form of writing

$$\Phi(t_j, \eta_j, h) = b_1 K_1, \qquad K_1 = J\nabla H(\eta_j + h a_{11} K_1). \tag{5.7}$$

For simplicity, take $n = 1$, and write

$$\eta_j = (\varrho_j, \pi_j)^\top, \qquad K_1 = (K_{1,1}, K_{1,2})^\top,$$

	a_{11}
	b_1

Table 5.3: The easiest symplectic scheme

with ϱ_j approximating $q(t_j)$, and π_j approximating $p(t_j)$. Then the scheme can be written as

$$\varrho_{j+1} = \varrho_j + hb_1K_{1,1}, \qquad \pi_{j+1} = \pi_j + hb_1K_{1,2},$$
$$K_{1,1} = \frac{\partial H}{\partial p}(\varrho_j + ha_{11}K_{1,1}, \pi_j + ha_{11}K_{1,2}), \tag{5.8}$$
$$K_{1,2} = -\frac{\partial H}{\partial q}(\varrho_j + ha_{11}K_{1,1}, \pi_j + ha_{11}K_{1,2}). \tag{5.9}$$

Our goal is $\mathrm{d}\varrho_{j+1} \wedge \mathrm{d}\pi_{j+1} = \mathrm{d}\varrho_j \wedge \mathrm{d}\pi_j$. Now

$$\mathrm{d}\varrho_{j+1} = \mathrm{d}\varrho_j + hb_1\,\mathrm{d}K_{1,1}, \qquad \mathrm{d}\pi_{j+1} = \mathrm{d}\pi_j + hb_1\,\mathrm{d}K_{1,2},$$

and therefore

$$\begin{aligned}
\mathrm{d}\varrho_{j+1} \wedge \mathrm{d}\pi_{j+1} &= \mathrm{d}\varrho_j \wedge \mathrm{d}\pi_j \\
&\quad + hb_1\Big(\mathrm{d}\varrho_j \wedge \mathrm{d}K_{1,2} + \mathrm{d}K_{1,1} \wedge \mathrm{d}\pi_j\Big) \\
&\quad + h^2b_1^2\,\mathrm{d}K_{1,1} \wedge \mathrm{d}K_{1,2} \\
&= \mathrm{d}\varrho_j \wedge \mathrm{d}\pi_j \\
&\quad + hb_1\Big(\mathrm{d}(\varrho_j + ha_{11}K_{1,1}) \wedge \mathrm{d}K_{1,2} + \mathrm{d}K_{1,1} \wedge \mathrm{d}(\pi_j + ha_{11}K_{1,2})\Big) \qquad (5.10) \\
&\quad + h^2(b_1^2 - 2b_1a_{11})\,\mathrm{d}K_{1,1} \wedge \mathrm{d}K_{1,2}. \qquad (5.11)
\end{aligned}$$

Choosing $b_1 = 2a_{11}$ eliminates (5.11). To kill (5.10), we observe that (5.8) and (5.9) give

$$\mathrm{d}K_{1,1} = \frac{\partial^2 H}{\partial q \partial p}\,\mathrm{d}(\varrho_j + ha_{11}K_{1,1}) + \frac{\partial^2 H}{\partial p^2}\,\mathrm{d}(\pi_j + ha_{11}K_{1,2}),$$
$$\mathrm{d}K_{1,2} = -\frac{\partial^2 H}{\partial q^2}\,\mathrm{d}(\varrho_j + ha_{11}K_{1,1}) - \frac{\partial^2 H}{\partial q \partial p}\,\mathrm{d}(\pi_j + ha_{11}K_{1,2}),$$

and plugging this into (5.10) gives the desired cancellation, because of $\mathrm{d}(\varrho_j + ha_{11}K_{1,1}) \wedge \mathrm{d}(\varrho_j + ha_{11}K_{1,1}) = 0$, and $\mathrm{d}(\pi_j + ha_{11}K_{1,2}) \wedge \mathrm{d}(\pi_j + ha_{11}K_{1,2}) = 0$, and also $\mathrm{d}u \wedge \mathrm{d}v + \mathrm{d}v \wedge \mathrm{d}u = 0$, for arbitrary terms u and v.

Lemma 5.13. *The implicit Runge Kutta method (5.7) with $a_{11} = 1/2$ and $b_1 = 1$ (commonly called* Implicit Midpoint method*) is symplectic and consistent of order two.*

Proof. Only the consistency is not yet proved, and this is a wonderful exercise in Taylor expansions. □

Remark 5.14. *Another symplectic scheme is the* Symplectic Euler scheme*:*

$$\varrho_{j+1} = \varrho_j + h\frac{\partial H}{\partial p}(\varrho_j, \pi_{j+1}), \qquad \pi_{j+1} = \pi_j - h\frac{\partial H}{\partial q}(\varrho_j, \pi_j)$$

valid if the Hamiltonian splits into kinetic and potential energy in the sense of $H(q,p) = T(p) + V(q)$.

Concerning the energy conversation, Ge and Marsden have shown (1988) that, in the *general* case, a symplectic method can not conserve both the symplectic form and the Hamiltonian. However, typically one can find another Hamiltonian H_h, with $H - H_h = \mathfrak{O}(h)$, such that the symplectic scheme preserves H_h (except for exponentially small errors), over periods of length $\mathfrak{O}(h^{-1})$. This is much better than the standard Runge Kutta methods.

Chapter 6

Boundary Value Problems and Eigenvalues

> Young Irving Joshua Bush, who later took the name of Matrix... grew up a devout believer in the biblical prophecies of his parents' faith, and owing to a natural bent in mathematics, was particularly intrigued by the numerical aspects of those prophecies. At the age of seven he surprised his father by pointing out that there was 1 God, 2 testaments, 3 persons in the Trinity, 4 Gospels, 5 books of Moses, 6 days of creation, and 7 gifts of the Holy Spirit.
> "What about 8?" his father had asked.
> "It is the holiest number of all," the boy replied, "The other numbers with holes are 0, 6, and 9, and sometimes 4, but 8 has two holes, therefore it is the holiest."
>
> Martin Gardner [1]

6.1 Introduction

Consider a vibrating string[2] of length L. The elongation at position x shall be called $u(t,x)$, and of course the string is fixed at the end points. Then the differential equation is

$$\begin{cases} u_{tt}(t,x) - c^2 u_{xx}(t,x) = 0, & 0 < x < L, \quad 0 \le t < \infty, \\ u(t,0) = u(t,L) = 0, & 0 \le t < \infty, \\ u(0,x) = u_0(x), \qquad u_t(0,x) = u_1(x), & x \in (0,L), \end{cases} \tag{6.1}$$

with c as the sound speed on the string, u_0 as the initial elongation, and u_1 as the initial velocity.

This is a hard problem, but we wish to find at least some solutions, and make the ansatz $u(t,x) = a(t)v(x)$, which brings us to

$$\frac{v''(x)}{v(x)} = \frac{a''(t)}{c^2 a(t)} = \text{const.} = -\lambda,$$

because the left side does not depend on t, and therefore the right side *can not* depend on t.

Then we obtain an ODE for v:

$$v''(x) + \lambda v(x) = 0, \quad x \in (0,L), \qquad v(0) = v(L) = 0. \tag{6.2}$$

This is a second order differential equation together with an additional condition, with the following differences to what we have studied so far:

[1] *The Magic Numbers of Dr. Matrix*, New York: Prometheus, 1985, p.4.
[2] Saite

- we do not have prescribed initial values of $v(0)$ and $v'(0)$, but prescribed boundary values $v(0)$ and $v(L)$. Therefore this problem is called a *boundary value problem*[3] in contrast to *initial value problems*[4].
- there is an unknown parameter λ, which has not been determined yet. We have always the zero function $v \equiv 0$ as a solution to (6.2), but this function is boring. We will learn that interesting solutions v (which are defined to be not the zero function) exist only for special values of λ.

This theory which we will develop now has many similarities to the theory of eigenvalues and eigenvectors of a matrix $A \in \mathbb{C}^{n\times n}$ known from the second semester.

Now we discuss (6.2) in a more general setting: the function v and the parameter λ may take complex values. Our goal is to find all non-trivial solutions v. This means $v(x) \neq 0$ at least for some x, which we write as $v \not\equiv 0$.

Step 1: λ must be real and non-negative: to show this, we introduce the scalar product

$$\langle v, w\rangle_{L^2(0,L)} := \int_{x=0}^{L} v(x)\overline{w(x)}\,\mathrm{d}x.$$

Then we have

$$\begin{aligned}
\lambda \langle v, v\rangle_{L^2(0,L)} &= \lambda \int_{x=0}^{L} v(x)\overline{v(x)}\,\mathrm{d}x \\
&= -\int_{x=0}^{L} v''(x)\overline{v(x)}\,\mathrm{d}x \\
&= -v'(x)\overline{v(x)}\Big|_{x=0}^{x=L} + \int_{x=0}^{L} v'(x)\overline{v'(x)}\,\mathrm{d}x \\
&= 0 + \int_{x=0}^{L} v'(x)\overline{v'(x)}\,\mathrm{d}x \\
&= \langle v', v'\rangle_{L^2(0,L)},
\end{aligned}$$

which implies $\lambda = \frac{\langle v', v'\rangle}{\langle v, v\rangle}$. Note that the division is possible because of $v \not\equiv 0$. Now $\langle v', v'\rangle \in \mathbb{R}_{\geq 0}$ and $\langle v, v\rangle \in \mathbb{R}_{\geq 0}$ by definition of the scalar product, hence $\lambda \in \mathbb{R}_{\geq 0}$.

Step 2: λ can not be zero: assume λ were zero, then $\langle v', v'\rangle = 0$, hence $v' \equiv 0$, implying $v \equiv$ const., bringing us to $v \equiv 0$ due to the boundary condition. Contradiction.

Step 3: λ is positive: this can happen sometimes, and we discuss an explicit construction of the solution. All solutions to $v''(x) + \lambda v(x) = 0$ with positive λ are given as

$$v(x) = c_1 \cos\left(\sqrt{\lambda}x\right) + c_2 \sin\left(\sqrt{\lambda}x\right), \qquad c_1, c_2 \in \mathbb{C},$$

and the boundary condition $v(0) = 0$, $v(L) = 0$ give first $c_1 = 0$, and then $c_2 \sin(\sqrt{\lambda}L) = 0$, which has a non-trivial solution $c_2 \neq 0$ if and only if $\sqrt{\lambda}L \in \pi\mathbb{N}_+$, or

$$\lambda = \frac{\pi^2 k^2}{L^2}, \qquad k = 1, 2, 3, \ldots.$$

From this discussion we learn that introducing function vector spaces of L^2 type with appropriate choice of scalar product, together with partial integration, can give us some insights with little effort.

To get an overview, we compare linear initial value problems [5] and linear boundary value problems:

[3] Randwertproblem
[4] Anfangswertprobleme
[5] lineare Anfangswertprobleme

linear IVP	linear BVP
We look for functions $y = y(t)$ (scalar or vectorial) with t often being the time variable.	We look for functions $y = y(t)$ (scalar or vectorial) with t often being the space variable.
Scalar IVPs for a higher order equation can be transformed into a first order system.	Scalar BVPs for a higher order equation can be transformed into a first order system.
IVPs to first order systems are always uniquely solvable (Theorem 1.8).	BVPs to first order systems are often uniquely solvable, but they can also be unsolvable (Example 6.2), or there can be more than one solution (the string example from above).
You have an almost complete understanding of the behaviour of solutions if you have found the Fundamental Solution X.	You have an almost complete understanding of the behaviour of solutions (in the case of unique solvability) if you have found the Green's Function G (Proposition 6.3).
	The function vector space $L^2([a,b] \to \mathbb{C}^n)$ with its scalar product $\langle f, g\rangle_{L^2([a,b])} := \sum_{j=1}^{n} \int_a^b f(t)\overline{g(t)}\,\mathrm{d}t$ gives a deeper understanding of the BVP, even more so when you play with partial integration.

6.2 Solutions to First Order BVPs

We consider the boundary value problem to a vector-valued function y,

$$\begin{cases} y'(t) = F(t)y(t) + g(t), & a \le t \le b, \\ Ay(a) + By(b) = c, \end{cases} \tag{6.3}$$

where $F \in C([a,b] \to \mathbb{C}^{n\times n})$, $g \in C([a,b] \to \mathbb{C}^n)$, and $A, B \in \mathbb{C}^{n\times n}$, $c \in \mathbb{C}^n$-
From (3.10) we know that a solution $y = y(t)$ to $y'(t) = F(t)y(t) + g(t)$ must satisfy

$$y(t) = X(t,a)y(a) + \int_{s=a}^{t} X(t,s)g(s)\,\mathrm{d}s$$

with X as fundamental solution

$$\partial_t X(t,t_0) = F(t)X(t,t_0), \qquad X(t_0,t_0) = I, \qquad t, t_0 \in \mathbb{R}.$$

Proposition 6.1. *The following statements are equivalent:*

1. *the BVP (6.3) is uniquely solvable for any $g \in C([a,b] \to \mathbb{C}^n)$ and any $c \in \mathbb{C}^n$,*
2. *the characteristic matrix*

   $$C_X := A + BX(b,a)$$

   is invertible,
3. *the homogeneous BVP*

   $$y'(t) = F(t)y(t), \qquad Ay(a) + By(b) = 0$$

   possesses only the trivial solution.

Proof. The boundary condition $Ay(a) + By(b) = c$ is equivalent to

$$\begin{aligned} c &= Ay(a) + B\left(X(b,a)y(a) + \int_{s=a}^{b} X(b,s)g(s)\,\mathrm{d}s\right) \\ &= C_X y(a) + B\int_{s=a}^{b} X(b,s)g(s)\,\mathrm{d}s, \end{aligned}$$

or

$$C_X y(a) = c - B\int_{s=a}^{b} X(b,s)g(s)\,\mathrm{d}s. \tag{6.4}$$

(1) $\Longrightarrow$ (3): This is obvious since (3) is a special case of (1).

(3) $\Longrightarrow$ (2): We wish to show that $\ker C_X = \{0\}$. Let y_a be an element of $\ker C_X$. Define a function $y = y(t)$ by $y(t) = X(t,a)y_a$. Then y solves $y'(t) = F(t)y(t)$ and $Ay(a) + By(b) = 0$, by (6.4). However, by the assumption of (3), y must be the zero function. In particular, y must take the value zero at $t = a$. Hence $0 = y(a) = y_a$. Therefore $\ker C_X = \{0\}$, as desired.

(2) $\Longrightarrow$ (1): The function g and the vector c are given, then there is exactly one vector $y(a) \in \mathbb{C}^n$ that satisfies (6.4). Then $y(t) = X(t,a)y(a) + \int_a^t X(t,s)g(s)\,\mathrm{d}s$ solves (6.3). And (6.3) can not have another solution $z = z(t)$, because then also $z(a)$ solves (6.4), hence $C_X(y(a) - z(a)) = 0$, which gives us $y(a) - z(a) = 0$ by invertibility of C_X. But then $z(t) = y(t)$ for all t.

□

Example 6.2. *Show that the BVP*

$$\begin{gathered} y'(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} y(t), \qquad 0 \le t \le \pi, \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} y(0) + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} y(\pi) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{gathered}$$

is unsolvable.

Next we discuss a BVP with homogeneous boundary condition:

$$\begin{cases} y'(t) = F(t)y(t) + g(t), & a \le t \le b, \\ Ay(a) + By(b) = 0. \end{cases} \tag{6.5}$$

Proposition 6.3. *Suppose that C_X is invertible. Then there is at least one matrix-valued function*

$$G = G(t,s)\colon [a,b] \times [a,b] \to \mathbb{C}^{n\times n},$$

called Green's function[6] *with the following properties:*

1. *G is continuous on the triangle $\{(t,s)\colon a \le t < s \le b\}$, and G is continuous on the triangle $\{(t,s)\colon a \le s \le t \le b\}$,*
2. *on the diagonal of $(a,b) \times (a,b)$, G jumps:*

$$G(t+0,t) - G(t-0,t) = I_n, \qquad a < t < b,$$

3. *for each $g \in C([a,b] \to \mathbb{C}^n)$, the function $y = y(t)$ given by*

$$y(t) = \int_{s=a}^{b} G(t,s)g(s)\,\mathrm{d}s$$

is the unique solution to (6.5).

[6] George Green, 1793–1841, attended school for one year, was the first to present a mathematical theory of electricity and magnetism

Definition 6.4 (**Green's Function**). *Every function G with these three properties is called* Green's function *or* Green's matrix.

Proof. By Proposition 6.1, the unique solution y exists and is given by

$$\begin{aligned} y(t) &= \int_{s=a}^{t} X(t,s)g(s)\,\mathrm{d}s + X(t,a)y(a), \\ C_X y(a) &= -B\int_{s=a}^{b} X(b,s)g(s)\,\mathrm{d}s, \end{aligned}$$

which brings us to

$$\begin{aligned} y(t) &= \int_{s=a}^{t} X(t,s)g(s)\,\mathrm{d}s + X(t,a)C_X^{-1}\left(-B\int_{s=a}^{b} X(b,s)g(s)\,\mathrm{d}s\right) \\ &= \int_{s=a}^{t} X(t,s)g(s)\,\mathrm{d}s + \int_{s=a}^{b}\left(-X(t,a)C_X^{-1}BX(b,s)\right)g(s)\,\mathrm{d}s, \end{aligned}$$

and now it remains to choose

$$G(t,s) = \begin{cases} X(t,s) - X(t,a)C_X^{-1}BX(b,s) & : a \le s \le t \le b, \\ 0 - X(t,a)C_X^{-1}BX(b,s) & : a \le t < s \le b. \end{cases}$$

The jump relation (2) is now easily seen from $X(t,t) = I_n$. □

The formulae for G in both triangles look quite different, which is unaesthetic, because the times a and b should have equal rights. From

$$I_n = C_X^{-1}C_X = C_X^{-1}(A + BX(b,a)) = C_X^{-1}A + C_X^{-1}BX(b,a) \tag{6.6}$$

we conclude that

$$\begin{aligned} X(t,s) - X(t,a)C_X^{-1}BX(b,s) &= X(t,a)X(a,s) - X(t,a)C_X^{-1}BX(b,a)X(a,s) \\ &= X(t,a)\left(I_n - C_X^{-1}BX(b,a)\right)X(a,s) \\ &= X(t,a)C_X^{-1}AX(a,s), \end{aligned}$$

which gives us the more symmetric formula

$$G(t,s) = \begin{cases} X(t,a)C_X^{-1}AX(a,s) & : a \le s \le t \le b, \\ -X(t,a)C_X^{-1}BX(b,s) & : a \le t < s \le b. \end{cases} \tag{6.7}$$

Lemma 6.5. *If C_X is invertible then there is exactly one Green's function.*

Proof. Suppose that $\tilde{G}$ were another Green's function. Then both G and $\tilde{G}$ are continuous on the triangle $\{(t,s)\colon a \le t \le s \le b\}$, and they are both continuous on the other triangle. Then the same holds for the difference

$$H(t,s) = G(t,s) - \tilde{G}(t,s).$$

Moreover, H has no jump across the diagonal:

$$H(t+0,t) - H(t-0,t) = (G(t+0,t) - G(t-0,t)) - (\tilde{G}(t+0,t) - \tilde{G}(t-0,t)) = I_n - I_n = 0,$$

hence H is continuous on the square $(a,b)\times(a,b)$. We also have the representation

$$y(t) = \int_{s=a}^{b} G(t,s)g(s)\,\mathrm{d}s = \int_{s=a}^{b} \tilde{G}(t,s)g(s)\,\mathrm{d}s$$

for the unique solution y to (6.5), for each continuous g. Then

$$0 = \int_{s=a}^{b} H(t,s)g(s)\,\mathrm{d}s, \qquad a \le t \le b,$$

for each continuous function $g = g(s)$. Now choose a time t_*, and choose an index $l \in \{1, \dots, n\}$. Put $g(s) = (\overline{h_{l1}(t_*, s)}, \dots, \overline{h_{ln}(t_*, s)})^\top$ as the adjoint of the l-th row of $H(t_*, s)$. This choice gives us then

$$0 = \int_{s=a}^{b} \sum_{k=1}^{n} |h_{lk}(t_*, s)|^2 \, \mathrm{d}s,$$

which is possible only for $h_{lk}(t_*) \equiv 0$ for all k, because the matrix H is continuous. Now choose an arbitrary t_*, and an arbitrary l, and repeat.

Hence $G = \tilde{G}$, contradicting our assumption. □

Lemma 6.6. *Let G be a Green function. Then: for each fixed $s \in [a, b]$, the function $G \mapsto G(t, s)$ solves*

$$\begin{aligned} \partial_t G(t, s) &= F(t)G(t, s), \qquad \forall\, t \in [a, b] \setminus \{s\}, \\ AG(a, s) + BG(b, s) &= 0 \qquad \textit{provided } s \neq a, \quad s \neq b. \end{aligned}$$

Proof. By Lemma 6.5 and (6.7), we have

$$G(t, s) = X(t, a)P_{\pm}(s),$$

for a certain matrix function P_+ if $t \geq s$, and a certain matrix function P_- for $t < s$. But we know that $\partial_t X(t, a) = F(t)X(t, a)$.

Finally, for $s \notin \{a, b\}$ the following calculation is valid:

$$\begin{aligned} &AG(a, s) + BG(b, s) \\ &\quad = A\left(-X(a, a)C_X^{-1}BX(b, s)\right) + B\left(X(b, a)C_X^{-1}AX(a, s)\right) \\ &\quad = \left(-AC_X^{-1}BX(b, a) + BX(b, a)C_X^{-1}A\right)X(a, s) \qquad \Big| \quad \text{exploit (6.6)} \\ &\quad = \left(-A(I_n - C_X^{-1}A) + BX(b, a)C_X^{-1}A\right)X(a, s) \\ &\quad = \left(-A + (A + BX(b, a))C_X^{-1}A\right)X(a, s) \\ &\quad = \left(-A + C_X C_X^{-1}A\right)X(a, s) \\ &\quad = 0. \end{aligned}$$

□

All these discussions can be generalised a bit. Instead of the fundamental matrix X, we may take any matrix valued function $Y = Y(t)$ which solves $Y'(t) = A(t)Y(t)$, and whose values are invertible matrices. Remember that according to Proposition 3.5 it is enough to check that $\det Y(t) \neq 0$ for one special time t.

Then we have $Y(t) = X(t, a)Y(a)$, or $X(t, a) = Y(t)Y^{-1}(a)$, which implies $C_X = A + BY(b)Y^{-1}(a)$, and this is invertible if and only if

$$C_Y := AY(a) + BY(b)$$

is invertible. We also have $X(b, a) = Y(b)Y^{-1}(a)$ as well as $X(a, s) = Y(a)Y^{-1}(s)$, from which we obtain

$$G(t, s) = \begin{cases} Y(t)C_Y^{-1}AY(a)Y^{-1}(s) & : a \leq s \leq t \leq b, \\ -Y(t)C_Y^{-1}BY(b)Y^{-1}(s) & : a \leq t < s \leq b. \end{cases} \tag{6.8}$$

Whatever matrix Y we choose here, the Green's matrix G will always be the same. A good selection of Y might simplify the calculations, as we will see in the next section.

6.3 Second Order Scalar BVPs

Now we will apply the results of the previous section to

$$\begin{cases} u''(t) + a_1(t)u'(t) + a_0(t)u(t) = f(t), & a \leq t \leq b, \\ \alpha_0 u(a) + \alpha_1 u'(a) = c_a, \\ \beta_0 u(b) + \beta_1 u'(b) = c_b, \end{cases} \tag{6.9}$$

where we assume $(\alpha_0, \alpha_1) \neq (0,0)$ and $(\beta_0, \beta_1) \neq (0,0)$ to stay away from trivialities. For convenience, we abbreviate the differential equation and the boundary conditions as

$$\Lambda u = f, \qquad R_a u = c_a, \qquad R_b u = c_b. \tag{6.10}$$

Here $\Lambda = \frac{\mathrm{d}^2}{\mathrm{d}t^2} + a_1 \frac{\mathrm{d}}{\mathrm{d}t} + a_0$ is a *differential operator*. In this section we always assume that all the functions and boundary data are real-valued. Then we can omit the conjugation in the scalar products.

To obtain a first order system, we set

$$y(t) = \begin{pmatrix} u(t) \\ u'(t) \end{pmatrix},$$

and then we find

$$y'(t) = \begin{pmatrix} 0 & 1 \\ -a_0(t) & -a_1(t) \end{pmatrix} y(t) + \begin{pmatrix} 0 \\ f(t) \end{pmatrix} =: F(t)y(t) + g(t),$$
$$Ay(a) + By(b) := \begin{pmatrix} \alpha_0 & \alpha_1 \\ 0 & 0 \end{pmatrix} y(a) + \begin{pmatrix} 0 & 0 \\ \beta_0 & \beta_1 \end{pmatrix} y(b) = \begin{pmatrix} c_a \\ c_b \end{pmatrix} =: c,$$

compare (6.3). If u_1 and u_2 are two linearly independent solutions to $\Lambda u = 0$, then we define their Wronski matrix Y and their Wronski determinant W as

$$Y(t) = \begin{pmatrix} u_1(t) & u_2(t) \\ u_1'(t) & u_2'(t) \end{pmatrix}, \qquad W(t) = \det Y(t),$$

and we know that $W(t)$ is never zero.

The matrix C_Y, whose invertibility determines the solvability of (6.9), is

$$C_Y := AY(a) + BY(b) = \begin{pmatrix} \alpha_0 u_1(a) + \alpha_1 u_1'(a) & \alpha_0 u_2(a) + \alpha_1 u_2'(a) \\ \beta_0 u_1(b) + \beta_1 u_1'(b) & \beta_0 u_2(b) + \beta_1 u_2'(b) \end{pmatrix} = \begin{pmatrix} R_a u_1 & R_a u_2 \\ R_b u_1 & R_b u_2 \end{pmatrix}.$$

Lemma 6.7. *The problem (6.9) is uniquely solvable for all f, c_a, c_b if and only if* $\det C_Y \neq 0$. *This property does not depend on the choice of the Wronski matrix Y.*

There is nothing wrong with this approach, but it has some drawbacks:

- the Wronski determinant $W(t)$ is not easy to compute or understand,
- from the introduction we know that scalar products and partial integrations can be helpful in understanding what is going on. However, in our case we have

$$\begin{aligned} \langle \Lambda u, v \rangle &= \int_{t=a}^{b} \Big(u''(t) + a_1(t)u'(t) + a_0(t)u(t) \Big) v(t)\,\mathrm{d}t \\ &= \int_{t=a}^{b} u(t) \cdot \Big(v''(t) - (a_1(t)v(t))' + a_0(t)v(t) \Big)\,\mathrm{d}t + (\text{boundary terms}), \end{aligned}$$

  and the big parenthesis[7] in the last line looks quite different from Λv.
- if we intend to compute $G(t,s)$ from (6.8), we have to multiply five matrices, and two of them are the inverses of other matrices. This is an exercise which most people with a sense for beauty prefer to avoid.

In the following we assume that (6.9) is uniquely solvable, because otherwise there is not much we can do.

We should refine our approach a bit. Define

$$p(t) := \exp\left(\int_{s=a}^{t} a_1(s)\,\mathrm{d}s \right)$$

[7] Klammer ( )

and observe that $p'(t) = p(t)a_1(t)$. This function p is never zero, and multiplying with a non-zero function never gives trouble:

$$p(t)u''(t) + p(t)a_1(t)u(t) + p(t)a_0(t)u(t) = p(t)f(t),$$

but this is nothing else than

$$\Big(p(t)u'(t)\Big)' + p(t)a_0(t)u(t) = p(t)f(t).$$

We fix $m := pa_0$ and define a new differential operator L:

$$Lu := \frac{\mathrm{d}}{\mathrm{d}t}\left(p\frac{\mathrm{d}}{\mathrm{d}t}u\right) + mu,$$

and then we have to solve

$$Lu = pf, \qquad R_a u = c_a, \qquad R_b u = c_b.$$

By a simple shift in the unknown function u, we may suppose $c_a = c_b = 0$.

Concerning the question for the Wronski determinant $W(t)$, a quick calculation persuades us of the *Lagrange identity*

$$vLw - wLv = \frac{\mathrm{d}}{\mathrm{d}t}\left(p\,\det\begin{pmatrix} v & w \\ v' & w' \end{pmatrix}\right), \tag{6.11}$$

valid for all functions v and w (they need not be solutions of whatever equation). Choosing $v = u_1$ and $w = u_2$, we then get

$$p(t)W(t) = \text{const.}, \quad a \le t \le b,$$

in particular $p(t)W(t) = p(a)W(a)$. Note that $p(a) = 1$.

Concerning the question about the partial integrations, we remark that

$$\langle Lv, w\rangle = \langle v, Lw\rangle + (\text{boundary terms}),$$

for all functions v and w. This looks good.

And now we compute the Green's matrix. To this end and for reasons of computational beauty, we take carefully selected solutions u_1 and u_2: let u_1 be a solution with

$$\Lambda u_1 = 0, \qquad R_a u_1 = 0, \quad R_b u_1 \neq 0, \tag{6.12}$$

and let u_2 be a solution with

$$\Lambda u_2 = 0, \qquad R_a u_2 \neq 0, \quad R_b u_2 = 0. \tag{6.13}$$

Such functions u_1 and u_2 do exist, because of our assumption that (6.9) be uniquely solvable for all f and c. Then we find

$$C_Y = \begin{pmatrix} R_a u_1 & R_a u_2 \\ R_b u_1 & R_b u_2 \end{pmatrix} = \begin{pmatrix} 0 & R_a u_2 \\ R_b u_1 & 0 \end{pmatrix}, \qquad C_Y^{-1} = \frac{1}{\det C_Y}\begin{pmatrix} 0 & -R_a u_2 \\ -R_b u_1 & 0 \end{pmatrix}.$$

Exercise: *Recall the formula for the inverse of a 2×2 matrix.*

We also have

$$AY(a) = \begin{pmatrix} \alpha_0 & \alpha_1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_1(a) & u_2(a) \\ u_1'(a) & u_2'(a) \end{pmatrix} = \begin{pmatrix} 0 & R_a u_2 \\ 0 & 0 \end{pmatrix},$$
$$BY(b) = \begin{pmatrix} 0 & 0 \\ \beta_0 & \beta_1 \end{pmatrix}\begin{pmatrix} u_1(b) & u_2(b) \\ u_1'(b) & u_2'(b) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ R_b u_1 & 0 \end{pmatrix},$$

from which we gain

$$C_Y^{-1}AY(a) = \frac{1}{-(R_a u_2)(R_b u_1)} \begin{pmatrix} 0 & -R_a u_2 \\ -R_b u_1 & 0 \end{pmatrix} \begin{pmatrix} 0 & R_a u_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$
$$C_Y^{-1}BY(b) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and now (6.8) turns into

$$G(t,s) = \begin{cases} Y(t) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} Y^{-1}(s) & : a \le s \le t \le b, \\ -Y(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} Y^{-1}(s) & : a \le t < s \le b. \end{cases}$$

Before we simplify this further, note that this G gives us the solution y to the first order system $y'(t) = F(t)y(t) + g(t)$ via $y(t) = \int_{s=a}^{b} G(t,s)g(s)\,\mathrm{d}s$. But in our case,

$$y(t) = \begin{pmatrix} u(t) \\ u'(t) \end{pmatrix}, \qquad g(s) = \begin{pmatrix} 0 \\ f(s) \end{pmatrix},$$

and we are mainly interested in u which is the first component of y. Hence

$$u(t) = \int_{s=a}^{b} G_{12}(t,s)f(s)\,\mathrm{d}s,$$

and the other three entries of G are of limited interest only. The inverse $Y^{-1}(s)$ is found via

$$Y^{-1}(s) = \begin{pmatrix} u_1(s) & u_2(s) \\ u_1'(s) & u_2'(s) \end{pmatrix}^{-1} = \frac{1}{\det Y(s)} \begin{pmatrix} u_2'(s) & -u_2(s) \\ -u_1'(s) & u_1(s) \end{pmatrix} = \frac{1}{W(s)} \begin{pmatrix} u_2'(s) & -u_2(s) \\ -u_1'(s) & u_1(s) \end{pmatrix}.$$

The final result then is

$$G_{12}(t,s) = \begin{cases} \frac{u_2(t)u_1(s)}{W(s)} & : a \le s \le t \le b, \\ \frac{u_1(t)u_2(s)}{W(s)} & : a \le t \le s \le b, \end{cases} \tag{6.14}$$

and $W(s)$ can be computed via $W(s) = \frac{p(a)W(a)}{p(s)} = \frac{u_1(a)u_2'(a) - u_2(a)u_1'(s)}{p(s)}$.

We summarise:

Theorem 6.8. *Let a_0, $a_1 \in C([a,b] \to \mathbb{R})$ be given. Put $p(t) = \exp(\int_{s=a}^{t} a_1(s)\,\mathrm{d}s)$, and define Λ, R_a, R_b as in (6.10). Suppose that the fully homogeneous problem (6.9) with $f \equiv 0$, $c_a = c_b = 0$ has only the zero solution. Define u_1, u_2 by (6.12), (6.13), and G_{12} by (6.14).*

Then the unique solution u to the half-homogeneous problem (6.9) with $c_a = c_b = 0$ is given by

$$u(t) = \int_{s=a}^{b} G_{12}(t,s)f(s)\,\mathrm{d}s, \qquad a \le t \le b,$$

for each continuous function f. The Green function G_{12} (given in (6.14)) is continuous on $[a,b] \times [a,b]$. The term $W = u_1u_2' - u_2u_1'$ in (6.14) is the Wronskian of (u_1, u_2). The function $t \mapsto p(t)W(t)$ is constant. The first derivative with respect to t jumps with height one:

$$\partial_1 G_{12}(t+0,t) - \partial_1 G_{12}(t-0,t) = 1, \qquad a < t < b.$$

For fixed s, and $t \neq s$, G_{12} solves the ODE $\Lambda G_{12}(\cdot,s) = 0$ with respect to the variable t.

For fixed $s \notin \{a,b\}$, $G_{12}(\cdot,s)$ solves the boundary conditions $R_a G_{12}(\cdot,s) = 0$ and $R_b G_{12}(\cdot,s) = 0$.

We can also consider the BVP

$$\begin{cases} Lu(t) = f(t), & a \le t \le b, \\ R_a u = 0, \quad R_b u = 0, \\ Lu := \frac{\mathrm{d}}{\mathrm{d}t}\left(p\frac{\mathrm{d}u}{\mathrm{d}t}\right) + mu, \end{cases}$$

with $p \in C^1([a,b] \to \mathbb{R})$ taking only positive values on $[a,b]$, and m continuous on $[a,b]$. All coefficients are real-valued.

Then the (by assumption unique) solution u is given by

$$u(t) = \int_{s=a}^{b} G_{12}(t,s) f(s)\,\mathrm{d}s,$$

and now G_{12} is written as

$$G_{12}(t,s) = \begin{cases} \frac{u_2(t)u_1(s)}{p(a)W(a)} & : a \le s \le t \le b, \\ \frac{u_1(t)u_2(s)}{p(a)W(a)} & : a \le t \le s \le b, \end{cases}$$

with G_{12} continuous on the square $[a,b]\times[a,b]$, but $\partial_t G_{12}$ has a jump of height $1/p(t)$ across the diagonal.

We conclude this section with a small result on the zeros of solution to the homogeneous problem.

Proposition 6.9. *Let u_1 and u_2 be two linearly independent solutions to $Lu = 0$. If t_a and $t_b > t_a$ are two consecutive*[8] *zeros of u_1, then u_2 must have a zero between t_a and t_b.*

Proof. The Wronski determinant $W(t) = u_1(t)u_2'(t) - u_2(t)u_1'(t)$ is never zero, and it has no jumps. Hence W always has the same sign.

Then we have

$$W(t_a) = 0 - u_2(t_a)u_1'(t_a), \qquad W(t_b) = 0 - u_2(t_b)u_1'(t_b).$$

Because t_a and t_b are two *consecutive* zeros of u_1, the derivatives $u_1'(t_a)$ and $u_1'(t_b)$ must have different sign. On the other hand, $W(t_a)$ and $W(t_b)$ must have the same sign. This is only possible of $u_2(t_a)$ and $u_2(t_b)$ have different sign. □

6.4 Playing in Hilbert Spaces

Definition 6.10 (Hilbert Space). *A vector space $\mathcal{H}$ over the field $\mathbb{C}$ is called a* Hilbert space *if it has a scalar product*

$$\langle \cdot,\cdot \rangle_{\mathcal{H}} : \mathcal{H}\times\mathcal{H} \to \mathbb{C},$$

which generates a norm via $\|v\|_{\mathcal{H}} := \sqrt{\langle v,v\rangle_{\mathcal{H}}}$, under which $\mathcal{H}$ is a complete *normed space (each Cauchy sequence in $\mathcal{H}$ converges in $\mathcal{H}$).*

We start with an example.

Ground space: $\mathcal{H} = \mathbb{C}^n$ with the scalar product $\langle v,w\rangle := \sum_{j=1}^n v_j \overline{w_j}$.

Linear operator $A\colon z \mapsto Az$ with $A \in \mathbb{C}^{n\times n}$

Domain of A [9]: $D(A)$ contains all $z \in \mathcal{H}$ for which Az makes sense ($Az \in \mathcal{H}$). Hence $D(A) = \mathcal{H} = \mathbb{C}^n$.

Adjoint operator A^* and its domain: the vector space $D(A^*)$ contains all $w \in \mathcal{H}$ for which A^*w makes sense, and such that

$$\langle Az, w\rangle_{\mathcal{H}} = \langle z, A^*w\rangle_{\mathcal{H}}, \qquad \forall\, z \in D(A), \qquad \forall\, w \in D(A^*).$$

In our case, $D(A^*) = \mathcal{H}$ and $A^* = \overline{A^\top}$.

Self-adjoint operator: A is self-adjoint if $A = A^*$ and $D(A) = D(A^*)$.

[8] aufeinanderfolgend
[9] Definitionsbereich von A

Eigenvalues of self-adjoint A: if $Az = \lambda z$ with $z \neq 0$, then

$$\lambda \langle z, z \rangle_{\mathcal{H}} = \ldots = \overline{\lambda} \langle z, z \rangle_{\mathcal{H}},$$

hence $\lambda \in \mathbb{R}$ (fill in the gap yourself).

Eigenvectors to different eigenvalues for self-adjoint A: if $Az = \lambda z$ and $Aw = \mu w$ with $\lambda \neq \mu$ then $\langle z, w \rangle_{\mathcal{H}} = 0$ because of

$$\lambda \langle z, w \rangle_{\mathcal{H}} = \ldots = \mu \langle z, w \rangle_{\mathcal{H}}.$$

Spectral theorem: if A is self-adjoint, then an ONB $(u_1, \ldots, u_n)$ can be selected from the eigenvectors of A (the proof is a bit harder than the previous two). If $\mu_1 < \mu_2 < \cdots < \mu_m$ are the distinct eigenvalues of $A \in \mathbb{C}^{n \times n}$ (with $m \leq n$), and P_j is the orthogonal projector of $\mathcal{H}$ onto $\ker(A - \mu_j I)$ (the eigenspace to the eigenvalue μ_j), then

$$I = \sum_{j=1}^{m} P_j, \qquad A = \sum_{j=1}^{m} \mu_j P_j.$$

The first equation means that the eigenvectors span $\mathcal{H}$, or

$$\mathcal{H} = \ker(A - \mu_1 I) \oplus \ker(A - \mu_2 I) \oplus \ldots \oplus \ker(A - \mu_m I).$$

The second equation means that, on the eigenspace $\ker(A - \mu_j I)$, A acts as a multiplication by μ_j.

Next we consider boundary value problems (sometimes called Sturm[10]–Liouville BVPs):

$$\begin{cases} \Lambda u = \lambda u, \qquad a \leq t \leq b, \\ R_a u = 0, \quad R_b u = 0, \\ \Lambda = a_2(t) \dfrac{\mathrm{d}^2}{\mathrm{d}t^2} + a_1(t) \dfrac{\mathrm{d}}{\mathrm{d}t}, \end{cases}$$

and we assume that a_2 and a_1 are continuous and $\mathbb{R}$–valued, and a_2 is strictly positive. Note that the differential equation $\Lambda u = \lambda u$ can be equivalently rewritten as

$$\begin{aligned} Lu &= \lambda q u, \qquad \frac{1}{q} Lu = \lambda u, \\ Lu &:= \frac{\mathrm{d}}{\mathrm{d}t} \left(p(t) \frac{\mathrm{d}}{\mathrm{d}t} u \right), \\ p(t) &:= \exp \left(\int_{s=a}^{t} \frac{a_1(s)}{a_2(s)} \, \mathrm{d}s \right), \\ q(t) &:= \frac{p(t)}{a_2(t)}. \end{aligned}$$

The functions p and q are assumed to be strictly positive on the finite interval $[a, b]$. The differences to (6.9) are: a_2 is present (for greater generality), a_0 is dropped (for ease of notation), and u can be $\mathbb{C}$–valued.

Ground space: $\mathcal{H}$ contains all functions $u \colon [a, b] \to \mathbb{C}$ with

$$\int_{t=a}^{b} |u(t)|^2 q(t) \, \mathrm{d}t < \infty,$$

and $\mathcal{H}$ is equipped with the scalar product

$$\langle v, w \rangle_{\mathcal{H}} := \int_{t=a}^{b} v(t) \overline{w(t)} q(t) \, \mathrm{d}t.$$

[10] Jacques Charles François Sturm, 1803–1855

Linear operator: the most interesting linear operator is $\Lambda = \frac{1}{q}L$.

Domain of Λ: $D(\Lambda)$ contains all functions $u \in \mathcal{H}$ for which Λu makes sense (this means $\Lambda u \in \mathcal{H}$), and for which additionally $R_a u = 0$ and $R_b u = 0$. In our situation $D(\Lambda) \subsetneq \mathcal{H}$, and typically the elements u of $D(\Lambda)$ satisfy $a_2 u'' \in \mathcal{H}$, which means that the elements of $D(\Lambda)$ are smoother than those of $\mathcal{H}$.

Adjoint operator Λ^* and its domain: $D(\Lambda^*)$ contains all $w \in \mathcal{H}$ for which $\Lambda^* w \in \mathcal{H}$ and additionally

$$\langle \Lambda v, w\rangle_{\mathcal{H}} = \langle v, \Lambda^* w\rangle_{\mathcal{H}}\,, \qquad \forall v \in D(\Lambda), \qquad \forall w \in D(\Lambda^*).$$

By the Lagrange identity (6.11), we have

$$\begin{aligned}\langle \Lambda v, w\rangle_{\mathcal{H}} &= \int_{t=a}^{b} \frac{1}{q}(Lv)\overline{w} q \,\mathrm{d}t = \int_{t=a}^{b} (Lv)\overline{w}\,\mathrm{d}t = \int_{t=a}^{b} v\overline{Lw}\,\mathrm{d}t + \int_{t=a}^{b} \frac{\mathrm{d}}{\mathrm{d}t}\left(p\det\begin{pmatrix}\overline{w} & v\\ \overline{w'} & v'\end{pmatrix}\right)\mathrm{d}t \\ &= \int_{t=a}^{b} v\cdot\overline{\frac{1}{q}Lw}\cdot q\,\mathrm{d}t + p\det\begin{pmatrix}\overline{w} & v\\ \overline{w'} & v'\end{pmatrix}\Big|_{t=a}^{b} = \langle v, \Lambda w\rangle_{\mathcal{H}} + p\det\begin{pmatrix}\overline{w} & v\\ \overline{w'} & v'\end{pmatrix}\Big|_{t=a}^{b}.\end{aligned}$$

Now we have to make sure that the boundary term $p\det(\dots)|_{t=a}^{b}$ is zero. We know already that $R_a v = 0$ and $R_b v = 0$, because of the assumption $v \in D(\Lambda)$. The boundary conditions $R_a^* w = 0$ and $R_b^* w = 0$ for the operator Λ^* have to be designed in such a way that this boundary term vanishes.

Self-adjoint operators: we define that the operator Λ is self-adjoint if $\Lambda = \Lambda^*$ and $D(\Lambda) = D(\Lambda^*)$.

Eigenvalues of self-adjoint Λ: if $\Lambda u = \lambda u$ with $u \in D(\Lambda) = D(\Lambda^*)$ and $u \neq 0$ then

$$\lambda\,\langle u, u\rangle_{\mathcal{H}} = \ldots = \overline{\lambda}\,\langle u, u\rangle_{\mathcal{H}}\,,$$

hence $\lambda \in \mathbb{R}$. From now on, we assume that all functions are real-valued, and then the conjugation bars in the integrals can be omitted.

Eigenfunctions to different eigenvalues: if $\Lambda u = \lambda u$ and $\Lambda v = \mu v$ with $\lambda \neq \mu$, then $\langle u, v\rangle_{\mathcal{H}} = 0$ because of

$$\lambda\,\langle u, v\rangle_{\mathcal{H}} = \ldots = \mu\,\langle u, v\rangle_{\mathcal{H}}\,.$$

Eigenfunctions to the same eigenvalue: if $\Lambda u = \lambda u$ and $\Lambda v = \lambda v$, then u, v are linearly dependent. This means that each eigenvalue of Λ has multiplicity one. The reason is this: we know

$$\begin{aligned} R_a u = 0 \quad &\Longrightarrow \quad \alpha_0 u(a) + \alpha_1 u'(a) = 0,\\ R_a v = 0 \quad &\Longrightarrow \quad \alpha_0 v(a) + \alpha_1 v'(a) = 0,\end{aligned}$$

which can be reformulated as

$$\begin{pmatrix} u(a) & u'(a)\\ v(a) & v'(a)\end{pmatrix}\begin{pmatrix}\alpha_0\\ \alpha_1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix},$$

but $(\alpha_0, \alpha_1) \neq (0, 0)$, by assumption. Therefore $W(a) = 0$, with W as Wronski determinant of (u, v).

Spectral theorem: if Λ is a Sturm–Liouville operator, then it has eigenvalues $\lambda_1 > \lambda_2 > \ldots$ which approach $-\infty$, and the associated eigenfunctions u_1, u_2,... (scaled to norm one) form an ONB of $\mathcal{H}$: for each $f \in \mathcal{H}$, there are unique numbers γ_1, γ_2, $\ldots \in \mathbb{C}$ such that

$$f = \sum_{j=1}^{\infty} \gamma_j u_j, \qquad \|f\|_{\mathcal{H}}^2 = \sum_{j=1}^{\infty} |\gamma_j|^2.$$

The coefficients γ_j can be found via $\gamma_j = \langle f, u_j\rangle_{\mathcal{H}}$.

Unfortunately, the proof is beyond our reach, see §37–§40 of [13] for details.

The projector P_j of $\mathcal{H}$ onto $\ker(\Lambda - \lambda_j I) = \mathrm{span}(u_j)$ is given by

$$P_j g = \langle g, u_j\rangle_{\mathcal{H}}\, u_j.$$

Then we have

$$I = \sum_{j=1}^{\infty} P_j \qquad \text{(valid when applied to functions from } \mathcal{H}\text{)}, \tag{6.15}$$

$$\Lambda = \sum_{j=1}^{\infty} \lambda_j P_j \qquad \text{(valid when applied to functions from } D(\Lambda)\text{)}. \tag{6.16}$$

As a concrete example, we take $a_2 \equiv 1$ and $a_1 \equiv 0$, giving us $p = q = 1$:

$$\begin{cases} \dfrac{\mathrm{d}^2}{\mathrm{d}t^2} u(t) = \lambda u(t), & 0 \le t \le \pi, \\ u(0) = 0, \qquad u(\pi) = 0. \end{cases}$$

We have $\mathcal{H} = L^2((0,\pi))$ with the scalar product

$$\langle v, w \rangle_{\mathcal{H}} = \int_{t=0}^{\pi} v(t)\overline{w(t)}\,\mathrm{d}t.$$

The operator is $L = \Lambda = \frac{\mathrm{d}^2}{\mathrm{d}t^2}$, and its domain $D(\Lambda) = D(L)$ consists of all those functions $v \in \mathcal{H}$ with $v'' \in \mathcal{H} = L^2((0,\pi))$ and $v(0) = v(\pi) = 0$. Now we determine $D(\Lambda^*)$. A function $w \in \mathcal{H}$ belongs to $D(\Lambda^*)$ if $w'' \in \mathcal{H}$ and

$$\det \begin{pmatrix} \overline{w} & v \\ \overline{w}' & v' \end{pmatrix} \Big|_{t=0}^{\pi} = 0$$

provided that $v \in D(\Lambda)$. Since nothing is known about the values of v' on the boundary, we need $w(0) = w(\pi) = 0$. These are the same boundary conditions as for v, and therefore $D(\Lambda) = D(\Lambda^*)$, $\Lambda = \Lambda^*$, which makes Λ self-adjoint.

The spectral theorem says that the eigenfunctions of L form an ONB of $\mathcal{H}$. These eigenfunctions are

$$u_j(t) = \frac{\sin(jt)}{\|\sin(j\cdot)\|_{L^2((0,\pi))}}, \qquad j = 1, 2, \ldots, \tag{6.17}$$

and the associated eigenvalues are $\lambda_j = -j^2$.

Equation (6.15) says that each function $f \in \mathcal{H} = L^2((0,\pi))$ can be expanded like this:

$$f(t) = \sum_{j=1}^{\infty} \gamma_j \sin(jt), \qquad \gamma_j = \frac{\int_{t=0}^{\pi} f(t)\sin(jt)\,\mathrm{d}t}{\int_{t=0}^{\pi} \sin^2(jt)\,\mathrm{d}t}.$$

As a second example, we take again $p = q = 1$, but now other boundary conditions:

$$\begin{cases} \dfrac{\mathrm{d}^2}{\mathrm{d}t^2} u(t) = \lambda u(t), & 0 \le t \le \pi, \\ u'(0) = 0, \qquad u'(\pi) = 0. \end{cases}$$

We have again $\mathcal{H} = L^2((0,\pi))$ with the same scalar product. The domain $D(\Lambda)$ of Λ consists of those functions $v \in \mathcal{H}$ with $v'' \in L^2((0,\pi))$ and $v'(0) = v'(\pi) = 0$. By a similar computation as in the previous example, we find $D(\Lambda^*) = D(\Lambda)$ and $\Lambda = \Lambda^*$. Therefore, Λ is self-adjoint, and its eigenfunctions are the elements of an ONB of $\mathcal{H}$. These eigenfunctions are

$$u_j(t) = \frac{\cos(jt)}{\|\cos(j\cdot)\|_{L^2((0,\pi))}}, \qquad j = 0, 1, 2\ldots, \tag{6.18}$$

and the associated eigenvalues are $\lambda_j = -j^2$. Be careful: now j starts at zero instead of one.

According to (6.15), each function $f \in \mathcal{H} = L^2((0,\pi))$ can be expanded like this:

$$f(t) = \sum_{j=0}^{\infty} \gamma_j \cos(jt), \qquad \gamma_j = \frac{\int_{t=0}^{\pi} f(t)\cos(jt)\,\mathrm{d}t}{\int_{t=0}^{\pi} \cos^2(jt)\,\mathrm{d}t}.$$

6.5 Orthogonal Polynomials

- now we will consider special BVPs of "almost–Sturm–Liouville" type,
- the eigenfunctions of these BVPs are polynomials, which form an orthogonal family with respect to a certain scalar product,
- these polynomials give rise to an orthogonal basis of the ground space,
- and (of course !) we will find numerous applications in physics.

Take an interval $(a, b) \subset \mathbb{R}$. This interval might be of infinite length (recall that the intervals of traditional Sturm–Liouville BVPs are always bounded).

Choose a weight function[11] $q = q(t)$, which is real valued and positive on (a, b). On the end-points of the interval, p and q might take the value zero, and q might have a pole (this is prohibited for traditional Sturm–Liouville BVPs).

Consider the vector space $\mathcal{H}$ of all functions $u\colon (a, b) \to \mathbb{R}$ with

$$\int_{t=a}^{b} u^2(t)q(t)\,\mathrm{d}t < \infty.$$

We equip this space with the scalar product

$$\langle u, v\rangle_{\mathcal{H}} := \int_{t=a}^{b} u(t)v(t)q(t)\,\mathrm{d}t.$$

We wish that each polynomial u is a member of $\mathcal{H}$. Therefore, if (a, b) is an unbounded interval, the weight function q must decay fast for t going to infinity; otherwise the integral $\int_{t=a}^{b} u^2(t)q(t)\,\mathrm{d}t$ will not be finite.

We start with the infinite family $(t^0, t^1, t^2, \dots)$ of polynomials. These are linearly independent.

The Gram–Schmidt procedure then gives us a family of polynomials $(p_0, p_1, \dots)$ with

$$\langle p_j, p_k\rangle_{\mathcal{H}} = 0 \qquad \text{if } j \neq k,$$

and each p_j has degree j. We do not care whether these p_j have norm equal to one or not.

Each polynomial Q can be written as a linear combination of the p_j, and this linear combination involves only a finite number of the p_j.

It is unclear whether the family $(p_0, p_1, \dots)$ is an orthogonal basis of $\mathcal{H}$, because it might happen that these p_j span only a smaller sub-space of $\mathcal{H}$, but not the full $\mathcal{H}$. The next lemma tells us that we are lucky if (a, b) is an interval of finite length.

Lemma 6.11. *If (a, b) is a bounded interval, then $(p_0, p_1, \dots)$ is a complete orthogonal system in $\mathcal{H}$.*

Sketch of proof. Assume the opposite. Then $\overline{\operatorname{span}(p_0, p_1, \dots)} \subsetneq \mathcal{H}$, where $\operatorname{span}(p_0, p_1, \dots)$ is the set of all the linear combinations of the p_j (where each linear combination contains only a finite number of summands). And the over-line means that we take the topological closure[12]: we add all the cluster points[13] of the set to it.

Then there is an element of $\mathcal{H}$ which is not in $V := \overline{\operatorname{span}(p_0, p_1, \dots)}$. This element can be decomposed into a part parallel to V, and a part orthogonal to V (here we need that V is a closed set). Call the orthogonal part f. Clearly $f \neq 0$.

Then $\langle f, p_j\rangle_{\mathcal{H}} = 0$ for each polynomial p_j from the orthogonal system. We assume that f is continuous (this is the part where our proof has a gap. See [14] for how to close it).

[11] Gewichtsfunktion
[12] topologischer Abschluß
[13] Häufungspunkte

We recall the WEIERSTRASS approximation theorem from the second semester: if f is a continuous function on the bounded interval $[a,b]$, then, for each positive ε, we find a polynomial Q_ε which approximates f up to a uniform error of size ε:

$$|f(t) - Q_\varepsilon(t)| < \varepsilon, \qquad \forall\, t \in [a,b].$$

At this point we need that $[a,b]$ is bounded. This Q_ε can be written as a finite linear combination of the p_j. Therefore,

$$\langle f, Q_\varepsilon\rangle_{\mathcal{H}} = 0.$$

Then we can conclude that

$$0 < \langle f, f\rangle_{\mathcal{H}} = \langle f, f - Q_\varepsilon\rangle_{\mathcal{H}} \le \int_{t=a}^{b} |f(t)| \cdot |f(t) - Q_\varepsilon(t)| \cdot q(t)\,\mathrm{d}t < \varepsilon \int_{t=a}^{b} |f(t)| \cdot q(t)\,\mathrm{d}t.$$

Now send ε to zero to obtain a contradiction. □

Our approach is now coming from the other side: first we invent a family of polynomials $(p_0, p_1, \dots)$, and then we determine their differential equation.

Mysteriously, this approach leads to results applicable in physics. Just enjoy the show.

Take $(a,b) = (-1,1)$ as the interval.

The scalar product of the space $\mathcal{H}$ is $\langle u, v\rangle_{\mathcal{H}} = \int_{t=-1}^{1} u(t)v(t)q(t)\,\mathrm{d}t$. Set, for $n = 0, 1, 2, \dots,$

$$p_n(t) := \frac{1}{q(t)} \frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big(q(t)(1-t^2)^n\Big), \qquad r(t) := 1 - t^2.$$

Formulae of this kind are known as RODRIGUES' FORMULA[14].

Lemma 6.12. *This function p_n is $\mathcal{H}$-perpendicular to each polynomial of degree $< n$.*

Proof. We check by n–fold partial integration that

$$\big\langle t^l, p_n\big\rangle_{\mathcal{H}} = \int_{t=-1}^{1} t^l p_n(t) q(t)\,\mathrm{d}t = \int_{t=-1}^{1} t^l \frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big(q(t)(1-t^2)^n\Big)\,\mathrm{d}t = 0,$$

for $l = 0, 1, 2, \dots, n-1$. Boundary terms never appear because of $r(t) = 0$ for $t = \pm 1$. □

Next we try to find q such that each p_n is indeed a polynomial of degree n. Start with $n = 1$. Then

$$p_1(t) = \frac{1}{q(t)}(q(t)(1-t^2))' = -2t + \frac{q'(t)}{q(t)}(1-t^2) \stackrel{!}{=} \gamma_0 + \gamma_1 t,$$

for some $\gamma_0, \gamma_1 \in \mathbb{R}$, and this can be solved, giving us (after some time)

$$q(t) = (1-t)^\alpha (1+t)^\beta, \qquad \alpha > -1, \quad \beta > -1,$$

with α, β depending on γ_0, γ_1 somehow. We can check that then also the other functions p_2, p_3, … are polynomials of the correct degree.

For $\alpha = \beta = 0$, we get, after an additional normalisation step,

$$\mathsf{P}_n(t) := \frac{(-1)^n}{2^n n!} \frac{\mathrm{d}^n}{\mathrm{d}t^n}(1-t^2)^n,$$

the LEGENDRE[15] polynomials. The weight function is $q \equiv 1$, and $\mathsf{P}_n = \frac{(-1)^n}{2^n n!} p_n$. The purpose of the additional factor is $\mathsf{P}_n(1) = 1$.

For $\alpha = \beta = -1/2$, we get, again with an additional normalisation step,

$$T_n(t) = \frac{(-1)^n 2^n}{(2n)!}(1-t^2)^{1/2} \frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big((1-t^2)^{n-1/2}\Big) = \cos(n \arccos t), \qquad -1 \le t \le 1,$$

[14] BENJAMIN OLINDE RODRIGUES, 1795–1851, french banker, mathematician, social reformer
[15] ADRIEN-MARIE LEGENDRE, 1752–1833

the CHEBYSHEV[16] polynomials. The weight function is $q(t) = (1-t^2)^{-1/2}$, and $T_n = \frac{(-1)^n 2^n}{(2n)!} p_n$.

General α, β would bring us the JACOBI[17] polynomials $P_n^{\alpha,\beta}$, by the way.

Let us write the formula for p_n as

$$p_n = \frac{1}{q} \frac{\mathrm{d}^n}{\mathrm{d}t^n} \Big(q r^n \Big), \qquad r(t) = 1 - t^2, \quad n = 0, 1, 2, \dots .$$

Now we will guess the differential equation for p_n. We know already from Section 6.4 that a differential operator

$$\frac{1}{q} L = \frac{1}{q} \frac{\mathrm{d}}{\mathrm{d}t} \left(p \frac{\mathrm{d}}{\mathrm{d}t} \ \cdot \right)$$

could be useful, with an unknown function $p = p(t)$. In any case, we then have

$$\frac{1}{q} L p_n = \frac{1}{q} \Big(p p_n' \Big)' = \frac{1}{q} (p' p_n' + p p_n''),$$

and now it seems reasonable to try $p = qr$. The advantage of this choice is that $p' = (qr)' = q p_1$ is a known function. Then we have

$$\frac{1}{q} L p_n = \frac{1}{q} \Big(q p_1 p_n' + q r p_n'' \Big) = r p_n'' + p_1 p_n',$$

and this is a polynomial of degree n, because r has degree 1, p_n'' has degree $n-2$, and p_1 has degree 1, p_n' has degree $n-1$. Hence we can decompose (remember that $(p_0, p_1, \dots)$ is an orthogonal basis of $\mathcal{H}$):

$$\frac{1}{q} L p_n = \sum_{j=0}^{n} \alpha_j p_j, \qquad \alpha_j = \frac{\left\langle \frac{1}{q} L p_n, p_j \right\rangle_{\mathcal{H}}}{\langle p_j, p_j \rangle_{\mathcal{H}}}, \qquad j = 0, 1, \dots, n.$$

The coefficients α_j are not yet known. However, with zeroes as markers for vanishing boundary terms,

$$\begin{aligned}
\left\langle \frac{1}{q} L p_n, p_j \right\rangle_{\mathcal{H}} &= \int_{t=a}^{b} (p p_n')' p_j \, \mathrm{d}t = \int_{t=a}^{b} (r q p_n')' p_j \, \mathrm{d}t \qquad \Big| \ \text{observe } r(a) = r(b) = 0 \\
&= 0 - \int_{t=a}^{b} r q p_n' p_j' \, \mathrm{d}t = - \int_{t=a}^{b} p_n' (r q p_j') \, \mathrm{d}t \qquad \Big| \ \text{observe } r(a) = r(b) = 0 \\
&= 0 + \int_{t=a}^{b} p_n (r q p_j')' \, \mathrm{d}t = \int_{t=a}^{b} (p p_j')' p_n \, \mathrm{d}t = \left\langle \frac{1}{q} L p_j, p_n \right\rangle_{\mathcal{H}},
\end{aligned}$$

and this must be zero for $j < n$ because $\frac{1}{q} L p_j$ is a polynomial of degree j, which is smaller than n, and Lemma 6.12 can be applied.

This gives us

$$\frac{1}{q} L p_n = \alpha_n p_n,$$

with some unknown number α_n. To determine α_n, we spell out the differential equation:

$$\begin{aligned}
&\frac{1}{q(t)} \frac{\mathrm{d}}{\mathrm{d}t} \left(q(t)(1-t^2) \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{1}{q(t)} \frac{\mathrm{d}^n}{\mathrm{d}t^n} \left(q(t)(1-t^2)^n \right) \right) \right) \\
&\qquad = \alpha_n \frac{1}{q(t)} \frac{\mathrm{d}^n}{\mathrm{d}t^n} \left(q(t)(1-t^2)^n \right),
\end{aligned}$$

and now the easiest (harrumph) way of finding α_n is to compare the highest powers of t on both sides.

After that computation, we find:

[16] PAFNUTY LVOVICH CHEBYSHEV, 1821–1894
[17] CARL GUSTAV JACOB JACOBI, 1804–1851

Proposition 6.13. *The Legendre polynomial* P_n *(*$n \geq 0$*) solves the BVP*

$$\begin{cases} (1-t^2)\mathsf{P}_n''(t) - 2t\mathsf{P}_n'(t) + n(n+1)\mathsf{P}_n(t) = 0, & -1 \leq t \leq 1. \end{cases}$$

The differential equation can also be written as $r(t)\mathsf{P}_n''(t) - 2\mathsf{P}_1(t)\mathsf{P}_n'(t) + n(n+1)\mathsf{P}_n(t) = 0$.

The Legendre Polynomials P_0, P_1, ... *form an orthogonal family in the Hilbert space* $L^2((-1,1);\,\mathrm{d}t)$*:*

$$\int_{t=-1}^{1} \mathsf{P}_n(t)\mathsf{P}_m(t)\,\mathrm{d}t = \frac{2}{2n+1}\delta_{nm}.$$

The Chebyshev polynomial T_n *(*$n \geq 0$*) solves the BVP*

$$\begin{cases} (1-t^2)T_n''(t) - tT_n'(t) + n^2T_n(t) = 0, & -1 \leq t \leq 1. \end{cases}$$

The differential equation can also be written as $r(t)T_n''(t) - T_1(t)T_n'(t) + n^2T_n(t) = 0$.

These polynomials T_0, T_1, ... *form an orthogonal family in the Hilbert space* $L^2((-1,1);(1-t^2)^{-1/2}\,\mathrm{d}t)$*:*

$$\int_{t=-1}^{1} T_n(t)T_m(t)\frac{1}{\sqrt{1-t^2}}\,\mathrm{d}t = \begin{cases} 0 & : n \neq m, \\ \pi & : n = m = 0, \\ \pi/2 & : n = m \neq 0. \end{cases}$$

We do not need boundary conditions for P_n or T_n because the coefficient $a_2(t) = 1-t^2 = r(t)$ vanishes on the boundary of the interval. Because of $a_2(t=a) = a_2(t=b) = 0$, these BVPs are not of Sturm–Liouville type.

Lemma 6.14. *On the space* $\mathcal{H} = L^2((-1,1))$*, the Legendre differential operator*

$$\Lambda_{\mathsf{P}} := (1-t^2)\frac{\mathrm{d}^2}{\mathrm{d}t^2} - 2t\frac{\mathrm{d}}{\mathrm{d}t}$$

has only the eigenvalues $-n(n+1)$*, (*$n \in \mathbb{N}_0$*), and no others.*

On the space $\mathcal{H} = L^2((-1,1);(1-t^2)^{-1/2}\,\mathrm{d}t)$*, the Chebyshev differential operator*

$$\Lambda_T := (1-t^2)\frac{\mathrm{d}^2}{\mathrm{d}t^2} - t\frac{\mathrm{d}}{\mathrm{d}t}$$

has only the eigenvalues $-n^2$*, (*$n \in \mathbb{N}_0$*), and no others.*

Proof. Assume $\Lambda_{\mathsf{P}}u = \lambda u$ for $u \neq 0$ and another number λ. Then this function u must be $\mathcal{H}$–orthogonal to any Legendre polynomial P_n. But these polynomials $(\mathsf{P}_0, \mathsf{P}_1, \dots)$ form a *complete* orthogonal system in $\mathcal{H}$, by Lemma 6.11. This is a contradiction. □

Take $(a,b) = (0,\infty)$ **as the interval.**

The scalar product of the space $\mathcal{H}$ is $\langle u,v\rangle_{\mathcal{H}} = \int_{t=0}^{\infty} u(t)v(t)q(t)\,\mathrm{d}t$, and the weight function q is assumed to decay exponentially for $t \to \infty$, and all its derivatives also decay exponentially for $t \to \infty$. Set, for $n = 0, 1, 2, \dots,$

$$p_n(t) := \frac{1}{q(t)}\frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big(q(t)t^n\Big), \qquad r(t) := t.$$

Lemma 6.15. *This function* p_n *is* $\mathcal{H}$*-perpendicular to each polynomial of degree* $< n$*.*

Proof. We check by n–fold partial integration that

$$\big\langle t^l, p_n\big\rangle_{\mathcal{H}} = \int_{t=0}^{\infty} t^l p_n(t)q(t)\,\mathrm{d}t = \int_{t=0}^{\infty} t^l \frac{\mathrm{d}^n}{\mathrm{d}t^n}\Big(q(t)t^n\Big)\,\mathrm{d}t = 0,$$

for $l = 0, 1, 2, \dots, n-1$. Boundary terms never appear due to $r(0) = 0$ and the fast decay of q at ∞. □

Next we try to find q such that each p_n is indeed a polynomial of degree n. Take $n = 1$ first. Then

$$p_1(t) = \frac{1}{q(t)}(q(t)t)' = 1 + \frac{tq'(t)}{q(t)} \stackrel{!}{=} \gamma_0 + \gamma_1 t,$$

for some γ_0, $\gamma_1 \in \mathbb{R}$. The function q shall decay for $t \to \infty$, hence $q'(t) < 0$ for large t, which makes positive γ_1 impossible. We can also take the freedom to scale the t–variable by a positive factor (this scaling maps the interval $(0, \infty)$ onto itself). By suitable scaling, we can make sure that $\gamma_1 = -1$. Then we find after some computation that

$$q(t) = t^\alpha e^{-t} \qquad \alpha > -1,$$

with α depending on γ_0 somehow. We can check that then also the other functions p_2, p_3, ... are polynomials of the correct degree. After an additional normalisation step, we get the LAGUERRE[18] polynomials,

$$L_{n,\alpha}(t) = \frac{1}{n!} t^{-\alpha} e^t \frac{\mathrm{d}^n}{\mathrm{d}t^n} \left(e^{-t} t^{n+\alpha} \right),$$

with $L_{n,\alpha} = \frac{1}{n!} p_n$. This normalisation makes the leading coefficient equal to $(-1)^n/n!$.

Let us write the formula for p_n as

$$p_n = \frac{1}{q} \frac{\mathrm{d}^n}{\mathrm{d}t^n} \Big(q r^n \Big), \qquad r(t) = t, \quad n = 0, 1, 2, \ldots.$$

And by the same computation as for the interval $(a, b) = (-1, 1)$, we find the differential operator

$$\frac{1}{q} L = \frac{1}{q} \frac{\mathrm{d}}{\mathrm{d}t} \left(p \frac{\mathrm{d}}{\mathrm{d}t} \ \cdot \right), \qquad p(t) = q(t) r(t) = t^{1+\alpha} e^{-t},$$

and the differential equation

$$\frac{1}{q} L p_n = r p_n'' + p_1 p_n' = \alpha_n p_n,$$

with some unknown number α_n, which can be determined by a lengthy calculation.

Proposition 6.16. *The Laguerre polynomial $L_{n,\alpha}$ ($n \geq 0$) solves the BVP*

$$\begin{cases} t L_{n,\alpha}''(t) + (\alpha + 1 - t) L_{n,\alpha}'(t) + n L_{n,\alpha}(t) = 0, & 0 < t < \infty. \end{cases}$$

The Laguerre Polynomials $L_{0,\alpha}$, $L_{1,\alpha}$, ... form a complete *orthogonal family in the Hilbert space $L^2((0, \infty); q(t)\,\mathrm{d}t)$:*

$$\int_{t=0}^{\infty} L_{n,\alpha}(t) L_{m,\alpha}(t) t^\alpha e^{-t}\,\mathrm{d}t = \frac{\Gamma(n + \alpha + 1)}{n!} \delta_{nm},$$

with $\Gamma(z) := \int_{s=0}^{\infty} s^{z-1} e^{-s}\,\mathrm{d}s$ as the well-known Gamma function.

Take $(a, b) = (-\infty, \infty)$ as the interval.

The scalar product of the space $\mathcal{H}$ is $\langle u, v \rangle_{\mathcal{H}} = \int_{t=-\infty}^{\infty} u(t) v(t) q(t)\,\mathrm{d}t$, and the weight function q is assumed to decay exponentially for $|t| \to \infty$, and all its derivatives also decay exponentially for $|t| \to \infty$. Set, for $n = 0, 1, 2, \ldots,$

$$p_n(t) := \frac{1}{q(t)} \frac{\mathrm{d}^n}{\mathrm{d}t^n} q(t), \qquad r(t) := 1.$$

Lemma 6.17. *This function p_n is $\mathcal{H}$-perpendicular to each polynomial of degree $< n$.*

[18] EDMOND NICOLAS LAGUERRE, 1834–1886

Proof. We check by n–fold partial integration that $\langle t^l, p_n \rangle_{\mathfrak{H}} = 0$, for $l = 0, 1, 2, \ldots, n-1$. □

Next we try to find q such that each p_n is indeed a polynomial of degree n. Take $n = 1$ first. Then

$$p_1(t) = \frac{q'(t)}{q(t)} \stackrel{!}{=} \gamma_0 + \gamma_1 t,$$

for some γ_0, $\gamma_1 \in \mathbb{R}$. The function q shall decay for $t \to +\infty$, which makes positive γ_1 impossible. We can also take the freedom to scale the t–variable by a positive factor, and to shift the t–variable by a constant. These transformations map the interval $(-\infty, \infty)$ onto itself. Then we can make sure that $\gamma_1 = -2$ and $\gamma_0 = 0$, which gives us quickly

$$q(t) = e^{-t^2}.$$

We can check that then also the other functions p_2, p_3, ... are polynomials of the correct degree.

Then we get the HERMITE[19] polynomials,

$$H_n(t) = (-1)^n e^{t^2} \frac{\mathrm{d}^n}{\mathrm{d}t^n} e^{-t^2},$$

with $H_n = (-1)^n p_n$. The coefficient of the highest power is 2^n. Pay attention: in several books, the terms $\exp(\pm t^2)$ are replaced by $\exp(\pm t^2/2)$.

By the standard calculation, we then find:

Proposition 6.18. *The Hermite polynomial H_n ($n \geq 0$) solves the BVP*

$$\begin{cases} H_n''(t) - 2tH_n'(t) + 2nH_n(t) = 0, & -\infty < t < \infty. \end{cases}$$

The Hermite Polynomials H_0, H_1, ... form a complete *orthogonal family in the Hilbert space $L^2((-\infty, \infty); q(t)\,\mathrm{d}t)$:*

$$\int_{t=-\infty}^{\infty} H_n(t) H_m(t) e^{-t^2}\,\mathrm{d}t = 2^n n! \sqrt{\pi} \delta_{nm}.$$

6.6 Applications of Orthogonal Polynomials

We begin with some considerations about **electrostatics**. In the three-dimensional space, introduce the Cartesian coordinates (x, y, z) and the polar coordinates,

$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r \cos\theta.$$

Put a unit charge at position $(0, 0, 1)$. This charge generates an electric field, whose potential $U = U(x, y, z)$ is

$$U(x, y, z) = \frac{1}{\|(x, y, z) - (0, 0, 1)\|} = \frac{1}{\sqrt{1 - 2r\cos\theta + r^2}} = U(r, \theta, \varphi),$$

by the Cosine theorem from school. On the other hand, we have $\triangle U = 0$ if we are not in the point $(0, 0, 1)$. Now the Laplace operator written in polar coordinates then gives

$$\partial_r^2 U(r, \theta, \varphi) + \frac{2}{r} \partial_r U(r, \theta, \varphi) + \frac{1}{r^2 \sin\theta} \partial_\theta \Big(\sin\theta \cdot \partial_\theta U(r, \theta, \varphi) \Big) + \frac{1}{r^2 \sin^2\theta} \partial_\varphi^2 U(r, \theta, \varphi) = 0.$$

By cylindrical symmetry, U certainly does not depend on φ, hence $U = U(r, \theta)$. We transform a variable: $\cos\theta = t \in [-1, 1]$,

$$\begin{aligned} \partial_\theta U(r, \theta) &= \partial_t U(r, t) \frac{\mathrm{d}t}{\mathrm{d}\theta} = -\sin\theta \partial_t U(r, t), \\ \sin\theta \partial_\theta U(r, \theta) &= -\sin^2\theta \partial_t U(r, t) = -(1 - t^2) \partial_t U(r, t), \end{aligned}$$

[19] CHARLES HERMITE, 1822–1901. The Hermite polynomials had been found by Chebyshev a few years earlier.

and then the Laplace equation becomes

$$\partial_r^2 U(r,t) + \frac{2}{r}\partial_r U(r,t) + \frac{1}{r^2}\partial_t\Big((1-t^2)\partial_t U(r,t)\Big) = 0.$$

Integrating the function $|U(x,y,z)|^2$ over the small ball $\{(x,y,z)\colon x^2+y^2+z^2 \le 1/4\}$ should certainly give a bounded value:

$$\infty > \iiint_{x^2+y^2+z^2\le 1/4} U^2(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = 2\pi\int_{r=0}^{1/2}\int_{\theta=0}^{\pi} |U(r,\theta)|^2 r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta,$$

but $\mathrm{d}t = -\sin\theta\,\mathrm{d}\theta$, hence

$$\int_{r=0}^{1/2}\int_{t=-1}^{1} |U(r,t)|^2 r^2\,\mathrm{d}r\,\mathrm{d}t < \infty.$$

And integrating $|\nabla U(x,y,z)|^2$ over this small ball should also give a bounded value, since ∇U is simply the electric field, which brings us (after some thinking) the condition

$$\int_{r=0}^{1/2}\int_{t=-1}^{1} |U_r(r,t)|^2 r^2\,\mathrm{d}r\,\mathrm{d}t < \infty.$$

Since a **complete** orthogonal system in $L^2((-1,1))$ is given by the Legendre polynomials, we can decompose $U(r,\cdot)$ for each r:

$$U(r,t) = \sum_{n=0}^{\infty} \mathsf{P}_n(t) R_n(r),$$

with some unknown functions R_n. By orthogonality of the P_n, we have

$$\infty > \int_{r=0}^{1/2}\int_{t=-1}^{1} |U(r,t)|^2 r^2\,\mathrm{d}r\,\mathrm{d}t = \sum_{n=0}^{\infty}\frac{2}{2n+1}\int_{r=0}^{1/2} |R_n(r)|^2 r^2\,\mathrm{d}r,$$
$$\infty > \int_{r=0}^{1/2}\int_{t=-1}^{1} |U_r(r,t)|^2 r^2\,\mathrm{d}r\,\mathrm{d}t = \sum_{n=0}^{\infty}\frac{2}{2n+1}\int_{r=0}^{1/2} |R_n'(r)|^2 r^2\,\mathrm{d}r.$$

We write the differential equation as

$$\partial_t\Big((1-t^2)\partial_t U(r,t)\Big) = -r^2\left(\partial_r^2 U(r,t) + \frac{2}{r}\partial_r U(r,t)\right),$$

and plugging the orthogonal series into both sides then gives

$$-\sum_{n=0}^{\infty} n(n+1)\mathsf{P}_n(t)R_n(r) = -\sum_{n=0}^{\infty}\mathsf{P}_n(t)\Big(r^2 R_n''(r) + 2rR_n'(r)\Big),$$

and by the linear independence of the Legendre functions, we then deduce that

$$r^2 R_n''(r) + 2rR_n'(r) = n(n+1)R_n(r),$$

which has the general solution

$$R_n(r) = c_{1,n}r^n + \frac{c_{2,n}}{r^{n+1}}.$$

If $c_{2,n} \neq 0$, then R_n has a pole at $r=0$, which violates the above boundedness condition on the integral of $|R_n'|^2$. This brings us the identity

$$\frac{1}{\sqrt{1-2r\cos\theta+r^2}} = \sum_{n=0}^{\infty} c_{1,n}\mathsf{P}_n(\cos\theta)r^n,$$

and the $c_{1,n}$ are not yet known. Set $\theta = 0$ on both sides:

$$\frac{1}{1-r} = \sum_{n=0}^{\infty} c_{1,n}\mathsf{P}_n(1)r^n.$$

We know $\mathsf{P}_n(1) = 1$ for all n, and then the formula for the geometric series gives us $c_{1,n} = 1$ for all n.

Theorem 6.19 (**Generating Functions**). [20] *The Legendre polynomials* P_n *satisfy*

$$\frac{1}{\sqrt{1-2rt+r^2}} = \sum_{n=0}^{\infty} \mathsf{P}_n(t) r^n, \qquad -1 \le t \le 1, \qquad 0 \le r \le 1/2.$$

The Chebyshev polynomials T_n *satisfy*

$$\frac{1-r^2}{1-2rt+r^2} + 1 = 2\sum_{n=0}^{\infty} T_n(t) r^n, \qquad -1 < t < 1, \qquad |r| < 1.$$

The Laguerre polynomials $L_{n,\alpha}$ *satisfy*

$$(1-r)^{-\alpha-1} \exp\left(-\frac{tr}{1-r}\right) = \sum_{n=0}^{\infty} L_{n,\alpha}(t) r^n, \qquad 0 < t < \infty, \quad |r| < 1.$$

The Hermite polynomials H_n *satisfy*

$$\exp(2tr - r^2) = \sum_{n=0}^{\infty} \frac{1}{n!} H_n(t) r^n, \qquad -\infty < t, r < \infty.$$

Many important properties of orthogonal polynomials can be proved using such generating functions. As an example, the generating function $U(r,t) = (1-2rt+r^2)^{-1/2}$ of the Legendre polynomials solves the differential equation

$$(1-2rt+r^2)\partial_r U(r,t) + (r-t)U(r,t) = 0,$$

and plugging the power series $\sum_{n=0}^{\infty} \mathsf{P}_n(t) r^n$ into this equation, and equating corresponding powers of r then gives us the important recursion formula

$$(n+1)\mathsf{P}_{n+1}(t) - (2n+1)t\mathsf{P}_n(t) + n\mathsf{P}_{n-1}(t) = 0,$$

which permits us to find formulae for P_n with n large, avoiding the Rodrigues formula which quickly becomes inconvenient for larger n.

The next application comes from **quantum mechanics**. The momentum operator p is quantised as $p = \frac{\hbar}{\mathrm{i}}\nabla$, and then the Hamiltonian H turns into

$$H = \frac{p^2}{2m} + V(x) = -\frac{\hbar^2}{2m} \triangle + V(x).$$

The stationary (time independent) Schrödinger equation reads $H\psi = E\psi$, with the real number E as the energy level, and $\psi = \psi(x)$ as the wave function, whose square $|\psi(x)|^2$ describes the probability density to find the particle at the position $x \in \mathbb{R}^d$. Clearly,

$$\int_{x\in\mathbb{R}^d} |\psi(x)|^2 \,\mathrm{d}x = 1,$$

because the particle must be *somewhere*.

Now we simplify: $x \in \mathbb{R}^1$, $V(x) = x^2$, $\hbar^2/2m = 1$, and obtain the problem

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)\psi(x) = E\psi(x), \qquad -\infty < x < \infty.$$

The boundary conditions on ψ are contained implicitly in the restriction $\psi \in L^2((-\infty,\infty))$, which is only possible of $\psi(x)$ decays for $|x| \to \infty$.

[20] erzeugende Funktionen

The function ψ belongs to the non-weighted space $L^2((-\infty,\infty))$ if and only if the function $u = u(x) = \exp(x^2/2)\psi(x)$ belongs to the weighted space $L^2((-\infty,\infty), \exp(-x^2)\,\mathrm{d}x)$, because of

$$\int_{x=-\infty}^{\infty} |\psi(x)|^2\,\mathrm{d}x = \int_{x=-\infty}^{\infty} \left|e^{x^2/2}\psi(x)\right|^2 e^{-x^2}\,\mathrm{d}x.$$

Translating the differential equation for ψ into a differential equation for u, we find

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) &= \left(u(x)e^{-x^2/2}\right)'' = u''(x)e^{-x^2/2} + 2u'(x)\cdot(-x)e^{-x^2/2} + u(x)\cdot e^{-x^2/2}(x^2-1)\\ &\stackrel{!}{=} (x^2-E)\psi(x) = (x^2-E)e^{-x^2/2}u(x),\end{aligned}$$

and then the problem for u becomes

$$\begin{cases} u''(x) - 2xu'(x) + (E-1)u(x) = 0, & -\infty < x < \infty,\\ u \in L^2((-\infty,\infty), \exp(-x^2)\,\mathrm{d}x). \end{cases}$$

From the theory of Hermite polynomials we know that this problem has a non-zero solution u if and only if $E - 1 = 2n$ for some $n \in \mathbb{N}_0$, and then u is given by

$$u(x) = cH_n(x),$$

with $c \neq 0$ as a constant and H_n as the Hermite polynomial.

As a summary: only certain energy levels $E_n = 2n+1$ with $n \in \mathbb{N}_0$ are admissible for a quantum mechanical particle in the harmonic oscillator potential. In particular, the lowest energy (corresponding to $n = 0$) is *not* at the bottom of the potential V, in difference to the classical mechanics.

And finally, we wish to understand mathematically why there are at most two electrons in the s sub-shell, at most 6 electrons in the p sub-shell, at most 10 in the d sub-shell, and at most 14 in the f sub-shell.

Before we start, let us recall which orthogonal basises in function spaces of L^2 type we know: if we look at non-weighted spaces $L^2((a,b))$ over a finite interval, we know already

- a pure sine family (6.17),
- a pure cosine family (6.18),
- a family consisting of sine and cosine functions together (remember the theory of Fourier series from the second semester),
- the family of Legendre polynomials,
- the family $((1-t^2)^{-1/4}T_0, (1-t^2)^{-1/4}T_1, \dots)$, with T_j as the Chebyshev polynomial.

On the non-weighted space $L^2((-\infty,\infty))$, we have

- the family $(e^{-t^2/2}H_0(t), e^{-t^2/2}H_1(t), \dots)$, with H_j as the Hermite polynomial.

And on the half-unbounded interval $(0,\infty)$, we have a large number of basises (one for each $\alpha > -1$ via the Laguerre polynomials).

Recall that all these functions are the eigenfunctions to a self-adjoint second order differential operator.

At first glance, this does not help us so much when attacking the electrons in the atomic shells because intervals are one-dimensional objects, but the electron shells are not.

Define $S := \{(x,y,z) \in \mathbb{R}^3 \colon x^2+y^2+z^2 = 1\}$ and call it the *unit sphere*. The Laplace operator in polar coordinates is

$$\begin{aligned}\triangle U(r,\vartheta,\varphi) &= \frac{1}{r^2}\partial_r\Big(r^2\partial_r U(r,\vartheta,\varphi)\Big) + \frac{1}{r^2\sin\vartheta}\partial_\vartheta\Big(\sin\vartheta\partial_\vartheta U(r,\vartheta,\varphi)\Big) + \frac{1}{r^2\sin^2\vartheta}\partial_\varphi^2 U(r,\vartheta,\varphi)\\ &=: \triangle_r U(r,\vartheta,\varphi) + \frac{1}{r^2}\triangle_S U(r,\vartheta,\varphi),\end{aligned}$$

where $\triangle_r$ only contains derivatives with respect to r, and $\triangle_S$ only contains derivatives with respect to the angles (ϑ,φ).

Definition 6.20 (**Laplace–Beltrami operator**). *The operator*

$$\triangle_S := \frac{1}{\sin\vartheta}\partial_\vartheta\Big(\sin\vartheta\partial_\vartheta \quad \cdot\Big) + \frac{1}{\sin^2\vartheta}\partial_\varphi^2$$

is called the Laplace–Beltrami[21] operator.

This is an operator which differentiates functions which live on the unit sphere S.

In the same way as in Section 6.4, we then perform the following steps:

- call $\mathcal{H} := L^2(S)$ the ground space,
- define the domain $D(\triangle_S)$ as the set of all those functions $u \in \mathcal{H}$ for which $\triangle_S u \in \mathcal{H}$,
- prove that $D(\triangle_S^*) = D(\triangle_S)$ and $\triangle_S = \triangle_S^*$,
- prove that eigenfunctions of $\triangle_S$ to different eigenvalues are orthogonal with respect to the scalar product in $\mathcal{H} = L^2(S)$ (this is quite easy),
- prove that the eigenfunctions of $\triangle_S$ form a *complete* orthogonal system of $L^2(S)$ (this is really hard).

This orthogonal system will then be extremely helpful in understanding the behaviour of the electrons. Therefore we need to understand how the eigenfunctions of $\triangle_S$ look like.

We start with $\triangle_S u(\vartheta,\varphi) = \lambda u(\vartheta,\varphi)$ for a real number λ. Clearly, the function u is 2π–periodic with respect to φ, which makes a Fourier expansion possible:

$$u(\vartheta,\varphi) = \sum_{m\in\mathbb{Z}} e^{\mathrm{i}m\varphi}\Theta_m(\vartheta).$$

Plugging this into $\lambda u = \triangle_S u$ then gives us

$$\lambda\sum_{m\in\mathbb{Z}} e^{\mathrm{i}m\varphi}\Theta_m(\vartheta) = \lambda u(\vartheta,\varphi) = \triangle_S u(\vartheta,\varphi) = \sum_{m\in\mathbb{Z}} e^{\mathrm{i}m\varphi}\left(\frac{1}{\sin\vartheta}\partial_\vartheta\Big(\sin\vartheta\partial_\vartheta\Theta_m(\vartheta)\Big) - \frac{m^2}{\sin^2\vartheta}\Theta_m(\vartheta)\right),$$

$$\lambda\Theta_m(\vartheta) = \frac{1}{\sin\vartheta}\partial_\vartheta\Big(\sin\vartheta\partial_\vartheta\Theta_m(\vartheta)\Big) - \frac{m^2}{\sin^2\vartheta}\Theta_m(\vartheta) \qquad \Big| \quad t = \cos\vartheta, \quad \Theta_m(\vartheta) =: T_m(t),$$

$$\lambda T_m(t) = \partial_t\Big((1-t^2)\partial_t T_m(t)\Big) - \frac{m^2}{1-t^2}T_m(t),$$

$$(1-t^2)T_m''(t) - 2tT_m'(t) + \left(-\lambda - \frac{m^2}{1-t^2}\right)T_m(t) = 0, \qquad -1 < t < 1,$$

and the side condition is that $T_m \in L^2((-1,1))$.

Using methods from *complex analysis*[22] one can show that the function T_m is without poles if and only if

$$\lambda = -l(l+1), \qquad l \in \mathbb{N}_0, \qquad m \in \{-l, -l+1, \ldots, l-1, l\},$$

and in such a case, the solution T_m is a multiple of the *associated Legendre polynomial*:

$$T_m(t) = c\mathsf{P}_l^m(t) := c\begin{cases}(-1)^m(1-t^2)^m\frac{\mathrm{d}^m}{\mathrm{d}t^m}\mathsf{P}_l(t) & : m \geq 0,\\ (-1)^m\frac{(l+m)!}{(l-m)!}\mathsf{P}_l^{-m}(t) & : m < 0.\end{cases}$$

Therefore, we have shown that

$$u(\vartheta,\varphi) = \sum_{m=-l}^{l} e^{\mathrm{i}m\varphi}c_m\mathsf{P}_l^m(\cos\vartheta),$$

where $l \in \mathbb{N}_0$ corresponds to λ via $\lambda = -l(l+1)$.

Hence we have proved:

[21] Eugenio Beltrami, 1835–1900

[22] Funktionentheorie

Lemma 6.21. *The eigenfunctions to $\triangle_S$ are the* spherical harmonics[23]

$$Y_{lm}(\vartheta,\varphi) = e^{\mathrm{i}m\varphi}\mathsf{P}_l^m(\cos\vartheta), \qquad l = 0,1,2,\ldots, \qquad m = -l,-l+1,\ldots,l-1,l.$$

and we have $\triangle_S Y_{lm} = -l(l+1)Y_{lm}$ as well as

$$\langle Y_{lm}, Y_{l'm'}\rangle_{L^2(S)} = \delta_{ll'}\delta_{mm'}c_{lm},$$

with a positive constant c_{lm} (which differs from book to book). These eigenfunctions form an orthogonal basis of $L^2(S)$.

After these preparations, we can now attack the (heavily simplified) Schrödinger equation of an electron in the Coulomb potential generated by the atomic nucleus:

$$\left(-\triangle - \frac{2}{\|x\|}\right)\psi(x) = E\psi(x),$$

where $x \in \mathbb{R}^3$, E is the (negative) energy, and $\psi \in L^2(\mathbb{R}^3)$ is the wave function.

Introduce polar coordinates:

$$\left(-\triangle_r - \frac{1}{r^2}\triangle_S - \frac{2}{r}\right)\psi(r,\vartheta,\varphi) = E\psi(r,\vartheta,\varphi).$$

For each fixed $r > 0$, we can decompose ψ using the spherical harmonics:

$$\psi(r,\vartheta,\varphi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_{lm}(\vartheta,\varphi)R_{lm}(r),$$

with unknown functions R_{lm}. The condition

$$\int_{\mathbb{R}^3} |\psi(x)|^2\,\mathrm{d}x = 1 < \infty$$

then turns into (by orthogonality of the Y_{lm})

$$\sum_{l,m} c_{lm}\int_{r=0}^{\infty} |R_{lm}(r)|^2 r^2\,\mathrm{d}r < \infty,$$

for some constants c_{lm} coming from the normalisation convention of the Y_{lm}. From this we learn that R_{lm} can not have a strong pole at $r = 0$.

Plugging the decomposition of ψ into the Schrödinger equation then gives

$$\sum_{l,m} Y_{lm}\left(-\triangle_r R_{lm} + \frac{1}{r^2}l(l+1)R_{lm} - \frac{2}{r}R_{lm}\right) = \sum_{l,m} Y_{lm}ER_{lm},$$

(commuting $\triangle$ and $\sum$ is permitted if U has sufficiently many continuous derivatives with respect to x,y,z), and comparing corresponding Y_{lm} then implies

$$-\frac{1}{r^2}\partial_r\Big(r^2\partial_r R_{lm}(r)\Big) + \frac{1}{r^2}l(l+1)R_{lm} - \frac{2}{r}R_{lm}(r) = ER_{lm}(r).$$

By the physical assumption $E < 0$, we can write $E = -K^2$ for some positive K, and we obtain

$$R''_{lm}(r) + \frac{2}{r}R'_{lm}(r) + \left(-K^2 + \frac{2}{r} - \frac{l(l+1)}{r^2}\right)R_{lm}(r) = 0.$$

Using (advanced) methods from complex analysis, one can show that R_{lm} *must* behave like r^l for $r \to 0$. Hence we can set $R_{lm}(r) = r^l Q_{lm}(r)$, for some function Q_{lm} which is bounded near $r = 0$. Then we find

$$\begin{aligned}
R'_{lm}(r) &= r^l Q'_{lm}(r) + lr^{l-1}Q_{lm}(r),\\
R''_{lm}(r) &= r^l Q''_{lm}(r) + 2lr^{l-1}Q'_{lm}(r) + l(l-1)r^{l-2}Q_{lm}(r),\\
R''_{lm}(r) + \frac{2}{r}R'_{lm}(r) &= r^l Q''_{lm}(r) + (2l+2)r^{l-1}Q'_{lm}(r) + l(l+1)r^{l-2}Q_{lm}(r),
\end{aligned}$$

[23] Kugelflächenfunktionen

and the differential equation turns into

$$Q''_{lm}(r) + \frac{2l+2}{r} Q'_{lm}(r) + \left(-K^2 + \frac{2}{r}\right) Q_{lm}(r) = 0.$$

For large r, the relevant terms in the left-hand side are $Q''_{lm} - K^2 Q_{lm}$ (this conclusion is mathematically not very rigorous, but we do not have the (advanced) tools from complex analysis to do it better), and therefore Q_{lm} is expected to behave like $c_+ e^{+Kr} + c_- e^{-Kr}$ for large r. Then $\psi \in L^2(\mathbb{R}^3)$ is possible only if $c_+ = 0$. Then the electron in the field of the hydrogen atom can only have special amounts of energy, but never an energy between these energy levels. Therefore we split off a factor e^{-Kr}:

$$\begin{aligned}
Q_{lm}(r) &=: e^{-Kr} Z_{lm}(r), \\
Q'_{lm}(r) &= e^{-Kr} Z'_{lm}(r) - K e^{-Kr} Z_{lm}(r), \\
Q''_{lm}(r) &= e^{-Kr} Z''_{lm}(r) - 2K e^{-Kr} Z'_{lm}(r) + K^2 e^{-Kr} Z_{lm}(r),
\end{aligned}$$

which brings us to

$$Z''_{lm}(r) + \left(\frac{2l+2}{r} - 2K\right) Z'_{lm}(r) + \left(\frac{2l+2}{r} \cdot (-K) + \frac{2}{r}\right) Z_{lm}(r) = 0.$$

We need polynomial solutions to this equation. One can show that if Z_{lm} is a non-polynomial solution, then for large r the terms $Z''_{lm} - 2K Z'_{lm}$ are the main terms, giving us a behaviour $Z_{lm}(r) \sim \exp(2Kr)$ for large r, which makes the function Q_{lm} explode exponentially for $r \to \infty$. This violates the condition that $\psi \in L^2(\mathbb{R}^3)$.

We need just one last scaling step to obtain a Laguerre differential equation:

$$\begin{aligned}
r &=: \frac{s}{2K}, \qquad Z_{lm}(r) =: S_{lm}(s), \\
Z'_{lm}(r) &= 2K S'_{lm}(s), \\
Z''_{lm}(r) &= 4K^2 S''_{lm}(s),
\end{aligned}$$

$$\begin{aligned}
S''_{lm}(s) + \left(\frac{2l+2}{s} - 1\right) S'_{lm}(s) + \frac{2(1-(l+1)K)}{2Ks} S_{lm}(s) = 0, \\
s S''_{lm}(s) + (2l+2-s) S'_{lm}(s) + \left(\frac{1}{K} - (l+1)\right) S_{lm}(s) = 0.
\end{aligned}$$

The theory of Laguerre polynomials tells us that a polynomial solution $L_{j,\alpha}$ with

$$j = \frac{1}{K} - (l+1), \qquad \alpha = 2l+1$$

exists only if j is a natural number, which brings us to $\frac{1}{K} = j + l + 1$ for $j \in \mathbb{N}_0$. Then the admissible energy levels are

$$E = -\frac{1}{(j+l+1)^2}.$$

Now the quantum numbers are chosen as follows:

- first a *principal quantum number*[24] $n \in \mathbb{N}_+$ is selected (this corresponds to $j + l + 1$),
- then the *azimuthal quantum number*[25] $j \in \mathbb{N}_0$ is selected, subject to the restriction $0 \le j \le n-1$ (this determines $l \in \mathbb{N}_0$ via $l = n - j - 1$),
- then the *magnetic quantum number*[26] $m \in \mathbb{Z}$ can be chosen, subject to $-l \le m \le l$.

[24] Hauptquantenzahl
[25] Nebenquantenzahl
[26] Magnetquantenzahl

The energy depends only on n. The (s, p, d, f) sub-shells correspond to $l \in (0, 1, 2, 3)$. Answering the question from the beginning is now only a counting exercise. Do not forget the spin. For fixed n, all the states have the same energy $-\frac{1}{n^2}$.

At the very end of this part, we comment on the mysterious condition $E < 0$. This condition describes the outer radius of the *atomic shell*[27]. An electron with energy $E > 0$ is no longer bound to the atomic nucleus, can move around freely, and can have any amount of energy. An electron with negative energy E can only possess special amounts of energy, namely $-1/n^2$ for some $n \in \mathbb{N}_+$.

This corresponds mathematically to the Hamiltonian operator $H = -\triangle - \frac{2}{\|x\|}$ as follows:

$E < 0$, and $E = -1/n^2$: then the operator $H - E\colon D(H) \to L^2(\mathbb{R}^3)$ is not injective (all these numbers E form the *discrete spectrum*[28] of the operator E),

$E < 0$, and $E \neq -1/n^2$ for all n: then the operator $H - E\colon D(H) \to L^2(\mathbb{R}^3)$ is bijective, continuous, and its inverse map is also continuous (all these numbers E, and also the non-real numbers E, form the *resolvent set*[29] of the operator H),

$E > 0$: then the operator $H - E\colon D(H) \to L^2(\mathbb{R}^3)$ is injective, but not surjective (all these numbers E form the *continuous spectrum*[30] of the operator H).

Recall that in a finite-dimensional vector space U, a linear map $A\colon U \to U$ is injective if and only if it is surjective, by the dimension formula for linear maps. Therefore, for maps of a finite-dimensional space U into itself, the adjectives "injective" and "surjective" are logically equivalent (although they mean different things, of course).

As you can see, this is no longer true for maps in spaces of infinite dimension.

[27] Atomhülle
[28] diskretes Spektrum
[29] Resolventenmenge
[30] stetiges Spektrum

Part II

Complex Analysis (Funktionentheorie)

Chapter 7

Holomorphic Functions

7.1 Back to the Roots

Soon we will see that complex differentiable functions behave in a completely different manner when compared to real differentiable functions, and in order to understand the deeper reason, we ask what complex numbers are.

In the first semester, complex numbers z had been defined as pairs of real numbers, $z = (x, y)$, and an addition, multiplication had been specified like this:

$$(x, y) +_{\mathbb{C}} (u, v) := (x + u, y + v),$$
$$(x, y) \cdot_{\mathbb{C}} (u, v) := (x \cdot u - y \cdot v, x \cdot v + y \cdot u).$$

The addition looks like the addition of vectors in the vector space $\mathbb{R}^2$, and to find an interpretation of the multiplication, we write

$$z = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad w = \begin{pmatrix} u \\ v \end{pmatrix}, \qquad z \cdot_{\mathbb{C}} w = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u & -v \\ v & u \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

which brings us to the second way of writing complex numbers: instead of $z = (x, y)$ or $z = \binom{x}{y}$, we write

$$z = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}, \tag{7.1}$$

and then adding and multiplying in $\mathbb{C}$ correspond to adding and multiplying matrices. Note that

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} + \begin{pmatrix} u & -v \\ v & u \end{pmatrix} = \begin{pmatrix} (x+u) & -(y+v) \\ (y+v) & (x+u) \end{pmatrix},$$
$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \cdot \begin{pmatrix} u & -v \\ v & u \end{pmatrix} = \begin{pmatrix} (xu - yv) & -(yu + xv) \\ (yu + xv) & (xu - yv) \end{pmatrix},$$

and the right-hand sides are again matrices of the correct structure.

And the third way of writing complex numbers is, of course, $z = x + \mathrm{i}y$ with $\mathrm{i}^2 = -1$ instead of $z = (x, y)$.

We may identify $\mathbb{C} \simeq \mathbb{R}^2$, but the complex multiplication introduces an additional structure into $\mathbb{R}^2$.

The set $\mathbb{C}$ can be seen as a two-dimensional vector space $\mathbb{R}^2$ over the field $K = \mathbb{R}$, with canonical basis

$$\Big\{1_{\mathbb{C}} := (1_{\mathbb{R}}, 0_{\mathbb{R}}), \quad \mathrm{i} := (0_{\mathbb{R}}, 1_{\mathbb{R}})\Big\},$$

or as a one-dimensional vector space $\mathbb{C}^1$ over the field $K = \mathbb{C}$, with the canonical basis

$$\Big\{1_{\mathbb{C}}\Big\}.$$

Here $1_{\mathbb{C}}$ means the number one, understood as a complex number.

Recall that a map $T\colon U \to U$ of a K–vector space U into itself is *linear* if

- it is additive: $T(u_1 + u_2) = T(u_1) + T(u_2)$, for all u_1, $u_2 \in U$,
- it is K–homogeneous: $T(\lambda u) = \lambda T(u)$ for all $u \in U$ and all $\lambda \in K$.

Lemma 7.1.

- *Each $\mathbb{C}$–linear map from $\mathbb{C}$ into $\mathbb{C}$ is also $\mathbb{R}$–linear.*
- *Each $\mathbb{R}$–linear map from $\mathbb{C}$ into $\mathbb{C}$ with $T(\mathrm{i}) = \mathrm{i}T(1_\mathbb{C})$ is also $\mathbb{C}$–linear.*

Proof.

- This is clear since each $\mathbb{C}$–homogeneous map is $\mathbb{R}$–homogeneous, because $\mathbb{R} \subset \mathbb{C}$.
- Let $T\colon \mathbb{C} \to \mathbb{C}$ be $\mathbb{R}$–linear. Then T is additive, and it remains to show that $T(\lambda u) = \lambda T(u)$ for all $\lambda \in \mathbb{C}$ and all $u \in \mathbb{C}$. To this end, we show $T(z) = zT(1_\mathbb{C})$ for all $z \in \mathbb{C}$. Write $z = x + \mathrm{i}y$ with real x, y. Then

$$\begin{aligned} T(z) &= T(x + \mathrm{i}y) && \mid\ T \text{ is additive} \\ &= T(x) + T(\mathrm{i}y) = T(x \cdot 1_\mathbb{C}) + T(y \cdot \mathrm{i}) && \mid\ T \text{ is } \mathbb{R}\text{–linear} \\ &= x \cdot T(1_\mathbb{C}) + y \cdot T(\mathrm{i}) && \mid\ \text{assumption} \\ &= x \cdot T(1_\mathbb{C}) + y\mathrm{i} \cdot T(1_\mathbb{C}) = (x + \mathrm{i}y)T(1_\mathbb{C}) = z \cdot T(1_\mathbb{C}). \end{aligned}$$

  Now we can argue like this:

$$T(\lambda \cdot u) = T((\lambda \cdot u) \cdot 1_\mathbb{C}) = (\lambda \cdot u) \cdot T(1_\mathbb{C}) = \lambda \cdot (u \cdot T(1_\mathbb{C})) = \lambda \cdot T(u),$$

  for all $\lambda \in \mathbb{C}$ and all $u \in \mathbb{C}$.

□

From the first semester, we know that each K–linear map $T\colon K^n \to K^n$ can be represented by a matrix from $K^{n\times n}$. Therefore, each $\mathbb{R}$–linear map $T\colon \mathbb{C} \to \mathbb{C}$ can be written using a 2×2 matrix,

$$T(z) = T(x + \mathrm{i}y) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad a_{jk} \in \mathbb{R}.$$

We compute some values:

$$T(1_\mathbb{C}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = a_{11} + \mathrm{i}a_{21}, \qquad T(\mathrm{i}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = a_{12} + \mathrm{i}a_{22}.$$

Such a map is additionally $\mathbb{C}$–linear if $T(\mathrm{i}) = \mathrm{i}T(1_\mathbb{C})$, which implies $a_{12} + \mathrm{i}a_{22} = \mathrm{i}(a_{11} + \mathrm{i}a_{21})$, or

$$a_{11} = a_{22}, \quad a_{12} = -a_{21},$$

which corresponds nicely to the special matrix structure from (7.1). We summarise:

Lemma 7.2. *A map $T\colon \mathbb{C} \to \mathbb{C}$ is $\mathbb{C}$–linear if and only if $T(z) = A\binom{x}{y}$ (where $z = x + \mathrm{i}y = \binom{x}{y}$ with real x, y) for a matrix $A \in \mathbb{R}^{2\times 2}$ with $a_{11} = a_{22}$ and $a_{12} = -a_{21}$.*

Next we recall polar coordinates. Each $z \in \mathbb{C}$ can be represented as

$$z = r(\cos\varphi + \mathrm{i}\sin\varphi), \qquad r \geq 0, \qquad \varphi \in \mathbb{R},$$

and the angle φ is called an *argument* of z, written as

$$\varphi = \arg z.$$

The argument of z is not uniquely determined; you can always add or subtract multiples of 2π.

The formula of DE MOIVRE[1] is

$$\Big(r(\cos\varphi + \mathrm{i}\sin\varphi)\Big)^n = r^n\big(\cos(n\varphi) + \mathrm{i}\sin(n\varphi)\big), \qquad n \in \mathbb{N}_+,$$

and this tells us how to compute the n–th roots of $w \in \mathbb{C}$: a number $z \in \mathbb{C}$ is an n–th root of w if and only if

$$|z| = \sqrt[n]{|w|}, \qquad \arg z \in \frac{1}{n}\arg w + \frac{2\pi}{n}\mathbb{Z},$$

where $\arg w$ is one special argument of w.

To each $w \neq 0$, there are exactly n distinct numbers $z \in \mathbb{C}$ with $z^n = w$, called *the n–th roots of* w, and each of these roots is obtained from another one of these roots by multiplication by a suitably chosen n–th root of the number one.

Finally, we come to the exponential function. For $z = x + \mathrm{i}y$ with real x, y, we have

$$\exp(z) = \exp(x)\exp(\mathrm{i}y) = e^x(\cos y + \mathrm{i}\sin y),$$

hence exp maps the horizontal strip S of width 2π,

$$S := \{(x,y) \in \mathbb{C}\colon\ -\pi < y \leq \pi\},$$

bijectively onto $\mathbb{C}\setminus\{0\}$. The upper boundary $\{(x,y)\colon y = \pi\}$ is mapped onto the northern riverbank of $\mathbb{R}_- := (-\infty, 0)\times\{0\}$, and the lower boundary $\{(x,y)\colon y = -\pi\}$ is mapped onto the southern riverbank of $\mathbb{R}_-$.

Definition 7.3 (Principal branch of complex logarithm[2]). *The inverse function of* $\exp\colon S \to \mathbb{C}\setminus\{0\}$ *is called* Ln, *the* principal branch of the complex logarithm. *If* $w = |w|(\cos\varphi + \mathrm{i}\sin\varphi)$ *with* $w \neq 0$ *and* $-\pi < \varphi \leq \pi$, *then*

$$\operatorname{Ln} w = \ln|w| + \mathrm{i}\varphi,$$

where $\ln\colon \mathbb{R}_+ \to \mathbb{R}$ *is the traditional logarithm from school.*

Note that $\operatorname{Ln}\colon \mathbb{C}\setminus\{0\} \to S$ has a jump of height $2\pi\mathrm{i}$ when we cross $\mathbb{R}_-$.

Another possibility of defining a logarithm function on $\mathbb{C}\setminus\{0\}$ is to select an angle α, e.g., $\alpha = \pi/137$, to define

$$S_\alpha := \{(x,y) \in \mathbb{C}\colon\ -\pi + \alpha < y \leq \pi + \alpha\},$$

and to each $w \neq 0$, there is exactly one number $\varphi \in (-\pi + \alpha, \pi + \alpha]$ with $\arg w = \varphi$. Then we may define

$$\ln w := \ln|w| + \mathrm{i}\varphi.$$

Observe that this function $\ln\colon \mathbb{C}\setminus\{0\} \to S_\alpha$ has a jump of height $2\pi\mathrm{i}$ when we cross the ray $\{w\colon \arg w = \pi + \alpha\}$.

Definition 7.4 (Principal branch of complex root function). *If* $w = |w|(\cos\varphi + \mathrm{i}\sin\varphi)$ *with* $-\pi < \varphi \leq \pi$, *then we define*

$$\sqrt{w} := \sqrt{|w|}\cdot(\cos(\varphi/2) + \mathrm{i}\sin(\varphi/2)),$$

with $\sqrt{|w|}$ *as the traditional (nonnegative, of course) root of a nonnegative number, as in school.*

Warning: *In general,* $\operatorname{Ln}(z_1 z_2) \neq \operatorname{Ln} z_1 + \operatorname{Ln} z_2$ *and* $\sqrt{z_1 z_2} \neq \sqrt{z_1}\sqrt{z_2}$. *Find examples yourself.*

[1] ABRAHAM DE MOIVRE, 1667–1754

[2] Hauptzweig des komplexen Logarithmus

7.2 Differentiation

We recall from the first year how to define a derivative of a function $f\colon K^m \to K^n$:

- the function $f\colon K^m \to K^n$ is differentiable at a point $x_0 \in K^m$ if a matrix $A \in K^{n\times m}$ exists with

  $$f(x) = f(x_0) + A(x - x_0) + \mathfrak{o}(\|x - x_0\|)$$

  for $x \to x_0$,
- if $m = 1$: the function $f\colon K^1 \to K^n$ is differentiable at a point $x_0 \in K^1$ if the limit

  $$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

  exists, and then this limit equals $A \in K^{n\times 1}$ from the first $\bullet$.

Now we will follow both approaches (keeping in mind that $\mathbb{C} \simeq \mathbb{R}^2$), and comparing the results will then bring us new insights.

Definition 7.5. *Let $\Omega \subset \mathbb{C}$ be an open, non-empty set.*

- *A function $f\colon \Omega \to \mathbb{C}$ is* complex differentiable *in $z_0 \in \Omega$ if the limit*

  $$f'(z_0) = \lim_{z\to z_0} \frac{f(z) - f(z_0)}{z - z_0} \in \mathbb{C}$$

  exists.
- *A function $f\colon \Omega \to \mathbb{C}$ is* holomorphic in Ω[3] *if f is complex differentiable at each $z_0 \in \Omega$.*
- *A function $f\colon \Omega \to \mathbb{C}$ is* holomorphic in $z_0 \in \Omega$ *if an open neighbourhood U of z_0 exists ($z_0 \in U \subset \Omega$) such that f (restricted to U) is holomorphic in U.*
- *A function $f\colon \mathbb{C} \to \mathbb{C}$ is an* entire function[4] *if it is holomorphic in $\mathbb{C}$.*

Lemma 7.6. *Each complex differentiable function at a point z_0 is continuous at z_0.*

The rules for differentiating sums, products, compositions of differentiable functions hold in $\mathbb{C}$ as they do in $\mathbb{R}$.

Because the proofs are the same in $\mathbb{C}$ as in $\mathbb{R}$.

Lemma 7.7. *If f is given as a power series,*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_*)^n, \qquad a_n \in \mathbb{C},$$

which converges in the open ball

$$B_R(z_*) := \{z \in \mathbb{C}\colon |z - z_*| < R\},$$

then f is holomorphic in the ball $B_R(z_*)$, the derivative is found by term-wise differentiation,

$$f'(z) = \sum_{n=0}^{\infty} a_n n (z - z_*)^{n-1}, \qquad z \in B_R(z_*),$$

and this power series converges in $B_R(z_*)$.

The proof was already given in the first year.

A bad example might be salubrious:

[3] holomorph
[4] ganze Funktion

Lemma 7.8. *The function $f\colon \mathbb{C} \to \mathbb{C}$, given by $f(z) = z\overline{z} = |z|^2$, is nowhere holomorphic.*

Proof. If $z_0 = 0$, then

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{z\overline{z}}{z} = \overline{z} \to 0 \qquad (\text{if } z \to z_0),$$

hence f is complex differentiable at $z_0 = 0$.

If $z_0 \neq 0$, then

$$\frac{f(z) - f(z_0)}{z - z_0} = \frac{z\overline{z} - z_0\overline{z_0}}{z - z_0} = \frac{(z - z_0)\overline{z} + z_0(\overline{z} - \overline{z_0})}{z - z_0} = \overline{z} + z_0 \cdot \frac{\overline{z} - \overline{z_0}}{z - z_0},$$

and this has no limit for $z \to z_0$, because we can write $z = z_0 + \varepsilon e^{\mathrm{i}\varphi}$, and then

$$\frac{\overline{z} - \overline{z_0}}{z - z_0} = \frac{\varepsilon \exp(-\mathrm{i}\varphi)}{\varepsilon \exp(\mathrm{i}\varphi)} = \exp(-2\mathrm{i}\varphi),$$

and now each ray $\{z\colon \arg(z - z_0) = \varphi\}$ has its own limit for $\frac{\overline{z} - \overline{z_0}}{z - z_0}$.

Therefore, f is not complex differentiable at $z_0 \neq 0$. □

Theorem 7.9 (**Cauchy–Riemann differential equations**). *Let $\Omega \subset \mathbb{C}$ be non-empty and open, $f\colon \Omega \to \mathbb{C}$ be a function with $f = u + \mathrm{i}v = \binom{u}{v}$, and $z_0 = x_0 + \mathrm{i}y_0 \in \Omega$. Here u, v, x_0, y_0 are real. Then the following are equivalent:*

1. *f is complex differentiable at z_0,*
2. *$f = f(x, y)$ is real differentiable at $(x_0, y_0) \in \mathbb{R}^2$, and the Jacobi matrix $f'(x_0, y_0)$ generates a $\mathbb{C}$–linear map,*
3. *$f = f(x, y)$ is real differentiable at $(x_0, y_0) \in \mathbb{R}^2$, and the* Cauchy–Riemann differential equations *hold at the point (x_0, y_0):*

$$u_x(x_0, y_0) = v_y(x_0, y_0), \qquad u_y(x_0, y_0) = -v_x(x_0, y_0).$$

Proof. We exploit Lemma 7.2:

f is complex differentiable at z_0

$\iff$ $\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$ exists

$\iff$ there is a number $a \in \mathbb{C}$ with

$$f(z) = f(z_0) + a(z - z_0) + \mathfrak{o}(z - z_0), \qquad z \to z_0,$$

$\iff$ there is a $\mathbb{C}$–linear map $T\colon \mathbb{C} \to \mathbb{C}$ with

$$f(z) = f(z_0) + T(z - z_0) + \mathfrak{o}(z - z_0), \qquad z \to z_0,$$

$\iff$ there is a matrix $A \in \mathbb{R}^{2\times 2}$ with $a_{11} = a_{22}$ and $a_{12} = -a_{21}$ with

$$\begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix} = \begin{pmatrix} u(x_0, y_0) \\ v(x_0, y_0) \end{pmatrix} + A \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} + \mathfrak{o}\left(\left\| \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \right\| \right), \qquad \begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x_0 \\ y_0 \end{pmatrix},$$

$\iff$ the function $\binom{u}{v}\colon \Omega \to \mathbb{R}^2$ is differentiable at $\binom{x_0}{y_0}$, and the Jacobi matrix

$$\begin{pmatrix} u \\ v \end{pmatrix}'(x_0, y_0) = \begin{pmatrix} u_x(x_0, y_0) & u_y(x_0, y_0) \\ v_x(x_0, y_0) & v_y(x_0, y_0) \end{pmatrix}$$

satisfies $u_x(x_0, y_0) = v_y(x_0, y_0)$ and $u_y(x_0, y_0) = -v_x(x_0, y_0)$. □

Another approach to the Cauchy–Riemann differential equations is the following:

$$
\begin{aligned}
& f \quad \text{is complex differentiable at } z_0 \\
\iff \quad & \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0} \quad \text{exists,} \\
\implies \quad & \lim_{x\to x_0} \frac{f(x+\mathrm{i}y_0)-f(x_0+\mathrm{i}y_0)}{(x+\mathrm{i}y_0)-(x_0+\mathrm{i}y_0)} \quad \text{and} \quad \lim_{y\to y_0} \frac{f(x_0+\mathrm{i}y)-f(x_0+\mathrm{i}y_0)}{(x_0+\mathrm{i}y)-(x_0+\mathrm{i}y_0)} \\
& \text{both exist and are equal} \\
\implies \quad & \lim_{x\to x_0} \frac{(u(x,y_0)+\mathrm{i}v(x,y_0))-(u(x_0,y_0)+\mathrm{i}v(x_0,y_0))}{x-x_0} \\
& \qquad = \lim_{y\to y_0} \frac{(u(x_0,y)+\mathrm{i}v(x_0,y))-(u(x_0,y_0)+\mathrm{i}v(x_0,y_0))}{\mathrm{i}(y-y_0)} \\
\iff \quad & u_x(x_0,y_0)+\mathrm{i}v_x(x_0,y_0) = \frac{1}{\mathrm{i}}(u_y(x_0,y_0)+\mathrm{i}v_y(x_0,y_0)),
\end{aligned}
$$

and from this computation, we learn that

$$f'(z_0) = \partial_x f(z_0) = \frac{1}{\mathrm{i}}\partial_y f(z_0),$$

which has as consequence that

$$f'(z_0) = \frac{1}{2}(\partial_x - \mathrm{i}\partial_y) f(z_0).$$

We can express this as

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right).$$

Moreover, the Cauchy–Riemann equations can be compressed into

$$0 = \left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)(u(x_0,y_0)+\mathrm{i}v(x_0,y_0)).$$

Check this !

We take z and $\overline{z}$ as two new complex variables instead of the real variables x and y: by the chain rule,

$$
\begin{aligned}
\frac{\partial}{\partial x} &= \frac{\partial z}{\partial x}\frac{\partial}{\partial z} + \frac{\partial \overline{z}}{\partial x}\frac{\partial}{\partial \overline{z}} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \overline{z}}, \\
\frac{\partial}{\partial y} &= \frac{\partial z}{\partial y}\frac{\partial}{\partial z} + \frac{\partial \overline{z}}{\partial y}\frac{\partial}{\partial \overline{z}} = \mathrm{i}\frac{\partial}{\partial z} - \mathrm{i}\frac{\partial}{\partial \overline{z}},
\end{aligned}
$$

and therefore

$$
\begin{aligned}
0 &= \left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)(u(x_0,y_0)+\mathrm{i}v(x_0,y_0)) \\
&= \left(\left(\frac{\partial}{\partial z} + \frac{\partial}{\partial \overline{z}}\right) + \mathrm{i}\left(\mathrm{i}\frac{\partial}{\partial z} - \mathrm{i}\frac{\partial}{\partial \overline{z}}\right)\right)(u(x_0,y_0)+\mathrm{i}v(x_0,y_0)) \\
&= 2\frac{\partial}{\partial \overline{z}} f(z).
\end{aligned}
$$

Hence we have shown:

Lemma 7.10. *A function $f\colon \Omega \to \mathbb{C}$ is complex differentiable at $z_0 \in \Omega$ if and only if*

$$\frac{\partial}{\partial \overline{z}} f(z_0) = 0.$$

Going back to the example $f = f(z) = z\overline{z}$ from Lemma 7.8, we have $\partial_{\overline{z}} f = z$, which vanishes only at the origin, but nowhere else.

Exercise: *Check that the real part and the imaginary part of the principal branch of the complex logarithm solve the Cauchy–Riemann differential equations.*

7.3 Conclusions and Applications

Lemma 7.11. *If $f = u + \mathrm{i}v = \binom{u}{v}$ (where u and v are real) is holomorphic in Ω, with u and v being twice continuously differentiable, then $\triangle u = 0$ and $\triangle v = 0$ in Ω.*

Proof. Wonderful exercise. □

Lemma 7.12. *Let $\Omega \subset \mathbb{C}$ be non-empty, open, and connected (also called a* domain[5] *in $\mathbb{C}$). Let $f\colon \Omega \to \mathbb{C}$ be holomorphic in $\mathbb{C}$.*

- *if $f'(z) = 0$ everywhere in Ω, then $f \equiv$ const. in Ω,*
- *if f takes real values everywhere in Ω, then $f \equiv$ const. in Ω,*
- *if $|f(z)| = 1$ for all $z \in \Omega$, then $f \equiv$ const. in Ω.*

Proof. Write $f = u + \mathrm{i}v$ as usual, with u and v being real. Similarly we split $z = x + \mathrm{i}y$ with real x, y.

- We have

  $$0 = f'(z) = \frac{\partial}{\partial z} f(z) = \frac{\partial}{\partial x} f(z) = u_x(z) + \mathrm{i}v_x(z), \qquad \forall\, z \in \Omega,$$

  and then, by the Cauchy–Riemann differential equations,

  $$u_y(z) = -v_x(z) \equiv 0, \qquad v_y(z) = u_x(z) \equiv 0,$$

  hence $\nabla u \equiv 0$, $\nabla v \equiv 0$ in Ω. Then $u \equiv$ const. and $v \equiv$ const. in Ω, because Ω is connected.
- Now we have $v \equiv 0$ in Ω, and consequently $\nabla u \equiv 0$ in Ω.
- We know $u^2(z) + v^2(z) = 1$ for all $z \in \Omega$, hence

  $$uu_x + vv_x \equiv 0, \qquad uu_y + vv_y \equiv 0,$$

  and, by the Cauchy–Riemann differential equations, $uv_x = -uu_y = vv_y = vu_x$,

  $$0 \equiv u \cdot (uu_x + vv_x) = u^2 u_x + uvv_x = u^2 u_x + v^2 u_x = (u^2 + v^2)u_x = 1 \cdot u_x.$$

  Similarly, we show $v_x \equiv 0$, which brings us to $f' \equiv 0$. Now apply the first •.

□

Definition 7.13. *A map $T\colon \mathbb{C} \to \mathbb{C}$ is called* angle-preserving[6] *if it is $\mathbb{R}$–linear, injective, and if*

$$\frac{\langle T(w), T(z)\rangle}{|T(w)| \cdot |T(z)|} = \frac{\langle w, z\rangle}{|w| \cdot |z|}$$

for all w, $z \in \mathbb{C} \setminus \{0\}$, with $\langle p, q\rangle = p_1 q_1 + p_2 q_2$ being the usual scalar product in $\mathbb{R}^2$.

Lemma 7.14. *An injective $\mathbb{R}$–linear map $T\colon \mathbb{C} \to \mathbb{C}$ is angle-preserving if and only if there is a number $a \in \mathbb{C} \setminus \{0\}$ with $T(z) = az$ for all $z \in \mathbb{C}$, or $T(z) = a\bar{z}$ for all $z \in \mathbb{C}$.*

Proof. We consider z and w from $\mathbb{C}$ as vectors from $\mathbb{R}^2$. Recall that $\langle z, w\rangle = |z| \cdot |w| \cdot \cos(\sphericalangle(z, w))$.

In $\mathbb{R}^2 \simeq \mathbb{C}$, we take the triangle with corners $O = 0_{\mathbb{C}} = (0,0)^\top$, $P = 1_{\mathbb{C}} = (1,0)^\top$, and $Q = \mathrm{i} = (0,1)^\top$. The linear map T maps this triangle OPQ to the triangle $O'P'Q'$, with $O' = O$, by linearity. The linear map T is realised by a matrix $A \in \mathbb{R}^{2\times 2}$, and the entries in the columns of A are equal to the coordinates of the images of the basis vectors, hence

$$P' = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \qquad Q' = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}.$$

[5] Gebiet
[6] winkeltreu

The vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ are perpendicular, and T is angle-preserving. Therefore also the vectors $\overrightarrow{OP'}$ and $\overrightarrow{OQ'}$ must be perpendicular, hence

$$\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} \perp \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} \qquad \Longrightarrow \qquad \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \alpha \begin{pmatrix} -a_{21} \\ a_{11} \end{pmatrix} \qquad (\exists\, \alpha \in \mathbb{R}).$$

Now the triangle OPQ has an angle $\pi/4$ at P, and then also the triangle $OP'Q'$ must have an angle $\pi/4$ at P', because T preserves the (modulus of the) angles. This means $|\overrightarrow{OP'}| = |\overrightarrow{OQ'}|$, hence

$$\left| \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} \right| = \left| \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} \right| \qquad \Longrightarrow \qquad |\alpha| = 1.$$

If $\alpha = 1$, then we have

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{21} \\ a_{21} & a_{11} \end{pmatrix}$$

and the product $T(z) = A\binom{x}{y}$ becomes $(a_{11} + \mathrm{i}a_{21})(x + \mathrm{i}y)$, which equals az for $a = a_{11} + \mathrm{i}a_{21}$.
And if $\alpha = -1$, then we have

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{21} & -a_{11} \end{pmatrix},$$

and then the product $T(z) = A\binom{x}{y}$ becomes $(a_{11} + \mathrm{i}a_{21})(x - \mathrm{i}y)$, which equals $a\overline{z}$ for $a = a_{11} + \mathrm{i}a_{21}$. □

Lemma 7.15. *If $\gamma_1 = \gamma_1(t)$ and $\gamma_2 = \gamma_2(t)$ with $a_1 \le t \le b_1$ and $a_2 \le t \le b_2$ are two differentiable curves in $\mathbb{C}$ which intersect at $z_0 \in \mathbb{C}$ with intersection angle α, and if $f = f(z)$ is a holomorphic map with $f'(z_0) \neq 0$, then the image curves $f(\gamma_1)$ and $f(\gamma_2)$ intersect at $f(z_0)$, again with intersection angle α.*

Proof. Let t_1 and t_2 be such that $z_0 = \gamma_1(t_1)$ and $z_0 = \gamma_2(t_2)$. Then the tangential vectors on the curves γ_1 and γ_2 are $\gamma_1'(t_1)$ and $\gamma_2'(t_2)$. The images of these tangential vectors under the map f are $f'(z_0)\gamma_1'(t_1)$ and $f'(z_0)\gamma_2'(t_2)$. Now observe that the multiplication with the complex number $f'(z_0)$ constitutes an angle-preserving map, because $f'(z_0) \neq 0$. □

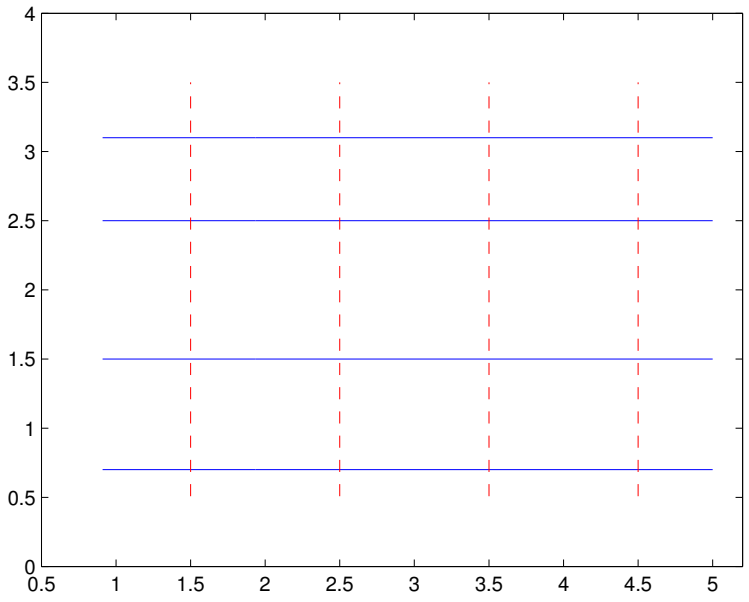


Figure 7.1: A grid in the z–plane

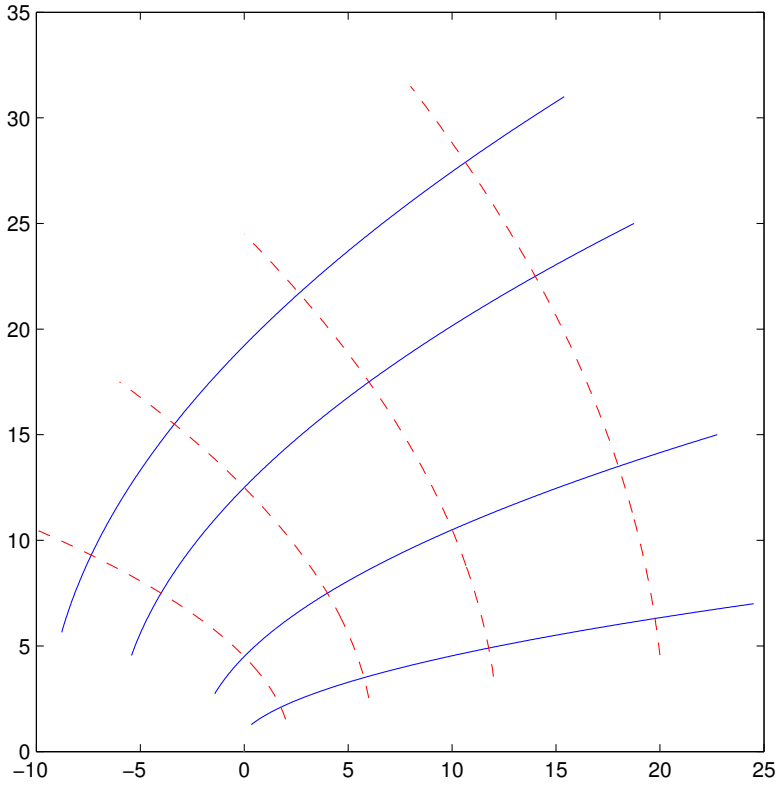


Figure 7.2: The image of the grid from Figure 7.1 under the map $z \mapsto z^2$. Observe that the red and blue lines intersect orthogonally in both figures.

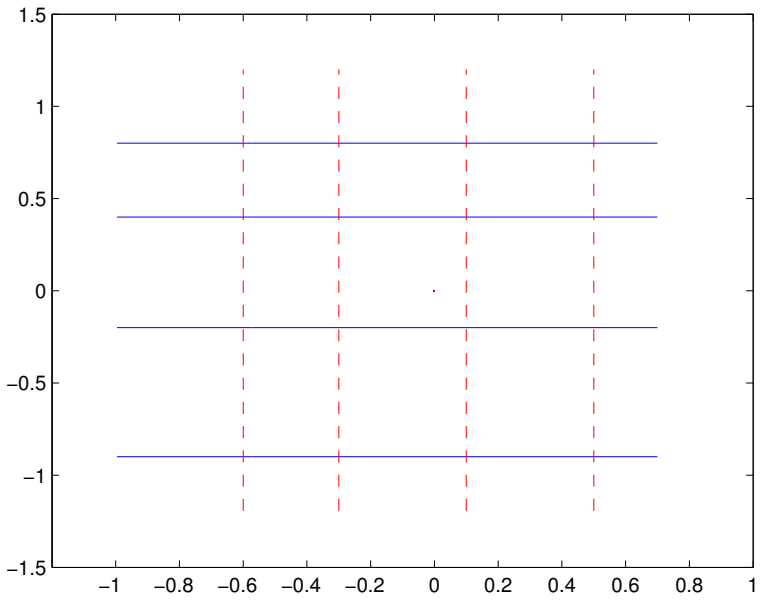


Figure 7.3: One more grid in the z–plane

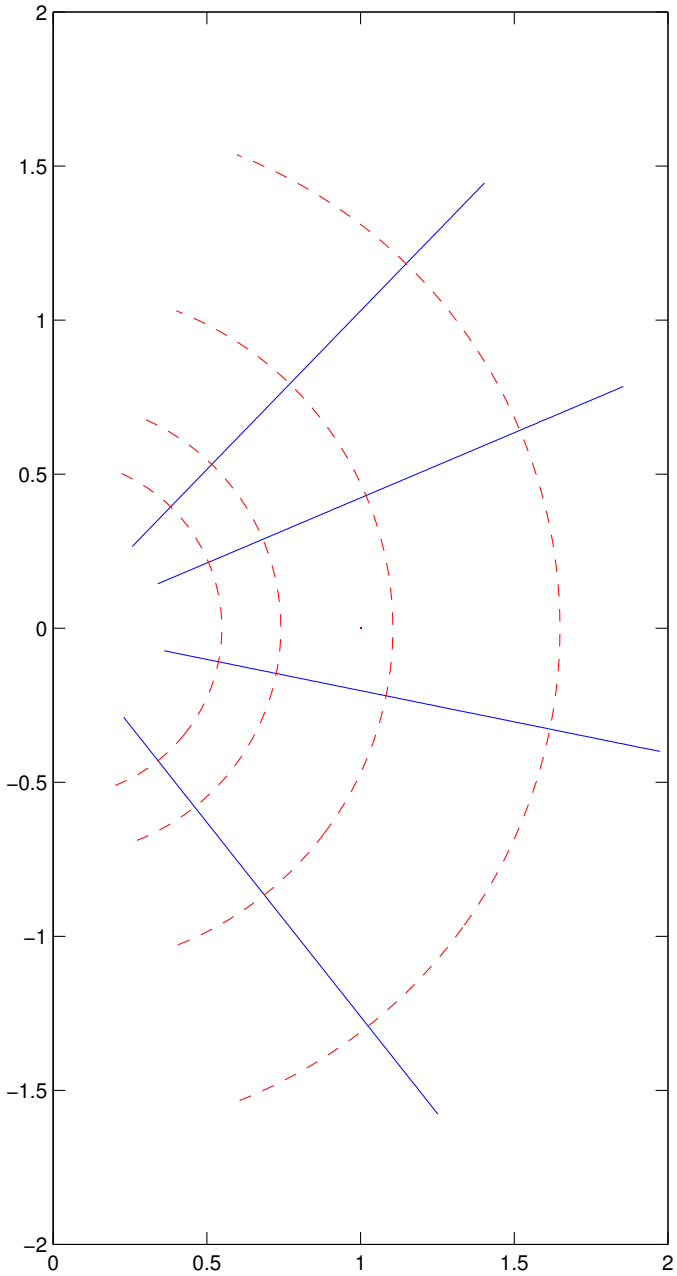


Figure 7.4: The image of the grid from Figure 7.3 under the map $z \mapsto \exp(z)$

We come to an application in **fluid dynamics**. In a two-dimensional world, let some stream of water flow in the plane, around an obstacle. We assume the flow to be independent of time, and the velocity vector at a point $(x, y) \in \mathbb{R}^2$ shall be denoted by $(w_1(x, y), w_2(x, y))$. The obstacle shall be the unit ball,

$$B = \{(x, y) \in \mathbb{R}^2 \colon x^2 + y^2 \leq 1\}.$$

Our physical assumptions are the following:

- the flow is free of rotations (curl-free), hence

  $$\operatorname{rot}(w_1, w_2) \equiv 0 \quad \text{in } \mathbb{R}^2 \setminus B,$$

  which means $w_{1,y} - w_{2,x} \equiv 0$.

  This does **not** imply that (w_1, w_2) possesses a potential, because $\mathbb{R}^2 \setminus B$ is a doubly connected set, but not a simply connected set.

- there exists a scalar potential V for (w_1, w_2):

  $$w_1(x, y) = V_x(x, y), \qquad w_2(x, y) = V_y(x, y), \qquad \forall\, (x, y) \in \mathbb{R}^2 \setminus B.$$

- the fluid is incompressible and homogeneous (the density is the same everywhere), and there are neither sources nor sinks:

  $$\operatorname{div}(w_1, w_2) \equiv 0,$$

  which means $w_{1,x} + w_{2,y} = 0$.

For a discussion of the validity of these assumptions, see Chapter 40 "The flow of dry water" in Feynman's lecture notes[7]. Some conclusions are:

- $\triangle V = \operatorname{div} \operatorname{grad} V = \operatorname{div}(w_1, w_2) \equiv 0$ in $\mathbb{R}^2 \setminus B$,
- The velocity field is perpendicular to the level sets[8] $\{(x, y) \colon V(x, y) = \text{const.}\}$. (These level sets are called *potential lines*[9].)
- on the boundary ∂B, the vector ∇V must be tangential, because otherwise the fluid would enter the obstacle, or come out of the obstacle.

The last conclusion is typical for dry water, because wet water will have ∇V completely equal to zero at ∂B, for reasons of friction between water and obstacle.

Our next assumption is:

- in the upper half-plane $\{(x, y) \colon y > 0\}$, the picture of the flow is the reflected picture from the lower half-plane.

In particular, the flow is horizontal on the real axis. Then we can consider $B \cup \{(x, y) \colon y \leq 0\}$ as the new obstacle, and only care about the points with $y > 0$. The new interesting domain

$$\Omega := \{(x, y) \colon y > 0 \text{ and } x^2 + y^2 > 1\}$$

is then simply connected.

In this set Ω, we are looking for a real-valued function $W = W(x, y)$ with

$$V_x = W_y, \qquad V_y = -W_x,$$

[7] Richard P. Feynman, Robert B. Leighton, Matthew Sands. The Feynman Lectures on Physics, The Definitive Edition Volume 2 (2nd Edition). Addison Wesley, 2005.

[8] Niveaulinien

[9] Potentiallinien

and then we put $Z(x,y) := V(x,y) + \mathrm{i}W(x,y)$, which will be a holomorphic function in Ω, because the Cauchy–Riemann differential equations are satisfied. Expressed in another way,

$$\nabla W = \begin{pmatrix} W_x \\ W_y \end{pmatrix} \perp \begin{pmatrix} V_x \\ V_y \end{pmatrix} = \nabla V,$$

and therefore the curves $\{(x,y)\colon W(x,y) = \text{const.}\}$ intersect the curves $\{(x,y)\colon V(x,y) = \text{const.}\}$ orthogonally. The curves along which W is constant are called *stream lines*[10] because the particles travel along these lines (this is true because the flow is independent of time).

The boundary $\partial\Omega$ of Ω consists of three parts: the interval $(-\infty,-1]$, the upper half circle $\partial_+ B := \{(x,y)\colon x^2+y^2=1, y>0\}$, and the interval $[1,\infty)$. The flow must not cross any of these parts, hence it must flow along $\partial\Omega$, and therefore W must be constant along $\partial\Omega$.

Additionally, far away from the obstacle, the flow should not feel that an obstacle even exists. The flow should be horizontal with speed one there:

$$(w_1(x,y), w_2(x,y)) \approx (1,0) \qquad \text{if } x^2+y^2 \gg 1.$$

Then it is reasonable to expect that $V(x,y) \approx x$ for $x^2+y^2 \gg 1$, and also

$$\nabla W = \begin{pmatrix} W_x \\ W_y \end{pmatrix} = \begin{pmatrix} -V_y \\ V_x \end{pmatrix} = \begin{pmatrix} -w_2 \\ w_1 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

far away from the obstacle, which makes $W(x,y) \approx y$ plausible. This gives us $Z(z) \approx z$ for $z = x + \mathrm{i}y$ with $|z| \gg 1$.

The goal is now to find a function Z on Ω with the following conditions:

- Z is holomorphic on Ω,
- $Z(z) \approx z$ for $|z| \gg 1$,
- $\Im Z(z)$ is constant on $\partial\Omega$. Combined with the second condition, $\Im Z(z)$ should be zero on $\partial\Omega$.

The strategy is now: the shape of the set Ω is quite awkward. So maybe we can find $Z(z) = (G \circ H)(z)$, with G and H both holomorphic, and H maps Ω onto another domain which is more beautiful, in comparison to Ω. Then we only have to find G, in a second step.

After some playing around with various functions, we come to $H(z) = z + \frac{1}{z}$. Observe that

- $H\colon [1,\infty) \to [2,\infty)$,
- $H\colon \partial_+ B \to (-2,2)$,
- $H\colon (-\infty,-1] \to (-\infty,-2]$,

and therefore H maps Ω onto the upper half-plane $\mathbb{C}_{\mathrm{i},+} := \{(x,y)\colon y>0\}$. The function G shall have the following properties:

- G is holomorphic on $\mathbb{C}_{\mathrm{i},+}$,
- $G(\zeta) \approx \zeta$ for $|\zeta| \gg 1$,
- $\Im G(\zeta)$ is zero on the real axis $\mathbb{R} = \partial\mathbb{C}_{\mathrm{i},+}$.

The most natural choice is $G(\zeta) = \zeta$. Hence we have found $Z(z) = z + \frac{1}{z}$, and therefore

$$\begin{aligned} Z(x,y) = V(x,y) + \mathrm{i}W(x,y) &= (x+\mathrm{i}y) + \frac{1}{x+\mathrm{i}y} = x + \mathrm{i}y + \frac{x-\mathrm{i}y}{x^2+y^2} \\ &= x\left(1 + \frac{1}{x^2+y^2}\right) + \mathrm{i}y\left(1 - \frac{1}{x^2+y^2}\right) \end{aligned}$$

[10] Stromlinien

which gives us

$$W(x,y) = y\left(1 - \frac{1}{x^2+y^2}\right),$$

and the lines $\{(x,y)\colon W(x,y) = \text{const.}\}$ are good candidates for stream lines.

The big open question is: is this the only solution ? Perhaps there are also other solutions, and there is one more criterion (which we have not found yet) that tells us which of these solutions is the physically correct one ?

The answer (which we can not justify) is that the above constructed function W does indeed describe the stream lines of dry water, see also the figure, which seems physically believable (at least if we do not look to closely at the boundary between obstacle and water).

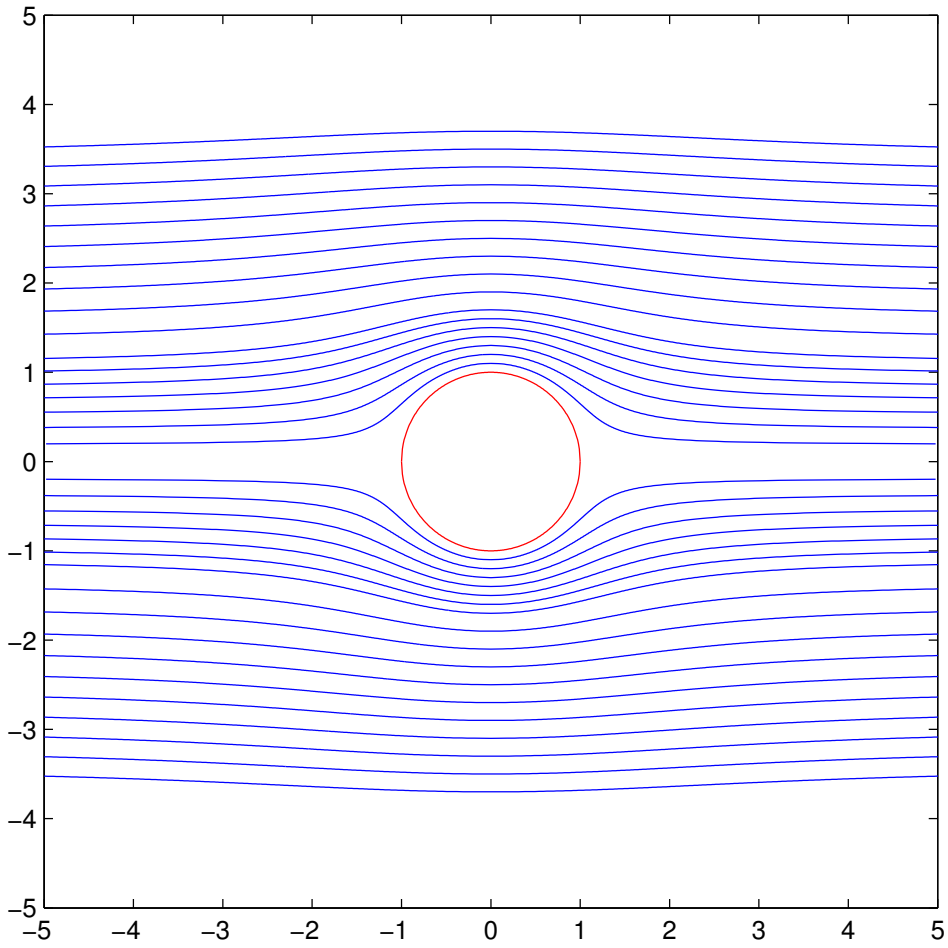


Figure 7.5: The lines $\{(x,y)\colon W(x,y) = \text{const.}\}$

[speaking about the scientific fields of chemical kinetics and fluid dynamics:][11] *Neither subject had yet reached the dignified status of a science in the nineteenth century, when as Sir Cyril Hinshelwood has observed, chemical reactions were classified mainly into those that go and those that do not go, and when fluid dynamicists were divided into hydraulic engineers who observed things that could not be explained and mathematicians who explained things that could not be observed.*

[11] *Physics of gas flow at very high speeds*, Nature, **4529**, 1956, 343–345

Chapter 8

Integration

8.1 Definition and Simple Properties

We start with a curve $\Gamma \subset \mathbb{C}$ and its parametrisation γ, which is supposed to fulfil the following conditions:

- $\gamma \in C^1([t_{\text{start}}, t_{\text{end}}] \to \mathbb{C})$, for a certain time interval $[t_{\text{start}}, t_{\text{end}}] \subset \mathbb{R}$,
- $\gamma'(t)$ is never zero, for $t_{\text{start}} \le t \le t_{\text{end}}$,
- the image Γ intersects itself only a finite number of times.

The second condition makes sure that the image Γ has no "corner" points. Call the endpoints of the image $\Gamma_{\text{start}} = \gamma(t_{\text{start}})$ and $\Gamma_{\text{end}} = \gamma(t_{\text{end}})$.

To approximate complex curve integrals, we choose a large natural number N, split the time interval into N parts:

$$t_{\text{start}} = t_0 < t_1 < \ldots < t_N = t_{\text{end}},$$

give a name to the associated points on the image:

$$z_j := \gamma(t_j), \qquad 0 \le j \le N,$$

and then we imagine a complex curve integral as something which can be approximated like this:

$$\int_\gamma f(z)\,\mathrm{d}z \approx \sum_{j=0}^{N-1} f(\gamma(\tau_j))(z_{j+1} - z_j),$$

where τ_j are arbitrary times from the corresponding sub-interval: $t_j \le \tau_j \le t_{j+1}$.

With this interpretation in mind, the following definition becomes natural:

Definition 8.1 (**Curve integral**). *Let $f\colon \Gamma \to \mathbb{C}$ be continuous, and suppose the above conditions on γ and Γ. Then we set*

$$\int_\gamma f(z)\,\mathrm{d}z := \int_{t=t_{\text{start}}}^{t_{\text{end}}} f(\gamma(t))\gamma'(t)\,\mathrm{d}t.$$

We split f and γ into real part and imaginary part:

$$\begin{aligned}
f &= u + \mathrm{i}v, \qquad u, v \in \mathbb{R}, \\
\gamma(t) &= \alpha(t) + \mathrm{i}\beta(t), \qquad \alpha(t), \beta(t) \in \mathbb{R}, \qquad \gamma'(t) = \alpha'(t) + \mathrm{i}\beta'(t), \\
\int_\gamma f(z)\,\mathrm{d}z &= \int_{t=t_{\text{start}}}^{t_{\text{end}}} (u(\gamma(t)) + \mathrm{i}v(\gamma(t)))(\alpha'(t) + \mathrm{i}\beta'(t))\,\mathrm{d}t \\
&= \int_{t=t_{\text{start}}}^{t_{\text{end}}} (u\alpha' - v\beta')\,\mathrm{d}t + \mathrm{i}\int_{t=t_{\text{start}}}^{t_{\text{end}}} (v\alpha' + u\beta')\,\mathrm{d}t \\
&= \int_{\Gamma_{\text{start}}}^{\Gamma_{\text{end}}} (u\,\mathrm{d}x - v\,\mathrm{d}y) + \mathrm{i}\int_{\Gamma_{\text{start}}}^{\Gamma_{\text{end}}} (v\,\mathrm{d}x + u\,\mathrm{d}y),
\end{aligned}$$

which are curve integrals of second kind, as we have studied them in the first year. Recalling the results from the past, we directly get:

Proposition 8.2. *If f is continuous on Γ, then we have:*

- *the integral $\int_\gamma f(z)\,\mathrm{d}z$ does not depend on the parametrisation of Γ, as long as the orientation of Γ is not inverted (otherwise a factor -1 appears),*
- *the following estimate holds:*

$$\left|\int_\gamma f(z)\,\mathrm{d}z\right| \le \|f\|_{L^\infty(\Gamma)} \cdot \mathrm{length}(\Gamma), \tag{8.1}$$

- *the functional $f \mapsto \int_\gamma f(z)\,\mathrm{d}z$ is a homomorphism from $C(\Gamma \to \mathbb{C})$ to $\mathbb{C}$.*

By the first •, we can now write $\int_\Gamma f(z)\,\mathrm{d}z$ instead of $\int_\gamma f(z)\,\mathrm{d}z$.

Considering a sequence of curves, for which the end point of one curve is the starting point of the next curve, we can define curve integrals along curves with a finite number of "corners".

Example 8.3. *Take $f = f(z) = z^n$ with $n \in \mathbb{N}_0$, and Γ has the parametrisation $\gamma(t) = Re^{\mathrm{i}t}$ for $0 \le t \le 2\pi$. Then Γ is a closed curve with counter-clockwise orientation, and we have*

$$\int_\Gamma f(z)\,\mathrm{d}z = \int_{t=0}^{2\pi} (Re^{\mathrm{i}t})^n \cdot R\mathrm{i}e^{\mathrm{i}t}\,\mathrm{d}t = \mathrm{i}R^{n+1}\int_{t=0}^{2\pi} e^{\mathrm{i}t(n+1)}\,\mathrm{d}t = \mathrm{i}R^{n+1}\cdot\frac{1}{\mathrm{i}(n+1)}e^{\mathrm{i}t(n+1)}\Big|_{t=0}^{t=2\pi} = 0.$$

Example 8.4. *Take $f = f(z) = z^n$ with $n \in \{-2,-3,-4,\dots\}$ and Γ as before. Then we have*

$$\int_\Gamma f(z)\,\mathrm{d}z = \cdots = 0,$$

by the same computation.

Example 8.5. *Take $f = f(z) = 1/z$, and again Γ as a circle around the origin of radius R, run counter-clockwise. Then*

$$\int_\Gamma f(z)\,\mathrm{d}z = \int_{t=0}^{2\pi} \frac{1}{Re^{\mathrm{i}t}} \cdot R\mathrm{i}e^{\mathrm{i}t}\,\mathrm{d}t = \int_{t=0}^{2\pi} \mathrm{i}\,\mathrm{d}t = 2\pi\mathrm{i}.$$

Note that in all three examples, the final result does not depend on the radius R of the circle.

Proposition 8.6. *If Ω is a domain in $\mathbb{C}$ (which means: non-empty, open, connected), and $f\colon \Omega \to \mathbb{C}$ is continuous, with a holomorphic function $F\colon \Omega \to \mathbb{C}$ such that $F'(z) = f(z)$ for all $z \in \Omega$, then*

$$\int_\Gamma f(z)\,\mathrm{d}z = F(\Gamma_{\mathrm{end}}) - F(\Gamma_{\mathrm{start}}).$$

Proof. By direct computation and the chain rule:

$$\int_\Gamma f(z)\,\mathrm{d}z = \int_{t=t_{\mathrm{start}}}^{t_{\mathrm{end}}} f(\gamma(t))\gamma'(t)\,\mathrm{d}t = \int_{t=t_{\mathrm{start}}}^{t_{\mathrm{end}}} \frac{\mathrm{d}}{\mathrm{d}t}F(\gamma(t))\,\mathrm{d}t = F(\gamma(t_{\mathrm{end}})) - F(\gamma(t_{\mathrm{start}})).$$

□

Going back to Example 8.3, we have $f(z) = z^n$, and we easily check that $F = F(z) = \frac{1}{n+1}z^{n+1}$ is a *primitive function*[1] of f (which means $F' = f$), and then it is clear that the integral over Γ gives the value zero, because Γ is a loop.

In case of Example 8.4, we can argue in the same manner.

Exercise: *Check that the complex logarithm* Ln, *defined on $\Omega := \mathbb{C} \setminus \mathbb{R}_-$, is a primitive function to $f = f(z) = \frac{1}{z}$:*

$$(\mathrm{Ln}(z))' \stackrel{?}{=} \frac{1}{z}, \qquad \forall z \in \mathbb{C} \setminus \mathbb{R}_-.$$

[1] Stammfunktion

Recall that the complex logarithm has a jump of height $2\pi\mathrm{i}$ when we cross $\mathbb{R}_-$, which explains the result from Example 8.5.

In the situation of Proposition 8.6, the value of the curve integral $\int_\Gamma f(z)\,\mathrm{d}z$ depends only on the location of the start point Γ_{start} and on the location of the end point Γ_{end}, but not on the path connecting these two points. This can be expressed in another way:

Lemma 8.7. *Let Ω be a domain in $\mathbb{C}$ (a multiply-connected Ω is allowed), and $f\colon \Omega \to \mathbb{C}$ be a continuous function. Then the following two statements are equivalent:*

- *for each curve Γ in Ω, the value of the integral $\int_\Gamma f(z)\,\mathrm{d}z$ depends only on the start point Γ_{start} and the end point Γ_{end},*
- *for each loop Γ in Ω, the curve integral $\oint_\Gamma f(z)\,\mathrm{d}z$ is zero.*

Idea of proof. Take two points A and B in Ω, and take two curves Γ_1 and Γ_2 which join these two points, running from A to B. If you invert the orientation of one of these curves and join them together, you obtain a loop. □

Proposition 8.8. *Let Ω be a domain in $\mathbb{C}$ (a multiply-connected Ω is allowed), and $f\colon \Omega \to \mathbb{C}$ be continuous. Then the following are equivalent:*

1. *there is a function F, holomorphic on Ω, with $F'(z) = f(z)$ for all $z \in \Omega$,*
2. *for each loop Γ in Ω, we have $\oint_\Gamma f(z)\,\mathrm{d}z = 0$.*

Proof.

$1 \Longrightarrow 2$**:** have a look at Proposition 8.6 and Lemma 8.7.

$2 \Longrightarrow 1$**:** we construct a function F as follows. Pick a point $z_* \in \Omega$, and for each $z \in \Omega$, let Γ_z be a curve inside Ω from z_* to z. Then define

$$F(z) := \int_{\Gamma_z} f(\zeta)\,\mathrm{d}\zeta.$$

By Lemma 8.7, the value of $F(z)$ does not depend on the choice of the curve connecting z_* and z. We choose a point $z_0 \in \Omega$, and we wish to show $F'(z_0) = f(z_0)$, which means

$$\lim_{z\to z_0} \frac{F(z)-F(z_0)}{z-z_0} = f(z_0) \quad\Longleftrightarrow\quad \lim_{z\to z_0} \left| \frac{F(z)-F(z_0)}{z-z_0} - f(z_0) \right| = 0.$$

We are allowed to take a special curve Γ_z, namely $\Gamma_z := \Gamma_{z_0} \cup S(z_0 \to z)$, where $S(z_0 \to z)$ stands for the straight line from z_0 to z. Then it follows that

$$\begin{aligned}
F(z) - F(z_0) &= \int_{S(z_0\to z)} f(\zeta)\,\mathrm{d}\zeta, \\
\left| \frac{F(z)-F(z_0)}{z-z_0} - f(z_0) \right| &= \left| \frac{1}{z-z_0} \int_{S(z_0\to z)} f(\zeta)\,\mathrm{d}\zeta - f(z_0) \right| \\
&= \left| \frac{1}{z-z_0} \int_{t=0}^{1} f(z_0 + t(z-z_0)) \cdot (z-z_0)\,\mathrm{d}t - f(z_0) \right| \\
&= \left| \int_{t=0}^{1} f(z_0 + t(z-z_0))\,\mathrm{d}t - f(z_0) \right| \le \int_{t=0}^{1} \left| f(z_0 + t(z-z_0)) - f(z_0) \right| \mathrm{d}t \\
&\le \sup_{\zeta \in S(z_0\to z)} |f(\zeta) - f(z_0)|,
\end{aligned}$$

and this goes to zero for $z \to z_0$, because f is continuous. This proves $F'(z_0) = f(z_0)$, for an arbitrary $z_0 \in \Omega$.

□

A direct consequence then is that the function $f = f(z) = 1/z$ can not have a primitive function on $\Omega = \mathbb{C} \setminus \{0\}$. But it does have a primitive function on $\Omega = \mathbb{C} \setminus \mathbb{R}_-$, namely the complex logarithm.

On the other hand, the function $f = f(z) = -1/z^2$ *does* possess a primitive function on the doubly-connected set $\Omega = \mathbb{C} \setminus \{0\}$, namely $F = F(z) = 1/z$.

8.2 The Cauchy Integral Theorem

In the previous section, f was merely continuous on Γ or on Ω. Now we make the assumptions on f stronger: it shall be holomorphic on Ω, and this will bring us a Great Beautification of the theory.

We recall a result from the first year:

Proposition 8.9. *Let $\Omega \subset \mathbb{R}^n$ be a domain that is simply-connected. Assume that $g\colon \Omega \to \mathbb{R}^n$ is continuous, with continuous derivative g', and that g satisfies the integrability conditions:*

$$\frac{\partial g_j}{\partial x_k}(x) = \frac{\partial g_k}{\partial x_j}(x), \qquad \forall\, j,k, \quad \forall\, x \in \Omega.$$

Let Γ be a curve inside Ω.

Then the value of the curve integral (of second kind) $\int_\Gamma \vec{g}\,\mathrm{d}\vec{x}$ depends only on the location of the start point Γ_{start} and of the end point Γ_{end}, but not on the path connecting these points.

Each curve integral $\oint_\Gamma \vec{g}\,\mathrm{d}\vec{x}$ over a loop Γ vanishes.

The condition that g' be continuous was really needed in the proof from the first year.

In our situation, we have $\mathbb{C} \simeq \mathbb{R}^2$, and the complex curve integral splits into two real curve integrals:

$$\int_\Gamma f(z)\,\mathrm{d}z = \int_\Gamma (u\,\mathrm{d}x + (-v)\,\mathrm{d}y) + \mathrm{i}\int_\Gamma (v\,\mathrm{d}x + u\,\mathrm{d}y).$$

Here we have split $z = x + \mathrm{i}y$ with real x and y, and also $f = u + \mathrm{i}v$ with real u and v. Now we apply that result from the first year to $g = (u, -v)$ and to $g = (v, u)$, and then we quickly get:

Proposition 8.10. *Let $\Omega \subset \mathbb{R}^2 \simeq \mathbb{C}$ be a domain that is simply-connected. Assume that $f\colon \Omega \to \mathbb{C} \simeq \mathbb{R}^2$ with $f = u + \mathrm{i}v$ (where u and v are real) is continuous, with continuous derivatives ∇u, ∇v, and that satisfies the integrability conditions*

$$u_y \equiv (-v)_x, \qquad v_y \equiv u_x.$$

Then each curve integral $\oint_\Gamma f(z)\,\mathrm{d}z$ over a loop Γ inside Ω vanishes.

The formulation is not quite satisfactory: we mention u and v too often, and the result will be nicer if we express everything in terms of f:

$$\begin{aligned}
\partial_x &= \partial_z + \partial_{\overline{z}}, \\
\partial_x(u + \mathrm{i}v) = \partial_x f = \partial_z f + \partial_{\overline{z}} f \quad &\Longrightarrow \qquad u_x = \Re(\partial_z f + \partial_{\overline{z}} f), \qquad v_x = \Im(\partial_z f + \partial_{\overline{z}} f), \\
\partial_y &= \mathrm{i}\partial_z - \mathrm{i}\partial_{\overline{z}}, \\
\partial_y(u + \mathrm{i}v) = \partial_y f = \mathrm{i}\partial_z f - \mathrm{i}\partial_{\overline{z}} f \quad &\Longrightarrow \qquad u_y = \Re(\mathrm{i}\partial_z f - \mathrm{i}\partial_{\overline{z}} f), \qquad v_y = \Im(\mathrm{i}\partial_z f - \mathrm{i}\partial_{\overline{z}} f),
\end{aligned}$$

and therefore we conclude that:

$$\nabla u,\ \nabla v \quad \text{are continuous} \quad \Longleftrightarrow \quad \partial_z f,\ \partial_{\overline{z}} f \quad \text{are continuous.}$$

We recall that the holomorphy of a function f can be expressed as $\partial_{\overline{z}} f \equiv 0$ and obtain our final result:

Theorem 8.11 (Cauchy **Integral Theorem**). *Let $\Omega \subset \mathbb{C}$ be a domain that is simply-connected. Assume that $f\colon \Omega \to \mathbb{C}$ is holomorphic, with continuous derivative f'.*

Then each curve integral $\oint_\Gamma f(z)\,\mathrm{d}z$ over a loop Γ inside Ω vanishes.

Remark 8.12. *We discuss the assumptions.*

- *The assumption "Ω simply connected" is indispensable, as the example $\Omega = \mathbb{C} \setminus \{0\}$, $f(z) = 1/z$ shows.*
- *The conclusion of the Cauchy Integral Theorem also holds without the condition "f' continuous" (which was only needed for applying Proposition 8.9 from the second semester), as can be shown by a completely different proof, see [9], [16], or the classical booklet [17].*

Example 8.13. *Take numbers* $0 < r_1 < r_2 < r_3 < r_4$ *and* $z_* \in \mathbb{C}$*, and choose* Ω *as an annulus*[2]*,*

$$\Omega = \{z \in \mathbb{C} \colon r_1 < |z - z_*| < r_4\}.$$

If f *is a holomorphic function on* Ω*, then*

$$\oint_{|z-z_*|=r_2} f(z)\,\mathrm{d}z = \oint_{|z-z_*|=r_3} f(z)\,\mathrm{d}z,$$

provided that both circles have the same orientation. The reason is that the two circles can be connected by a radial line, and the integration variable is walking along this radial line inward and outward, and walking along both circles.

This is an example of a general principle: we can deform carefully the path of a curve integral without changing the value of this integral, as long as we stay in the domain where the integrand is holomorphic.

Again, have a look at the Examples 8.3–8.5, in which the result was independent of the radius R.

We will need a generalisation of the Cauchy Integral Theorem:

Proposition 8.14. *Let* $\Omega \subset \mathbb{C}$ *be a simply-connected domain,* p *a point in* Ω*, and* $f \colon \Omega \setminus \{p\} \to \mathbb{C}$ *holomorphic, and* f *bounded near* p*.*

Then each curve integral $\oint_\Gamma f(z)\,\mathrm{d}z$ *over a loop* Γ *inside* Ω *vanishes.*

The key difficulty here is that $\Omega \setminus \{p\}$ is no longer simply-connected.

Idea of proof. The following is mathematically not very precise, but geometrically hopefully clear.

Case 1: Γ does not "revolve around" p: Then we can replace Ω by a smaller domain Ω_{new} that contains Γ, but not p, and then we apply the Cauchy Integral Theorem to Ω_{new} instead of Ω.

Case 2: Γ does "revolve around" p, perhaps several times: By the idea from Example 8.13, we can deform Γ, until we obtain another curve Γ_{new} which also revolves around p (the same number of times as Γ does), but which is much shorter. We remember that f is bounded near p, and have a look back to the integral estimate (8.1). Observe that the curve Γ_{new} can be made as short as we wish.

□

Of course, there may be several exceptional points in Ω like p, not only one.

We need some concepts. Imagine G as $\mathbb{C}$ minus a collection of some closed curves which may intersect (but any open non-empty subset G of $\mathbb{C}$ would also be good).

Definition 8.15. *Let* $G \subset \mathbb{C}$ *be non-empty and open, not necessarily connected. Then two points* z_1*,* $z_2 \in G$ *are called* path-equivalent[3] *if there is a curve in* G *that connects* z_1 *and* z_2*. If this holds, we write* $z_1 \sim_G z_2$*.*

We quickly check:

$$\begin{aligned}
&z \sim_G z && && \forall\, z \in G && \text{(reflexivity)},\\
&z \sim_G w \quad \Longrightarrow \quad w \sim_G z && && \forall\, z, w \in G && \text{(symmetry)},\\
&z \sim_G w, \quad w \sim_G \zeta \quad \Longrightarrow \quad z \sim_G \zeta && && \forall\, z, w, \zeta \in G && \text{(transitivity)},
\end{aligned}$$

which are the three conditions for an *equivalence relation*.

We take the opportunity to **start** an **excursion into algebra**.

Other examples of equivalence relations are:

- straight lines in the plane can be considered equivalent when they are parallel,

[2] Kreisring
[3] weg–äquivalent

- triangles in the plane can be considered equivalent when they are congruent,
- integers a, $b \in \mathbb{Z}$ are "congruent modulo 2" (written as $a \equiv b \bmod 2$) if $2|(b-a)$,
- if U and V are vector spaces and $f\colon U \to V$ a homomorphism, then we can define that $u_1 \equiv_f u_2$ if $u_1 - u_2 \in \ker f$,
- two students can be considered equivalent if they started their studies in the same year,
- screws are defined to be equivalent if they are of the same size,
- two integrable functions $f, g\colon [a,b] \to \mathbb{R}$ are equivalent if $\int_{x=a}^{b} |f(x) - g(x)|\,\mathrm{d}x = 0$ (imagine that f and g coincide almost everywhere, except a finite number of points in $[a,b]$).

Definition 8.16. *Let $\mathcal{M}$ be an arbitrary set with an equivalence relation $\sim$. For an $x \in \mathcal{M}$, the set*

$$[x] := \{y \in \mathcal{M}\colon y \sim x\} \subset \mathcal{M}$$

is called the equivalence class *of the element x.*

As an example, take $\mathcal{M} = \mathbb{Z}$ with the equivalence relation being: $a \sim b$ if and only if 2 divides $(b-a)$, as above. Then $[3]$ is the set of odd integers, and $[2]$ is the set of even integers. A nice property is that the relation $\sim$ respects the arithmetical operations: if $a_1 \sim b_1$ and $a_2 \sim b_2$, and if $\diamond$ is one of the symbols $+$, $-$, $\cdot$, then $(a_1 \diamond a_2) \sim (b_1 \diamond b_2)$. Of course, $[1] = [3] = [5] = \dots$ are the same set, and in order to represent the interests of this set at some other place, you can send a *representative*[4] (which is a member of $[1] = [3] = \dots$) to that other place. In this particular cases, such a representative could be -43 or 991 or any odd number (see below).

Lemma 8.17. *Let $\mathcal{M}$ be a (countable[5]) set with an equivalence relation $\sim$. Then $\mathcal{M}$ decomposes into disjoint subsets,*

$$\mathcal{M} = M_1 \cup M_2 \cup \dots,$$

with $M_j \cap M_k = \emptyset$ for $j \neq k$, and two members of $\mathcal{M}$ belong to the same M_j if and only if these two members are equivalent. Each M_j is an equivalence class of the relation $\sim$.

And if $\mathcal{M}$ is uncountable[6] (like every vector space over $\mathbb{R}$), then $\mathcal{M}$ decomposes as $\mathcal{M} = \cup_{\alpha \in A} M_\alpha$ with a possibly uncountable index set A, and each M_α is an equivalence class of the relation $\sim$.

The gain is: quite often, the equivalence classes correspond to interesting mathematical objects.

- an equivalence class of parallel straight lines defines a "direction" in the plane,
- the equivalence classes of the relation $\equiv_f$ form a vector space which is isomorphic to $\operatorname{img} f$.

Assume we only know integers from $\mathbb{Z}$, how can we define (in a logically precise manner) rational numbers from $\mathbb{Q}$? First we define ordered pairs (n,d) with $n \in \mathbb{Z}$ and $d \in \mathbb{Z} \setminus \{0\}$. Think of n as numerator[7] and d as denominator[8]. Define addition and multiplication via

$$(n_1,d_1) + (n_2,d_2) := (n_1d_2 + n_2d_1, d_1d_2), \qquad (n_1,d_1) \cdot (n_2,d_2) := (n_1n_2, d_1d_2).$$

Define an equivalence relation as $(n_1,d_1) \equiv (n_2,d_2)$ if and only if $n_1d_2 = n_2d_1$. Check that this is indeed an equivalence relation. Check that the relation $\equiv$ respects the arithmetical operations (in the sense explained above). Then we specify: each rational number is defined as an equivalence class of such pairs. The addition of rational numbers is defined via $[(n_1,d_1)] + [(n_2,d_2)] := [(n_1,d_1) + (n_2,d_2)]$. This means: to add two equivalence classes, you take a representative of each class, add these representatives, and then you take the equivalence class corresponding to this sum. It does not matter who are the two representatives because $\equiv$ respects the addition. Similarly for the multiplication. Finally we check that these rational numbers form a field.

[4] Vertreter, Stellvertreter, Abgeordneter, Delegierter
[5] abzählbar
[6] überabzählbar
[7] Zähler
[8] Nenner

Assume we only know rational numbers from $\mathbb{Q}$, how can we define (in a logically precise manner) real numbers from $\mathbb{R}$? First we define Cauchy sequences $(r_1, r_2, r_3, \dots)$ of rational numbers. Define addition and multiplication component-wise:

$$(r_1, r_2, r_3, \dots) + (s_1, s_2, s_3, \dots) := (r_1 + s_1, r_2 + s_2, r_3 + s_3, \dots),$$
$$(r_1, r_2, r_3, \dots) \cdot (s_1, s_2, s_3, \dots) := (r_1 \cdot s_1, r_2 \cdot s_2, r_3 \cdot s_3, \dots),$$

and check that the right-hand sides are Cauchy sequences again. Define two such Cauchy sequences of rational numbers to be equivalent if their difference sequence converges to the rational number zero. Check that this is indeed an equivalence relation. Check that this equivalence relation respects the arithmetical operations (in the sense explained above). Then we specify: each real number is defined as an equivalence class of such sequences. The addition and multiplication of real numbers are defined via representatives. Finally we check that these real numbers form a field.

Let us **end** here the **excursion into algebra** and return to complex analysis.

Let $G \subset \mathbb{C}$ be non-empty and open, not necessarily connected. By the equivalence relation $\sim_G$, G splits into disjoint subsets (whose number might be infinite),

$$G = C_1 \cup C_2 \cup \ldots \cup C_K,$$

and two points of G belong to the same subset if and only they are path equivalent.

Definition 8.18. *These sets C_j are called* connected components of G[9].

For instance, the bready subset of a cheeseburger comprises exactly two connected components (upper and lower part), at least initially.

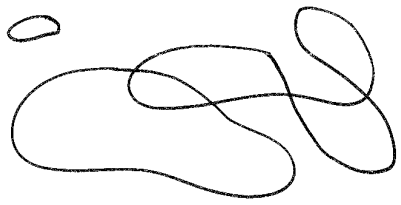


Figure 8.1: This set G has 7 connected components, one of them unbounded.

Definition 8.19. *Let Γ be a loop in $\mathbb{C}$. For $z \in \mathbb{C} \setminus \Gamma$, set*

$$\operatorname*{Ind}_\Gamma(z) := \frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{1}{\zeta - z} \, \mathrm{d}\zeta,$$

and call it the winding number[10] *of Γ with respect to z.*

Lemma 8.20. *For Γ a loop in $\mathbb{C}$ and $z \in \mathbb{C} \setminus \Gamma$, the winding number $\operatorname{Ind}_\Gamma(z)$ is an integer, and it is constant on each connected component of $G := \mathbb{C} \setminus \Gamma$. On the unbounded component, $\operatorname{Ind}_\Gamma(z)$ is zero.*

Proof. Take $z \in G$ fixed. Parametrise Γ with $\gamma \colon [a, b] \to \mathbb{C}$. Then we have

$$\operatorname*{Ind}_\Gamma(z) = \frac{1}{2\pi \mathrm{i}} \int_{t=a}^{b} \frac{\gamma'(t)}{\gamma(t) - z} \, \mathrm{d}t. \tag{8.2}$$

For $s \in [a, b]$, we set

$$\varphi(s) := \exp\left(\int_{t=a}^{s} \frac{\gamma'(t)}{\gamma(t) - z} \, \mathrm{d}t \right),$$

[9] Zusammenhangskomponenten von G
[10] Umlaufzahl, Windungszahl, Index

and then we get

$$\frac{\varphi'(s)}{\varphi(s)} = \frac{\gamma'(s)}{\gamma(s) - z}$$

for all $s \in [a, b]$, except (possibly) a finite number of points where γ' does not exist because Γ has a corner there. Then we get

$$\frac{\mathrm{d}}{\mathrm{d}s} \frac{\varphi(s)}{\gamma(s) - z} = \frac{\varphi'(s)(\gamma(s) - z) - \gamma'(s)\varphi(s)}{(\gamma(s) - z)^2} = 0,$$

almost everywhere on $[a, b]$. Now $\frac{\varphi(s)}{\gamma(s)-z}$ is clearly a continuous function of s, and therefore a constant, hence

$$\frac{\varphi(b)}{\gamma(b) - z} = \frac{\varphi(a)}{\gamma(a) - z} = \frac{1}{\gamma(a) - z}.$$

From $\gamma(a) = \gamma(b)$, we then find $\varphi(b) = 1$, or

$$1 = \varphi(b) = \exp\left(\int_{t=a}^{b} \frac{\gamma'(t)}{\gamma(t) - z}\,\mathrm{d}t\right) = \exp\left(2\pi\mathrm{i}\operatorname*{Ind}_{\Gamma}(z)\right),$$

which is only possible for $\operatorname{Ind}_\Gamma(z) \in \mathbb{Z}$.

In the second semester, we have learned that integrals may be differentiated with respect to a real parameter under the integral symbol. Let this parameter be x or y:

$$\partial_x \operatorname*{Ind}_{\Gamma}(z) = \frac{1}{2\pi\mathrm{i}} \partial_x \int_{t=a}^{b} \frac{\gamma'(t)}{\gamma(t) - z}\,\mathrm{d}t = \frac{1}{2\pi\mathrm{i}} \int_{t=a}^{b} \partial_x \frac{\gamma'(t)}{\gamma(t) - z}\,\mathrm{d}t,$$
$$\partial_y \operatorname*{Ind}_{\Gamma}(z) = \frac{1}{2\pi\mathrm{i}} \int_{t=a}^{b} \partial_y \frac{\gamma'(t)}{\gamma(t) - z}\,\mathrm{d}t,$$

and then also $\partial_{\overline{z}} \operatorname{Ind}_\Gamma(z) = \frac{1}{2}\partial_x \operatorname{Ind}_\Gamma(z) + \frac{\mathrm{i}}{2}\partial_y \operatorname{Ind}_\Gamma(z) = 0$, because

$$\partial_{\overline{z}} \frac{\gamma'(t)}{\gamma(t) - z} = 0$$

since the function $z \mapsto \frac{\gamma'(t)}{\gamma(t)-z}$ is holomorphic. Then also $z \mapsto \operatorname{Ind}_\Gamma(z)$ is holomorphic on Γ, hence continuous. Therefore $\operatorname{Ind}_\Gamma(z)$ must be constant on each connected component of G, since the values of $\operatorname{Ind}_\Gamma(z)$ are integers.

If we send z to ∞ in (8.2), we get $|\gamma(t) - z| \to 0$ uniformly for $t \in [a, b]$, and this is the reason why $\operatorname{Ind}_\Gamma(z) = 0$ for z from the unbounded connected component of G. □

Now we have all tools for the next important result.

Theorem 8.21 (Cauchy Integral Formula). *Let Ω be a simply connected domain, $\Gamma \subset \Omega$ a loop, and $z \in \Omega \setminus \Gamma$. Then, for f holomorphic on Ω,*

$$f(z) \operatorname*{Ind}_{\Gamma}(z) = \frac{1}{2\pi\mathrm{i}} \oint_{\Gamma} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta.$$

In particular: if $\Gamma = \partial B(a, r)$ is a circle of radius r and centre a (oriented counter-clockwise), and $z \in B(a, r)$, then

$$f(z) = \frac{1}{2\pi\mathrm{i}} \oint_{\Gamma} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta.$$

Proof. Keep $z \in \Omega \setminus \Gamma$ fixed, and define

$$g(\zeta) = \begin{cases} \frac{f(\zeta)-f(z)}{\zeta-z} & : \zeta \neq z, \\ f'(z) & : \zeta = z. \end{cases}$$

Then this function g is holomorphic on $\Omega \setminus \{z\}$, and bounded for ζ near z (even continuous at z). By Proposition 8.14,

$$0 = \oint_\Gamma g(\zeta)\,\mathrm{d}\zeta = \oint_\Gamma \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - f(z)\oint_\Gamma \frac{1}{\zeta - z}\,\mathrm{d}\zeta = \oint_\Gamma \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - 2\pi\mathrm{i}\operatorname*{Ind}_\Gamma(z) f(z),$$

which completes the proof. □

Corollary 8.22. *If f is holomorphic on Ω (multiply-connected is allowed), then also f' is holomorphic on Ω.*

Proof. Choose a ball $B(a,r) \subset \Omega$ and a point $z \in B(a,r)$. Then, with $\Gamma = \partial B(a,r)$, we have

$$f(z) = \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta,$$

because $B(a,r)$ is simply connected, and differentiating under the integral symbol (as we have introduced it in the second semester) gives

$$f'(z) = \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \partial_z \left(\frac{f(\zeta)}{\zeta - z}\right) \mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \frac{f(\zeta)}{(\zeta - z)^2}\,\mathrm{d}\zeta.$$

Now f' is holomorphic if and only if $\partial_{\overline{z}} f'(z) \equiv 0$. To check this, we differentiate once again under the integral symbol:

$$\partial_{\overline{z}} f'(z) = \frac{1}{2\pi\mathrm{i}}\partial_{\overline{z}} \oint_\Gamma \frac{f(\zeta)}{(\zeta - z)^2}\,\mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \partial_{\overline{z}} \frac{f(\zeta)}{(\zeta - z)^2}\,\mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \oint_\Gamma 0\,\mathrm{d}\zeta = 0,$$

because $z \mapsto \frac{f(\zeta)}{(\zeta - z)^2}$ is holomorphic. □

Corollary 8.23. *If f is holomorphic on Ω, then f is infinitely differentiable.*

The next result can be understood as a converse to the Cauchy Integral Theorem.

Theorem 8.24 (**Morera's Theorem**[11])**.** *Let Ω be a domain, possibly multiply-connected. If f is continuous on Ω with $\oint_\Gamma f(z)\,\mathrm{d}z = 0$ for each loop Γ in Ω, then f is holomorphic on Ω.*

Proof. By Proposition 8.8, there is a holomorphic function F with $F'(z) = f(z)$ for all $z \in \Omega$. Now apply Corollary 8.22 to the function F instead of f. □

By repeated differentiation under the integral, we deduce that

$$(\partial_z^n f)(z) \cdot \operatorname*{Ind}_\Gamma(z) = \frac{n!}{2\pi\mathrm{i}} \oint_\Gamma \frac{f(\zeta)}{(\zeta - z)^{n+1}}\,\mathrm{d}\zeta$$

if Γ is a loop in a simply connected domain Ω. We may choose $\Gamma = \partial B(a,r)$ (positively oriented) and $z \in B(a,r)$. Then $\operatorname{Ind}_\Gamma(z) = 1$ and

$$(\partial_z^n f)(z) = \frac{n!}{2\pi\mathrm{i}} \oint_{|\zeta - a| = r} \frac{f(\zeta)}{(\zeta - z)^{n+1}}\,\mathrm{d}\zeta, \qquad |z - a| < r. \tag{8.3}$$

It is even possible to expand f into a converging power series. To this end, we consider

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta - a) - (z - \alpha)} = \left((\zeta - a)\left(1 - \frac{z - a}{\zeta - a}\right)\right)^{-1} = \frac{1}{\zeta - a}\sum_{n=0}^{\infty}\left(\frac{z - a}{\zeta - a}\right)^n$$

by the summation rule of the geometric series, which is applicable because of

$$\left|\frac{z - a}{\zeta - a}\right| = \frac{|z - a|}{r} < \frac{r}{r} = 1.$$

[11] Giacinto Morera, 1856–1907

The convergence of this series is uniform if $|z-a|$ is strictly less than r. Then we can proceed as follows:

$$\begin{aligned} f(z) &= \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r} \frac{f(\zeta)}{\zeta - a} \sum_{n=0}^{\infty} \left(\frac{z-a}{\zeta-a}\right)^n \mathrm{d}\zeta \\ &= \frac{1}{2\pi\mathrm{i}} \sum_{n=0}^{\infty} \oint_{|\zeta-a|=r} \frac{f(\zeta)}{\zeta - a} \left(\frac{z-a}{\zeta-a}\right)^n \mathrm{d}\zeta \\ &= \frac{1}{2\pi\mathrm{i}} \sum_{n=0}^{\infty} \left(\oint_{|\zeta-a|=r} \frac{f(\zeta)}{(\zeta-a)^{n+1}}\,\mathrm{d}\zeta \right) \cdot (z-a)^n. \end{aligned}$$

Commuting $\oint$ and $\sum_n$ was possible by the uniform convergence of $\sum_n$. We put

$$c_n := \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r} \frac{f(\zeta)}{(\zeta-a)^{n+1}}\,\mathrm{d}\zeta$$

and find the power series expansion

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \qquad \text{if } |z-a| < r.$$

Observe that (8.1) gives

$$|c_n| \le \frac{1}{2\pi} \max_{\zeta\in\Gamma} \frac{|f(\zeta)|}{r^{n+1}} \cdot \mathrm{length}(\Gamma) = \frac{\max_{\zeta\in\Gamma} |f(\zeta)|}{r^n},$$

giving us then

$$|c_n (z-a)^n| \le \left|\frac{z-a}{r}\right|^n \max_{\zeta\in\Gamma} |f(\zeta)|,$$

which tells us that the power series $\sum_{n=0}^{\infty} c_n (z-a)^n$ indeed converges for all z with $|z-a| < r$. Comparing with (8.3) then brings us

$$c_n = \frac{1}{n!} (\partial_z^n f)(a).$$

We summarise:

Lemma 8.25. *Let $\Omega \subset \mathbb{C}$ be a domain (multiply-connected allowed), and choose a point $a \in \Omega$. If f is holomorphic on Ω, then f can be expanded into a power series at the point a,*

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{n!} (\partial_z^n f(a)) \cdot (z-a)^n,$$

and this power series converges in the biggest ball $B(a,r)$ that is contained in the closure $\overline{\Omega}$ of the open set Ω.

Geometrically: the distance from the point a to the nearest singularity of f determines the radius of convergence. Here "singularity" is defined as a point of $\mathbb{C}$ where f is not holomorphic.

Exercise: *Determine the radii of convergence for the functions*

$$\frac{1}{1-z}, \quad \frac{1}{1+z^2}, \quad \sin(z), \quad \tan(z), \quad \mathrm{Ln}(1+z),$$

all of them expanded at the point $a = 0$.

The following is impossible for a function f:

- $f = f(z)$ is complex differentiable for $|z| < 7$,
- $f(z) = 0$ for all z with $|z| \le 1$,
- $f(z) \neq 0$ for all z with $1 < |z| < 7$.

And the reason is this: first we note that f is holomorphic on the ball $B(0,7)$, and therefore f is infinitely differentiable at each point in this ball. Pick a point z_* with $|z_*| = 1$. Then a sequence $(z_1, z_2, \dots)$ exists with $|z_k| < 1$ for all k and $\lim_{k\to\infty} z_k = z_*$. By continuity of all the derivatives,

$$(\partial_z^n f)(z_*) = \lim_{k\to\infty} (\partial_z^n f)(z_k) = \lim_{k\to\infty} 0 = 0,$$

and then the power series expansion of f at the point z_* reads

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{n!} (\partial_z^n f)(z_*) \cdot (z - z_*)^n = \sum_{n=0}^{\infty} \frac{1}{n!} \cdot 0 \cdot (z - z_*)^n = 0,$$

and this expansion is valid for all z with $|z - z_*| < 6$, because $B(z_*, 6) \subset B(0,7)$, and f is holomorphic in $B(0,7)$. But the last assumption was that $f(z) \neq 0$ for $1 < |z| < 7$, which is a contradiction.

The situation is totally different for real differentiable functions. There are $f\colon \mathbb{R}^2 \to \mathbb{R}$ such that

- $f = f(x)$ is real differentiable for $|x| < 7$,
- $f(x) = 0$ for all $x \in \mathbb{R}^2$ with $|x| \leq 1$,
- $f(x) \neq 0$ for all x with $1 < |x| < 7$,

for instance

$$f(x) = \begin{cases} 0 & : |x| \leq 1, \\ (x_1^2 + x_2^2 - 1)^{100} & : |x| > 1. \end{cases}$$

Definition 8.26 (**Analytic function**). *A function on a domain of $\mathbb{C}$ that can be expanded into a converging power series (at each point of its domain of definition) with positive radius of convergence is called* analytic function[12].

In the first year, we had learned that each analytic function is infinitely differentiable, and the power series can be differentiated term-wise. And in this semester, we have found that each holomorphic function is analytic. Therefore "analytic" and "holomorphic" are equivalent concepts in $\mathbb{C}$.

Theorem 8.27 (**Liouville's Theorem**). *Each entire bounded function is constant.*

Proof. Call this function f. Since f is entire, the power series of f at $a = 0$ converges on all of $\mathbb{C}$, hence

$$f(z) = \sum_{n=0}^{\infty} c_n z^n, \qquad \forall\, z \in \mathbb{C},$$

and

$$c_n = \frac{1}{2\pi \mathrm{i}} \oint_{|\zeta|=r} \frac{f(\zeta)}{\zeta^{n+1}}\, \mathrm{d}\zeta,$$

where the radius r of $\Gamma = \partial B(0,r)$ is arbitrary. We know already that

$$|c_n| \leq \frac{\max_{\zeta\in\Gamma} |f(\zeta)|}{r^n}, \qquad \forall\, r > 0, \quad \forall\, n \in \mathbb{N}_0.$$

Because f is bounded on $\mathbb{C}$, we have $\sup_{\zeta\in\mathbb{C}} |f(\zeta)| \leq M$ for some M, hence $|c_n| \leq M r^{-n}$ for all r, and therefore $0 = c_1 = c_2 = \dots$. □

Theorem 8.28 (**Fundamental Theorem of Algebra**). *Each polynomial of degree $n \geq 1$ possesses n zeroes in $\mathbb{C}$ (counted according to their multiplicity).*

[12] analytische Funktion

Proof. Take $P(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$ with $a_n \neq 0$. We show that P has at least one zero in $\mathbb{C}$. For large $|z|$, the term $a_n z^n$ is the biggest contribution to $P(z)$:

$$\begin{aligned} |a_{n-1} z^{n-1} + \ldots + a_1 z + a_0| &= |z|^n \cdot \left| \frac{a_{n-1}}{z} + \cdots + \frac{a_1}{z^{n-1}} + \frac{a_0}{z^n} \right| \\ &= |a_n z^n| \cdot \left| \frac{a_{n-1}}{a_n z} + \cdots + \frac{a_1}{a_n z^{n-1}} + \frac{a_0}{a_0 z^n} \right| \\ &\leq |a_n z^n| \cdot \frac{1}{3} \quad \text{if } |z| \geq R_0 \end{aligned}$$

for some suitably chosen number R_0. Then P can not have a zero outside $B(0, R_0)$ because otherwise we had

$$\begin{aligned} 0 = |P(z)| &= |a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0| \geq |a_n z^n| - |a_{n-1} z^{n-1} + \ldots + a_1 z + a_0| \\ &\geq |a_n z^n| - |a_n z^n| \cdot \frac{1}{3} = \frac{2}{3} |a_n z^n| > 0. \end{aligned}$$

Now assume that P has no zero in $\mathbb{C}$. The set $\{z \in \mathbb{C} \colon |z| \leq R_0\}$ is compact, and P has no zero there. But the function $z \mapsto |P(z)|$ is continuous and real-valued, and such functions attain their infimum on compact sets (a first year result):

$$\exists\, z_* \in \overline{B(0, R_0)} \colon \quad |P(z_*)| = \inf_{|z| \leq R_0} |P(z)|.$$

Then $|P(z_*)|$ must be positive, and there is a small number $\varepsilon > 0$ with $|P(z)| \geq \varepsilon$ for all z with $|z| \leq R_0$. Therefore we can define

$$Q(z) := \frac{1}{P(z)}, \qquad z \in \mathbb{C},$$

and this is holomorphic on $\mathbb{C}$ because we never divide by zero. But Q is also bounded:

$$|Q(z)| \leq \begin{cases} \frac{1}{\varepsilon} & : |z| \leq R_0, \\ \frac{3}{2|a_n z^n|} & : |z| > R_0, \end{cases}$$

and then Q must be constant, by Liouville's Theorem. Then $P = 1/Q$ is also a constant function. But P was a polynomial of degree ≥ 1. Contradiction.

Therefore P has a zero z_1. Then we can divide polynomials, i.e., find another polynomial P_1 with

$$P(z) = (z - z_1) P_1(z), \qquad \forall\, z \in \mathbb{C}.$$

The degree of P_1 is $n - 1$. If $n - 1 \geq 1$, repeat the reasoning from above. □

Chapter 9

Zeroes, Singularities, Residues

9.1 Zeroes of Holomorphic Functions

Proposition 9.1. *Let f be a holomorphic function in the open ball $B(a,r)$, with $f(a)=0$. Then exactly one of the following two cases occurs:*

Case 1: *f has a zero of finite order $K \in \mathbb{N}_+$ at the point a, which means*

$$0 = f(a) = \partial_z f(a) = \ldots = (\partial_z^{K-1} f)(a), \quad (\partial_z^K f)(a) \neq 0,$$

and moreover, there is a small positive ε such that f has no zero in $B(a,\varepsilon)$ except the centre a,

Case 2: *$f \equiv 0$ in $B(a,r)$.*

Proof. The function f has a representation as a converging power series,

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n, \qquad |z-a| < r,$$

and $f(a)=0$ implies $c_0 = 0$. Now exactly two cases are possible.

Case 1: at least one coefficient c_n is non-zero: Then one of these non-zero coefficients has the smallest index, call it c_K with $K \in \mathbb{N}_+$. Then

$$f(z) = \sum_{n=K}^{\infty} c_n (z-a)^n = (z-a)^K \sum_{m=0}^{\infty} c_{K+m} (z-a)^m =: (z-a)^K g(z),$$

with $g(z) := \sum_{m=0}^{\infty} \ldots$ as a holomorphic function on the ball $B(a,r)$, and $g(a) = c_K \neq 0$. Because g is continuous, there is a small radius ε such that $g(z) \neq 0$ for all $z \in B(a,\varepsilon)$.

Case 2: all c_n are zero: then f is the zero-function.

□

We wish to extend this result to more general subsets of $\mathbb{C}$, namely domains, instead of open balls $B(a,r)$. The most elegant proof takes the route of topology, so we list a few topological concepts, most of them should be already known to you.

Definition 9.2. *A set $M \subset \mathbb{R}^n$ is* open *if for each $a \in M$, a ball $B(a,r)$ exists with $B(a,r) \subset M$ (and $r > 0$, of course).*

A point $x^* \in \mathbb{R}^n$ is a cluster point of M[1] *if in each small ball $B(x^*,\varepsilon)$ an element x_ε of M exists with $x_\varepsilon \neq x_*$. (For the definition, it does not matter whether $x^* \in M$ or not.)*

Let $\Omega \subset \mathbb{R}^n$ be open. A subset $M \subset \Omega$ is called closed in Ω[2] *if each point $x^* \in \Omega$ that is a cluster point of M belongs to M.*

The empty set $M = \emptyset$ is open, and it is closed in Ω.

[1] Häufungspunkt von M
[2] abgeschlossen in Ω

Example 9.3. *The interval $M = (0, 1]$ is closed in $\Omega = (0, \infty) \subset \mathbb{R}^1$.*

The interval $M = (0, 2)$ is closed in $\Omega = (0, 2) \subset \mathbb{R}^1$.

Lemma 9.4. *Let $\Omega \subset \mathbb{R}^n$ be open, and M be any subset of Ω. Write $M_{\text{c.p.}}$ for the set of cluster points of M that belong to Ω.*

Then $M_{\text{c.p.}}$ is closed in Ω.

Proof. We have to show: if $x^* \in \Omega$ is a cluster point of $M_{\text{c.p.}}$, then $x^* \in M_{\text{c.p.}}$.

This is equivalent to: if $x^* \in \Omega$ is a cluster point of $M_{\text{c.p.}}$, then it is a cluster point of M.

In other words: if $x^* \in \Omega$ is a cluster point of $M_{\text{c.p.}}$, then for each $\varepsilon > 0$ a point $x_\varepsilon \in M$ exists with $x^* \neq x_\varepsilon$ and $|x^* - x_\varepsilon| < \varepsilon$.

We know: for each $\varepsilon > 0$, there is an element $y_\varepsilon \in M_{\text{c.p.}}$ with $x^* \neq y_\varepsilon$ and $|x^* - y_\varepsilon| < \varepsilon/12$, because x^* is a cluster point of $M_{\text{c.p.}}$. This y_ε is a cluster point of M because y_ε is a member of $M_{\text{c.p.}}$.

We also know: there is an $x_\varepsilon \in M$ with $|y_\varepsilon - x_\varepsilon| < \frac{1}{10}|x^* - y_\varepsilon|$, because y_ε is a cluster point of M.

Then $x^* = x_\varepsilon$ is impossible, because of $|x^* - x_\varepsilon| > \frac{9}{10}|x^* - y_\varepsilon| > 0$ (draw a picture !).

Finally, we have

$$|x^* - x_\varepsilon| \leq |x^* - y_\varepsilon| + |y_\varepsilon - x_\varepsilon| < \left(1 + \frac{1}{10}\right)|x^* - y_\varepsilon| < \frac{11}{10} \cdot \frac{\varepsilon}{12} < \varepsilon,$$

as desired. Therefore x^* is a cluster point of M, hence $x^* \in M_{\text{c.p.}}$. □

Lemma 9.5. *Let $\Omega \subset \mathbb{R}^n$ be open, non-empty and connected. Assume $M \subset \Omega$ be open and closed in Ω. Then either $M = \emptyset$ or $M = \Omega$.*

Proof. Beautiful exercise. You will need that Ω is connected (otherwise there are counter-examples). □

In the following, $\Omega \subset \mathbb{C}$ is a domain (open, non-empty, connected), and $f \colon \Omega \to \mathbb{C}$ is a holomorphic function. The set of zeroes of f is $N^{(f)}$,

$$N^{(f)} := \{z \in \Omega \colon f(z) = 0\},$$

and the cluster points of $N^{(f)}$ form $N^{(f)}_{\text{c.p.}}$,

$$N^{(f)}_{\text{c.p.}} := \left\{ z \in \Omega \colon \ z \text{ is a cluster point of } N^{(f)} \right\}.$$

Lemma 9.6. *Either $N^{(f)}_{\text{c.p.}} = \emptyset$ or $N^{(f)}_{\text{c.p.}} = \Omega$.*

Proof. By Lemma 9.4, $N^{(f)}_{\text{c.p.}}$ is closed in Ω.

We are done if we can show that $N^{(f)}_{\text{c.p.}}$ is open. Take a point $a \in N^{(f)}_{\text{c.p.}}$. Then $a \in \Omega$, and Ω is open, hence a ball $B(a, r)$ exists with $r > 0$ and $B(a, r) \subset \Omega$.

Since a is a cluster point of $N^{(f)}$, a sequence $(z_1, z_2, \dots) \subset \Omega$ of zeroes of f exists with $\lim_{j\to\infty} z_j = a$. Then this sequence must enter the ball $B(a, r)$ for large index j, and approach the centre a. Then Case 2 in Proposition 9.1 occurs, and $f \equiv 0$ in $B(a, r)$. Then each $z \in \mathbb{C}$ with $|z - a| < r$ is a zero of f, and therefore each such z is a cluster point of $N^{(f)}$, whence $B(a, r) \subset N^{(f)}_{\text{c.p.}}$.

We have shown: if $a \in N^{(f)}_{\text{c.p.}}$ then an $r > 0$ exists with $B(a, r) \subset N^{(f)}_{\text{c.p.}}$. This is the very definition of $N^{(f)}_{\text{c.p.}}$ being open. □

We begin to approach the highlight of this section:

Proposition 9.7. *Let f be holomorphic on the domain Ω, and suppose that a sequence of zeroes of f converges to a point a* in the domain Ω.

Then $f \equiv 0$ in Ω.

Proof. Because the point a belongs to Ω, we have $a \in N_{\text{c.p.}}^{(f)}$, which makes $N_{\text{c.p.}}^{(f)} = \emptyset$ impossible, hence $N_{\text{c.p.}}^{(f)} = \Omega$. Now repeat the middle of the proof of Lemma 9.6. □

Warning: *The conclusion is wrong if the limit point a of the zeroes of f sits on the boundary $\partial\Omega$. For instance, take $\Omega = \{z \in \mathbb{C} \colon \Re z > 0\}$ as the open right half-plane, and $f(z) = \sin(1/z)$. Then f possesses a sequence of zeroes that approaches the limit point $a = 0$, which does not belong to Ω because Ω is open. But $f(z)$ is certainly not equal to zero everywhere in Ω.*

Theorem 9.8 (Identity Theorem). *If two functions g and h are holomorphic on the domain Ω, and they coincide on a set which has a cluster point in Ω, then $g \equiv h$ in Ω.*

Proof. Apply Proposition 9.7 to the function $f := g - h$. □

We come back to an example from the last chapter: there is no function $g = g(z)$ which is holomorphic for $|z| < 7$, takes the value zero for $|z| \le 1$, and takes only non-zero values for $1 < |z| < 7$. Now we have a second proof: take $h = h(z)$ as the zero-function on the ball $B(0,7)$. Then both g and h are holomorphic in $\Omega = B(0,7)$, and they coincide in the ball $B(0,1)$, which certainly has a cluster point in Ω. Then the identity theorem says that $g \equiv h$ on $B(0,7)$, and therefore g must be the zero-function also outside of $B(0,1)$. Contradiction.

Corollary 9.9. *There is exactly one holomorphic function f on a neighbourhood Ω of $\mathbb{R}$ that coincides on the real axis with the sine function:*

$$f(z) = \sin z \qquad \text{if } z \in \mathbb{R}.$$

In other words: there is only one way for extending the sine function from $\mathbb{R}$ into $\mathbb{C}$ without losing holomorphy.

Lemma 9.10. *For $z, w \in \mathbb{C}$, it holds*

$$\sin(z+w) = \sin(z)\cos(w) + \cos(z)\sin(w). \tag{9.1}$$

Proof. Keep $w = w_0 \in \mathbb{R}$ fixed. Then the function on the left-hand side, $z \mapsto \sin(z + w_0)$, is holomorphic in $\mathbb{C}$, and the function on the right-hand side, $z \mapsto \sin(z)\cos(w_0) + \cos(z)\sin(w_0)$, is also holomorphic on $\mathbb{C}$. Both functions coincide for $z \in \mathbb{R}$, and by the identity theorem, they then coincide for all $z \in \mathbb{C}$.

This proves (9.1) for $z \in \mathbb{C}$ and $w \in \mathbb{R}$. Now keep $z = z_0 \in \mathbb{C}$ fixed, let w run, and apply the identity theorem again. □

Warning: *This technique can not be used to show*

$$\operatorname{Ln}(z \cdot w) = \operatorname{Ln}(z) + \operatorname{Ln}(w) \tag{9.2}$$

for all $z, w \in \mathbb{C} \setminus \mathbb{R}_-$ and Ln *as the principal branch of the complex logarithm, because (9.2) is wrong for such general z, w. However, (9.2) holds for all z, w with positive real part.*

As an additional example, we take the Gamma function

$$\Gamma(z) = \int_{t=0}^{\infty} e^{-t} t^{z-1} \, \mathrm{d}t, \qquad z \in \mathbb{R}_+, \tag{9.3}$$

with integration along the half-axis $\mathbb{R}_+$. For positive t and real z, we have

$$t^{z-1} = e^{(z-1)\ln t},$$

which (for each fixed $t \in \mathbb{R}_+$) is an entire function of $z \in \mathbb{C}$. Now we find all $z \in \mathbb{C}$ for which the integral in (9.3) exists. Put $z = x + \mathrm{i}y$, then

$$\left|t^{z-1}\right| = \left|e^{(x-1+\mathrm{i}y)\ln t}\right| = e^{(x-1)\ln t} = t^{x-1},$$

and then we can show that (9.3) exists for all $z = x + \mathrm{i}y$ with $x > 0$ and $y \in \mathbb{R}$. And if (x, y) varies in a compact subset of the open right half-plane, then the convergence $\lim_{R\to\infty} \int_{t=0}^{R} \ldots \, \mathrm{d}t$ is uniform.

Next we determine a power series expansion of Γ. We can not expand at $z_0 = 0$ because Γ has a pole there. Take $z_0 = x_0 + iy_0$ instead, with $x_0 > 0$. Then

$$t^{z-1} = e^{(z_0-1)\ln t} \cdot e^{(z-z_0)\ln t} = t^{z_0-1} \sum_{n=0}^{\infty} \frac{1}{n!} (\ln t)^n (z-z_0)^n,$$

and therefore

$$\begin{aligned} \Gamma(z) &= \int_{t=0}^{\infty} e^{-t} t^{z_0-1} \left(\sum_{n=0}^{\infty} \frac{1}{n!} (\ln t)^n (z-z_0)^n \right) \mathrm{d}t \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_{t=0}^{\infty} e^{-t} t^{z_0-1} (\ln t)^n \, \mathrm{d}t \right) (z-z_0)^n, \end{aligned}$$

and this series converges for all $z \in \mathbb{C}$ with $|z - z_0| < \Re z_0$. We know this because the Γ function as defined in (9.3) is holomorphic in the open right half-plane, and the power series converges in any ball that lies in the domain of holomorphy.

Next we have the induction formula $z\Gamma(z) = \Gamma(z+1)$ for $z \in \mathbb{R}_+$, which quickly follows from partial integration on (9.3). Then we use

$$\Gamma(z) := \frac{\Gamma(z+1)}{z}$$

for an *analytic continuation*[3] of the Gamma function to the domain $\{z \in \mathbb{C} \colon \Re z > -1\} \setminus \{0\}$, with a pole at zero. By repeated application of this formula, the Gamma function can be defined as a holomorphic function on $\mathbb{C} \setminus \{0, -1, -2, \dots\}$.

The Gamma function never takes the value zero (we do not prove this), and then $z \mapsto 1/\Gamma(z)$ is an entire function (without proof). Then one can show that the Bessel function $J_\nu = J_\nu(z)$, defined via

$$J_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+\nu+1)} \left(\frac{z}{2}\right)^{2m},$$

compare (2.4), is, for each fixed $\nu \in \mathbb{C}$, a holomorphic function of z on the domain $\mathbb{C} \setminus \mathbb{R}_-$ (because $z \mapsto z^\nu$ is delicate for $\nu \notin \mathbb{Z}$). This is easy to show since each power series is a holomorphic function of its argument in its ball of convergence. And the Bessel function $J_\nu(z)$ is, for each fixed $z \in \mathbb{C} \setminus \{0\}$, an entire function of the variable ν. This is non-trivial because an infinite sum of entire functions need not be entire.

Our next concept is the *maximum modulus principle*[4]. If you have a real differentiable function $f \colon \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^n$ and ask at which point $x^* \in \overline{\Omega}$ the function $x \mapsto |f(x)|$ attains its maximal value, the answer is: $x^* \in \overline{\Omega}$ can be anywhere. For instance, in case of $\Omega = B(0,3) \subset \mathbb{R}^2$ and $f(x) = 7 - (x_1^2 + x_2^2)$, $|f|$ attains its maximum at $x = (0,0)$.

The situation is different for complex differentiable functions.

Proposition 9.11 (Maximum modulus principle). *On a bounded domain $\Omega \subset \mathbb{C}$, a holomorphic function f which is continuous on $\overline{\Omega}$ attains its maximum at the boundary:*

$$\max_{z \in \overline{\Omega}} |f(z)| = \max_{z \in \partial\Omega} |f(z)|.$$

Note that we can write max instead of sup since $\overline{\Omega}$ and $\partial\Omega$ are compact.

Proof. Assume the opposite: there is an interior point $z^* \in \Omega$ with

$$\max_{z \in \overline{\Omega}} |f(z)| = |f(z^*)| > \max_{z \in \partial\Omega} |f(z)|.$$

[3] analytische Fortsetzung
[4] modulus = Betrag

Then f can not be a constant function. Because any point of the open set Ω is an interior point of Ω, hence there is a small ball $B(z^*, r)$ contained in Ω, and we have

$$|f(z)| \le |f(z^*)| \qquad \forall\, z \in B(z^*, r).$$

By the Cauchy Integral Formula,

$$f(z^*) = \frac{1}{2\pi \mathrm{i}} \oint_{|\zeta - z^*| = \varrho} \frac{f(\zeta)}{\zeta - z^*}\, \mathrm{d}\zeta, \qquad \forall\, \varrho \in (0, r).$$

On the circle $\Gamma = \partial B(z^*, \varrho)$, we have, exploiting (8.1),

$$\max_{z \in \Gamma} |f(z)| \le |f(z^*)| \le \frac{1}{2\pi} \max_{|\zeta - z^*| = \varrho} \frac{|f(\zeta)|}{|\zeta - z^*|} \cdot 2\pi\varrho = \max_{\zeta \in \Gamma} |f(\zeta)|,$$

which is only possible if $|f(\zeta)| = |f(z^*)|$ for each $\zeta \in \partial B(z^*, \varrho)$. But the radius ϱ was chosen arbitrarily between 0 and r, and consequently

$$|f(z)| = |f(z^*)| \quad \forall\, z \in B(z^*, r).$$

By the same method as in Lemma 7.12, we then can prove that $f \equiv$ const. in the ball $B(z^*, r)$. The identity theorem then implies $f \equiv$ const. in Ω, which we had excluded in the beginning. □

9.2 Singularities

Definition 9.12 (Isolated singularity). *Let $\Omega \subset \mathbb{C}$ be open. A point $a \in \Omega$ is called* isolated singularity[5] *of f it f is undefined at the point a, but holomorphic on a punctured*[6] *ball $B(a, \varepsilon) \setminus \{a\}$.*

If f can be extended to a function that is holomorphic on $B(a, \varepsilon)$, the singularity is called removable[7].

For example, the function $f = f(z) = (\sin z)/z$ has a removable singularity at $a = 0$ (by defining the value of f as 1 there).

Proposition 9.13. *If $\Omega \subset \mathbb{C}$ is open, and f is holomorphic on $\Omega \setminus \{a\}$, and bounded near the point a, then the singularity is removable.*

Proof. We define a new function

$$g(z) := \begin{cases} (z-a)^2 f(z) & : z \ne a, \\ 0 & : z = a, \end{cases}$$

and, clearly, g is complex differentiable for $z \ne a$. And for $z = a$, we have

$$g'(a) = \lim_{z \to a} \frac{g(z) - g(a)}{z - a} = \lim_{z \to a} (z - a) f(z) = 0,$$

because f is bounded near a. Then g is complex differentiable everywhere, hence holomorphic on Ω, and therefore a Taylor expansion of g is available:

$$g(z) = \sum_{n=0}^{\infty} c_n (z-a)^n, \quad \forall\, z \in B(a, \varepsilon) \quad (\exists\, \varepsilon > 0),$$

with $c_0 = 0$ because of $g(a) = 0$, and $c_1 = 0$ because of $g'(a) = 0$. Then we have, for $z \in B(a, \varepsilon) \setminus \{a\}$,

$$f(z) = \frac{g(z)}{(z-a)^2} = \sum_{m=0}^{\infty} c_{m+2} (z-a)^m,$$

and we can extend f to Ω by defining $f(a) := c_2$. □

[5] isolierte Singularität
[6] gelocht. puncture = Reifenpanne
[7] hebbar

Definition 9.14 (**Pole of order** m). *A function f with isolated singularity at the point $a \in \Omega$ has a* pole of order $m \in \mathbb{N}_+$ *if complex numbers $c_{-1}, \ldots, c_{-m}$ with $c_{-m} \neq 0$ exist such that*

$$f - \sum_{n=1}^{m} c_{-n}(z-a)^{-n}$$

has a removable singularity at the point a.

Lemma 9.15 (**Partial fraction decomposition**). *Let $f(z) = p(z)/q(z)$ be a rational function (quotient of two polynomials), with $\deg p < \deg q$; and q has zeroes $z_1, \ldots, z_K$ of multiplicities $m_1, \ldots, m_K$.*

Then numbers $c_{j,l} \in \mathbb{C}$ exist such that

$$f(z) = \sum_{j=1}^{K} \sum_{l=-m_j}^{-1} c_{j,l}(z-z_j)^l, \qquad \forall\, z \in \mathbb{C} \setminus \{z_1, \ldots, z_K\}.$$

Proof. By the definition of poles of order m, there are numbers $c_{j,l}$ such that $f - \sum_l c_{j,l}(z-z_j)^l$ has a removable singularity at z_j. Then also $f - \sum_{j=1}^{K}\sum_l c_{j,l}(z-z_j)^l$ has only removable singularities in $\mathbb{C}$, hence it is an entire function, and it goes to zero for $|z| \to \infty$. Now apply the Liouville theorem. □

Definition 9.16 (**Essential singularity**). *A function f has an* essential singularity[8] *at a point $a \in \Omega$ if this point is an isolated singularity of f, but neither removable nor a pole.*

Theorem 9.17 (**Casorati–Weierstraß Theorem**[9]). *If f has an essential singularity at $a \in \Omega$, then, for each small ε, the set $f(B(a,\varepsilon) \setminus \{a\})$ is dense*[10] *in $\mathbb{C}$ (this means that every complex number is a cluster point of $f(B(a,\varepsilon) \setminus \{a\})$, for each $\varepsilon > 0$).*

Proof. Suppose the opposite: for some $\varepsilon_0 > 0$, the set $f(B(a,\varepsilon_0) \setminus \{a\})$ is not dense in $\mathbb{C}$. Then some $w \in \mathbb{C}$ is not a cluster point of $f(B(a,\varepsilon_0) \setminus \{a\})$, and consequently a positive δ exists with

$$|f(z) - w| > \delta \qquad \forall\, z \in B(a,\varepsilon_0) \setminus \{a\}.$$

For such z, define $g(z) := 1/(f(z)-w)$, which is holomorphic on $B(a,\varepsilon) \setminus \{a\}$ and bounded, $|g(z)| < \delta^{-1}$. By Proposition 9.13, this singularity of g is removable, and g can be holomorphically extended to the ball $B(a,\varepsilon_0)$.

Case A: $g(a) \neq 0$**:** then f is bounded near a, and the singularity of f is removable. Contradiction.

Case B: $g(a) = 0$**:** Then we can exploit Proposition 9.1, and either g has a finite order zero (then f has a finite order pole — contradiction) or $g \equiv 0$ in $B(a,\varepsilon_0)$, which contradicts $g(z) = 1/(f(z)-w)$.

We always got a contradiction. □

A typical example is $f(z) = \exp(1/z)$ for $z \neq 0$. In each small ball $B(0,\varepsilon) \setminus \{0\}$, f assumes every value from $\mathbb{C} \setminus \{0\}$, and this is already the general case:

Theorem 9.18 (**Stronger Version of the Casorati–Weierstraß Theorem**). *If f has an essential singularity at $a \in \Omega$, then a number $w_0 \in \mathbb{C}$ exists such that, in each ball $B(a,\varepsilon) \setminus \{a\}$ (whatever small number ε is), f attains each number from $\mathbb{C} \setminus \{w_0\}$.*

Remark 9.19. *The functions* Ln *and $\sqrt{\cdot}$ have no isolated singularities at 0, because they are not holomorphic on the punctured ball $B(0,\varepsilon) \setminus \{0\}$, since they have a jump type discontinuity when crossing a ray that starts at 0.*

Theorem 9.20 (**Cauchy Integral Formula for Annular Domains**). *Let $a \in \mathbb{C}$, $0 \leq r_1 < r_2 < r_3 < r_4 \leq \infty$, and f be holomorphic on the annular domain*

$$\Omega := \{z \in \mathbb{C} \colon r_1 < |a-z| < r_4\}.$$

[8] wesentliche Singularität
[9] Felice Casorati, 1835 – 1890
[10] dicht

If $r_2 < |a - z| < r_3$, *then*

$$f(z) = \frac{1}{2\pi \mathrm{i}} \oint_{|\zeta - a| = r_3} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta - \frac{1}{2\pi \mathrm{i}} \oint_{|\zeta - a| = r_2} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta,$$

both circles oriented counter-clockwise.

Sketch of Proof. Connect the circles $\partial B(a, r_2)$ and $\partial B(a, r_3)$ by two radial lines. Then the annular domain between $\partial B(a, r_2)$ and $\partial B(a, r_3)$ decomposes into two parts, each of them looking like a horse-shoe[11], and domains of such a shape are simply connected. Now apply the Cauchy Integral Formula to each of the two loops that encircle the horse-shoe domains. □

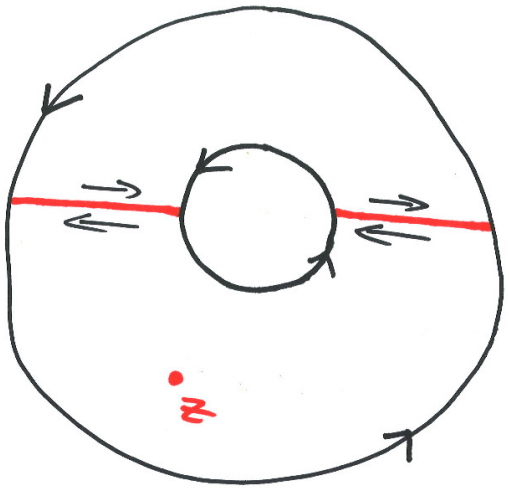


Figure 9.1: The Cauchy Integral Formula in annular domains

Now we are in a position to describe better how functions behave near essential singularities:

Theorem 9.21 (**Laurent series**[12])**.** *Let* $0 \le r < R \le \infty$ *and* f *be holomorphic on*

$$\Omega := \{z \in \mathbb{C} \colon r < |z - a| < R\}.$$

Then the expansion

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - a)^n, \qquad r < |z - a| < R,$$

*holds (*Laurent *series), and the coefficients are*

$$c_n = \frac{1}{2\pi \mathrm{i}} \oint_{|\zeta - a| = \varrho} \frac{f(\zeta)}{(\zeta - a)^{n+1}} \, \mathrm{d}\zeta, \qquad n \in \mathbb{Z},$$

with an arbitrary ϱ *between* r *and* R.
The two serieses $\sum_{n=-\infty}^{-1} c_n (z - a)^n$ *and* $\sum_{n=0}^{\infty} c_n (z - a)^n$ *converge uniformly on compact subsets of* Ω.

Proof. Fix z. Choose r_2 and r_3 with

$$r < r_2 < |z| < r_3 < R.$$

[11] Hufeisen
[12] Pierre Alphonse Laurent, 1813 – 1854

Then we can use Theorem 9.20 and find

$$\begin{aligned} f(z) &= F(z) + H(z), \\ F(z) &= \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_3} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta, \\ H(z) &= -\frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta, \end{aligned}$$

and F, H do not depend on r_2, r_3. The part F is holomorphic on $B(0,R)$; we have the power series expansion

$$F(z) = \sum_{n=0}^{\infty} c_n (z-a)^n, \qquad c_n = \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_3} \frac{f(\zeta)}{(\zeta-a)^{n+1}}\,\mathrm{d}\zeta,$$

and c_n is independent of r_3. Concerning the part H, we note that

$$\frac{1}{\zeta - z} = \frac{1}{(\zeta-a)-(z-a)} = \frac{-1}{z-a} \cdot \frac{1}{1 - \frac{\zeta-a}{z-a}} = -\frac{1}{z-a} \sum_{n=0}^{\infty} \left(\frac{\zeta-a}{z-a}\right)^n,$$

with convergence of the series because of $|\zeta - a| < |z-a|$. Then we find

$$\begin{aligned} H(z) &= -\frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{z-a} \sum_{n=0}^{\infty} \left(\frac{\zeta-a}{z-a}\right)^n \mathrm{d}\zeta \\ &= \sum_{n=0}^{\infty} \left(\frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{(\zeta-a)^{-n}}\,\mathrm{d}\zeta \right) \cdot (z-a)^{-1-n} \qquad \Big| \quad m := -1-n \\ &= \sum_{m=-\infty}^{-1} \left(\frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{(\zeta-a)^{m+1}}\,\mathrm{d}\zeta \right) \cdot (z-a)^{m} \\ &= \sum_{m=-\infty}^{-1} c_m (z-a)^m, \end{aligned}$$

where the coefficient

$$c_m = \frac{1}{2\pi\mathrm{i}} \oint_{|\zeta-a|=r_2} \frac{f(\zeta)}{(\zeta-a)^{m+1}}\,\mathrm{d}\zeta$$

is independent of $r_2 \in (r,R)$. □

In case of an isolated singularity, we may take $r = 0$, and then the expansion

$$f(z) = \sum_{m=-\infty}^{\infty} c_m (z-a)^m, \qquad 0 < |z-a| < R,$$

follows. Now three cases are possible:

Case 1: all c_m with $m < 0$ are zero: then f has a removable singularity at the point a.

Case 2: a finite number of c_m with $m < 0$ are non-zero: then f has a pole at the point a.

Case 3: an infinite number of c_m ($m < 0$) are non-zero: then f has an essential singularity at the point a.

Definition 9.22 (Singularities at infinity). *Suppose that f is holomorphic on $\Omega = \{z \in \mathbb{C} : r < |z-a| < \infty\}$. Then we say that f has a pole of order m at ∞ if $g(z) = f(1/z)$ has a pole of order m at 0, and f has an essential singularity at ∞ if g has an essential singularity at 0.*

For instance, $f(z) = z^2 - 1/z$ has a second order pole at ∞, and the sine function has an essential singularity at ∞.

Definition 9.23 (Meromorphic functions). *Let $\Omega \subset \mathbb{C}$ be open. A function f is called* meromorphic on Ω[13] *if a set $P^{(f)}$ (closed in Ω) exists such that f is holomorphic on $\Omega \setminus P^{(f)}$, and $P^{(f)}$ has no cluster points in Ω, and each element of $P^{(f)}$ is a pole of f or a removable singularity. The set $P^{(f)}$ is known as the* set of poles of f.

[13] meromorph auf Ω

9.3 The Residue Theorem

Lemma 9.24. *Let f be meromorphic on Ω, a a point in Ω, and assume that the Laurent series of f at a reads*

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z-a)^n, \qquad \forall\, z \text{ with } 0 < |z-a| < R.$$

If $0 < r < R$, then, with counter-clockwise orientation of the circle $\partial B(a,r)$,

$$\oint_{|z-a|=r} f(z)\,\mathrm{d}z = 2\pi \mathrm{i} c_{-1}.$$

Proof. By Theorem 9.21, the convergence of the Laurent series of f is uniform on compact subsets of the annular domain. A circle of radius r about a is compact, hence we can commute $\sum_n$ and $\oint$:

$$\oint_{|z-a|=r} f(z)\,\mathrm{d}z = \oint_{|z-a|=r} \sum_{n=-\infty}^{\infty} c_n (z-a)^n\,\mathrm{d}z = \sum_{n=-\infty}^{\infty} c_n \oint_{|z-a|=r} (z-a)^n\,\mathrm{d}z = 2\pi \mathrm{i} c_{-1},$$

following the computations from the Examples 8.3–8.5. □

Definition 9.25 (Residue[14]). *The coefficient c_{-1} of a Laurent series expansion of f at a point z_0 is called* residue of f at z_0[15]*,*

$$\operatorname*{res}_{z_0}(f) := c_{-1}.$$

The term with c_{-1} is the only one to survive the integration along the circle, all others disappear, compare the Examples 8.3–8.5.

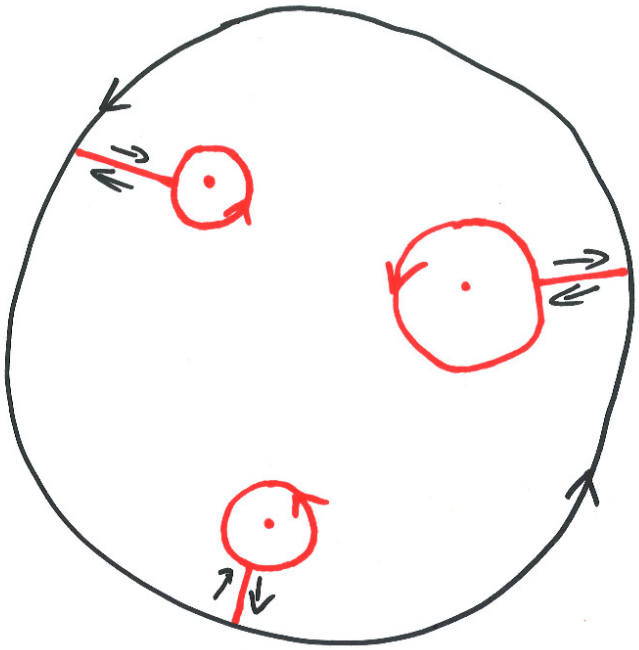


Figure 9.2: The residue theorem for a circle Γ

Theorem 9.26 (Residue Theorem for a circle). *Let f be meromorphic on Ω and $\Gamma \subset \Omega$ be a circle with counter-clockwise orientation, such that no singularity of f is on Γ. Let $z_1, \ldots, z_N$ denote the singularities of f in the bounded component of $\mathbb{C} \setminus \Gamma$.*

[14] residue = Nachlaß, Rest, Rückstand, Überbleibsel, Überrest
[15] Residuum of f at z_0

Then

$$\oint_\Gamma f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\sum_{k=1}^{N} \operatorname*{res}_{z_k} f.$$

Proof. Let $\Gamma_1, \ldots, \Gamma_N$ be circles with centres $z_1, \ldots, z_N$ and radii so small that none of the circles Γ, Γ_1, $\ldots$, Γ_N intersect. For each k, connect Γ and Γ_k by a straight line that does not intersect another circle or another such straight line. Then you can show, by the Cauchy integral theorem that

$$\oint_\Gamma f(z)\,\mathrm{d}z = \sum_{k=1}^{N} \oint_{\Gamma_k} f(z)\,\mathrm{d}z.$$

Now apply Lemma 9.24. □

Of course, the curve Γ need not be a circle, but in the general case, it is harder to describe how to connect Γ and the Γ_k by non-intersecting lines.

A more general version of the residue theorem (whose proof we skip) is:

Theorem 9.27 (General version of the Residue Theorem). *Let $\Omega \subset \mathbb{C}$ be a simply–connected domain, $\Gamma \subset \Omega$ a loop, and $f\colon \Omega \to \mathbb{C}$ meromorphic with the a finite set of poles $P^{(f)}$, and no pole of f is on Γ.*

Then

$$\oint_\Gamma f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\sum_{p\in P^{(f)}} \operatorname{Ind}_\Gamma(p)\cdot \operatorname*{res}_p(f).$$

As a **first example**, we consider the integral

$$I = \int_{x=-\infty}^{\infty} \frac{\mathrm{d}x}{2x^2+4x+20}, \qquad f(x) := \frac{1}{2x^2+4x+20}.$$

The poles of f are

$$z_{1,2} = -1 \pm 3\mathrm{i}, \qquad f(z) = \frac{1}{2(z-z_1)(z-z_2)}.$$

For $R \gg 1$, split I as follows:

$$I = I_1(R) + I_2(R) + I_3(R) := \int_{x=-\infty}^{-R} f(x)\,\mathrm{d}x + \int_{x=-R}^{R} f(x)\,\mathrm{d}x + \int_{x=R}^{\infty} f(x)\,\mathrm{d}x.$$

If R is large, then (compare the proof of Theorem 8.28)

$$|4x+20| \le \frac{1}{3}|2x^2| \qquad \text{for } |x| > R,$$

hence

$$|I_3(R)| \le \int_{x=R}^{\infty} \frac{\mathrm{d}x}{\frac{2}{3}\cdot 2x^2} = \mathfrak{O}(R^{-1}),$$

and also $|I_1(R)| = \mathfrak{O}(R^{-1})$. This gives $I = \lim_{R\to\infty} I_2(R)$. Call Γ_2 the straight line from $-R$ to R, and Γ_4 the half-circle parametrised by $\gamma(t) = Re^{\mathrm{i}t}$ with $0 \le t \le \pi$. Then $\Gamma_2 \cup \Gamma_4$ is a loop encircling z_1 but not z_2. The residue theorem then implies

$$\oint_{\Gamma_2\cup\Gamma_4} f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\operatorname*{res}_{z_1}(f).$$

On the other hand,

$$|f(z)| \le \frac{1}{\frac{2}{3}|2z^2|} = \frac{3}{4R^2} \qquad \text{for } z \in \Gamma_4,$$

which gives is

$$\left|\int_{\Gamma_4} f(z)\,\mathrm{d}z\right| \le \max_{z\in\Gamma_4}|f(z)|\cdot\pi R = \mathfrak{O}(R^{-1}),$$

and now we can argue like this:

$$\begin{aligned} I &= \lim_{R\to\infty} I_2(R) = \lim_{R\to\infty}\left(\oint_{\Gamma_2\cup\Gamma_4} f(z)\,\mathrm{d}z - \int_{\Gamma_4} f(z)\,\mathrm{d}z\right) = \lim_{R\to\infty}\left(2\pi\mathrm{i}\operatorname*{res}_{z_1}(f) - \mathfrak{O}(R^{-1})\right) \\ &= 2\pi\mathrm{i}\operatorname*{res}_{z_1}(f). \end{aligned}$$

To compute the residue of f at z_1, we remember that

$$\frac{1}{a+\varepsilon} = \frac{1}{a} + \mathfrak{O}(\varepsilon)$$

for all ε, $a\in\mathbb{C}$ with $|\varepsilon| \ll |a|$, and then the expansion for $z \approx z_1$ becomes

$$\begin{aligned} f(z) &= \frac{1}{2(z-z_1)(z-z_2)} = \frac{1}{z-z_1}\cdot\frac{1}{2(z-z_2)} = \frac{1}{z-z_1}\cdot\frac{1}{2(z_1-z_2)+2(z-z_1)} \\ &= \frac{1}{z-z_1}\cdot\left(\frac{1}{2(z_1-z_2)} + \mathfrak{O}(z-z_1)\right) = \frac{1}{z-z_1}\cdot\frac{1}{2(z_1-z_2)} + \mathfrak{O}(1), \end{aligned}$$

and then the residue is

$$\operatorname*{res}_{z_1}(f) = c_{-1} = \frac{1}{2(z_1-z_2)},$$

which results in

$$I = 2\pi\mathrm{i}\frac{1}{2\cdot 6\mathrm{i}} = \frac{\pi}{6}.$$

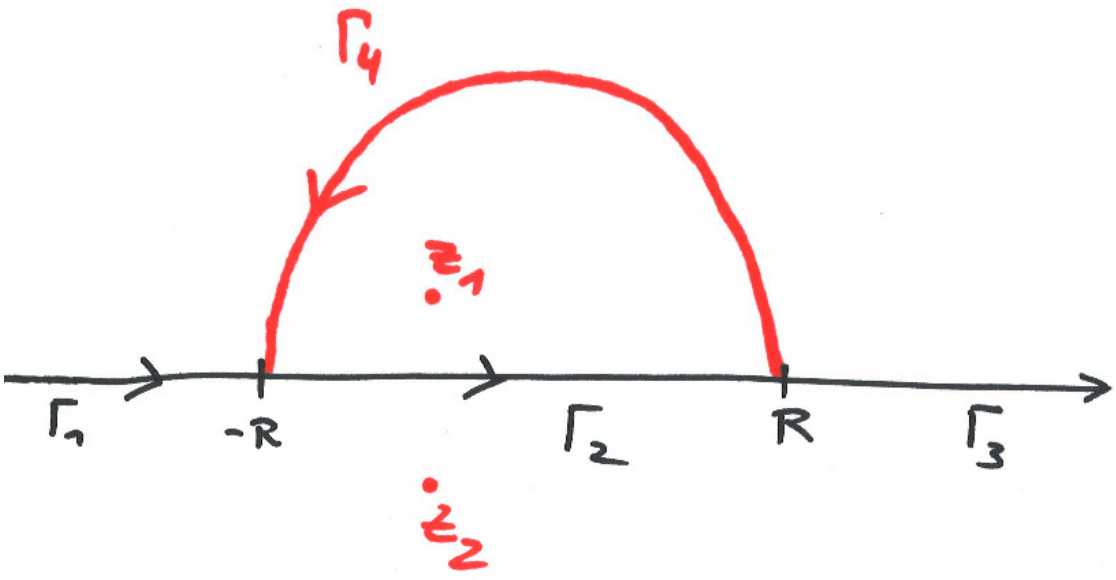


Figure 9.3: A first example to the residue theorem

This methods is applicable to rational functions of x where the degree of the polynomial in the numerator is at most equal to the degree of the polynomial in the denominator minus two (and the real axis must not contain a pole). Of course, we could have evaluated the integral using traditional methods like partial fraction decomposition, too.

A **second example** which is not accessible to traditional methods is

$$I = \int_{\mathbb{R}} \frac{e^{\mathrm{i}x}}{x-\mathrm{i}}\,\mathrm{d}x, \qquad f(x) := \frac{e^{\mathrm{i}x}}{x-\mathrm{i}},$$

with the integration along the real axis to be understood as $I := \lim_{R\to\infty} \int_{x=-R}^{x=R} \ldots\,\mathrm{d}x$. We extend the line $\Gamma_1 = [-R, R]$ to a square in the complex plane by the following curves:

$$\begin{aligned}
&\Gamma_2 \colon z = R + \mathrm{i}y, && 0 \le y \le 2R,\\
&\Gamma_3 \colon z = -x + 2\mathrm{i}R, && -R \le x \le R,\\
&\Gamma_4 \colon z = -R - \mathrm{i}y, && -2R \le y \le 0.
\end{aligned}$$

Then $\Gamma := \Gamma_1 \cup \ldots \cup \Gamma_4$ forms a loop with counter-clockwise orientation, and we have

$$\oint_\Gamma f(z)\,\mathrm{d}z = 2\pi\mathrm{i} \operatorname*{res}_{\mathrm{i}}(f),$$

because i is the only pole of the function f in $\mathbb{C}$. Now we estimate the integrals over the three new lines. On Γ_2, we have

$$\left|\frac{e^{\mathrm{i}z}}{z-\mathrm{i}}\right| \le \frac{e^{-y}}{R},$$

and therefore

$$\left|\int_{\Gamma_2} f(z)\,\mathrm{d}z\right| \le \int_{y=0}^{2R} \frac{1}{R} e^{-y}\,\mathrm{d}y \le \frac{1}{R},$$

and the same bound holds for the integral over Γ_4. And concerning Γ_3, we have

$$\left|\frac{e^{\mathrm{i}z}}{z-\mathrm{i}}\right| \le \frac{e^{-2R}}{R} \qquad \Longrightarrow \qquad \left|\int_{\Gamma_3} f(z)\,\mathrm{d}z\right| \le \frac{e^{-2R}}{R} \cdot 2R = 2e^{-R}.$$

The result then is

$$I = \lim_{R\to\infty} \int_{\Gamma_1} f(z)\,\mathrm{d}z = 2\pi\mathrm{i} \operatorname*{res}_{\mathrm{i}}(f),$$

and the residue of f at the point i can be evaluated via

$$\frac{e^{\mathrm{i}z}}{z-\mathrm{i}} = \frac{e^{\mathrm{i}\cdot\mathrm{i}}}{z-\mathrm{i}} \cdot e^{\mathrm{i}(z-\mathrm{i})} = \frac{e^{-1}}{z-\mathrm{i}} \cdot (1 + \mathfrak{O}(\mathrm{i}(z-\mathrm{i}))) = \frac{e^{-1}}{z-\mathrm{i}} + \mathfrak{O}(1), \qquad z \to \mathrm{i},$$

or $\operatorname{res}_{\mathrm{i}}(f) = e^{-1}$, which gives us eventually $I = 2\pi e^{-1}\mathrm{i}$.

Our **third example** is more delicate:

$$I = \int_{x=-\infty}^{\infty} \frac{\sin x}{x}\,\mathrm{d}x := \lim_{R\to\infty} \int_{-R}^{R} \frac{\sin x}{x}\,\mathrm{d}x = \pi.$$

Observe that the integrand $\frac{\sin x}{x}$ has no pole in $\mathbb{C}$, and moreover, the complex sine function

$$\sin z = \frac{1}{2\mathrm{i}}(e^{\mathrm{i}z} - e^{-\mathrm{i}z})$$

explodes exponentially for $\Im z \to +\infty$, because of the term $e^{-\mathrm{i}z}$. We overcome this difficulty exploiting the odd symmetry of the sine function:

$$\begin{aligned}
\int_{x=-R}^{-\varepsilon} \frac{\sin x}{x}\,\mathrm{d}x + \int_{x=\varepsilon}^{R} \frac{\sin x}{x}\,\mathrm{d}x &= 2\int_{x=\varepsilon}^{R} \frac{\sin x}{x} = \frac{1}{\mathrm{i}} \int_{x=\varepsilon}^{R} \frac{e^{\mathrm{i}x} - e^{-\mathrm{i}x}}{x}\,\mathrm{d}x\\
&= \frac{1}{\mathrm{i}} \int_{x=\varepsilon}^{R} \frac{e^{\mathrm{i}x}}{x}\,\mathrm{d}x + \frac{1}{\mathrm{i}} \int_{x=\varepsilon}^{R} \frac{e^{-\mathrm{i}x}}{-x}\,\mathrm{d}x \qquad \Big|\ \tilde{x} := -x\\
&= \frac{1}{\mathrm{i}} \int_{x=\varepsilon}^{R} \frac{e^{\mathrm{i}x}}{x}\,\mathrm{d}x + \frac{1}{\mathrm{i}} \int_{\tilde{x}=-R}^{-\varepsilon} \frac{e^{\mathrm{i}\tilde{x}}}{\tilde{x}}\,\mathrm{d}\tilde{x}\\
&= \frac{1}{\mathrm{i}} \int_{[-R,R]\setminus[-\varepsilon,\varepsilon]} \frac{e^{\mathrm{i}z}}{z}\,\mathrm{d}z.
\end{aligned}$$

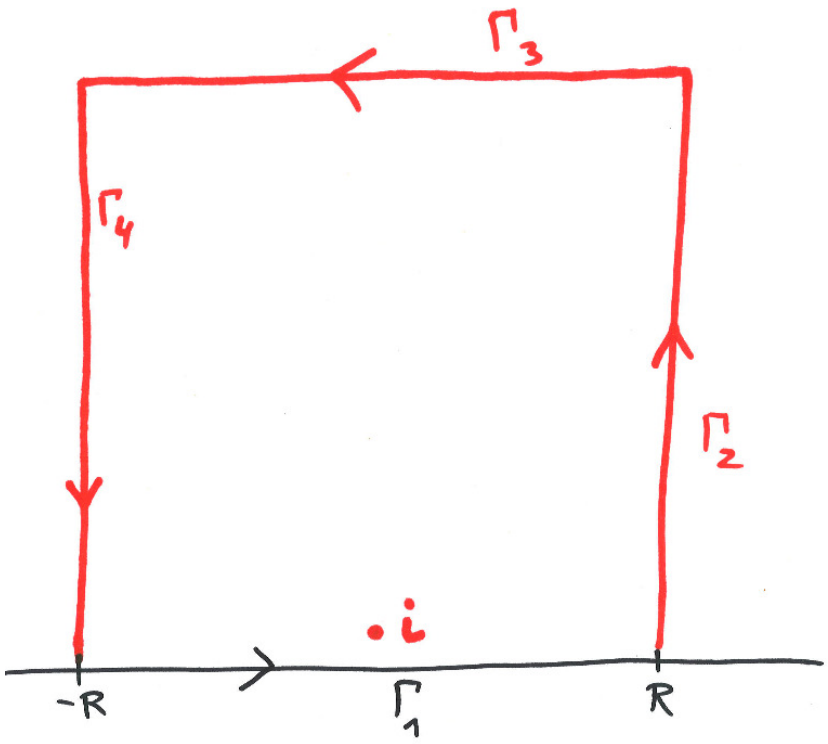


Figure 9.4: The second example to the residue theorem

Now we take a chain of six curves:

$$\Gamma_1 = [-R, -\varepsilon],$$
$$\Gamma_2 = \{\varepsilon e^{\mathrm{i}(\pi - t)} \colon 0 \le t \le \pi\},$$
$$\Gamma_3 = [\varepsilon, R],$$
$$\Gamma_4 = \{R + \mathrm{i}y \colon 0 \le y \le 2R\},$$
$$\Gamma_5 = \{-x + 2\mathrm{i}R \colon -R \le x \le R\},$$
$$\Gamma_6 = \{-R - \mathrm{i}y \colon -2R \le y \le 0\},$$

compare the figure. For brevity of notation, define $I_k = \int_{\Gamma_k} f(z)\,\mathrm{d}z$ with $f(z) = \frac{\exp(\mathrm{i}z)}{\mathrm{i}z}$. Then we have on the one hand

$$I = \lim_{(\varepsilon, R) \to (0, \infty)} (I_1 + I_3),$$

and on the other hand, from the Cauchy Integral Theorem,

$$I_1 + \cdots + I_6 = 0.$$

As in the second example, we show that

$$|I_4| + |I_5| + |I_6| = \mathfrak{O}(R^{-1}),$$

which brings us to

$$I = -\lim_{\varepsilon \to 0} I_2(\varepsilon).$$

By direct calculation, we have

$$I_2 = \int_{\Gamma_2} \frac{e^{\mathrm{i}z}}{\mathrm{i}z}\,\mathrm{d}z = \int_{t=0}^{\pi} \frac{\exp(\mathrm{i}\varepsilon e^{\mathrm{i}(\pi - t)})}{\mathrm{i}\varepsilon e^{\mathrm{i}(\pi - t)}} \cdot \varepsilon(-\mathrm{i})e^{\mathrm{i}(\pi - t)}\,\mathrm{d}t = -\int_{t=0}^{\pi} \exp(\mathrm{i}\varepsilon e^{\mathrm{i}(\pi - t)})\,\mathrm{d}t,$$

which converges to $-\pi$ for $\varepsilon \to 0$, and then our final result is

$$\int_{x=-\infty}^{\infty} \frac{\sin x}{x}\,\mathrm{d}x = \pi.$$

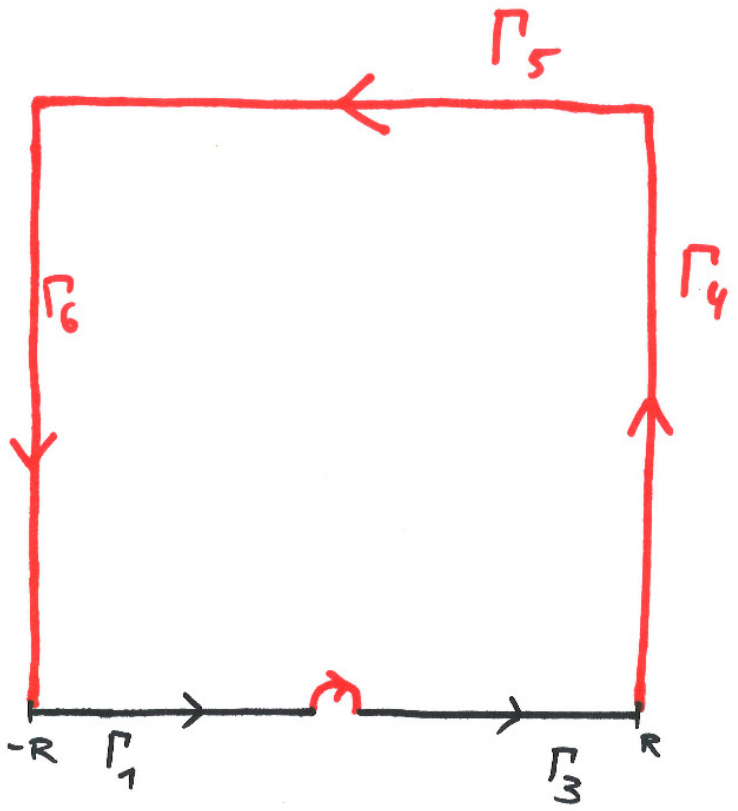


Figure 9.5: The third example to the residue theorem

Finally, we give some hints how to compute residues. Note that a function h with a zero of order m at z_0 possesses the Taylor expansion

$$h(z) = \sum_{k=m}^{\infty} \frac{1}{k!}(\partial_z^k h)(z_0) \cdot (z - z_0)^k = \frac{(\partial_z^m h)(z_0)}{m!}(z - z_0)^m \cdot (1 + \mathfrak{O}(z - z_0)).$$

- if $f = g/h$ and h has a single zero at z_0 then

$$f(z) = \frac{g(z_0) + g'(z_0)(z - z_0) + \dots}{h'(z_0) \cdot (z - z_0)(1 + \mathfrak{O}(z - z_0))} = \frac{g(z_0)}{h'(z_0) \cdot (z - z_0)} + \mathfrak{O}(1), \qquad z \to z_0,$$

  and then the residue is $\operatorname{res}_{z_0}(f) = \frac{g(z_0)}{h'(z_0)} = \lim_{z \to z_0}(z - z_0)f(z)$,

- if $f = g/(z - z_0)^m$ them

$$f(z) = \frac{g(z_0) + g'(z_0)(z - z_0) + \dots}{(z - z_0)^m} = \dots + \frac{\frac{1}{(m-1)!}g^{(m-1)}(z_0)}{z - z_0} + \dots, \qquad z \to z_0,$$

- if $f = g/h$ and h has an m–th order zero at z_0, then $h(z) = (z - z_0)^m w(z)$ with $w(z_0) \neq 0$, and we find

$$f(z) = \frac{g(z) \cdot (w(z))^{-1}}{(z - z_0)^m} \qquad \Longrightarrow \qquad \operatorname*{res}_{z_0}(f) = \frac{1}{(m-1)!}\left(\frac{g}{w}\right)^{(m-1)}(z_0),$$

- if $u = u(z)$ is holomorphic near z_0 and $v = v(z)$ has a single pole at z_0, then

$$\operatorname*{res}_{z_0}(uv) = u(z_0) \cdot \operatorname*{res}_{z_0}(v),$$

- if $u = u(z)$ is holomorphic near z_0 and $v = v(z)$ has a pole (of whatever order) at z_0, then

$$\operatorname*{res}_{z_0}(u + \alpha v) = \alpha \operatorname*{res}_{z_0}(v).$$

Chapter 10

Applications of Complex Analysis

10.1 Behaviour of Functions

In this section, Γ is always a loop that does not intersect itself, with counter-clockwise orientation. Let G denote the (unique) bounded component of $\mathbb{C} \setminus \Gamma$.

Suppose $f\colon \Omega \to \mathbb{C}$ is meromorphic, Γ contained in Ω, and you know from somewhere that f has no poles in the domain G, and f has no zero on the curve Γ.

How can you determine the number of zeroes of f in G ?

Lemma 10.1. *In the situation described above, the number of zeroes of f in G (counted according to multiplicity) is*

$$N = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z.$$

Proof. Let z_0 be an m–th order zero of f. Then, by Proposition 9.1, $f(z) = (z - z_0)^m g(z)$ with a holomorphic g that does not vanish near z_0, and

$$\frac{f'(z)}{f(z)} = \frac{m(z-z_0)^{m-1}g(z) + (z-z_0)g'(z)}{(z-z_0)^m g(z)} = \frac{m}{z-z_0} + \frac{g'(z)}{g(z)} \qquad \Longrightarrow \qquad \operatorname*{res}_{z_0}\left(\frac{f'}{f}\right) = m.$$

Now apply the residue theorem 9.27. □

For a general meromorphic function, the integral $\frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z$ equals $N - P$, the difference between the number of zeroes and the number of poles in G, both counted according to their multiplicities.

Corollary 10.2 (**Fundamental Theorem of Algebra**)**.** *Each polynomial of degree n possesses exactly n complex roots.*

Proof. Take $f = f(z) = a_n z + \cdots + a_1 z + a_0$ with $a_n \neq 0$. Choose $\Gamma = \partial B(0, R)$ with $R \gg 1$. Then (as we know already), f can not have zeroes outside the ball $B(0, R)$. On the circle Γ, we have (with appropriate numbers b_j and c_j which we do not need to compute)

$$\frac{f'(z)}{f(z)} = \frac{na_n z^{n-1} + \cdots + a_1}{a_n z^n + \cdots + a_0} = \frac{na_n z^{n-1}(1 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_{n-1} z^{-(n-1)})}{a_n z^n (1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_n z^{-n})}.$$

Put $q := -(c_1 z^{-1} + \cdots + c_n z^{-n})$. Then the summation formula for the geometric series gives

$$\frac{f'(z)}{f(z)} = \frac{n}{z}\left(1 + b_1 z^{-1} + \cdots + b_{n-1} z^{-(n-1)}\right)\left(1 + q + q^2 + \dots\right) = \frac{n}{z} + \sum_{j=2}^{\infty} \gamma_j z^{-j},$$

with new coefficients γ_j. This is a Laurent series in the annular domain $\Omega = \{z \in \mathbb{C}\colon R < |z| < \infty\}$, with uniform convergence of the series. Then

$$\oint_\Gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z = \oint_\Gamma \left(\frac{n}{z} + \sum_{j=2}^{\infty} \gamma_j z^{-j}\right)\mathrm{d}z = \oint_\Gamma \frac{n}{z}\,\mathrm{d}z + \sum_{j=2}^{\infty} \gamma_j \oint_\Gamma z^{-j}\,\mathrm{d}z = 2\pi \mathrm{i} n,$$

which was our goal. □

Now assume that we know from somewhere that f has exactly one zero in G, but no pole. How to find it ?

Lemma 10.3. *In this situation, the zero of f can be computed as*

$$z_0 = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{zf'(z)}{f(z)}\,\mathrm{d}z.$$

Proof. We have $f(z) = (z - z_0)g(z)$, and for z near z_0, g does not vanish. The function $z \mapsto z$ is holomorphic near z_0, and therefore

$$\operatorname*{res}_{z_0}\left(z \cdot \frac{f'(z)}{f(z)}\right) = z_0 \operatorname*{res}_{z_0}\left(\frac{f'}{f}\right) = z_0 \cdot 1,$$

as in the proof of lemma 10.1. □

As an application, consider a vibrating system. Then the eigenfrequencies are typically the eigenvalues of a certain matrix, and the entries of the matrix depend on parameters of the system (and on the errors in your measurements).

How do the eigenvalues depend on the perturbations of the coefficients ?

Lemma 10.4. *Let $a_{jk} = a_{jk}(\varepsilon)$ depend holomorphically on the parameter ε, for $1 \le j, k \le N$. Let $\lambda_1(0)$ be an eigenvalue of $A(0) = (a_{jk}(0))_{j,k=1,\dots,N}$ of algebraic multiplicity one.*

Then λ_1 depends analytically on ε.

Proof. Each eigenvalue of a matrix depends continuously on ε, because eigenvalues are zeroes of the characteristic polynomial, to which we can apply Lemma 10.1, with the consequence that no eigenvalue jumps if ε varies.

Now let Γ be a small circle about $\lambda_1(0)$, such that all the other eigenvalues $\lambda_2(0), \dots, \lambda_N(0)$ are outside Γ. Then

$$\lambda_1(0) = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{zf'(z)}{f(z)}\,\mathrm{d}z, \qquad f(z, \varepsilon) = \det(A(\varepsilon) - zI),$$

because f is clearly a holomorphic function of $z \in \mathbb{C}$. Now λ_1 depends holomorphically on $\varepsilon \in \mathbb{C}$, due to

$$\frac{\partial}{\partial \overline{\varepsilon}}\lambda_1(\varepsilon) = \frac{1}{2\pi \mathrm{i}} \frac{\partial}{\partial \overline{\varepsilon}} \oint_\Gamma \frac{zf'(z,\varepsilon)}{f(z,\varepsilon)}\,\mathrm{d}z = \frac{1}{2\pi \mathrm{i}} \oint_\Gamma \frac{\partial}{\partial \overline{\varepsilon}} \frac{zf'(z,\varepsilon)}{f(z,\varepsilon)}\,\mathrm{d}z = 0.$$

And each holomorphic function is analytic, therefore λ_1 can be expanded into a power series of ε. □

Warning: *The assumption of λ_1 to have algebraic multiplicity one is crucial. Take*

$$A(\varepsilon) = \begin{pmatrix} 0 & 1 \\ \varepsilon & 0 \end{pmatrix}.$$

Then $A(0)$ has the double eigenvalue zero, but

$$\lambda_1(\varepsilon) = -\sqrt{\varepsilon}, \qquad \lambda_2(\varepsilon) = +\sqrt{\varepsilon},$$

and this has no Taylor expansion at the point $\varepsilon_0 = 0$.

With a view towards applications, we note that a little investment (change the parameter from zero to ε) gives a larger gain (the difference of the eigenvalues changed from zero to $\sqrt{\varepsilon}$). A similar phenomenon occurs in the case of the Poincaré–Andronov–Hopf bifurcation: changing one parameter by ε gives rise to a stable periodic orbit of diameter $\sim \sqrt{\varepsilon}$. This can be observed in case of a variant of the VAN DER POL oscillator, which has several applications in electronics. Details can be found in [12].

Now imagine the following situation: you wish to know how many zeroes a function f has in a domain encircled by Γ, but f is complicated and evaluating the integral appearing in Lemma 10.1 is infeasible. But f is "approximated" by another function g, which is less complicated. Under which conditions have f and g the same number of zeroes in the domain inside Γ ?

Proposition 10.5 (**Theorem of Rouché**[1]). *Let Ω be a domain in $\mathbb{C}$, and $\Gamma \subset \Omega$, and f, $g\colon \Omega \to \mathbb{C}$ holomorphic. Suppose*

$$|g(z) - f(z)| < |f(z)| \qquad \forall\, z \in \Gamma.$$

Then the numbers of zeroes inside Γ of f and g coincide, $N(f) = N(g)$.

Proof. Let ω be a *tubular neighbourhood*[2] of Γ. If the "width" of ω is sufficiently small, then the inequality $|g(z) - f(z)| < |f(z)|$ holds also in ω. Then we have, for all $z \in \omega$:

$$\frac{g(z)}{f(z)} = 1 + \frac{g(z) - f(z)}{f(z)} \quad \text{with} \quad \left|\frac{g(z) - f(z)}{f(z)}\right| < 1 \quad \Longrightarrow \quad \Re\frac{g(z)}{f(z)} > 0,$$

hence the fraction $\frac{g(z)}{f(z)}$ only takes values in the right half-plane $\mathbb{C}_+$ of $\mathbb{C}$. Now define

$$u(z) := \frac{\mathrm{d}}{\mathrm{d}z}\operatorname{Ln}\left(\frac{g(z)}{f(z)}\right), \qquad z \in \omega.$$

Since $\frac{g(z)}{f(z)}$ never leaves $\mathbb{C}_+$, we can compute like this:

$$u(z) = \frac{1}{\frac{g(z)}{f(z)}} \cdot \left(\frac{g(z)}{f(z)}\right)' = \frac{f(z)}{g(z)} \cdot \frac{g'(z)f(z) - f'(z)g(z)}{f^2(z)} = \frac{g'(z)}{g(z)} - \frac{f'(z)}{f(z)}, \qquad z \in \omega.$$

The function u possesses a primitive function on ω, namely $\operatorname{Ln}(g/f)$. By Proposition 8.8, we have

$$\oint_\Gamma u(z)\,\mathrm{d}z = 0.$$

(We can not utilise the Cauchy Integral Theorem because ω is doubly–connected.) However, $u(z) = \frac{g'(z)}{g(z)} - \frac{f'(z)}{f(z)}$ in ω. Now apply Lemma 10.1. □

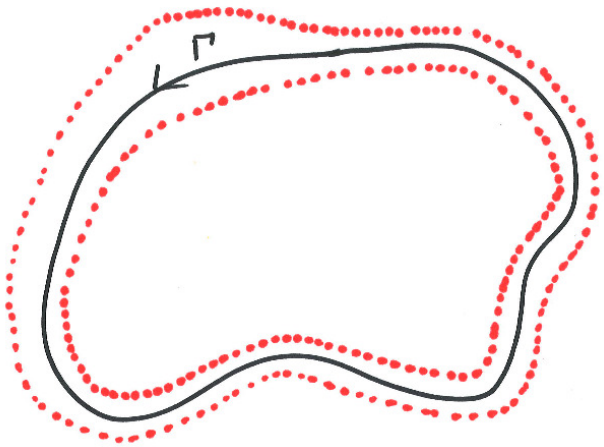


Figure 10.1: A tubular neighbourhood ω of the curve Γ

[1] Eugène Rouché, 1832 – 1910
[2] schlauchförmige Umgebung

10.2 The Laplace Transform

Definition 10.6 (**Exponential growth order and Laplace transform**). *For a function $f\colon [0,\infty) \to \mathbb{C}$ that is piecewise continuous and has* exponential growth order *M,*

$$\exists\, C_f\colon\ |f(t)| \le C_f e^{Mt}, \qquad \forall\, t \in [0,\infty),$$

we define the Laplace transform $F = \mathcal{L}\{f\}$ as

$$F(z) := \int_{t=0}^{\infty} e^{-tz} f(t)\,\mathrm{d}t, \qquad z \in \mathbb{C}_{\Re>M} := \{z \in \mathbb{C}\colon \Re z > M\}.$$

The following examples can be found by direct computation:

$f = f(t)$	$F = F(z)$
$t^n, \quad n \in \mathbb{N}_0$	$n!/z^{n+1}, \quad \Re z > 0$
$t^\alpha, \quad \alpha > -1$	$\Gamma(\alpha+1)/z^{\alpha+1}, \quad \Re z > 0$
$e^{\alpha t}, \quad \alpha \in \mathbb{C}$	$1/(z-\alpha), \quad \Re z > \Re\alpha$
$\sin(\omega t), \quad \omega \in \mathbb{R}$	$\omega/(z^2+\omega^2), \quad \Re z > 0$
$\cos(\omega t), \quad \omega \in \mathbb{R}$	$z/(z^2+\omega^2), \quad \Re z > 0$

Lemma 10.7. *The Laplace transform F of the function f with exponential growth order M is holomorphic on $\mathbb{C}_{\Re>M}$.*

Proof. Exercise. □

Proposition 10.8 (**Inverse Laplace transform**). *Let $f\colon [0,\infty) \to \mathbb{C}$ be continuous except a finite number of jumps, and with exponential growth order M. Then f can be obtained from $F = \mathcal{L}\{f\}$ by the formula*

$$\frac{1}{2\pi\mathrm{i}} \lim_{R\to\infty} \int_{\gamma-\mathrm{i}R}^{\gamma+\mathrm{i}R} e^{zt} F(z)\,\mathrm{d}z = \begin{cases} f(t) & : f \text{ is continuous at } t \\ \frac{1}{2}(f(t+0)+f(t-0)) & : f \text{ jumps at } t \\ \frac{1}{2}f(0) & : t = 0, \end{cases}$$

where $\gamma > M$ can be chosen freely, and the integration is performed along the vertical straight line from $\gamma - \mathrm{i}R$ to $\gamma + \mathrm{i}R$.

Proof. For $\gamma > M$, define

$$f_\gamma(t) = \begin{cases} 0 & : t < 0, \\ e^{-\gamma t} f(t) & : t \ge 0 \end{cases}$$

which belongs to $L^1(\mathbb{R}^1)$, because of $\gamma > M$. The Fourier transform $\hat{f}_\gamma$ of f_γ is

$$\begin{aligned} \hat{f}_\gamma(\tau) &= \int_{t=-\infty}^{\infty} e^{-\mathrm{i}t\tau} f_\gamma(t)\,\mathrm{d}t = \int_{t=0}^{\infty} e^{-\mathrm{i}t\tau} e^{-\gamma t} f(t)\,\mathrm{d}t = \int_{t=0}^{\infty} e^{-(\gamma+\mathrm{i}\tau)t} f(t)\,\mathrm{d}t \\ &= \mathcal{L}\{f\}(\gamma+\mathrm{i}\tau), \qquad \tau \in \mathbb{R}. \end{aligned}$$

In the appendix, Lemma A.16 presents the inversion formula for the Fourier transform, which reads (for a point t where f is continuous)

$$\begin{aligned} f_\gamma(t) &= \frac{1}{2\pi} \lim_{R\to\infty} \int_{\tau=-R}^{R} e^{\mathrm{i}t\tau} \hat{f}_\gamma(\tau)\,\mathrm{d}\tau = \frac{1}{2\pi} \lim_{R\to\infty} \int_{\tau=-R}^{R} e^{\mathrm{i}t\tau} F(\gamma+\mathrm{i}\tau)\,\mathrm{d}\tau \qquad \Big| \quad z := \gamma+\mathrm{i}\tau \\ &= \frac{1}{2\pi\mathrm{i}} \lim_{R\to\infty} \int_{\gamma-\mathrm{i}R}^{\gamma+\mathrm{i}R} e^{(z-\gamma)t} F(z)\,\mathrm{d}z, \end{aligned}$$

and for $t > 0$, we have

$$f(t) = e^{\gamma t} f_\gamma(t) = \frac{1}{2\pi\mathrm{i}} \lim_{R\to\infty} \int_{\gamma-\mathrm{i}R}^{\gamma+\mathrm{i}R} e^{zt} F(z)\,\mathrm{d}z.$$

□

Warning: *Some care is necessary here. Call $\mathcal{M}$ the set of all functions $F = F(z)$ that are holomorphic in a half-plane $\mathbb{C}_{\Re>M}$, and for which the integrals $\int_{\gamma-\mathrm{i}\infty}^{\gamma+\mathrm{i}\infty} e^{tz}F(z)\,\mathrm{d}z$ give finite values, for all $\gamma > M$ and all $t \geq 0$. This set $\mathcal{M}$ contains all those functions to which the inversion formula can be reasonably applied.*

Then the Laplace transform is ***not*** *a surjective map onto $\mathcal{M}$. This means: there are functions $F = F(z)$ to which you can apply the inversion formula, and this inversion formula gives you a function $f = f(t)$ with exponential growth, but $F \neq \mathcal{L}\{f\}$ because F is the Laplace transform of nobody.*

A positive result is the next one.

Lemma 10.9. *Let F be a holomorphic function for all $z \in \mathbb{C}$ with $\Re z > M$, and assume that*

- *for any $\delta > 0$, $F(z)$ converges uniformly to zero for $z \to \infty$, where $\Re z \geq M + \delta$,*
- *for any $\delta > 0$, the line integral $\int_{M+\delta-\mathrm{i}\infty}^{M+\delta+\mathrm{i}\infty} |F(z)|\,\mathrm{d}z$ is bounded.*

Then F is the Laplace transform of a function f which can be found by the inversion formula.

Proof. This is Satz 3 in Chapter 7 of [7]. See also the other two volumes of that magnum opus. □

To perform the inversion, the following procedure can be helpful in the case that F is a meromorphic function on $\mathbb{C}$ with $|F(z)| \leq C|z|^{-\varepsilon}$ for some positive ε, for all large $|z|$. Define $I_1 := \int_{\Gamma_1} e^{tz}F(z)\,\mathrm{d}z$ with Γ_1 being the straight line from $\gamma - \mathrm{i}R$ to $\gamma + \mathrm{i}R$. Then draw a rectangle as in Figure 10.2, with $\Gamma_2 \cup \Gamma_3$ being being the upper edge, Γ_2 in the right half-plane, Γ_3 in the left half-plane. Similarly $\Gamma_5 \cup \Gamma_6$ is the lower edge, and Γ_4 is the left edge. Define $I_k = \int_{\Gamma_k} e^{tz}F(z)\,\mathrm{d}z$.

The integrals $I_3, \ldots, I_5$ can be discussed as in Figure 9.4, and we find $|I_3|+|I_4|+|I_5| \leq CR^{-\varepsilon}$. Concerning I_2, we remember (8.1), hence

$$|I_2| \leq Ce^{t\gamma}R^{-\varepsilon} \cdot \gamma,$$

which approaches zero for $R \to \infty$, and therefore I_1 is approximated by $2\pi\mathrm{i}$ times the sum of the residues of $z \mapsto e^{tz}F(z)$ in the rectangle.

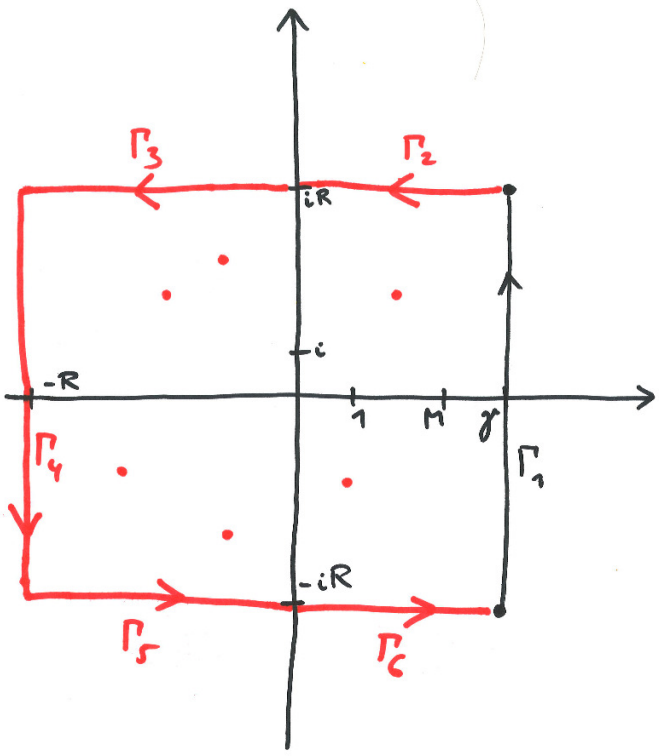


Figure 10.2: Computing the inverse Laplace transform

Proposition 10.10 (**Laplace transform of a derivative**). *Let the function $f\colon [0,\infty) \to \mathbb{C}$ be continuous, piecewise differentiable, and of exponential growth order M. Then*

$$\mathcal{L}\{f'\}(z) = z \cdot \mathcal{L}\{f\}(z) - f(0), \qquad \forall\, z \in \mathbb{C}_{\Re > M}.$$

Proof. Partial integration in the formula of $\mathcal{L}\{f'\}$. □

This can be remembered as follows:

If $f(0) = 0$, then differentiation in the t–world
corresponds to multiplication by z in the z–world.

We generalise a bit:

Lemma 10.11. *Let the function $f\colon [0,\infty) \to \mathbb{C}$ be continuous with one jump at $t_* > 0$, piecewise differentiable, and of exponential growth of order M. Then*

$$\mathcal{L}\{f'\}(z) = z \cdot \mathcal{L}\{f\}(z) - f(0) - (f(t_* + 0) - f(t_* - 0)) \cdot e^{-t_* z} \qquad \forall\, z \in \mathbb{C}_{\Re > M}.$$

Proof. Careful partial integration in the formula of $\mathcal{L}\{f'\}$. □

It is no surprise that we have also $\mathcal{L}\{f''\}(z) = z^2 \cdot \mathcal{L}\{f\}(z) - z \cdot f(0) - f'(0)$, and similar formulae for higher order derivatives.

Example: *To find the solution to $y''(t) + y(t) = \sin(3t)$ with $y(0) = 2$ and $y'(0) = 7$, we define $Y = Y(z)$ as the Laplace transform of y, write $f(t) = \sin(3t)$, and it follows that*

$$\begin{aligned}
(z^2 Y(z) - z \cdot 2 - 7) + Y(z) = F(z) &= \frac{3}{z^2+9}, \\
(z^2+1)Y(z) &= \frac{3}{z^2+9} + 2z + 7, \\
Y(z) &= \frac{3}{(z^2+1)(z^2+9)} + \frac{2z+7}{z^2+1} = \frac{\alpha}{z-\mathrm{i}} + \frac{\beta}{z+\mathrm{i}} + \frac{\gamma}{z-3\mathrm{i}} + \frac{\delta}{z+3\mathrm{i}},
\end{aligned}$$

with some easily computable constants $\alpha, \ldots, \delta$, and then we can directly express y:

$$y(t) = \alpha e^{\mathrm{i}t} + \beta e^{-\mathrm{i}t} + \gamma e^{3\mathrm{i}t} + \delta e^{-3\mathrm{i}t}.$$

We could have solved this initial value problem also by classical methods, of course. The real power of the Laplace transform becomes visible when we study *linear time invariant systems*. Think of an electronic amplifier, or an electronic devices which adds an echo to an acoustical signal. For such a device, you have an input signal which gets transformed into an output signal. The assumptions are:

- the output depends linearly on the input,
- the system does not depend itself on the time variable,
- the system obeys causality: the output can not depend on values of the input from the future. The present output can depend on the present input and the input of the past, of course[3].

Then the output signal must depend on the input signal via a *convolution*[4]:

Definition 10.12. *Let $u, v\colon [0,\infty) \to \mathbb{C}$ be piecewise continuous. Then the convolution $(u*v)\colon [0,\infty) \to \mathbb{C}$ is defined as*

$$(u * v)(t) := \int_{s=0}^{t} u(t-s)v(s)\,\mathrm{d}s, \qquad 0 \le t < \infty.$$

[3] An extreme example is a bottle of ketchup. The input signal are the motions of the bottle, and the output signal is the viscosity of the ketchup. This system is certainly time invariant, but it seems to be nonlinear.

[4] Faltung

Think of u as input, v as *impulse response function*[5] (or conversely, because the convolution is commutative), and $u * v$ as the output signal.

To give examples of the convolution, we first define $E = E(t)$ as that function that has the value one everywhere.

- $\int_0^t u(s)\,\mathrm{d}s = (E * u)(t)$,
- if $v(t) = \exp(at)$, then the solution $y = y(t)$ to the problem

  $$y'(t) = ay(t) + f(t), \quad y(0) = y_0,$$

  is given as $y(t) = v(t) \cdot y_0 + (v * f)(t)$, compare (3.3),
- if $v = v(t)$ solves

  $$v''(t) + av(t) = 0, \qquad v(0) = 0, \qquad v'(0) = 1,$$

  then the solution $y = y(t)$ to the problem

  $$y''(t) + ay(t) = f(t), \qquad y(0) = y_0, \qquad y'(0) = y_1$$

  is given by

  $$y(t) = v'(t) \cdot y_0 + v(t) \cdot y_1 + (v * f)(t). \tag{10.1}$$

Lemma 10.13. *If u and v are piecewise continuous and of exponential growth orders M_u and M_v, then $u * v$ is continuous and of exponential growth order $\max(M_u, M_v) + \varepsilon$, for an arbitrary $\varepsilon > 0$.*

Proof. The continuity of $u * v$ follows from the boundedness of u and v. We know $|u(t)| \le C_u e^{M_u t}$ and $|v(t)| \le C_v e^{M_v t}$, and therefore

$$\begin{aligned} |(u * v)(t)| &\le C_u C_v \int_{s=0}^{t} \exp(\max(M_u, M_v) \cdot (t - s)) \cdot \exp(\max(M_u, M_v) \cdot s)\,\mathrm{d}s \\ &= C_u C_v \exp(\max(M_u, M_v) \cdot t) \int_{s=0}^{t} 1\,\mathrm{d}s. \end{aligned}$$

Now exploit $t \le C_\varepsilon e^{\varepsilon t}$ for all positive ε. □

Proposition 10.14 (**Laplace transform of a convolution**). *If u and v are piecewise continuous and of exponential growth, then*

$$\mathcal{L}\{u * v\} = \mathcal{L}\{u\} \cdot \mathcal{L}\{v\}.$$

Proof. Exercise for Fubini's theorem. Be careful with the limits of the integrals. □

This can be remembered as follows:

A convolution in the t–world corresponds to a multiplication in the z–world.

This enables us to handle differential equations like

$$y''(t) + 4y(t) = \int_{s=0}^{t} \exp(-2(t - s))y(s)\,\mathrm{d}s, \qquad y(0) = 0, \qquad y'(0) = 1.$$

We come to an application. Consider an infinite chain of atoms, and each atom is connected to its two neighbours by harmonic oscillators. The deviation of the n–th atom from its resting position is $u_n = u_n(t)$, and Newton's Law $ma = F$ then brings us to

$$u_n''(t) = u_{n-1}(t) - 2u_n(t) + u_{n+1}(t), \qquad n \in \mathbb{Z}, \quad t \geq 0, \tag{10.2}$$

where the interaction constants and the masses have been normalized to one, and we have the initial conditions

$$u_n(0) = u_n^{(0)}, \qquad u_n'(0) = u_n^{(1)}, \qquad n \in \mathbb{Z}. \tag{10.3}$$

Our goal is an explicit solution formula.

We will show:

Lemma 10.15. *Assume that only a finite number of the initial values $u_n^{(0)}$ and $u_n^{(1)}$ is not zero. Then a solution to (10.2), (10.3) is given by*

$$u_n(t) = \sum_{m\in\mathbb{Z}} \left(u_m^{(0)} \cdot J_{2(n-m)}(2t) + u_m^{(1)} \cdot \left(\sum_{l=|n-m|}^{\infty} J_{2l+1}(2t) \right) \right), \qquad n \in \mathbb{Z},$$

where J_ν, $\nu \in \mathbb{C}$ are the Bessel functions:

$$J_\nu(t) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!\Gamma(m+\nu+1)} \left(\frac{t}{2}\right)^{2m+\nu}, \qquad t \in \mathbb{C}.$$

Some remarks are in order:

- it is a nice exercise (comparing the coefficients of all power series involved) that

  $$J_{2(n-m)}(2t) = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{l=|n-m|}^{\infty} J_{2l+1}(2t),$$

  and this corresponds to (10.1),
- we can read n and m as spatial variables, and then $\sum_m u_m^{(0)} J_{2(n-m)}(2t)$ can be understood as a convolution in the space $\mathbb{Z}$ where the spatial variables live,
- The bessel function $J_k(s)$ is exponentially small for $s < |k|/2$, it has the biggest contribution for $s \sim |k|$, and it decays slowly for $s > |k|$ (with decay rate as $s^{-1/2}$, but it possesses additionally some oscillations). Compare (2.5) and (2.6) and the graphs of several Bessel functions in that chapter. If we assume (for instance) that the initial values are non-zero only for $|n| \leq 3$, then the solution (at time t) is concentrated at the location $|n| \approx t$ (just find all those n where $2|n-m| \approx 2t$). This means that two waves are travelling through the oscillator chain (one wave to the left and one wave to the right), and the wave speed is approximately equal to one. Higher velocities are only possible at the price of an exponential damping, but smaller velocities can occur. The wave speed depends on the spatial frequency, which is known as dispersion in physics.

There are several ways to prove Lemma 10.15:

- The mathematically most correct proof is: just check that the functions u_n given in the above formula are indeed solutions to (10.2) and (10.3), using formulas as e.g.

  $$\begin{aligned} \cos(z\sin\theta) &= J_0(z) + 2\sum_{l=1}^{\infty} J_{2l}(z)\cos(2l\theta), \\ J_{-n}(z) &= (-1)^n J_n(z), \\ J_{\nu-1}(z) - J_{\nu+1}(z) &= 2J_\nu'(z), \end{aligned}$$

  which can perhaps be shown via comparison of all the coefficients in the power series (or via comparison of the Laplace transforms).

[5] Impulsantwort

- Another approach is maybe more helpful for a deeper understanding of the solution formula: we show, that each solution **must** have this form. A rigorous proof of this fact is beyond our reach (because we miss the tools from functional analysis and the time), so we settle for something less and accept a reduced logical precision. This is no logical problem since we can afterwards follow the methods of the first •.

Remark 10.16. *We note that the oscillator chain system possesses a mechanical energy. We define*

$$\mathcal{E}(t) := \frac{1}{2}\sum_{n\in\mathbb{Z}} \left((u_n'(t))^2 + (u_n(t) - u_{n-1}(t))^2\right),$$

and here we assume that $u_n(t)$ and $u_n'(t)$ decay for $|n| \to \infty$ at such a high rate that the series converges. Then we have (assuming that the u_n form a solution)

$$\mathcal{E}'(t) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\sum_{n\in\mathbb{Z}} \left((u_n'(t))^2 + (u_n(t) - u_{n-1}(t))^2\right) \overset{\spadesuit}{=} \frac{1}{2}\sum_{n\in\mathbb{Z}} \frac{\mathrm{d}}{\mathrm{d}t}\left((u_n'(t))^2 + (u_n(t) - u_{n-1}(t))^2\right) = 0,$$

and here we have supposed that the series decays so fast that it allows the step ♠. (In the case which is relevant for us, the terms $u_n(t)$ and $u_n'(t)$ decay (for fixed t) exponentially in $|n|$, and then ♠ is definitely no problem).

*Then we get $\mathcal{E}(t) = \mathcal{E}(0)$, and from this we deduce that vanishing initial data (which means $u_n^{(0)} = u_n^{(1)} = 0$ $(\forall\, n)$) also implies $u_n(t) = 0$ for all t and all n, and therefore the solutions to the problem (10.2), (10.3) are **unique**, assuming that only such solutions are considered for which ♠ is valid.*

The motivation of the solution formula in Lemma 10.15 needs some preparations:

Exercise: *Consider a function $F = F(t,z)\colon (\mathbb{C}\setminus\{0\})\times\mathbb{C}$ defined by*

$$F(t,z) := \exp\left(\frac{z}{2}\left(t - \frac{1}{t}\right)\right) = e^{\frac{zt}{2}}\cdot e^{\frac{-z}{2t}}$$

This function is holomorphic in each of its two variables. Let $c_n(z)$ denote its Laurent coefficients with respect to the variable t:

$$F(t,z) =: \sum_{n=-\infty}^{\infty} c_n(z)t^n, \qquad z\in\mathbb{C}, \quad t\in\mathbb{C}\setminus\{0\}.$$

Use the integral formula of the Laurent coefficients to prove that

$$c_n(z) = \frac{1}{\pi}\int_{\theta=0}^{\pi} \cos(n\theta - z\sin\theta)\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{\theta=-\pi}^{\pi} e^{-\mathrm{i}(n\theta - z\sin\theta)}\,\mathrm{d}\theta.$$

Use a power series expansion for $e^{zt/2}$ and $e^{-z/(2t)}$ to show that

$$c_n(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(n+k)!}\left(\frac{z}{2}\right)^{2k+n}.$$

*In particular, you get $c_n \equiv J_n$ for all $n\in\mathbb{Z}$. We say that F is the **generating function**[6] of the Bessel functions J_n.*

Lemma 10.17. *The Laplace transform of the function*

$$K_{a,\nu}\colon t\mapsto a^\nu J_\nu(at), \qquad a>0, \qquad \nu>-1,$$

with J_ν being the Bessel function, is

$$\mathcal{L}\{K_{a,\nu}\}\colon z\mapsto \frac{\left(\sqrt{z^2+a^2}-z\right)^\nu}{\sqrt{z^2+a^2}}, \qquad \Re z>0,$$

with $\sqrt{\cdot}$ and $(\cdot)^\nu$ as the principal branches.

[6] Erzeugende Funktion

Beautiful proof in case of $\nu = n \in \mathbb{Z}$. By the previous exercise, we have

$$J_n(at) = \frac{1}{2\pi}\int_{\tau=-\pi}^{\pi} e^{-\mathrm{i}(n\tau - at\sin\tau)}\,\mathrm{d}\tau,$$

and this then brings us to

$$\begin{aligned}
\mathcal{L}\{K_{a,n}\}(z) &= \frac{1}{2\pi}\int_{t=0}^{\infty} e^{-tz}\left(a^n\int_{\tau=-\pi}^{\pi} e^{-\mathrm{i}(n\tau - at\sin\tau)}\,\mathrm{d}\tau\right)\mathrm{d}t \\
&= \frac{a^n}{2\pi}\int_{\tau=-\pi}^{\pi} e^{-\mathrm{i}n\tau}\left(\int_{t=0}^{\infty} e^{-tz}e^{\mathrm{i}at\sin\tau}\,\mathrm{d}t\right)\mathrm{d}\tau = \frac{a^n}{2\pi}\int_{\tau=-\pi}^{\pi} e^{-\mathrm{i}n\tau}\frac{1}{z-\mathrm{i}a\sin\tau}\,\mathrm{d}\tau \\
&= \frac{1}{2\pi}\int_{\tau=-\tau}^{\tau} \left(ae^{-\mathrm{i}\tau}\right)^n \frac{1}{z-\frac{a}{2}(e^{\mathrm{i}\tau}-e^{-\mathrm{i}\tau})}\,\mathrm{d}\tau \\
&= \frac{1}{2\pi}\int_{\tau=-\tau}^{\tau} \left(ae^{-\mathrm{i}\tau}\right)^n \frac{2\left(ae^{-\mathrm{i}\tau}\right)}{2z\cdot ae^{-\mathrm{i}\tau} - a^2 + \left(ae^{-\mathrm{i}\tau}\right)^2}\,\mathrm{d}\tau.
\end{aligned}$$

Here we substitute $w = ae^{-\mathrm{i}\tau}$ and $\mathrm{d}w = -\mathrm{i}w\,\mathrm{d}\tau$. We end up with a circle ∂B of radius a about the origin, with clockwise orientation. Consequently

$$\mathcal{L}\{K_{a,n}\}(z) = \frac{1}{2\pi}\oint_{\partial B} w^n \frac{2w}{2z\cdot w - a^2 + w^2}\frac{1}{-\mathrm{i}w}\,\mathrm{d}w = \frac{-1}{2\pi\mathrm{i}}\oint_{\partial B}\frac{2w^n}{w^2+2zw-a^2}\,\mathrm{d}w.$$

The zeros of the denominator are $w_{1,2} = -z \pm \sqrt{z^2+a^2}$, and now we suppose temporarily that $z \in \mathbb{R}$ with $z \gg a > 0$. Then ∂B encircles only $w_1 = -z + \sqrt{z^2+a^2}$, and we obtain

$$\mathcal{L}\{K_{a,n}\}(z) = \operatorname*{res}_{w=w_1}\left\{\frac{2w^n}{(w-w_1)(w-w_2)}\right\} = \frac{2w_1^n}{(w_1-w_2)} = \frac{\left(-z+\sqrt{z^2+a^2}\right)^n}{\sqrt{z^2+a^2}}.$$

This is our claim in case of $z \in \mathbb{R}$ with $z \gg a$. Now apply the identity theorem for analytic functions. □

Corollary 10.18. *Put $P_n(t) := J_{2n-1}(2t) + J_{2n+1}(2t)$ for $n \in \mathbb{N}_+$. Then we have*

$$P_n * P_m = P_{n+m}, \qquad n, m \in \mathbb{N}_+.$$

Try to prove it without Laplace transformation !

Proof. It suffices to show that

$$\mathcal{L}\{P_{n+m}\} = \mathcal{L}\{P_n\}\cdot\mathcal{L}\{P_n\}, \qquad n, m \in \mathbb{N}_+,$$

but this is a routine calculation:

$$\begin{aligned}
\mathcal{L}\{P_n\}(z) &= \frac{(\sqrt{z^2+4}-z)^{2n-1}}{2^{n-1}\sqrt{z^2+4}}\left(1+\frac{(\sqrt{z^2+4}-z)^2}{4}\right), \\
1+\frac{(\sqrt{z^2+4}-z)^2}{4} &= 1+\frac{(z^2+4)-2z\sqrt{z^2+4}+z^2}{4} = \frac{1}{2}\left((z^2+4)-z\sqrt{z^2+4}\right) \\
&= \frac{1}{2}\sqrt{z^2+4}\left(\sqrt{z^2+4}-z\right), \\
\mathcal{L}\{P_n\}(z) &= \left(\frac{\sqrt{z^2+4}-z}{2}\right)^{2n} = \left(\sqrt{1+\frac{z^2}{4}}-\frac{z}{2}\right)^{2n}.
\end{aligned}$$

□

Lemma 10.19. *One solution to (10.2) with the special initial conditions*

$$\begin{cases} u_n(0) = 0, & n \in \mathbb{Z}, \\ u_n'(0) = \begin{cases} 1 & : n = 0, \\ 0 & : n \neq 0. \end{cases} \end{cases} \tag{10.4}$$

is given by

$$u_n(t) = \sum_{l=|n|}^{\infty} J_{2l+1}(2t), \qquad n \in \mathbb{Z}, \tag{10.5}$$

Logically correct proof. Just check the formula. This is a nice exercise in doing computations without getting lost in them. □

The next considerations are of heuristic nature and shall explain where the solution formula comes from.

Pseudo–proof. It seems physically reasonable that two waves emanate from the atom number zero: one wave travels to the right, the other wave travels to the left.

By uniqueness of the solution: if the infinite vector $(\dots, u_{-3}, u_{-2}, u_{-1}, u_0, u_1, u_2, u_3, \dots)(t)$ solves (10.2) and the special initial conditions (10.4), then also the reflected vector $(\dots, u_3, u_2, u_1, u_0, u_{-1}, u_{-2}, u_{-3}, \dots)(t)$ solves (10.2) and (10.4). The reflection simply means to turn around the crystal by 180°. Because of the uniqueness, both solutions must coincide, which means in particular that $u_1(t) = u_{-1}(t)$, for all t. This is physically reasonable: the atom number zero moves to the right for small times because of $u_0'(t=0) = 1$, hence it pushes the atom number one to the right, and it pulls the atom number -1 to the right. This means that the atoms with the numbers -1 and $+1$ both move to the right for small times. Hence $u_{-1} \equiv u_{+1}$ is physically plausible.

Consider the atom number n with $n \geq 1$. This atom and its neighbour at position $(n+1)$ are influenced only from a wave coming from the left. Because n is positive, the information here is travelling from left to right, but not from right to left. Therefore, we expect a **linear time invariant system**, and we conjecture a relation

$$u_{n+1}(t) = (Q * u_n)(t), \qquad t \geq 0,$$

with a function $Q = Q(t)$ not yet known. And the atoms are indistinguishable, hence we expect

$$u_n(t) = (Q * u_{n-1})(t), \qquad t \geq 0,$$

with the same function Q. Then we have

$$u_n'(t) = \frac{\mathrm{d}}{\mathrm{d}t} \int_{s=0}^{t} Q(t-s) u_{n-1}(s)\, \mathrm{d}s = Q(0) u_{n-1}(t) + (Q' * u_{n-1})(t),$$

and now we make the **assumption** $Q(0) = 0$. Taking one time derivative more yields

$$u_n''(t) = Q'(0) u_{n-1}(t) + (Q'' * u_{n-1})(t).$$

On the other hand, we know

$$\begin{aligned} u_n''(t) &= u_{n-1}(t) - 2u_n(t) + u_{n+1}(t) \\ &= u_{n-1}(t) - 2(Q * u_{n-1})(t) + (Q * Q * u_{n-1})(t). \end{aligned}$$

We equate both representations of u_n'':

$$Q'(0) u_{n-1}(t) + (Q'' * u_{n-1})(t) = u_{n-1}(t) - 2(Q * u_{n-1})(t) + (Q * Q * u_{n-1})(t).$$

Now it is physically reasonable that $Q(t)$ grows for $t \to \infty$ at most at exponential speed. Then the Laplace transform is applicable, and we arrive at

$$Q'(0) \hat{u}_{n-1}(z) + (Q'')\hat{}(z) \cdot \hat{u}_{n-1}(z) = \hat{u}_{n-1}(z) - 2\hat{Q}(z) \cdot \hat{u}_{n-1}(z) + \left(\hat{Q}(z)\right)^2 \hat{u}_{n-1}(z).$$

We may divide by $\hat{u}_{n-1}$:

$$(Q'')\hat{}(z) + Q'(0) = 1 - 2\hat{Q}(z) + \left(\hat{Q}(z)\right)^2.$$

Now we have the general rule $(Q'')\hat{}(z) = z^2 \hat{Q}(z) - zQ(0) - Q'(0)$, hence

$$z^2 \hat{Q}(z) = 1 - 2\hat{Q}(z) + \left(\hat{Q}(z)\right)^2,$$

and this equation has the two solutions

$$\hat{Q}(z) = \frac{2+z^2}{2} \pm \sqrt{\left(\frac{2+z^2}{2}\right)^2 - 1}.$$

We think about the behaviour of $\hat{Q}(z)$ for $\mathbb{R} \ni z \to +\infty$:

$$\lim_{z\to+\infty} \hat{Q}(z) = \lim_{z\to+\infty} \int_{t=0}^{\infty} e^{-tz} Q(t)\,\mathrm{d}t \overset{\heartsuit}{=} \int_{t=0}^{\infty} \lim_{z\to+\infty} e^{-tz} \cdot Q(t)\,\mathrm{d}t = \int_{t=0}^{\infty} 0 \cdot Q(t)\,\mathrm{d}t = 0.$$

(The step $\heartsuit$ deserves a justification !) This consideration excludes the $+$ in the $\pm$ symbol, hence we get

$$\begin{aligned}
\hat{Q}(z) &= \frac{2+z^2}{2} - \sqrt{\left(\frac{2+z^2}{2}\right)^2 - 1} = \frac{2+z^2}{2} - \frac{1}{2}\sqrt{4+4z^2+z^4-4} \\
&= \frac{2+z^2}{2} - z\sqrt{1+\frac{z^2}{4}} = \frac{z^2}{4} - 2\cdot\frac{z}{2}\cdot\sqrt{1+\frac{z^2}{4}} + \left(1+\frac{z^2}{4}\right) \\
&= \left(\frac{z}{2} - \sqrt{1+\frac{z^2}{4}}\right)^2 = \left(\frac{z}{2} + \sqrt{1+\frac{z^2}{4}}\right)^{-2},
\end{aligned}$$

and from this we conclude that

$$Q(t) = P_1(t) = J_1(2t) + J_3(2t) = \frac{2}{t} J_2(2t),$$

where the relation

$$J_{\nu-1}(s) + J_{\nu+1}(s) = \frac{2\nu}{s} J_\nu(s),$$

has been used (a nice exercise).

Now we determine $u_0(t)$, the position of the 0–th atom at time t. We have already shown that $u_1 = u_{-1}$, hence

$$u_0''(t) = u_1 + u_{-1} - 2u_0 = 2(Q * u_0)(t) - 2u_0(t), \qquad u_0(0) = 0, \quad u_0'(0) = 1,$$

and then the Laplace transform implies

$$(u_0'')\hat{}(z) = z^2 \hat{u}_0(z) - z u_0(0) - u_0'(0) = z^2 \hat{u}_0(z) - 1 = 2\hat{Q}(z)\cdot\hat{u}_0(z) - 2\hat{u}_0(z),$$

and the last equality can be condensed into

$$\left(z^2 - 2\hat{Q}(z) + 2\right)\hat{u}_0(z) = 1,$$

and this simplifies as follows:

$$\begin{aligned}
\left(z^2 - (2 + z^2 - z\sqrt{z^2+4}) + 2\right)\hat{u}_0(z) &= 1, \\
z\sqrt{z^2+4}\cdot\hat{u}_0(z) &= 1, \\
z\hat{u}_0(z) - 0 &= \frac{1}{\sqrt{z^2+4}}, \qquad \Big| \quad u_0(0) = 0, \\
\mathcal{L}\{u_0'\}(z) &= \frac{1}{\sqrt{z^2+4}}, \\
u_0'(t) &= J_0(2t),
\end{aligned}$$

and then u_0 is determined as

$$u_0(t) = \int_{s=0}^{t} J_0(2s)\,\mathrm{d}s = \sum_{l=0}^{\infty} J_{2l+1}(2t) = P_1(t) + P_3(t) + P_5(t) + \ldots,$$

because of $J_{\nu-1}(\sigma) - J_{\nu+1}(\sigma) = 2J_\nu'(\sigma)$ and $\lim_{\nu\to\infty} J_\nu(\sigma) = 0$ for each fixed σ.

Now we can compute u_1, u_2, …, via

$$u_1 = Q * u_0 = P_1 * (P_1 + P_3 + P_5 + \ldots) = P_2 + P_4 + P_6 + \cdots = \sum_{l=1}^{\infty} J_{2l+1}(2t),$$

$$u_2 = Q * u_1 = P_1 * (P_2 + P_4 + P_6 + \ldots) = P_3 + P_5 + P_7 + \cdots = \sum_{l=2}^{\infty} J_{2l+1}(2t),$$

and so on. This finishes the pseudo–proof of (10.5). □

Pseudo–proof of Lemma 10.15. First we consider more general initial conditions:

$$u_n(0) = 0, \qquad u_n'(0) = u_n^{(1)}, \qquad n \in \mathbb{Z}.$$

By superposition, we find

$$u_n(t) = \sum_{m\in\mathbb{Z}} u_m^{(1)} \cdot \left(\sum_{l=|n-m|}^{\infty} J_{2l+1}(2t) \right),$$

and this holds under the assumption $\sum_{m\in\mathbb{Z}} |u_m^{(1)}|^2 < \infty$, which corresponds to $\mathcal{E}(0) < \infty$.

Second we consider the general initial data (10.3). Appealing again to (10.1), we expect the solution as

$$\begin{aligned} u_n(t) &= \sum_{m\in\mathbb{Z}} \left(u_m^{(0)} \cdot \left(\frac{\mathrm{d}}{\mathrm{d}t} \sum_{l=|n-m|}^{\infty} J_{2l+1}(2t) \right) + u_m^{(1)} \cdot \left(\sum_{l=|n-m|}^{\infty} J_{2l+1}(2t) \right) \right) \\ &= \sum_{m\in\mathbb{Z}} \left(u_m^{(0)} \cdot J_{2(n-m)}(2t) + u_m^{(1)} \cdot \left(\sum_{l=|n-m|}^{\infty} J_{2l+1}(2t) \right) \right), \end{aligned}$$

due to $J_{-n}(z) = (-1)^n J_n(z)$. This is the solution formula we wanted to find, and the natural condition on the initial values is encoded into the finiteness of the energy, $\mathcal{E}(0) < \infty$. □

We also need a justification why the time derivatives (in the second step of the pseudo–proof of Lemma 10.15) can be commuted with the sums $\sum_m$ and $\sum_l$. For fixed $z \in \mathbb{C}$ and $\nu \to \infty$, $J_\nu(z)$ decays exponentially in the sense of

$$J_\nu(z) \sim \frac{1}{\sqrt{2\pi\nu}} \left(\frac{ez}{2\nu} \right)^\nu,$$

which means that the quotient of left-hand side and right-hand side goes to one. This takes care of $\sum_l$.

And concerning $\sum_m$, we argue like this: Call $\mathcal{H}_\mathcal{E}$ the vector space of all the initial data for which the energy is finite: $\mathcal{E}(0) < \infty$. Equip this space with the norm induced by the energy. This is the physically relevant space.

Call $\mathcal{H}_{\text{fin}}$ the vector space of all the initial data $(u_m^{(0)}, u_m^{(1)})$ for which only a finite number of entries is non-zero. Equip this space also with the energy norm. This is the mathematically easy space because then $\sum_m$ contains only a finite number of terms.

Now it is just a computation to show that the above solution candidate formula indeed produces a solution if the initial data come from $\mathcal{H}_{\text{fin}}$. Moreover, the map that transports the initial data to the solution at time t is continuous in the energy norm (because the energy is even conserved), and $\mathcal{H}_{\text{fin}}$ is a dense vector subspace of $\mathcal{H}_\mathcal{E}$. Therefore the solution formula holds also for initial data from $\mathcal{H}_\mathcal{E}$.

Remark 10.20. *Related results (and a completely different proof of the solution formula) can be found in [8], available in the KOPS of the University of Konstanz.*

10.3 Outlook: Maxwell's Equations in the Vacuum

Our notation follows [10], Chapter IV.

The Maxwell Equations read

$$\begin{aligned}
\operatorname{div} D &= 4\pi\varrho,\\
\operatorname{rot} E &= -\frac{1}{c}B_t,\\
\operatorname{div} B &= 0,\\
\operatorname{rot} H &= \frac{4\pi}{c}j + \frac{1}{c}D_t,
\end{aligned}$$

where the vector fields are connected via

$$D = E + 4\pi P, \qquad B = H + 4\pi M,$$

with P as the polarisation field, M the magnetisation field. In an isotropic medium that is normally polarisable and magnetisable, we have

$$D = \varepsilon E, \qquad B = \mu H,$$

and in the vacuum, we even have $\varepsilon = \mu = 1$ which we suppose from now on.

By $\operatorname{rot}\operatorname{rot} = \operatorname{grad}\operatorname{div} - \triangle$, we find

$$\begin{aligned}
E_{tt} &= D_{tt}\\
&= \partial_t(c\operatorname{rot} H - 4\pi j)\\
&= c\operatorname{rot} H_t - 4\pi j_t\\
&= -c^2 \operatorname{rot}\operatorname{rot} E - 4\pi j_t\\
&= -c^2 \operatorname{grad}\operatorname{div} E + c^2 \triangle E - 4\pi j_t\\
&= c^2 \triangle E - (4\pi c^2 \operatorname{grad}\varrho + 4\pi j_t),
\end{aligned}$$

or

$$(\partial_t^2 - c^2\triangle)E = -(4\pi c^2 \operatorname{grad}\varrho + 4\pi j_t).$$

We consider the right-hand side as known and wish to find the electric field $E = E(t,x)$.

More generally, we study

$$(\partial_t^2 - c^2\triangle)u(t,x) = f(t,x), \qquad (t,x) \in \mathbb{R}\times\mathbb{R}^3,$$

with known f and unknown u. For a philosophical reason named *causality*, we consider only those solutions reasonable, whose value $u(t,x)$ does **not** depend on values $f(s,y)$ with $s > t$. Instead, the value $u(t,x)$ can depend only on $f(s,y)$ with $y \in \mathbb{R}^3$ and $s \in (-\infty, t]$. Now the interval $(-\infty,t]$ is too long to perform the following computations (and also the solutions are certainly not uniquely determined), and therefore we temporarily settle for something less:

To solve is (by an understandable solution formula)

$$\begin{cases}
(\partial_t^2 - c^2\triangle)u(t,x) = f(t,x),\\
u(0,x) = u_0(x),\\
u_t(0,x) = u_1(x),
\end{cases}$$

with given f, u_0, u_1, and the function u is searched. We demand that $u(t,x)$ depends only on u_0, u_1, and $f = f(s,y)$ with $(s,y) \in [0,t]\times\mathbb{R}^3$.

To solve it, we want to apply the Fourier transform, and this is easier if we assume that u_0, u_1 and f decay fast for $|x| \to \infty$:

$$u_0, \quad u_1 \in \mathcal{S}(\mathbb{R}^3), \qquad f \in C([0,\infty), \mathcal{S}(\mathbb{R}^3)).$$

Then we set

$$\hat{u}(t,\xi) := \mathcal{F}_{x\to\xi}\{u(t,\cdot)\}(\xi) := \int_{x\in\mathbb{R}^3} e^{-\mathrm{i}x\xi} u(t,x)\,\mathrm{d}x,$$

and we have $\mathcal{F}\{\triangle u(t,\cdot)\}(\xi) = -|\xi|^2\hat{u}(t,\xi)$ with $|\xi|^2 = \xi_1^2+\xi_2^2+\xi_3^2$ as usual. This gives us

$$\begin{cases} \partial_t^2\hat{u}(t,\xi) + c^2|\xi|^2\hat{u}(t,\xi) = \hat{f}(t,\xi), \\ \hat{u}(0,\xi) = \hat{u}_0(\xi), \\ \hat{u}_1(0,\xi) = \hat{u}_1(\xi), \end{cases}$$

and this is an ODE in the variable t with the parameter ξ. To write down the solution formula, we define

$$\hat{U}(t,\xi) := \begin{cases} \frac{\sin(c|\xi|t)}{c|\xi|} & : \xi \neq 0, \\ t & : \xi = 0, \end{cases}$$

and then $\hat{u}(t,\xi)$ is given by

$$\hat{u}(t,\xi) = \hat{U}_t(t,\xi)\hat{u}_0(\xi) + \hat{U}(t,\xi)\hat{u}_1(\xi) + \int_{s=0}^{t} \hat{U}(s,\xi)\hat{f}(t-s,\xi)\,\mathrm{d}s.$$

The fundamental solution $\hat{U} = \hat{U}(t,\xi)$ solves

$$\begin{cases} \left(\partial_t^2 + c^2|\xi|^2\right)\hat{U}(t,\xi) = 0, \\ \hat{U}(0,\xi) = 0, \\ \hat{U}_t(0,\xi) = 1, \end{cases}$$

and $\hat{U}$ is the Fourier transform of $U = U(t,x)$:

$$U(t,x) = \mathcal{F}^{-1}_{\xi\to x}\{\hat{U}(t,\xi)\}(x) := \int_{\mathbb{R}^3_\xi} e^{\mathrm{i}x\xi}\hat{U}(t,\xi)\,đ\xi, \qquad đ\xi := \frac{\mathrm{d}\xi}{(2\pi)^3}.$$

The function which we are really interested in is $u = u(t,x)$:

$$\begin{aligned} u(t,x) &= \mathcal{F}^{-1}_{\xi\to x}\left(\hat{U}_t(t,\xi)\hat{u}_0(\xi) + \hat{U}(t,\xi)\hat{u}_1(\xi) + \int_{s=0}^{t}\hat{U}(t-s,\xi)\hat{f}(s,\xi)\,\mathrm{d}s\right) \\ &= \int_{\mathbb{R}^3_y} U_t(t,x-y)u_0(y)\,\mathrm{d}y + \int_{\mathbb{R}^3_y} U(t,x-y)u_1(y)\,\mathrm{d}y \\ &\qquad + \int_{s=0}^{t}\int_{\mathbb{R}^3_y} U(t-s,x-y)f(s,y)\,\mathrm{d}y\,\mathrm{d}s, \end{aligned}$$

and this last equality holds if U were a function (soon we will learn that this is not the case).

Therefore it seems advantageous to get a deeper understanding of the integral kernel $U = U(t,x)$. We will find an explicit formula of U by: first a Laplace transform for the time variable, then an inverse Fourier transform for the space variable, and finally an inverse Laplace transform for the time variable (see [10], Chapter 20, for a different approach).

$$U(t,x) = \mathcal{L}^{-1}_{\tau\to t}\mathcal{F}^{-1}_{\xi\to x}\{\mathcal{L}_{t\to\tau}\hat{U}(t,\xi)\},$$

and now we evaluate the transformations one after the other. The first one can be read from our table:

$$\begin{aligned} \mathcal{L}_{t\to\tau}\left\{\hat{U}(t,\xi)\right\}(\tau) &= \int_{t=0}^{\infty} e^{-\tau t}\hat{U}(t,\xi)\,\mathrm{d}t \\ &= \int_{t=0}^{\infty} e^{-\tau t}\frac{\sin(c|\xi|t)}{c|\xi|}\,\mathrm{d}t = \frac{1}{c|\xi|}\cdot\frac{c|\xi|}{\tau^2+(c|\xi|)^2} \\ &= \frac{1}{\tau^2+c^2|\xi|^2}, \qquad \Re\tau > 0. \end{aligned}$$

For $\xi\in\mathbb{R}^3$ and $\Re\tau > 0$, we never divide by zero here.

Next we replace ξ by x, via the inverse Fourier transform.

Lemma 10.21. *The Fourier transform and the inverse Fourier transform map functions with rotational symmetry to functions with rotational symmetry.*

Proof. Rotational symmetry for a function $\hat{\psi} = \hat{\psi}(\xi)$ means $\hat{\psi}(A\xi) = \hat{\psi}(\xi)$ for all $\xi \in \mathbb{R}^n$ and for each rotation matrix $A \in SO(n)$.
Then $A^\top = A^{-1}$ and $\det A = 1$, hence

$$\begin{aligned}
\psi(Ax) &= \int_{\mathbb{R}^n_\xi} e^{\mathrm{i}(Ax)\xi} \hat{\psi}(\xi)\, đ\xi \\
&= \int_{\mathbb{R}^n_\xi} \exp(\mathrm{i}\,\langle Ax, \xi\rangle) \hat{\psi}(\xi)\, đ\xi \\
&= \int_{\mathbb{R}^n_\xi} \exp(\mathrm{i}\,\langle x, A^*\xi\rangle) \hat{\psi}(\xi)\, đ\xi \quad \Big| \quad \xi =: A\eta, \\
&= \int_{\mathbb{R}^n_\eta} \exp(\mathrm{i}\,\langle x, \eta\rangle) \hat{\psi}(A\eta)\, đ\eta \\
&= \int_{\mathbb{R}^n_\eta} e^{\mathrm{i}x\eta} \hat{\psi(\eta)}\, đ\eta \\
&= \psi(x).
\end{aligned}$$

The computation for the forward Fourier transform runs similarly. □

In our case, the function $\xi \mapsto \frac{1}{\tau^2 + c^2|\xi|^2}$ is rotationally symmetric. This function does not belong to $L^1(\mathbb{R}^3_\xi)$, but to $L^2(\mathbb{R}^3_\xi)$, and therefore it suffices to choose $x = (0, 0, x_3)^\top$ with $x_3 = |x| > 0$, and compute

$$\begin{aligned}
\mathcal{F}^{-1}_{\xi\to x}\left\{\frac{1}{\tau^2 + c^2|\xi|^2}\right\}(x) &= \lim_{R\to\infty} \int_{|\xi|<R} e^{\mathrm{i}x_3\xi_3} \frac{1}{\tau^2 + c^2|\xi|^2} \frac{\mathrm{d}\xi}{(2\pi)^3} \\
&= \frac{1}{(2\pi)^3} \lim_{R\to\infty} \int_{r=0}^{R} \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} \exp(\mathrm{i}x_3 r\cos\theta) \frac{2}{\tau^2 + c^2 r^2} \cdot r^2 \sin\theta\, \mathrm{d}\varphi\, \mathrm{d}\theta\, \mathrm{d}r \quad \Big| \quad \cos\theta = s, \\
&= \frac{1}{(2\pi)^2} \lim_{R\to\infty} \int_{r=0}^{R} \int_{s=-1}^{1} \exp(\mathrm{i}r|x|s) \frac{1}{\tau^2 + c^2 r^2} \cdot r^2\, \mathrm{d}s\, \mathrm{d}r \\
&= \frac{1}{(2\pi)^2} \lim_{R\to\infty} \int_{r=0}^{R} \frac{r^2}{\tau^2 + c^2 r^2} \left(\int_{s=-1}^{1} \exp(\mathrm{i}r|x|s)\, \mathrm{d}s \right) \mathrm{d}r \\
&= \frac{1}{(2\pi)^2} \lim_{R\to\infty} \int_{r=0}^{R} \frac{r^2}{\tau^2 + c^2 r^2} \left(2\int_{s=0}^{1} \cos(sr|x|)\, \mathrm{d}s \right) \mathrm{d}r \\
&= \frac{1}{(2\pi)^2} \lim_{R\to\infty} \int_{r=0}^{R} \frac{r^2}{\tau^2 + c^2 r^2} \cdot \frac{2}{r|x|} \sin(r|x|)\, \mathrm{d}r \\
&= \frac{2}{(2\pi)^2|x|} \lim_{R\to\infty} \int_{r=0}^{R} \frac{r}{\tau^2 + c^2 r^2} \sin(r|x|)\, \mathrm{d}r \\
&= \frac{2}{(2\pi)^2|x|} \lim_{R\to\infty} \int_{r=0}^{R} \frac{r}{\tau^2 + c^2 r^2} \cdot \frac{1}{2\mathrm{i}} \left(e^{\mathrm{i}r|x|} - e^{-\mathrm{i}r|x|} \right) \mathrm{d}r \\
&= \frac{1}{(2\pi)^2|x|\mathrm{i}} \lim_{R\to\infty} \int_{r=0}^{R} \left(\frac{re^{\mathrm{i}r|x|}}{\tau^2 + c^2 r^2} + \frac{(-r)e^{-\mathrm{i}r|x|}}{\tau^2 + c^2(-r)^2} \right) \mathrm{d}r \\
&= \frac{1}{(2\pi)^2|x|\mathrm{i}} \lim_{R\to\infty} \int_{r=-R}^{R} \frac{re^{\mathrm{i}r|x|}}{\tau^2 + c^2 r^2}\, \mathrm{d}r.
\end{aligned}$$

And now the residue theorem comes in handy. We suppose that $\tau \in \mathbb{R}_+$ is real, and then the denominator has zeroes at

$$r_1 = +\mathrm{i}\frac{\tau}{c}, \quad r_2 = -\mathrm{i}\frac{\tau}{c},$$

and we have

$$\tau^2 + c^2 r^2 = c^2 (r - r_1)(r - r_2).$$

By expanding the integration into the upper half-plane, we then find

$$\begin{aligned}\mathcal{F}^{-1}_{\xi\to x}\left\{\frac{1}{\tau^2+c^2|\xi|^2}\right\}(x) &= \frac{1}{(2\pi)^2|x|\mathrm{i}}\cdot 2\pi\mathrm{i}\cdot\operatorname*{res}_{r_1}\left(\frac{re^{\mathrm{i}r|x|}}{c^2(r-r_1)(r-r_2)}\right)\\ &= \frac{1}{2\pi|x|}\cdot\frac{r_1e^{\mathrm{i}r_1|x|}}{c^2(r_1-r_2)}\\ &= \frac{1}{2\pi|x|}\cdot\frac{r_1e^{\mathrm{i}r_1|x|}}{c^2\cdot 2r_1}\\ &= \frac{1}{4\pi c^2}\cdot\frac{\exp\left(-\frac{|x|\tau}{c}\right)}{|x|}.\end{aligned}$$

We have proved this formula for real $\tau\in\mathbb{R}_+$, but by the identity theorem, it holds also for general τ from the right half-plane of $\mathbb{C}$.

Observe that this function has a pole at $x=0$, hence it does not belong to $L^\infty(\mathbb{R}^3_x)$. This is no surprise: because if $1/(\tau^2+c^2|\xi|^2)$ were a member of $L^1(\mathbb{R}^3_\xi)$, then its Fourier transform would belong to $L^\infty(\mathbb{R}^3_x)$.

The inverse Laplace transform applied to this function is supposed to give $U=U(t,x)$. However, if we do the computation, it turns out that we get a function of t that takes everywhere the value zero, except one value of t where it explodes. To find the exact coefficient in front of the Delta distribution, we introduce an artifical factor $1/(1+\varepsilon\tau)$, and later we send $\varepsilon\to+0$:

$$\begin{aligned}\mathcal{L}^{-1}\left\{\frac{1}{4\pi c^2}\cdot\frac{\exp\left(-\frac{|x|\tau}{c}\right)}{|x|(1+\varepsilon\tau)}\right\}(t) &= \frac{1}{2\pi\mathrm{i}}\cdot\frac{1}{4\pi c^2|x|}\lim_{R\to\infty}\int_{\tau=\gamma-\mathrm{i}R}^{\gamma+\mathrm{i}R}e^{t\tau}\frac{\exp\left(-\frac{|x|\tau}{c}\right)}{1+\varepsilon\tau}\,\mathrm{d}\tau\\ &= \frac{1}{4\pi c^2|x|\varepsilon}\cdot\frac{1}{2\pi\mathrm{i}}\lim_{R\to\infty}\int_{\tau=\gamma-\mathrm{i}R}^{\gamma+\mathrm{i}R}\exp\left(\left(t-\frac{|x|}{c}\right)\tau\right)\frac{1}{\tau-(-\varepsilon^{-1})}\,\mathrm{d}\tau\\ &= \frac{1}{4\pi c^2|x|\varepsilon}\cdot\mathcal{L}^{-1}\left\{\frac{1}{\tau-(-\varepsilon^{-1})}\right\}\left(t-\frac{|x|}{c}\right).\end{aligned}$$

Now recall that $\tau\mapsto 1/(\tau-\alpha)$ is the Laplace transform of that function that is defined as $e^{\alpha t}$ for positive t, and defined as 0 for negative t. Then it follows that

$$U_\varepsilon(t,x) := \frac{1}{4\pi c^2|x|\varepsilon}\cdot\mathcal{L}^{-1}\left\{\frac{1}{\tau-(-\varepsilon^{-1})}\right\}\left(t-\frac{|x|}{c}\right) = \begin{cases}\frac{1}{4\pi c^2|x|\varepsilon}\cdot\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right) & : t>\frac{|x|}{c},\\ \frac{1}{8\pi c^2|x|\varepsilon}\cdot\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right) & : t=\frac{|x|}{c},\\ 0 & : t<\frac{|x|}{c}.\end{cases}$$

The set

$$\{(t,x)\in\mathbb{R}^{1+3}\colon t\geq 0,\quad ct=|x|\}$$

is called *(forward) light cone*[7], and we expect that the above function U_ε is concentrated near the light cone for $\varepsilon\approx 0$. To describe this phenomenon more in detail, we choose a smooth test function $\varphi=\varphi(t,x)$, and ask for

$$\lim_{\varepsilon\to 0}\int_{t=0}^{\infty}\int_{\mathbb{R}^3_x}U_\varepsilon(t,x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t.$$

We may assume that $\varphi(t,x)=0$ for $t>\frac{|x|}{c}+1$, because such points give no relevant contribution to the

[7] Vorwärtslichtkegel

integral for small ε. Then, by partial integration,

$$\begin{aligned}
\int_{t=0}^{\infty}\int_{\mathbb{R}^3_x} U_\varepsilon(t,x)\varphi(t,x)\,\mathrm{d}x\,\mathrm{d}t &= \int_{\mathbb{R}^3_x}\left(\int_{t=\frac{|x|}{c}}^{\frac{|x|}{c}+2}\frac{1}{4\pi c^2|x|\varepsilon}\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right)\varphi(t,x)\,\mathrm{d}t\right)\mathrm{d}x\\
&= \frac{-1}{4\pi c^2}\int_{\mathbb{R}^3_x}\frac{1}{|x|}\left(\int_{t=\frac{|x|}{c}}^{\frac{|x|}{c}+2}\left[\frac{\partial}{\partial t}\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right)\right]\varphi(t,x)\,\mathrm{d}t\right)\mathrm{d}x\\
&= \frac{-1}{4\pi c^2}\int_{\mathbb{R}^3_x}\frac{1}{|x|}\left(\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right)\varphi(t,x)\right)\Big|_{t=\frac{|x|}{c}}^{t=\frac{|x|}{c}+2}\mathrm{d}x\\
&\qquad+\frac{1}{4\pi c^2}\int_{\mathbb{R}^3_x}\frac{1}{|x|}\left(\int_{t=\frac{|x|}{c}}^{\frac{|x|}{c}+2}\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right)\frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}t\right)\mathrm{d}x\\
&= \frac{1}{4\pi c^2}\int_{\mathbb{R}^3_x}\frac{1}{|x|}\varphi\left(\frac{|x|}{c},x\right)\mathrm{d}x-0\\
&\qquad+\frac{1}{4\pi c^2}\int_{\mathbb{R}^3_x}\frac{1}{|x|}\left(\int_{t=\frac{|x|}{c}}^{\frac{|x|}{c}+2}\exp\left(-\varepsilon^{-1}\left(t-\frac{|x|}{c}\right)\right)\frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}t\right)\mathrm{d}x,
\end{aligned}$$

and the last integral disappears for $\varepsilon\to 0$. Summing up, we learn that one causal solution to $(\partial_t^2-c^2\triangle)u=f$ is given by

$$u(t,x)=\int_{s=-\infty}^{t}\int_{\mathbb{R}^3_y}U(t-s,x-y)f(s,y)\,\mathrm{d}y\,\mathrm{d}s=\frac{1}{4\pi c^2}\int_{\mathbb{R}^3_y}\frac{1}{|x-y|}f\left(t-\frac{|x-y|}{c},y\right)\mathrm{d}y,$$

and here we see nicely that $u(t,x)$ depends only on those values $f(s,y)$ with $t-s=|x-y|/c$, which corresponds to the propagation of electromagnetic waves of speed c. And these waves have no "backside", which means that $U(t,x)$ is zero inside the light cone. This is a special effect of the three-dimensional space, because in two dimensions the situation is different. You can observe this when you let a pebble fall into a silent pond: it will generate not only one wave, but a large number of waves, which form concentric circles on the surface of the pond.

This is a good moment to conclude this course. There is so much more to find out, and we are confident that you are now proficient to continue on your own.

Enjoy doing physics !

Appendix A

The Fourier Transform

The purpose of this appendix is to explain what the Fourier transform is when applied to functions which do *not* decay at infinity.

A.1 Some Function Spaces and Distribution Spaces

Pseudo-Definition A.1. *A function is called* LEBESGUE[1]–MEASURABLE[2] *if it has "not too many discontinuities".*

Giving a rigorous definition of measurability would require several weeks, and for this reason we just present some examples:

- a function on $\mathbb{R}$ with a countable number of jumps is measurable,
- the function $f = f(x) = 1/x^{1492}$ is measurable on $\mathbb{R}^1$,
- the function on $\mathbb{R}$ which takes the value one for rational numbers, and the value zero for irrational numbers, is measurable.

Pseudo-Definition A.2. *Let Ω be a domain in $\mathbb{R}^n$. A function f belongs to the Lebesgue space $L^1(\Omega)$ if it is measurable, and if the Lebesgue integral $\int_\Omega |f(x)|\,\mathrm{d}x$ is finite.*

Again we can not give a precise definition of the Lebesgue integral, but only note that each function which is *properly integrable*[3] (in the sense of the first year) is also Lebesgue integrable.

Warning A.3. *Take $\Omega = (1,\infty) \subset \mathbb{R}^1$. Then the function $f = f(x) = (\sin x)/x$ does not belong to $L^1(\Omega)$ because $\int_{x=1}^{\infty} |\sin x|/x\,\mathrm{d}x = \infty$, but it is* improperly integrable *(in the sense of the first year) since $\lim_{R\to\infty} \int_{x=1}^{R} (\sin x)/x\,\mathrm{d}x$ exists.*

More generally, we can say:

Pseudo-Definition A.4. *For $1 \le p < \infty$, a function f belongs to the Lebesgue space $L^p(\Omega)$ if f is measurable, and the integral $\int_\Omega |f(x)|^p\,\mathrm{d}x$ is finite.*

The space $L^p(\Omega)$ is a vector space, and we equip it with the norm

$$\|f\|_{L^p(\Omega)} := \left(\int_{x\in\Omega} |f(x)|^p\,\mathrm{d}x \right)^{1/p}.$$

[1] HENRI LEBESGUE, 1875–1941
[2] Lebesgue–meßbar
[3] eigentlich integrierbar

Then $L^p(\Omega)$ is a Banach space (at this point the Lebesgue integration concept is really needed — our integration concept from the first year would not be strong enough to turn $L^p(\Omega)$ into a *complete* space), and $L^2(\Omega)$ even has a scalar product:

$$\langle f, g\rangle_{L^2(\Omega)} := \int_{x\in\Omega} f(x)\overline{g(x)}\,\mathrm{d}x.$$

For $\Omega = \mathbb{R}^1$, consider the functions

$$u(x) = \frac{1}{1+|x|}, \qquad v(x) = \begin{cases} \frac{1}{\sqrt{|x|}} & : |x| \le 1, \\ 0 & : |x| > 1. \end{cases}$$

Then $u \in L^2(\Omega)$ but $u \notin L^1(\Omega)$. On the other hand, $v \in L^1(\Omega)$ but $v \notin L^2(\Omega)$. From this example we learn that, for unbounded Ω, neither of the spaces $L^1(\Omega)$ and $L^2(\Omega)$ is contained in the other space.

Definition A.5 (Schwartz space). *The* Schwartz[4] space $\mathcal{S}(\mathbb{R}^n)$ *consists of all those functions* $f \in C^\infty(\mathbb{R}^n)$ *with*

$$p_{k,\alpha}(f) := \sup_{x\in\mathbb{R}^n} \left(1+|x|^k\right) |\partial_x^\alpha f(x)| < \infty,$$

for all $k \in \mathbb{N}_0$ *and all* $\alpha \in \mathbb{N}^n$.

These Schwartz functions are infinitely smooth, they decay at infinity faster than all powers of $|x|^{-1}$, and all their derivatives decay at infinity faster than all powers of $|x|^{-1}$, too.

Definition A.6 (Convergence in $\mathcal{S}(\mathbb{R}^n)$). *We say that a sequence* $(\varphi_1, \varphi_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ *converges to* $\varphi \in \mathcal{S}$ in the topology of $\mathcal{S}(\mathbb{R}^n)$ *if* $\lim_{j\to\infty} p_{k,\alpha}(\varphi_j - \varphi) = 0$ *for all* k, α. *We write*

$$\varphi_j \xrightarrow{\mathcal{S}} \varphi \qquad (j \to \infty)$$

for this convergence.

This means that the sequence $(\varphi_1, \varphi_2, \dots)$ converges to φ uniformly, and all the sequences of derivatives enjoy uniform convergence, too. The convergence in the topology of $\mathcal{S}$ is extremely strong and powerful.

Definition A.7 (Schwartz distributions). *A map* $T\colon \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ *is called a* Schwartz distribution *if it is linear and continuous. Here continuity means that*

$$\text{if} \quad \varphi_j \xrightarrow{\mathcal{S}} \varphi \quad \text{then} \quad \lim_{j\to\infty} T(\varphi_j) = T(\varphi).$$

The set (it is even a vector space) of all Schwartz distributions is denoted by $\mathcal{S}'(\mathbb{R}^n)$.

Example A.8. *The Delta distribution located at a point* $x_0 \in \mathbb{R}^n$ *is a Schwartz distribution:*

$$\delta_{x_0}(\varphi) := \varphi(x_0), \qquad \varphi \in \mathcal{S}(\mathbb{R}^n).$$

Example A.9. *Each function* f *which is piecewise continuous and grows at infinity at most polynomially generates a Schwartz distribution* T_f*:*

$$T_f(\varphi) := \int_{\mathbb{R}^n} f(x)\varphi(x)\,\mathrm{d}x, \qquad \varphi \in \mathcal{S}(\mathbb{R}^n).$$

The Delta distribution located at the point x_0 can be approximated by a distribution T_f whose generating function f has a sharp peak at x_0, and $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 1$.

It is common practice (but confusing) to not distinguish between f (which is a function that grows not too fast at infinity) and the associated distribution T_f.

[4] Laurent Schwartz, 1915–2002, french mathematician, inventor of the distributions (independent of Sobolev)

A.2 The Fourier Transform on $L^1(\mathbb{R}^n)$, $\mathcal{S}(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$

Definition A.10 (**Fourier**[5] **transform on** $L^1(\mathbb{R}^n)$)**.** *For $f \in L^1(\mathbb{R}^n)$, we define its Fourier transform $\mathcal{F}f = \hat{f}$ by*

$$(\mathcal{F}f)(\xi) := \int_{\mathbb{R}^n} e^{-\mathrm{i}x\xi} f(x)\,\mathrm{d}x, \qquad \xi \in \mathbb{R}^n, \quad x\xi := x_1\xi_1 + \cdots + x_n\xi_n.$$

We introduce the notations

$$D := \frac{1}{\mathrm{i}}\nabla, \qquad \text{đ}\xi := \frac{\mathrm{d}\xi}{(2\pi)^n}.$$

Proposition A.11. *The Fourier transform has the following properties:*

$$\begin{aligned}
f \in L^1(\mathbb{R}^n) &\implies |\hat{f}(\xi)| \le \|f\|_{L^1(\mathbb{R}^n)} \quad \forall\, \xi \in \mathbb{R}^n,\\
f \in \mathcal{S}(\mathbb{R}^n) &\implies \hat{f} \in \mathcal{S}(\mathbb{R}^n_\xi),\\
f_j \xrightarrow{\mathcal{S}} f &\implies \hat{f}_j \xrightarrow{\mathcal{S}(\mathbb{R}^n_\xi)} \hat{f},\\
f \in \mathcal{S}(\mathbb{R}^n) &\implies (\mathcal{F}(D_x^\alpha f))(\xi) = \xi^\alpha (\mathcal{F}f)(\xi), \qquad \forall\, \alpha \in \mathbb{N}^n,\\
f, g \in \mathcal{S}(\mathbb{R}^n) &\implies \int_{\mathbb{R}^n_x} f(x)\overline{g(x)}\,\mathrm{d}x = \int_{\mathbb{R}^n_\xi} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,\text{đ}\xi,\\
f, g \in \mathcal{S}(\mathbb{R}^n) &\implies (f\cdot g)\check{}(\xi) = (\hat{f} * \hat{g})(\xi),
\end{aligned}$$

with $*$ as the convolution operator:

$$(\hat{f} * \hat{g})(\xi) := \int_{\mathbb{R}^n_\eta} \hat{f}(\eta)\cdot\hat{g}(\xi-\eta)\,\text{đ}\eta. \tag{A.1}$$

The proof is a wonderful exercise for the calculus with integrals !

The penultimate property yields the Parseval[6] *identity*:

$$\|f\|_{L^2(\mathbb{R}^n_x)} = \left\|\hat{f}\right\|_{L^2(\mathbb{R}^n_\xi)} := \left(\int_{\mathbb{R}^n_\xi} |\hat{f}(\xi)|^2\,\text{đ}\xi\right)^{1/2}, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n).$$

Example A.12. *Take $f = f(x) = \exp(-x^2/2)$ on $\mathbb{R}^1$. Then*

$$\begin{aligned}
\partial_\xi \hat{f}(\xi) &= \partial_\xi \int_{\mathbb{R}_x} e^{-\mathrm{i}x\xi} f(x)\,\mathrm{d}x = \int_{\mathbb{R}_x} (-\mathrm{i}x) e^{-\mathrm{i}x\xi} f(x)\,\mathrm{d}x = \mathrm{i}\int_{\mathbb{R}_x} e^{-\mathrm{i}x\xi}\partial_x f(x)\,\mathrm{d}x\\
&= -\int_{\mathbb{R}_x} e^{-\mathrm{i}x\xi}(D_x f)(x)\,\mathrm{d}x = -(D_x f)\check{}(\xi) = -\xi\hat{f}(\xi),
\end{aligned}$$

and this ODE has the solutions

$$\hat{f}(\xi) = c\exp(-\xi^2/2),$$

with an unknown constant c which can be found by

$$c = \hat{f}(0) = \int_{\mathbb{R}_x} f(x)\,\mathrm{d}x = \int_{\mathbb{R}_x} e^{-x^2/2}\,\mathrm{d}x = (2\pi)^{1/2}.$$

Example A.13. *Take $f = f(x) = \exp(-|x|^2/2)$ on $\mathbb{R}^n$. Then*

$$\begin{aligned}
\hat{f}(\xi) &= \int_{\mathbb{R}^n_x} e^{-\mathrm{i}(x_1\xi_1+\cdots+x_n\xi_n)} f(x)\,\mathrm{d}x = \prod_{k=1}^n \left(\int_{\mathbb{R}^1_t} e^{-\mathrm{i}t\xi_k}\exp(-t^2/2)\,\mathrm{d}t\right)\\
&= \prod_{k=1}^n (2\pi)^{1/2}\exp(-\xi_k^2/2) = (2\pi)^{n/2}\exp(-|\xi|^2/2).
\end{aligned}$$

[5] Jean Baptiste Joseph Fourier, 1768–1830, french mathematician and physicist, famous for his law on the heat conduction

[6] Marc–Antoine Parseval des Chênes, 1755–1836

By repeated substitution we find that $f_\varepsilon(x) = \exp(-|\varepsilon x|^2/2)$ *has the Fourier transform* $\hat f_\varepsilon(\xi) = \varepsilon^{-n}(2\pi)^{n/2}\exp(-|\xi/\varepsilon|^2/2)$. *This function* $\hat f_\varepsilon$ *has a peak at* $\xi = 0$, *and we know that* $\int_{\mathbb{R}^n_\xi} \hat f_\varepsilon(\xi)\,\mathrm{d}\xi = (2\pi)^n$.

Proposition A.14 (Inverse Fourier transform on $\mathcal{S}(\mathbb{R}^n)$). *The Fourier transform is an isomorphism from* $\mathcal{S}(\mathbb{R}^n_x)$ *onto* $\mathcal{S}(\mathbb{R}^n_\xi)$, *and the inverse Fourier transform is given by*

$$f(x) = \int_{\mathbb{R}^n_\xi} e^{+\mathrm{i}x\xi}\hat f(\xi)\,\text{đ}\xi, \qquad x \in \mathbb{R}^n, \quad \hat f \in \mathcal{S}(\mathbb{R}^n). \tag{A.2}$$

Proof. Let $f \in \mathcal{S}(\mathbb{R}^n_x)$ be given. Then

$$\begin{aligned}
\int_{\mathbb{R}^n_\xi} e^{+\mathrm{i}x\xi}\hat f(\xi)\,\text{đ}\xi &= \int_{\mathbb{R}^n_\xi} \lim_{\varepsilon\to 0} e^{\mathrm{i}x\xi}e^{-|\varepsilon\xi|^2/2}\hat f(\xi)\,\text{đ}\xi = \lim_{\varepsilon\to 0}\int_{\mathbb{R}^n_\xi} e^{\mathrm{i}x\xi}e^{-|\varepsilon\xi|^2/2}\hat f(\xi)\,\text{đ}\xi \qquad (^\infty)\\
&= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^n_\xi}\left(\int_{\mathbb{R}^n_y} e^{\mathrm{i}x\xi}e^{-|\varepsilon\xi|^2/2}e^{-\mathrm{i}y\xi}f(y)\,\mathrm{d}y\right)\text{đ}\xi = \lim_{\varepsilon\to 0}\int_{\mathbb{R}^n_y}\left(\int_{\mathbb{R}^n_\xi} e^{\mathrm{i}x\xi}e^{-|\varepsilon\xi|^2/2}e^{-\mathrm{i}y\xi}\,\text{đ}\xi\right)f(y)\,\mathrm{d}y\\
&= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^n_y} \varepsilon^{-n}\exp(-|(x-y)/\varepsilon|^2/2)(2\pi)^{-n/2}f(y)\,\mathrm{d}y\\
&= \lim_{\varepsilon\to 0}\int_{\mathbb{R}^n_y} G_\varepsilon(x-y)f(y)\,\mathrm{d}y = \lim_{\varepsilon\to 0}(G_\varepsilon * f)(x),
\end{aligned}$$

with G_ε as a GAUSS bell shaped function having a peak at zero, and $\int_{\mathbb{R}^n_z} G_\varepsilon(z)\,\mathrm{d}z = 1$. Then the limit of $\varepsilon \to 0$ indeed gives $f(x)$. □

Example A.15. *Take* $f = f(x) \in L^1(\mathbb{R}^1)$ *with* $f(x) = 1$ *for* $0 \le x \le 1$, *and* $f(x) = 0$ *for all other* x. *Then*

$$\hat f(\xi) = \int_{x=-\infty}^{\infty} e^{-\mathrm{i}x\xi}f(x)\,\mathrm{d}x = \int_{x=0}^{1} e^{-\mathrm{i}x\xi}\,\mathrm{d}x = \frac{1-e^{-\mathrm{i}\xi}}{\mathrm{i}\xi} \qquad (\xi \neq 0), \qquad \hat f(0) = 1,$$

and this is a continuous function on $\mathbb{R}^1$, *but* $\hat f$ *does not belong to* $L^1(\mathbb{R}^1_\xi)$ *because it does not decay fast enough for* $\xi \to \infty$. *Unfortunately, the inversion formula (A.2) has no meaning as an integral in* $L^1(\mathbb{R}^1)$. *However, the next lemma will give a positive result.*

Lemma A.16 (Inverse Fourier transform for some non-smooth functions). *Let* $f \in L^1(\mathbb{R}^1)$ *be continuous, except a finite number of jumps. Then*

$$\frac{1}{2\pi}\lim_{R\to\infty}\int_{\xi=-R}^{R} e^{\mathrm{i}x\xi}\hat f(\xi)\,\mathrm{d}\xi = \begin{cases} f(x) & : \ f \text{ is continuous at } x,\\ \frac{1}{2}(f(x+0)+f(x-0)) & : \ f \text{ jumps at } x.\end{cases}$$

Proof. This is quoted from [4], Satz 8.2. □

The space $L^2(\mathbb{R}^n)$ is of great physical importance because it has a scalar product, in contrast to $L^1(\mathbb{R}^n)$. The Fourier transform is not yet defined on $L^2(\mathbb{R}^n)$ because there are functions in $L^2(\mathbb{R}^n)$ which are not in $L^1(\mathbb{R}^n)$. The definition of $\mathcal{F}$ on $L^2(\mathbb{R}^n)$ will be made possible by $\mathcal{S}(\mathbb{R}^n)$ being *dense*[7] in $L^2(\mathbb{R}^n)$: for each $f \in L^2(\mathbb{R}^n)$, there is a sequence $(f_1, f_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ with $\lim_{j\to\infty}\|f_j - f\|_{L^2(\mathbb{R}^n)} = 0$.

Definition A.17 (Fourier transform on $L^2(\mathbb{R}^n)$). *For* $f \in L^2(\mathbb{R}^n)$, *let* $(f_1, f_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ *be a sequence approximating* f. *Then we define*

$$\hat f(\xi) := \lim_{j\to\infty}\hat f_j(\xi).$$

This limit is independent of the choice of the sequence $(f_1, f_2, \dots)$. The convergence of the sequence $(\hat f_1, \hat f_2, \dots)$ in the norm of $L^2(\mathbb{R}^n_\xi)$ follows from the Parseval formula.

We draw an intermediate summary: the difference between the Fourier transform $\mathcal{F}$ and the inverse transform $\mathcal{F}^{-1}$ are the exchange of $\exp(+\mathrm{i}x\xi)$ against $\exp(-\mathrm{i}x\xi)$, and an additional factor $(2\pi)^{-n}$. Then

[7] dicht

it is no surprise that similar rules as given in Proposition A.11 hold also for the inverse transform, and in particular we mention

$$\mathcal{F}^{-1}_{\xi\to x}(\hat{f}\cdot\hat{g})(x) = (f * g)(x),$$

with the convolution in the x–world defined as

$$(f * g)(x) := \int_{\mathbb{R}^n_y} f(y)\cdot g(x-y)\,\mathrm{d}y.$$

Note that the differential is now $\mathrm{d}y$, instead of $đ\eta$ as in (A.1).

A.3 The Fourier Transform on $\mathcal{S}'(\mathbb{R}^n)$, Mathematically and Approximately

We still want to explain what the Fourier transform of $\sin(x)$ or x^2 is. Unfortunately, these functions belong neither to $L^1(\mathbb{R}^n)$, nor to $L^2(\mathbb{R}^n)$ or $\mathcal{S}(\mathbb{R}^n)$. At least, they have at most polynomial growth for $|x|\to\infty$.

We now follow the (at the beginning confusing) convention of writing a slowly growing function f and its associated Schwartz distribution T_f by the same symbol f. Then the expression $f = f(\cdot)$ becomes ambiguous (the dot could be an x or a test function φ), and we resolve this equivocation by writing $\langle f,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}$ when we apply the distribution $T_f\in\mathcal{S}'$ to the Schwartz function $\varphi\in\mathcal{S}$.

Note that Schwartz functions f, g satisfy

$$\int_{\mathbb{R}^n}\hat{f}g\,\mathrm{d}x = \int_{\mathbb{R}^n} f\hat{g}\,\mathrm{d}x,$$

which can be written as $\left\langle \hat{f}, g\right\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle f,\hat{g}\rangle_{\mathcal{S}'\times\mathcal{S}}$ for such functions f and g.

Keep in mind that a distribution from $\mathcal{S}'(\mathbb{R}^n)$ need not be a function; it could be also a Delta distribution.

Definition A.18 (Fourier transform on $\mathcal{S}'(\mathbb{R}^n)$). *For a distribution $T\in\mathcal{S}'(\mathbb{R}^n)$, we define its Fourier transform $\hat{T}\in\mathcal{S}'(\mathbb{R}^n)$ via*

$$\left\langle \hat{T},\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}} := \langle f,\hat{\varphi}\rangle_{\mathcal{S}'\times\mathcal{S}}\,, \qquad \forall\,\varphi\in\mathcal{S}(\mathbb{R}^n).$$

Example A.19. *What is $\hat{\delta}$?*

$$\left\langle \hat{\delta},\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}} := \langle\delta,\hat{\varphi}\rangle_{\mathcal{S}'\times\mathcal{S}} = \hat{\varphi}(0) = \int_{\mathbb{R}^n} 1\cdot\varphi(x)\,\mathrm{d}x = \langle 1,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}\,.$$

Answer: the Fourier transform of Dirac's delta is that function which is one everywhere.

This definition is mathematically correct, but certainly not easy to understand. Perhaps an approximate transformation, which we develop now, is less complicated.

Definition A.20 (Convergence in $\mathcal{S}'(\mathbb{R}^n)$). *A sequence $(T_1,T_2,\dots)\subset\mathcal{S}'(\mathbb{R}^n)$ converges to a distribution $T\in\mathcal{S}'(\mathbb{R}^n)$ if $\lim_{j\to\infty}\langle T_j,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}$ for every $\varphi\in\mathcal{S}(\mathbb{R}^n)$. This convergence shall be written as*

$$T_j \xrightarrow{\mathcal{S}'} T.$$

Proposition A.21. *If $T_j\xrightarrow{\mathcal{S}'}T$, then $\hat{T}_j\xrightarrow{\mathcal{S}'}\hat{T}$.*

We can also say that $\mathcal{F}$ is a continuous isomorphism between $\mathcal{S}'$ and $\mathcal{S}'$.

Proof. We know that $\langle T_j,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}\to\langle T,\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}$ for each test function $\varphi\in\mathcal{S}$, and we want to show that $\left\langle\hat{T}_j,\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}}\to\left\langle\hat{T},\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}}$. But $\hat{\varphi}\in\mathcal{S}$, and therefore

$$\left\langle\hat{T}_j,\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}} := \langle T_j,\hat{\varphi}\rangle_{\mathcal{S}'\times\mathcal{S}}\to\langle T,\hat{\varphi}\rangle_{\mathcal{S}'\times\mathcal{S}} =: \left\langle\hat{T},\varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}}\,.$$

□

And this continuity of $\mathcal{F}$ makes the following recipe of an approximate Fourier transform possible:

- take a function like $f = \sin x$ or $f(x) = x^2$, of which you want to know the Fourier transform $\hat{f}$,
- then we typically have $f \in \mathcal{S}'(\mathbb{R}^1)$,
- choose a small positive ε and set $f_\varepsilon = f_\varepsilon(x) := \exp(-|\varepsilon x|^2/2) f(x)$, which is a function from the Schwartz space $\mathcal{S}(\mathbb{R}^1)$, or at least from $L^1(\mathbb{R}^1)$ if f had some jumps,
- then $\lim_{\varepsilon\to 0} f_\varepsilon = f$ in the topology of $\mathcal{S}'$,
- compute the Fourier transform $\mathcal{F}f_\varepsilon$ as usual. This is no problem since f_ε has fast decay for $|x| \to \infty$,
- do **not** perform the limit $\varepsilon \to 0$. Just keep in mind that the number of protons in the universe is estimated as 10^{80}, choose $\varepsilon = 10^{-800}$, and stop here. Instead, think about how the function $\mathcal{F}f_\varepsilon$ looks like.

Some examples may elucidate this. Put $h_\varepsilon(x) = \exp(-|\varepsilon x|^2/2)$.

To compute the approximate Fourier transform of the one-function (which takes the value one for each $x \in \mathbb{R}^n$), we write

$$(\mathcal{F}1)(\xi) \overset{\mathcal{S}'}{\approx} (\mathcal{F}(1 \cdot h_\varepsilon))(\xi) = \hat{h}_\varepsilon(\xi) = \varepsilon^{-n}(2\pi)^{n/2} \exp(-|\xi/\varepsilon|^2/2),$$

and this is a function with a peak at the origin $\xi = 0$, and the volume under this peak is $(2\pi)^n$:

$$\int_{\mathbb{R}^n_\xi} \hat{h}_\varepsilon(\xi)\, \mathrm{d}\xi = (2\pi)^n.$$

The precise Fourier transform is $(\mathcal{F}1) = (2\pi)^n \delta$.

Next we choose a function $g(x) = x^\alpha$ on $\mathbb{R}^n$, and we look for $\mathcal{F}g$:

$$\begin{aligned}(\mathcal{F}g)(\xi) &\overset{\mathcal{S}'}{\approx} (\mathcal{F}(gh_\varepsilon))(\xi) = \int_{\mathbb{R}^n_x} e^{-\mathrm{i}x\xi} x^\alpha h_\varepsilon(x)\, \mathrm{d}x = (-1)^{|\alpha|} \int_{\mathbb{R}^n_x} \left(D^\alpha_\xi e^{-\mathrm{i}x\xi}\right) h_\varepsilon(x)\, \mathrm{d}x \\ &= (-1)^{|\alpha|} D^\alpha_\xi \hat{h}_\varepsilon(\xi) = \mathrm{i}^{|\alpha|} \partial^\alpha_\xi \hat{h}_\varepsilon(\xi).\end{aligned}$$

Compare the figures for $n = 1$.

Next we consider a planar wave: $g(x) = \exp(\mathrm{i}x \cdot k)$ on $\mathbb{R}^n$, for some fixed wave vector $k \in \mathbb{R}^n$. Then

$$(\mathcal{F}g)(\xi) \overset{\mathcal{S}'}{\approx} (\mathcal{F}(gh_\varepsilon))(\xi) = \int_{\mathbb{R}^n_x} e^{-\mathrm{i}x\xi} g(x) h_\varepsilon(x)\, \mathrm{d}x = \int_{\mathbb{R}^n_x} e^{-\mathrm{i}x(\xi-k)} h_\varepsilon(x)\, \mathrm{d}x = \hat{h}_\varepsilon(\xi - k),$$

which is a peak of volume $(2\pi)^n$ located at the point $k \in \mathbb{R}^n$.

And finally, on $\mathbb{R}^1$ we take $g(x) = \cos(\omega x)$ for some $\omega \in \mathbb{R}$, $\omega \neq 0$. By $g(x) = (\exp(\mathrm{i}\omega x) + \exp(-\mathrm{i}\omega x))/2$ we see that $\hat{g}$ is approximately a collection of two peaks located at $+\omega$ and $-\omega$, each of volume $(2\pi)^n/2$.

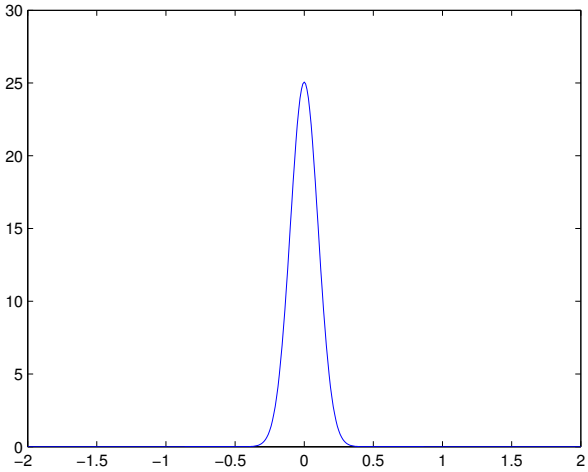


Figure A.1: $\hat{h}_\varepsilon(\xi) = (2\pi)^{1/2}\varepsilon^{-1}\exp(-\xi^2/(2\varepsilon^2))$ with $\varepsilon = 0.1$,

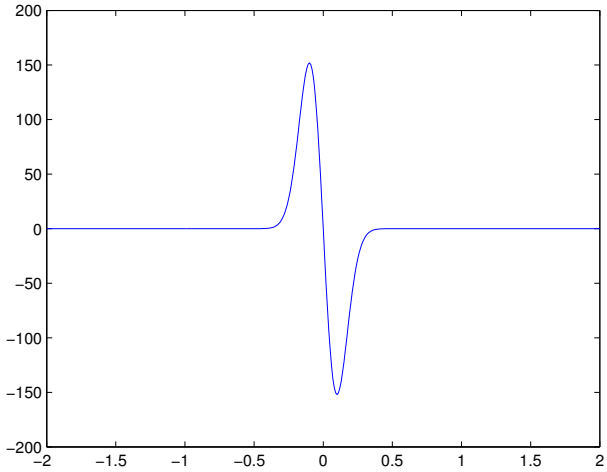


Figure A.2: $\hat{h}'_\varepsilon(\xi) = -\hat{h}_\varepsilon(\xi)\cdot\xi/\varepsilon^2$ with $\varepsilon = 0.1$

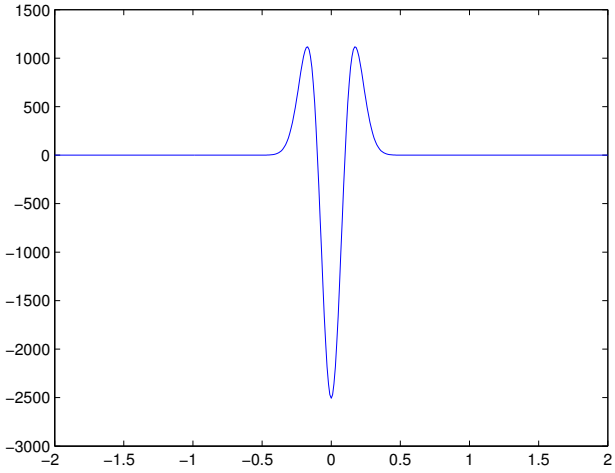


Figure A.3: $\hat{h}''_\varepsilon(\xi) = \hat{h}_\varepsilon(\xi)\cdot(\xi^2-\varepsilon^2)/\varepsilon^4$ with $\varepsilon = 0.1$

Appendix B

Core Components

Ten chapters is quite a lot of stuff, so we now give a list of the key results you should understand (with their proofs, of course). The concepts not mentioned here are also important, but they are mostly conclusions from the core components mentioned now.

Chapter 1: transforming a higher order equation into a first order system, the Theorem of Picard and Lindelöf,

Chapter 2: the drum example, and how it is related to the Fourier series expansion from the second semester,

Chapter 3: Duhamel's formula (3.3) for single equations, Wronski determinants, fundamental solutions for the case of general coefficients, fundamental solutions for the case of constant coefficients, how to compute them,

Chapter 4: stationary points, stable/unstable points, asymptotically stable points, and how they are related to the eigenvalues of the Jacobian,

Chapter 5: one method, order of consistency, purpose of implicit methods,

Chapter 6: solution behaviour and rôle of the characteristic matrix C_X, all of Section 6.4 and Section 6.5 (because this is the heart of our three semester course), and study again the drum example from Chapter 2,

Chapter 7: complex logarithm, complex root function, Cauchy–Riemann equations,

Chapter 8: definition of curve integrals, Cauchy integral theorem, Cauchy integral formula, analytic function, equivalence of holomorphy and analyticity, the several formulae for the coefficients of the Taylor expansion, Liouville's theorem,

Chapter 9: identity theorem, Theorem on the Laurent series, residue theorem.

Concerning the Chapters 8 and 9, please invest the time to observe how each theorem builds upon the other theorems mentioned before. For instance, in an attempt to prove the Cauchy integral formula, you can not exploit that each holomorphic function can be expanded (locally) into a Taylor series, because this Taylor expansion is itself a conclusion of the Cauchy integral formula, and therefore you would prove the C.I.F. by means of the C.I.F., which is clearly philosophical nonsense.

Bibliography

[1] *Pocketbook of mathematical functions. Abridged edition of "Handbook of mathematical functions" by Milton Abramowitz and Irene A. Stegun. Material selected by Michael Danos and Johann Rafelski.* Thun: Verlag Harri Deutsch, 1984.

[2] Vladimir Igorevich Arnold. *Ordinary Differential Equations.* Springer, Third edition, 1992, translated from the Russian Edition 1984 by Roger Cooke.

[3] David Brewster. *Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton*, volume 2, chapter XXVII. Reprinted from the Edinburgh Edition 1855, Johnson Reprint Corp., 1965.

[4] Klemens Burg, Herbert Haf, and Friedrich Wille. *Höhere Mathematik für Ingenieure. Band 3. Gewöhnliche Differentialgleichungen, Distributionen, Integraltransformationen. 3., durchges. Aufl.* Stuttgart: Teubner, 1993.

[5] Richard Courant and Kurt Otto Friedrichs. *Supersonic flow and shock waves.* (Pure and Applied Mathematics. I) New York: Interscience Publ., Inc., 1948.

[6] Robert Denk. Skript zur Vorlesung Mathematik für Physiker 3, Teil 1. Universität Konstanz, Wintersemester 2007/8.

[7] Gustav Doetsch. *Handbuch der Laplace-Transformation. Band I: Die theoretischen Grundlagen der Laplace-Transformation.* Lehrbücher und Monographien. Math. Reihe. 14. Basel: Birkhäuser, 1950.

[8] Michael Dreher and Shaoqiang Tang. Time history interfacial conditions in multiscale computations of lattice oscillations. *Comput. Mech.*, 41(5):683–698, 2008.

[9] Wilhelm Forst and Dieter Hoffmann. *Funktionentheorie erkunden mit Maple.* Springer-Lehrbuch. Berlin: Springer, 2002.

[10] Walter Greiner. *Klassische Elektrodynamik.* 6. Auflage. Verlag Harri Deutsch, 2002.

[11] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations.* Springer Series in Computational Mathematics. 31. Berlin: Springer, 2002.

[12] Jack K. Hale and Hüseyin Koçak. *Dynamics and bifurcations.* Texts in Applied Mathematics. 3. Springer-Verlag, 1991.

[13] Harro Heuser. *Gewöhnliche Differentialgleichungen. Einführung in Lehre und Gebrauch. 3. Auflage.* Stuttgart: Teubner, 1995.

[14] Harry Hochstadt. *The functions of mathematical physics. (Reprint with corrections). With a foreword by Wilhelm Magnus.* New York: Dover Publications, Inc., 1986.

[15] Klaus Jänich. *Analysis für Physiker und Ingenieure. Funktionentheorie, Differentialgleichungen, spezielle Funktionen. Ein Lehrbuch für das zweite Studienjahr. 4. Aufl.* Springer-Lehrbuch. Berlin: Springer, 2001.

[16] Klaus Jänich. *Funktionentheorie. Eine Einführung. 6. Aufl.* Springer-Lehrbuch. Berlin: Springer, 2004.

[17] Konrad Knopp. *Funktionentheorie. I. Grundlagen der allgemeinen Theorie der analytischen Funktionen.* Neunte, neubearbeitete Auflage. Sammlung Göschen Bd. 668. Walter de Gruyter and Co., Berlin, 1957.

[18] Dilip Kondepudi and Ilya Prigogine. *Modern Thermodynamics. From Heat Engines to Dissipative Structures.* John Wiley and Sons, 1998.

[19] Jürgen Saal. Skript zur Vorlesung Mathematik für Physiker 3, Teil 2. Universität Konstanz, Wintersemester 2008/9.

[20] J.M. Sanz-Serna. Symplectic integrators for Hamiltonian problems: An overview. Acta Numerica 1992, 243–286, 1992.

[21] Josef Stoer and Roland Bulirsch. *Numerische Mathematik II. Eine Einführung – unter Berücksichtigung von Vorlesungen von F. L. Bauer. 3., verb. Aufl.* Springer-Lehrbuch. Berlin, 1990.

[22] Wolfgang Walter. *Gewöhnliche Differentialgleichungen.* Springer-Lehrbuch. Berlin: Springer, 1996.

[23] George Neville Watson. *A treatise on the theory of Bessel functions.* Cambridge: University press, 1922.

[24] Hermann Weyl. *The classical groups, their invariants and representations. 2nd ed.* Princeton Mathematical Series. 1. Princeton, NJ: Princeton University Press, 1946.