Vorlesungsnotizen Mathematik 3 für den Lehramtsstudiengang Sekundarstufe, Wintersemester 2022/23

Birgit Richter, Markus Röser; Version vom 25. Januar 2023

Fachbereich Mathematik der Universität Hamburg, Bundesstrasse 55, 20146 Hamburg

Inhaltsverzeichnis

Diese Notizen orientieren sich an dem Skript, das Sven-Ake Wegner im Wintersemester 2021/22 angefertigt hat; andere Teile stammen aus BRs Skript zur Linearen Algebra.

KAPITEL I

Determinanten

I.1. Wiederholung zur Linearen Algebra

Wiederholung I.1.1. Diese Kurzzusammenstellung umfasst natürlich nicht alle Details, die in der Mathematik 2 behandelt wurden. Die Wiederholungsteile werden in der Vorlesung *nicht* besprochen.

Vektorräume.

- Ein Vektorraum über einem Körper K (z.B. $\mathbb{R}, \mathbb{C}, \mathbb{F}_2, \ldots$) ist nach Definition eine Menge V mit Operationen $+\colon V \times V \to V$, $\cdot\colon K \times V \to V$ die gewisse Spielregeln erfüllen. Beispiele: K^n, $K[X]$, $\mathrm{Abb}([a,b], K)$ usw.
- Ein Untervektorraum $U \subset V$ ist nach Definition eine Teilmenge, die mit induzierten Operationen selbst wieder ein Vektorraum ist. Dies ist genau dann der Fall, wenn der Nullvektor von V in U liegt und U unter Vektoraddition und Skalarmultiplikation abgeschlossen ist. Beispiel: Es sei $X \subset V$ eine beliebige Teilmenge, dann ist
$$\mathsf{Span}_K X = \Big\{ \sum_{i=1}^{n} \lambda_i x_i \;\Big|\; \lambda_i \in K,\, x_i \in X,\, n \in \mathbb{N} \Big\}$$
ein Untervektorraum von V.
- Ein K-Vektorraum V wird *erzeugt von* $v_1, \ldots, v_n \in V$, falls $V = \mathsf{Span}_K\{v_1, \ldots, v_n\}$ gilt.
- Die Vektoren $v_1, \ldots, v_n \in V$ sind *linear unabhängig*, wenn für $\lambda_i \in K$ aus $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ folgt, dass $\lambda_1 = \cdots = \lambda_n = 0$ gilt.
- $\mathcal{B} = (b_1, \ldots, b_n)$ ist *eine Basis von* V, wenn V von den $b_1, \ldots, b_n$ erzeugt wird und die $b_1, \ldots, b_n$ linear unabhängig sind. Äquivalent dazu ist, dass die $b_1, \ldots, b_n$ maximal linear unabhängig bzw. minimal erzeugend sind.
- Ist $\mathcal{B}$ eine Basis, so hat jedes $x \in V$ eine eindeutige Darstellung $x = \sum_{i=1}^{n} \beta_i b_i$.
- Ist $\mathcal{B} = (b_1, \ldots, b_n)$ eine Basis, so ist $n = \dim V$ unabhängig von der gewählten Basis. Beachten Sie aber, dass es unendlich-dimensionale Vektorräume gibt, z.B. $K[X]$.

Matrizen.

- $A = \begin{pmatrix} a_{11} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ mit $a_{ij} \in K$ ist eine $m \times n$-Matrix.
- Wir bezeichnen die Menge aller $m \times n$-Matrizen mit $M(m \times n, K)$. Diese bildet unter der eintragsweisen definierten Addition und Streckung mit Skalaren aus K wieder einen Vektorraum. Außerdem können wir Matrizen bei geeignetem Format auch multiplizieren. Die Multiplikation ist nicht kommutativ und Inversen existieren nicht für alle Matrizen A. Andere Operationen, die wir schon kennen, sind das Transponieren A^t und wir können eintragsweise komplex konjugieren und $\bar{A}$ bilden, falls $K = \mathbb{C}$ ist.

Lineare Abbildungen.

- Eine Abbildung $\varphi\colon V \to W$ ist K-linear, wenn $\varphi(u+v) = \varphi(u) + \varphi(v)$ und $\varphi(\lambda u) = \lambda \varphi(u)$ gilt für alle $u, v \in V$ und $\lambda \in K$. Beispiele hierzu sind:

(1) Eine Matrix $A \in M(m \times n, K)$ induziert eine lineare Abbildung $K^n \to K^m$ durch Matrix-Vektor-Multiplikation, d.h.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

(2) Eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V induziert eine lineare Bijektion

$$V \xrightarrow{\sim} K^n, \quad \sum_{i=1}^{n} \beta_i v_i = x \mapsto x_{\mathcal{B}} := \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

d.h. jeder endlich-dimensionale K-Vektorraum 'sieht aus wie K^n' mit $n \in \mathbb{N}$, und in der Tat sieht auch jede lineare Abbildung aus wie Matrix-Vektor-Multiplikation, wie der folgende Hauptsatz besagt.

- Hauptsatz: Es sei $\varphi\colon V \to W$ K-linear, $\mathcal{B} = (v_1, \dots, v_n)$ eine Basis von V, $\mathcal{C} = (w_1, \dots, w_m)$ eine Basis von W. Dann haben wir

$$\begin{tikzcd}
 & x \arrow[rr, mapsto] & & \varphi(x) & \\
x \arrow[d, mapsto] & V \arrow[rr, "\varphi"] \arrow[d, "\sim"'] & & W \arrow[d, "\sim"] & y \arrow[d, mapsto] \\
x_{\mathcal{B}} & K^n \arrow[rr, dashed] & & K^m & y_{\mathcal{C}} \\
 & x_{\mathcal{B}} \arrow[rr, mapsto] & & \mathrm{M}_{\mathcal{C}\mathcal{B}}(\varphi) \cdot x_{\mathcal{B}}, &
\end{tikzcd}$$

wobei $\mathrm{M}_{\mathcal{C}\mathcal{B}}(\varphi) = \big[\varphi(v_1)_{\mathcal{C}} \ \cdots \ \varphi(v_n)_{\mathcal{C}}\big]$.

Determinanten.

- Die kleinsten nicht-trivialen Beispiele sind Determinanten von 2×2- bzw. 3×3-Matrizen.

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \quad \text{und} \quad \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + dhc + gbf - ceg - fha - ibd,$$

wobei die zweite Regel unter dem Namen Sarrus-Regel bekannt ist. Eine wichtige Anwendung ist der Satz, dass eine Matrix A genau dann invertierbar ist, wenn $\det A \neq 0$ gilt. Wenn letzteres gilt, kann die inverse Matrix A^{-1} mithilfe der Adjunkten berechnet werden. In dieser kommen auch wieder Determinanten vor.

- In der Mathematik 2 haben Sie die sogenannte Leibniz-Formel kennengelernt, mit der Determinanten für beliebige quadratische Marizen berechnet werden können, nämlich

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \sum_{\pi \in \Sigma_n} \operatorname{sign}\pi \prod_{i=1}^{n} a_{i,\pi(i)}$$

wobei Σ_n die Menge aller Bijektionen von $\{1, \dots, n\}$ in sich ist, und $\operatorname{sign}\pi$ das Vorzeichnen der Permutation π bezeichnet. In der Mathematik 2 war dies eine Definition und es wurden dann Eigenschaften der Determinante aus dieser Formel abgeleitet.

Im Folgenden diskutieren wir eine geometrische Anwendung von Determinanten und erklären, warum man die Determinante zwangsläufig so definieren muss, wir wir das getan haben.

I.2. Das Vektorprodukt des $\mathbb{R}^3$

Definition I.2.1. Die Abbildung

$$\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$$

$$\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right) = (x, y) \mapsto x \times y = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}$$

heißt das *Vektorprodukt oder das Kreuzprodukt auf dem* $\mathbb{R}^3$.

Beispiele I.2.2. Die Vektorprodukte der Standardbasisvektoren sind einfach zu berechnen:

$$e_1 \times e_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

und

$$e_2 \times e_3 = e_1, \quad e_3 \times e_1 = e_2.$$

Aber Sie erhalten auch:

$$e_2 \times e_1 = -e_3, \quad e_3 \times e_2 = -e_1, \quad e_1 \times e_3 = -e_2 \text{ und } e_1 \times e_1 = e_2 \times e_2 = e_3 \times e_3 = 0.$$

Zusammen mit dem Skalarprodukt im $\mathbb{R}^3$, welches gegeben ist durch

$$\langle -, - \rangle \colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}, \quad \left\langle \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right\rangle = x_1 y_1 + x_2 y_2 + x_3 y_3$$

erhalten wir folgende Rechenregeln:

Lemma I.2.3. *Für alle $u, x, y, x', y' \in \mathbb{R}^3$ und $\lambda, \mu \in \mathbb{R}$ gilt:*

(1) *Das Vektorprodukt ist bilinear:*

$$\begin{aligned} (\lambda x + \mu x') \times y &= \lambda(x \times y) + \mu(x' \times y), \\ x \times (\lambda y + \mu y') &= \lambda(x \times y) + \mu(x \times y'). \end{aligned}$$

(2) *Das Vektorprodukt ist antisymmetrisch:*

$$x \times y = -y \times x.$$

(3) *Es gilt die Graßmann-Identität:*

$$u \times (x \times y) = \langle u, y \rangle \cdot x - \langle u, x \rangle \cdot y.$$

(4) *Es gilt die Jacobi-Identität:*

$$u \times (x \times y) + x \times (y \times u) + y \times (u \times x) = 0.$$

(5) $\langle u \times x, y \rangle = \langle u, x \times y \rangle$.

Beweis. (1), (2) und (5) beweisen wir hier nicht, (3) ist eine Übungsaufgabe. Wir leiten (4) aus (3) her:

$$\begin{aligned} & u \times (x \times y) + x \times (y \times u) + y \times (u \times x) \\ =& \langle u, y \rangle \cdot x - \langle u, x \rangle \cdot y + \langle x, u \rangle \cdot y - \langle x, y \rangle \cdot u + \langle y, x \rangle \cdot u - \langle y, u \rangle \cdot x \\ =& (\langle u, y \rangle - \langle y, u \rangle) \cdot x + (\langle x, u \rangle - \langle u, x \rangle) \cdot y + (\langle y, x \rangle - \langle x, y \rangle) \cdot u \\ =& 0. \end{aligned}$$

□

Bemerkung I.2.4. Die Graßmann-Identität ist in der Physik auch als *bac-cab-Regel* bekannt, weil die rechte Seite sich so liest, wenn man die Vektoren $\vec{a}, \vec{b}, \vec{c}$ nennt und die Skalare von rechts an die Vektoren heranmultipliziert:

$$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b} \langle \vec{a}, \vec{c} \rangle - \vec{c} \langle \vec{a}, \vec{b} \rangle.$$

Wir setzen im Folgenden $||x||^2 = \langle x, x \rangle$.

Satz I.2.5. *Für alle $x, y \in \mathbb{R}^3$ gilt:*

$$||x||^2 \cdot ||y||^2 = ||x \times y||^2 + \langle x, y \rangle^2.$$

Beweis. Nach Definition ist $||x \times y||^2 = \langle x \times y, x \times y\rangle$ und mit Lemma I.2.3 (5) können wir dies umformen:

$$\langle x \times y, x \times y\rangle = \langle x, y \times (x \times y)\rangle.$$

Mit der Graßmann-Identität erhalten wir daraus

$$\begin{aligned}\langle x, y \times (x \times y)\rangle &= \langle x, \langle y, y\rangle \cdot x - \langle y, x\rangle \cdot y\rangle \\ &= \langle y, y\rangle\langle x, x\rangle - \langle y, x\rangle\langle x, y\rangle \\ &= ||y||^2||x||^2 - \langle y, x\rangle^2.\end{aligned}$$

□

Aus obigem Satz erhalten wir als wichtige Folgerung:

Korollar I.2.6. *Für alle $x, y \in \mathbb{R}^3$ gilt insbesondere die* Cauchy-Schwarzsche Ungleichung*:*

$$|\langle x, y\rangle| \leqslant ||x|| \cdot ||y||.$$

Bemerkung I.2.7. Die Cauchy-Schwarzsche Ungleichung stellt sicher, dass für $x, y \in \mathbb{R}^3 \setminus \{0\}$ gilt

$$\left|\frac{\langle x, y\rangle}{||x||||y||}\right| = \frac{|\langle x, y\rangle|}{||x||||y||} \leqslant 1.$$

Wir wissen also, dass $\frac{\langle x,y\rangle}{||x||||y||}$ Werte im Intervall $[-1, 1] := \{t \in \mathbb{R}, -1 \leqslant t \leqslant 1\}$ annimmt. Die Kosinusabbildung

$$\cos\colon [0, \pi] \to [-1, 1]$$

ist bijektiv und arccos bezeichne die Umkehrabbildung

$$\arccos\colon [-1, 1] \to [0, \pi],$$

die also einem $t \in [-1, 1]$ einen Winkel in $[0, \pi]$ zuordnet.

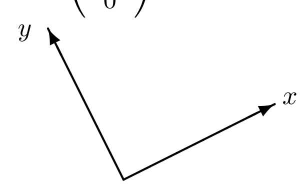
Definition I.2.8. Für $x, y \in \mathbb{R}^3 \setminus \{0\}$ heißt

$$\measuredangle(x, y) := \arccos\left(\frac{\langle x, y\rangle}{||x||||y||}\right)$$

der *Innenwinkel von x und y*.

Beispiele I.2.9.

- Ist $y = -x$, so erhalten wir
$$\langle x, y\rangle = -\langle x, x\rangle$$
und
$$\measuredangle(x, y) = \arccos\left(\frac{\langle x, y\rangle}{||x||||y||}\right) = \left(\frac{-\langle x, x\rangle}{||x||||x||}\right) = \arccos(-1) = \pi.$$
- Ist $x = y$, so gilt
$$\measuredangle(x, y) = \measuredangle(x, x) = \arccos(1) = 0.$$
- Ist $\langle x, y\rangle = 0$ für $x \neq 0 \neq y$, so ist $\arccos(0) = \frac{\pi}{2}$ entsprechend 90 Grad. Also stehen x und y senkrecht aufeinander. Wir sagen auch, dass x und y *orthogonal* zueinander sind. Zum Beispiel ist $x = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$ immer orthogonal zu $y = \begin{pmatrix} -x_2 \\ x_1 \\ 0 \end{pmatrix}$.



Bemerkung I.2.10. Wegen der Antisymmetrie gilt für alle $x \in \mathbb{R}^3$, dass

$$x \times x = -x \times x$$

und damit muss $x \times x = 0$ sein. Damit gilt aber ebenfalls

$$\langle x, x \times y \rangle = \langle x \times x, y \rangle = 0.$$

Dies bedeutet, dass $x \times y$ immer senkrecht auf x steht. Genauso zeigen Sie, dass $x \times y$ senkrecht auf y steht. Wir benutzen die Notation $x \perp y$ wenn x senkrecht auf y steht:

$$x \perp (x \times y) \perp y.$$

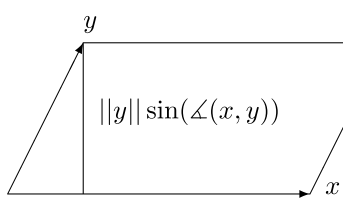
Welche Länge hat $x \times y$? Wir erhalten mit Satz I.2.5, dass

$$||x \times y||^2 = ||x||^2 ||y||^2 - \langle x, y \rangle^2$$

ist. Wir formen dies weiter um zu

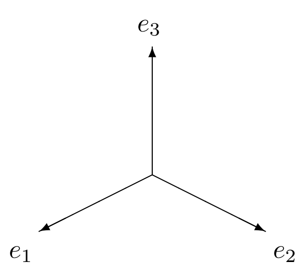
$$\begin{aligned} ||x \times y||^2 &= ||x||^2 ||y||^2 - \langle x, y \rangle^2 \\ &= ||x||^2 ||y||^2 (1 - \cos^2(\measuredangle(x, y))) \\ &= ||x||^2 ||y||^2 (\sin^2(\measuredangle(x, y))). \end{aligned}$$

Damit ist $||x \times y||$ der Flächeninhalt des von x und y aufgespannten Parallelograms.



Man kann sich merken, in welche Richtung $x \times y$ zeigt, indem man die rechte-Hand-Regel benutzt: Entspricht x dem Daumen und y dem Zeigefinger der rechten Hand, dann entspricht $x \times y$ dem Mittelfinger.

Wem das zu kompliziert ist, der kann sich auch lediglich den Fall $e_1 \times e_2 = e_3$ merken:



Definition I.2.11. Es sei (x, y, z) eine geordnete Basis des $\mathbb{R}^3$. Für ein $p \in \mathbb{R}^3$ heißt

$$P := \{p + \lambda x + \mu y + \nu z, 0 \leqslant \lambda, \mu, \nu \leqslant 1\}$$

das *von (x, y, z) aufgespannte Parallelotop (oder Spat) an p.*

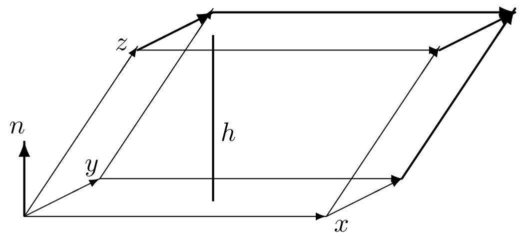
Satz I.2.12. *Das Volumen des von (x, y, z) aufgespannten Parallelotops ist*

$$\mathsf{Vol}(P) = |\langle x \times y, z \rangle|.$$

Das Vektorprodukt wird deshalb auch manchmal *Spatprodukt* genannt.

Beweis. Das Volumen von P ist natürlich unabhängig von p. Wir können also ohne Einschränkung annehmen, dass $p = 0$ ist. Ebenso können wir das Parallelotop im Raum drehen, damit es so liegt, wie in der Skizze unten. Das Volumen errechnet sich aus dem Produkt aus Grundfläche und Höhe. Die Grundfläche ist genau das Parallelogram, welches von x und y aufgespannt wird. Für die Fläche kennen wir den Flächeninhalt und der ist $||x \times y||$.

Wir betrachten $n := \frac{x \times y}{||x \times y||}$. Dieser Vektor hat Länge 1 und steht senkrecht auf x und y.



Damit ist die Höhe des Parallelotops $h = |\langle n, z\rangle|$ und wir erhalten insgesamt für das Volumen von P:

$$\mathsf{Vol}(P) = ||x \times y|| \cdot h = ||x \times y|| \cdot |\langle n, z\rangle| = ||x \times y|| \cdot \frac{|\langle x \times y, z\rangle|}{||x \times y||} = |\langle x \times y, z\rangle|.$$

□

Bemerkung I.2.13. Eine direkte Rechnung zeigt, dass das Volumen mit der Determinante wie folgt in Beziehung steht:

$$\mathsf{Vol}(P) = \left|\det \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix}\right|.$$

I.3. Axiomatische Beschreibung der Determinante

Wir hatten in der Mathematik 2 die Determinante einer quadratischen Matrix ad hoc über die Leibnizformel eingeführt. Wir sehen jetzt, dass dies nicht beliebig war.

Definition I.3.1. Es sei K ein beliebiger Körper und $n \in \mathbb{N}$. Eine Abbildung

$$\det\colon M(n \times n, K) \to K, \quad A \mapsto \det(A),$$

heißt eine *Determinantenabbildung*, falls gilt:

(D1) Die Abbildung det ist linear in jeder Zeile, das heißt

$$\det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ \lambda a_{i1} & \dots & \lambda a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \lambda \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

und

$$\det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1}+a'_{i1} & \dots & a_{in}+a'_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a'_{i1} & \dots & a'_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

(D2) Die Abbildung det ist *alternierend*, das heißt, dass $\det(A) = 0$, falls zwei Zeilen von A übereinstimmen.

(D3) Die Abbildung det ist normiert durch $\det(E_n) = 1$. Hierbei ist E_n die $n \times n$-Einheitsmatrix:

$$(E_n)_{ij} = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Satz I.3.2. *Ist $n \in \mathbb{N}$, K ein Körper und ist* $\det\colon M(n \times n, K) \to K$ *eine Determinantenabbildung, so gilt für alle $A, B \in M(n \times n, K)$ und alle $\lambda \in K$:*

(1) $\det(\lambda A) = \lambda^n \det(A)$.
(2) *Ist eine Zeile von A gleich 0, so ist $\det(A) = 0$.*
(3) *Entsteht B aus A durch das Vertauschen zweier Zeilen, so ist $\det(B) = -\det(A)$.*
(4) *Entsteht B aus A durch die Addition eines Vielfachen einer Zeile von A, so ist $\det(B) = \det(A)$.*
(5) *Ist A eine obere Dreiecksmatrix, also ist $A = (a_{ij})$ und $a_{ij} = 0$ für alle $i > j$, so ist $\det(A) = \prod_{i=1}^n a_{ii}$:*

$$\det\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} = \prod_{i=1}^n a_{ii}.$$

Beweis. Zu (1): Jede der n Zeilen von A ist mit λ multipliziert und daher erhalten wir mit (D1), dass $\det(\lambda A) = \lambda^n \det(A)$ ist.

Zu (2): Ist die i-te Zeile von A null, so wählen Sie $a_{i1} = \ldots = a_{in} = 1$. Dann ist

$$\det\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,n} \\ 0 & \dots & 0 \\ a_{i+1,1} & \dots & a_{i+1,n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \det\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,n} \\ 0\cdot a_{i1} & \dots & 0\cdot a_{in} \\ a_{i+1,1} & \dots & a_{i+1,n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = 0 \cdot \det(A) = 0.$$

Zu (3): Es sei $a_i = (a_{i1}, \ldots, a_{in})$ die i-te Zeile von A. Dann können wir A schreiben als

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Es sei B die Matrix, die aus A entsteht, indem man Zeile i und Zeile j vertauscht für $i < j$. Nach (D2) gilt

$$0 = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i + a_j \\ a_{i+1} \\ \vdots \\ a_{j-1} \\ a_i + a_j \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix}$$

und nach (D1) und (D2) können wir dies umformen zu

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_j \\ a_{i+1} \\ \vdots \\ a_{j-1} \\ a_i \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix} + \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i \\ a_{i+1} \\ \vdots \\ a_{j-1} \\ a_j \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix},$$

weil

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i \\ a_{i+1} \\ \vdots \\ a_{j-1} \\ a_i \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix} = 0 = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_j \\ a_{i+1} \\ \vdots \\ a_{j-1} \\ a_j \\ a_{j+1} \\ \vdots \\ a_n \end{pmatrix}$$

und wir erhalten $\det(B) + \det(A) = 0$.

Zu (4): Wir schreiben die Matrix B als $\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i+\lambda a_j \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}$ und erhalten wegen (D1):

$$\det B = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i+\lambda a_j \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_i \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} + \det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ \lambda a_j \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} = \det(A) + \lambda\det\begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_j \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}$$

aber im letzten Term kommt die Zeile a_j doppelt vor und daher bleibt nur $\det(A)$ übrig.

Zu (5): Ist $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$ und sind alle $a_{ii} \neq 0$, so kann man A durch wiederholte Addition von Vielfachen von Zeilen in die Matrix transformieren, die als nicht-triviale Einträge nur die Diagonaleinträge $a_{11}, \dots, a_{nn}$ hat:

$$\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}.$$

Aber det dieser Matrix ist nach (D1)

$$\det\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} = a_{11} \cdot \ldots \cdot a_{nn} \det(E_n)$$

und wegen der Normierung in (D3) ist dies gleich $\prod_{i=1}^n a_{ii}$.

Sind nicht alle $a_{ii} \neq 0$, so gibt es ein größtes i_0, sodass $a_{i_0 i_0} = 0$. Durch wiederholte Addition von Vielfachen der Zeilen $i_0+1, \dots, n$ kann man A transformieren zu einer Matrix, deren i_0-te Zeile nur aus Nullen besteht. Damit ist aber $\det(A)$ wegen (1) trivial. □

Lemma I.3.3. *Ist $A \in M(n \times n, K)$ und ist $\det\colon M(n \times n, K) \to K$ eine Determinantenabbildung. Dann ist $\det(A) = 0$ genau dann trivial, wenn $\mathsf{rg}(A) < n$ gilt.*

Beweis. Durch allgemeine Zeilenumformungen können wir A zu A' transformieren, wobei A' eine obere Dreiecksmatrix ist:

$$A' = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Hierbei steht $*$ für einen beliebigen Teil der Matrix A', der uns nicht weiter interessiert und 0 steht für ein unteres Dreieck voller Nullen. Damit ist

$$\det(A) = \pm\det(A') = \pm\prod_{i=1}^{n}\lambda_i$$

und die $\det(A)$ verschwindet genau dann, wenn es ein $i \in \{1, \dots, n\}$ gibt, sodass $\lambda_i = 0$. Dies ist genau dann der Fall, wenn A nicht invertierbar ist, und dies wiederum ist äquivalent dazu, dass A nicht vollen Rang hat. □

Korollar I.3.4 (Eindeutigkeit der Determinantenabbildung)**.** *Für jeden Körper K und für alle $n \geqslant 1$ gibt es höchstens eine Determinantenabbildung*

$$\det\colon M(n\times n, K) \to K.$$

Beweis. Jedes $A \in M(n \times n, K)$ kann durch elementare Zeilenumformungen auf obere Dreiecksgestalt

$$A' = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

gebracht werden und es gilt $\det(A) = \pm\det(A')$, wobei das Vorzeichen bestimmt ist durch die Anzahl der Zeilenvertauschungen. Damit ist

$$\det(A) = \pm\prod_{i=1}^{n}\lambda_i$$

eindeutig bestimmt. □

Satz I.3.5. *Für jeden Körper K und für alle $n \geqslant 1$ gibt es eine Determinantenabbildung*

$$\det = \det_n\colon M(n\times n, K) \to K.$$

Beweis. Wir beweisen die Behauptung mit vollständiger Induktion über n. Für $n = 1$ müssen wir $\det_1(a) = a$ setzen und diese Abbildung erfüllt die Axiome (D1), (D2) und (D3).

Für den Induktionsschritt definieren wir A'_{ij} als die Matrix, die aus A entsteht, wenn wir die i-te Zeile und die j-te Spalte streichen.

$$A'_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{nn} \end{pmatrix}$$

Damit ist $A'_{ij} \in M((n-1)\times(n-1), K)$ und $\det_{n-1} A'_{ij}$ existiert.

Wir wählen ein beliebiges aber festes $j \in \{1, \dots, n\}$ und definieren

$$\det_n(A) := \textstyle\sum_{i=1}^{n}(-1)^{i+j}a_{ij}\cdot\det_{n-1}(A'_{ij}).$$

Wir müssen (D1), (D2) und (D3) für dieses $\det_n$ nachrechnen.

(D1): Ist $\tilde{A} = (\tilde{a}_{ij})$ mit

$$\tilde{a}_{ij} = \begin{cases} \lambda a_{kj}, & \text{für } k = i, \\ a_{ij}, & \text{für } k \neq i. \end{cases}$$

Damit ist $A'_{kj} = \tilde{A}'_{kj}$ und für $i \neq k$ entsteht $\tilde{A}'_{ij}$ durch Multiplikation einer Zeile von A'_{ij} mit $\lambda \in K$. Damit ist $\det_{n-1}(\tilde{A}'_{ij}) = \lambda \det_{n-1}(A'_{ij})$ und für $\det_n$ erhalten wir

$$\begin{aligned}\det_n \tilde{A} &= \sum_{\substack{i=1\\ i\neq k}}^{n} (-1)^{i+j} \tilde{a}_{ij} \det_{n-1}(\tilde{A}'_{ij}) + (-1)^{k+j} \tilde{a}_{kj} \det_{n-1}(\tilde{A}'_{kj}) \\ &= \sum_{\substack{i=1\\ i\neq k}}^{n} (-1)^{i+j} \tilde{a}_{ij} \lambda \det_{n-1}(A'_{ij}) + (-1)^{k+j} \lambda a_{kj} \det_{n-1}(A'_{kj}) \\ &= \lambda \det_n(A).\end{aligned}$$

Die Additivität von $\det_n$ zeigen Sie analog.

Zu (D2): In A sei die k-te Zeile gleich der ℓ-ten für $k \neq \ell$.

Für $i \neq k$ oder $i \neq \ell$ hat A'_{ij} auch zwei übereinstimmende Zeilen und somit ist nach Induktionsannahme $\det_{n-1}(A'_{ij}) = 0$ für alle $i \neq k$ und alle $i \neq \ell$. Damit bleiben nur zwei Terme für $\det_n(A)$:

$$\det_n(A) = (-1)^{k+j} a_{k\ell} \det_{n-1}(A'_{kj}) + (-1)^{\ell+j} a_{\ell j} \det_{n-1}(A'_{\ell j}).$$

Es gilt $a_{kj} = a_{\ell j}$. Ohne Beschränkung der Allgemeinheit ist $k < \ell$. Die Matrix $A'_{\ell j}$ geht aus A'_{kj} durch $\ell - k - 1$ Zeilenvertauschungen hervor. Damit ist

$$\det_{n-1}(A'_{\ell j}) = (-1)^{\ell-k-1} \det_{n-1}(A'_{kj})$$

und wir erhalten

$$\det_n(A) = (-1)^{k+j} a_{kj} \det_{n-1}(A'_{kj}) + (-1)^{\ell+j} (-1)^{\ell-k-1} \det_{n-1}(A'_{kj}) = 0,$$

weil

$$(-1)^{\ell-k-1+\ell-j} = (-1)^{2\ell-k-j-1} = (-1)^{k+j+1}.$$

Zu (D3): Wir rechnen einfach nach:

$$\det_n(E_n) = \textstyle\sum_{i=1}^{n} (-1)^{i+j} (E_n)_{ij} \det_{n-1}((E_n)'_{ij})$$

Da $(E_n)_{ij} = \delta_{ij}$, bleibt nur der Term für $i = j$ übrig und die obige Summe ist

$$(-1)^{2j} \det_{n-1}(E_{n-1}) = 1.$$

□

Wir schreiben ab jetzt wieder det statt $\det_n$. Die Determinantenformel im Beweis von Satz I.3.5 sollte Ihnen aus der Mathematik 2 bekannt vorkommen. Wir erhalten sofort die Folgerung:

Korollar I.3.6 (Spaltenentwicklungssatz von Laplace)**.** *Für $A \in M(n \times n, K)$ ist*

(I.3.1)
$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A'_{ij})$$

für jedes beliebige aber fest gewählte $j \in \{1, \ldots, n\}$.

Da wir wissen, dass diese Formel für die Determinante die gleichen Werte hat wie die, die Sie mit der Leibnizformel erhalten, mussten wir also die Determinante so definieren, wie wir das in der Mathematik 2 getan haben.

KAPITEL II

Diagonalisierbare Abbildungen

Bemerkung II.0.1. Die Frage nach der Diagonalisierbarkeit ist die folgende: Gegeben sei eine lineare Abbildung $T\colon V \to V$ auf einem endlichdimensionalen K–Vektorraum V. Gibt es eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V, sodass $M_{\mathcal{B}}(T)$ möglichst einfache Gestalt hat?

Der Idealfall wäre:

$$M_{\mathcal{B}}(T) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

d.h., $M_{\mathcal{B}}(T)$ ist eine Diagonalmatrix. Wäre dies der Fall, so hätten wir

$$(Tv_k)_{\mathcal{B}} = M_{\mathcal{B}}(T) \cdot (v_k)_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = (\lambda_k v_k)_{\mathcal{B}}$$

also $Tv_k = \lambda_k v_k$ für jedes k.

II.1. Eigenwerte und Eigenvektoren

Wiederholung II.1.1. Es sei V ein endlichdimensionaler Vektorraum über K und $T\colon V \to V$ sei linear.

(i) Ein $\lambda \in K$ ist *Eigenwert* (EW) von T genau dann, wenn $0 \neq v \in V$ existiert mit $Tv = \lambda v$.

(ii) Ein v wie in (i) heißt *Eigenvektor* (EV); $\mathrm{E}_\lambda(T) := \{v \in V \mid v \text{ ist EV zum EW } \lambda\} \cup \{0\} = \{v \in V \mid Tv = \lambda v\} = \ker(T - \lambda)$ heißt *Eigenraum* von T zum EW λ.

Definition II.1.2. Wir nennen $\sigma(T) = \{\lambda \in K \mid \lambda \text{ ist EW von } T\}$ das *Spektrum* von T.

Bemerkungen II.1.3. Eigenvektoren zu verschiedenen Eigenwerten sind stets linear unabhängig; wir zeigen dies in einer Präsenzaufgabe.

Die Berechnung von Eigenwerten und Eigenvektoren einer linearen Abbildung $K^n \to K^n$, $x \mapsto A \cdot x$ mit $A \in M(n \times n, K)$ funktioniert nach folgendem Schema:

(1) Zuerst bestimmen wir das charakteristisches Polynom $P_A(\lambda) = \det(A - \lambda E_n)$. Beachten Sie, dass im letzten Semester $P_A(\lambda) = \det(\lambda E_n - A)$ definiert wurde, dass dies aber keinen Einfluss auf die Nullstellen hat.

(2) Wir bestimmen dann die Nullstellen von P_A. Dies liefert die EWe und deren *algebraische Vielfachheiten.*

(3) Für jeden EW λ lösen wir das LGS $(A - \lambda E_n)v = 0$. Dies liefert den Eigenraum (in Form einer Basis des Lösungsraums des LGS) und die *geometrische Vielfachheit*$(= \dim \mathrm{E}_\lambda(T))$ von λ.

Beispiel II.1.4. Es sei $A = \begin{pmatrix} -5 & 0 & 7 \\ 6 & 2 & -6 \\ -4 & 0 & 6 \end{pmatrix}$

(1) Dann ist $P_A(\lambda) = \det \begin{pmatrix} -5-\lambda & 0 & 7 \\ 6 & 2-\lambda & -6 \\ -4 & 0 & 6-\lambda \end{pmatrix} = \cdots = -\lambda^3 + 3\lambda^2 - 4.$

(b) Wir raten die Nullstelle $\lambda = -1$ und führen dann die Polynomdivision durch.

$$
\begin{array}{l}
\big(-\lambda^3+3\lambda^2 \qquad -4\big):\big(\lambda+1\big) = -\lambda^2+4\lambda-4 \\
\underline{\ \ \lambda^3 \ +\lambda^2\ \ } \\
\qquad\quad 4\lambda^2 \\
\qquad \underline{-4\lambda^2-4\lambda} \\
\qquad\qquad\quad -4\lambda-4 \\
\qquad\qquad\quad \underline{\ \ 4\lambda+4\ } \\
\qquad\qquad\qquad\quad 0
\end{array}
$$

und sehen, dass $\lambda = 2$ zweifache Nullstelle ist. Also ist $\sigma(A) = \{-1, 2\}$, wobei $\lambda = -1$ algebraische Vielfachheit 1 hat und $\lambda = 2$ algebraische Vielfachheit 2 hat.

(3) Wir lösen das LGS $(A-(-1))x = 0$ für $x = (x_1, x_2, x_3)^t$. Mit dem Gaußverfahren erhält man

$$
\begin{pmatrix} -4 & 0 & 7 \\ 6 & 3 & -6 \\ -4 & 0 & 7 \end{pmatrix} \begin{array}{l} \mid \cdot 3 \\ \mid \cdot 2 \ + \\ \end{array}
$$

$$
\begin{pmatrix} -4 & 0 & 7 \\ 0 & 6 & 9 \\ -4 & 0 & 7 \end{pmatrix} \begin{array}{l} \mid \cdot (-1) \\ \\ + \end{array}
$$

$$
\begin{pmatrix} -4 & 0 & 7 \\ 0 & 6 & 9 \\ 0 & 0 & 0 \end{pmatrix}
$$

also $x_1 = \frac{7}{4}x_3$, $x_2 = -\frac{3}{2}x_3$ und x_3 ist beliebig. Es folgt

$$
\mathrm{E}_{-1}(A) = \left\{ \begin{pmatrix} 7/4t \\ -3/2t \\ t \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} = \left\{ t \cdot \begin{pmatrix} 7 \\ -6 \\ 4 \end{pmatrix} \,\middle|\, t \in \mathbb{R} \right\} = \mathsf{Span}\left\{ \begin{pmatrix} 7 \\ -6 \\ 4 \end{pmatrix} \right\}.
$$

Für das LGS $(A-2)x = 0$ erhält man analog

$$
\begin{pmatrix} -7 & 0 & 7 \\ 6 & 0 & -6 \\ -4 & 0 & 4 \end{pmatrix} \begin{array}{l} \mid \cdot 1/7 \\ \mid \cdot 1/6 \\ \mid \cdot 1/4 \end{array}
$$

$$
\begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{array}{l} \qquad \mid \cdot (-1) \\ + \\ + \end{array}
$$

$$
\begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}
$$

d.h. $x_1 = x_3$ und x_2 und x_3 sind beliebig. Also

$$
\mathrm{E}_2(A) = \left\{ \begin{pmatrix} t \\ s \\ t \end{pmatrix} \,\middle|\, t,\, s \in \mathbb{R} \right\} = \mathsf{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}.
$$

Wir sehen also, dass $\lambda = -1$ die geometrische Vielfachheit 1 und $\lambda = 2$ die geometrische Vielfachheit 2 hat.

In der Tat hatten wir in der Mathematik 2 gesehen, dass die geometrische Vielfachheit immer kleiner gleich der algebraischen Vielfachheit. Es kann passieren, dass die echte Ungleichung vorliegt (z.B. $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, dann ist $\det(A-\lambda) = (1-\lambda)^2$, die algebraische Vielfachheit von $\lambda = 1$ ist also 2, aber $\mathrm{E}_1(A) = \left\{ \left(\begin{smallmatrix} t \\ 0 \end{smallmatrix}\right) \,\middle|\, t \in \mathbb{R} \right\}$ und damit hat $\lambda = 1$ geometrische Vielfacheit 1).

Sollen EWe und EVen für eine lineare Abbildung $T\colon V \to V$ auf einem beliebigen endlichdimensionalen Vektorraum V bestimmt werden, so kann obiges Verfahren auf die Matrix $A = M_{\mathcal{B}}(T)$ angewandt werden, wobei $\mathcal{B}$ eine Basis von V ist. Das charakteristische Polynom, und damit die Eigenwerte und ihre algebraischen Vielfachheiten, sind dann unabhängig von der Wahl von $\mathcal{B}$ und es macht Sinn P_T zu schreiben. Das Verfahren oben liefert die Eigenräume *in den Koordinaten der Basis* $\mathcal{B}$, insbesondere ändern sich also die Eigenräume wenn man die Basis ändert. Ihre Dimensionen, und damit die geometrischen Vielfachheiten der EWen, sind aber unabhängig von der Wahl der Basis. Das zeigen Sie in einer Übungsaufgabe.

II.2. Diagonalisierbarkeit

Definition II.2.1.

(1) Eine lineare Abbildung $T\colon V \to V$ auf einem n–dimensionalen Vektorraum V heißt *diagonalisierbar*, falls es eine Basis $\mathcal{B} = (b_1, \ldots, b_n)$ aus Eigenvektoren gibt, also genau dann, wenn $M_{\mathcal{B}}(T)$ eine Diagonalmatrix ist.

(2) Eine Matrix $A \in M(n \times n, K)$ heißt *diagonalisierbar*, wenn die induzierte lineare Abbildung $K^n \to K^n$, $x \mapsto A \cdot x$ diagonalisierbar ist.

Bemerkungen II.2.2.

- Ein $A \in M(n \times n, K)$ ist genau dann diagonalisierbar, wenn es eine invertierbare Matrix $S \in M(n \times n, K)$ gibt mit
$$SAS^{-1} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$
wobei $\lambda_1, \ldots, \lambda_n$ die EWe von A sind. Hierbei ist $S = M_{\mathcal{B}\mathcal{E}}(E)$ die darstellende Matrix des Basiswechsels von der Standardbasis $\mathcal{E} = (e_1, \ldots, e_n)$ zur Basis $\mathcal{B}$.
- Explizit: $SAS^{-1}e_k = \lambda e_k$ impliziert $AS^{-1}e_k = \lambda_i S^{-1}e_k$, d.h. $S^{-1}e_k$, die k–te Spalte von S^{-1}, ist ein Eigenvektor von A zum Eigenwert λ_k.

Satz II.2.3. *(Kriterium für Diagonalisierbarkeit) Es sei V ein n–dimensionaler Vektorraum über K. Eine lineare Abbildung $T\colon V \to V$ ist genau dann diagonalisierbar, wenn die zwei folgenden Bedingungen*[1] beide *erfüllt sind:*

(i) *Das charakteristisches Polynom zerfällt über K in Linearfaktoren.*

(ii) *Für jeden Eigenwert λ ist die algebraische Vielfachheit gleich der geometrischen Vielfachheit.*

Beweis. '$\Longrightarrow$' Es sei T diagonalisierbar, d.h. es existiert eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V und Skalare $\lambda_1, \ldots, \lambda_n \in K$ mit $Tv_k = \lambda v_k$ für $k = 1, \ldots, n$. Insbesondere ist
$$D := M_{\mathcal{B}}(T) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$
und $P_T(\lambda) = P_D(\lambda) = \det(D - \lambda) = \prod_{i=1}^{n} (\lambda_i - \lambda)$ zerfällt in Linearfaktoren, womit (i) gezeigt ist.

Es seien jetzt $\mu_1, \ldots, \mu_r$ die paarweise verschiedenen Nullstellen von P_T und $k_1, \ldots, k_r$ die zugehörigen Vielfachheiten. Durch eventuelles Umnummerieren der Basis erreichen wir
$$\begin{aligned} \lambda_1 = \cdots = \lambda_{k_1} &= \mu_1 \\ \lambda_{k_1+1} = \cdots = \lambda_{k_1+k_2} &= \mu_2 \\ \lambda_{k_1+k_2+1} = \cdots = \lambda_{k_1+k_2+k_3} &= \mu_3 \\ &\vdots \\ \lambda_{k_1+\cdots+k_{r-1}+1} = \cdots = \lambda_{k_1+\cdots+k_r} &= \mu_r \end{aligned}$$

[1] Beachten Sie, dass nach II.1.1(vii) beide Bedingungen unabhängig von der Basiswahl sind, wenn T durch die Matrix $A = M_{\mathcal{B}}(T)$ dargestellt wird, um das charakteristische Polynom und die Eigenräume zu bestimmen.

und müssen $\dim \mathrm{E}_{\mu_i}(T) = k_{\mu_i}$ für $i = 1, \ldots, r$ zeigen. Allerdings genügt es natürlich, dies für $k = 1$ zu zeigen, weil wir ansonsten einfach nochmal umnummerieren können. Durch die oben vorgenommene Umnummerierung und Umbenennung haben wir $\mathcal{B}$ jetzt so gewählt, dass

$$D = \begin{pmatrix} \mu_1 & & & & & \\ & \ddots & & & & \\ & & \mu_1 & & & \\ & & & \mu_2 & & \\ & & & & \ddots & \\ & & & & & \mu_2 \\ & & & & & & \ddots \end{pmatrix}$$

gilt, wobei μ_1 genau k_1–mal vorkommt, μ_2 genau k_2–mal usw. Damit gelten für $v \in V$ mit Koordinaten $x_1, \ldots, x_n$ bezüglich der Basis $\mathcal{B}$ die Äquivalenzen

$$\begin{aligned} Tv = \mu_1 v &\iff (Tv)_{\mathcal{B}} = (\mu_1 v)_{\mathcal{B}} \\ &\iff Dv_{\mathcal{B}} = \mu_1 v_{\mathcal{B}} \\ &\iff \begin{pmatrix} \mu_1 & & & & & \\ & \ddots & & & & \\ & & \mu_1 & & & \\ & & & \mu_2 & & \\ & & & & \ddots & \\ & & & & & \mu_2 \\ & & & & & & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{k_1} \\ x_{k_1+1} \\ \vdots \\ x_{k_1+k_2} \\ \vdots \end{pmatrix} = \mu_1 \begin{pmatrix} x_1 \\ \vdots \\ x_{k_1} \\ x_{k_1+1} \\ \vdots \\ x_{k_1+k_2} \\ \vdots \end{pmatrix} \\ &\iff \left\{ \begin{array}{l} \left.\begin{array}{c} \mu_1 x_1 = \mu_1 x_1 \\ \vdots \\ \mu_1 x_{k_1} = \mu_1 x_{k_1} \end{array}\right\} \text{Gilt für beliebige } x_1, \ldots, x_{k_1}. \\ \left.\begin{array}{c} \mu_2 x_{k_1+1} = \mu_1 x_{k_1+1} \\ \vdots \end{array}\right\} \begin{array}{l} \text{Da } \mu_1 \neq \mu_2,\ \mu_1 \neq \mu_3,\ \ldots, \text{ gilt dies} \\ \text{genau dann, wenn } x_i = 0 \text{ für } i > k_1. \end{array} \end{array} \right. \\ &\iff v_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_{k_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \text{ mit } x_1, \ldots, x_{k_1} \in K \text{ beliebig.} \end{aligned}$$

Daraus folgt, dass $\mathrm{E}_{\mu_1}(T) = \{v \in V \mid Tv = \mu_1 v\}$ Dimension k_1 haben muss, womit (ii) gezeigt ist.

'$\Longleftarrow$' Es seien $\mu_1, \ldots, \mu_r \in K$ die paarweise verschiedenen Eigenwerte von T und es seien $k_1, \ldots, k_r$ Ihre algebraischen/geometrischen Vielfachheiten, die nach (ii) gleich sind. Da nach (i) das charakteristische Polynom zerfällt, gilt $k_1 + \cdots + k_r = n$. Um eine Basis $\mathcal{B}$ aus EVen zu finden, wählen wir zunächst in jedem E_{μ_i} eine Basis $\mathcal{B}_i$ aus, die dann nach Definition aus Eigenvektoren besteht und nach obigem k_i-viele Elemente hat. Dann setzen wir

$$\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$$

und behaupten, dass diese Menge linear unabhängig ist: Angenommen, sie ist es nicht, dann gibt es eine Linearkombination der Elemente von $\mathcal{B}$

$$0 = \underbrace{\alpha_1 v_1 + \cdots + \alpha_{k_1} v_{k_1}}_{=: x_1 \in \mathrm{E}_{\mu_1}} + \underbrace{\alpha_{k_1+1} v_{k_1+1} + \cdots + \alpha_{k_1+k_2} v_{k_1+k_2}}_{=: x_2 \in \mathrm{E}_{\mu_2}} + \cdots =: (*),$$

in der mindestens ein Koeffizient nicht Null ist. Wir nehmen ohne Einschränkung an, dass dieser Koeffizient vor einem Vektor aus $\mathcal{B}_1$ auftritt (sonst nummerieren wir die Eigenwerte und Eigenräume geeignet um). Ist aber nun unter den $\alpha_1, \ldots, \alpha_{k_1}$ ein Koeffizient ungleich Null, dann muss $x_1 = \alpha_1 v_1 + \cdots + \alpha_{k_1} v_{k_1} \neq 0$ sein, weil die $v_1, \ldots, v_{k_1}$ als Basis von E_{μ_1} linear unabhängig sind. Jetzt lassen wir diejenigen x_i weg, die gleich Null sind und erhalten

$$0 = (*) = x_1 + \cdots + x_r = \sum_{\substack{i=1 \\ x_i \neq 0}}^{n} 1 \cdot x_i$$

eine nichttriviale Linearkombination der Null aus Eigenvektoren zu paarweise verschiedenen Eigenwerten im Widerspruch dazu, dass EVen zu paarweise verschiedenen EWen linear unabhängig sind, vgl. II.1.1(iv). $\square$

Beispiel II.2.4. (Fortsetzung des Beispiels in II.1.1(v)) Wir betrachten die Matrix

$$A = \begin{pmatrix} -5 & 0 & 7 \\ 6 & 2 & -6 \\ -4 & 0 & 6 \end{pmatrix},$$

für die wir bereits gezeigt haben, dass $\sigma(A) = \{-1, 2\}$ gilt, wobei $\lambda = -1$ algebraische Vielfachheit 1 und $\lambda = 2$ algebraische Vielfachheit 2 hat, sowie dass

$$\mathrm{E}_{-1}(A) = \mathsf{Span}\Big\{ \begin{pmatrix} 7 \\ -6 \\ 4 \end{pmatrix} \Big\} \text{ und } \mathrm{E}_2(A) = \mathsf{Span}\Big\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big\}$$

gelten. Insbesondere erfüllt die lineare Abbildung $K^n \to K^n$, $x \mapsto Ax$, die Eigenschaften (i) und (ii) des Diagonalisierbarkeitskriteriums in II.2.3 und wir können direkt ablesen, dass die Basis

$$\mathcal{B} = \Big\{ \begin{pmatrix} 7 \\ -6 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \Big\}$$

zur Darstellungsmatrix

$$M_{\mathcal{B}}(A) = \begin{pmatrix} -1 & & \\ & 2 & \\ & & 2 \end{pmatrix}$$

führt. Weiter sehen wir, dass

$$S^{-1} = \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix}$$

gewählt werden kann und dass jetzt nur noch $S = (S^{-1})^{-1}$ bestimmt werden muss. Dies kann mit dem Gaußverfahren gemacht werden oder via der Adjunkten. Wir machen hier Letzteres und bestimmen zunächst mit der Sarrus-Regel

$$\det S^{-1} = 0 + 0 + 4 - 0 - 7 - 0 = -3$$

und nun

$$\begin{aligned}
S &= \frac{1}{\det S^{-1}} \operatorname{adj} S^{-1} \\
&= -\tfrac{1}{3} \begin{pmatrix}
(-1)^2 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^3 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^4 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} \\
(-1)^3 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^4 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^5 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} \\
(-1)^4 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^5 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} & (-1)^6 \det \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix}
\end{pmatrix} \\
&= -\tfrac{1}{3} \begin{pmatrix} -1 & 0 & 1 \\ 4 & 0 & -7 \\ -6 & -3 & 6 \end{pmatrix}
\end{aligned}$$

Dies liefert in der Tat die gewünsche Zerlegung:

$$SAS^{-1} = -\tfrac{1}{3} \begin{pmatrix} -1 & 0 & 1 \\ 4 & 0 & -7 \\ -6 & -3 & 6 \end{pmatrix} \begin{pmatrix} -5 & 0 & 7 \\ 6 & 2 & -6 \\ -4 & 0 & 6 \end{pmatrix} \begin{pmatrix} 7 & 1 & 0 \\ -6 & 0 & 1 \\ 4 & 1 & 0 \end{pmatrix} = \cdots = \begin{pmatrix} -1 & & \\ & 2 & \\ & & 2 \end{pmatrix}.$$

Bemerkungen II.2.5.

(1) Zerfällt das charateristische Polynom einer linearen Abbildung und haben alle EWe algebraische Vielfachheit 1, so ist die Abbildung automatisch diagonalisierbar, weil die geometrische Vielfachheit eines jeden Eigenwertes mindestens 1 ist, also hier gleich 1.

(2) Nach II.2.3 ist nicht jede lineare Abbildung T diagonalisierbar; es bleibt also die Frage, wie eine möglichst einfach Darstellungsmatrix $M_{\mathcal{B}}(T)$ für solche T aussieht. Dies beantwortet der *Satz von der Jordan–Normalform*, den wir ohne Beweis notieren: Es sei K ein Körper, V ein Vektorraum über K und $T\colon V \to V$ sei linear. Falls P_T über K in Linearfaktoren zerfällt, so existiert eine Basis $\mathcal{B}$ von V, sodass

$$M_{\mathcal{B}}(T) = \begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & 1 & & \\ & \ddots & \ddots & \\ & & & 1 \\ & & & \lambda_1 \end{matrix}} & & \\ & \boxed{\begin{matrix} \lambda_2 & 1 & & \\ & \ddots & \ddots & \\ & & & 1 \\ & & & \lambda_2 \end{matrix}} & \\ & & \ddots \end{pmatrix}$$

Blockdiagonalform hat und die Blöcke sogenannte *Jordanblöcke* sind (d.h. in der Hauptdiagonalen des Blocks steht jeweils der gleiche Eintrag, in der ersten Nebendiagonalen stehen Einsen und alle restlichen Einträge sind Null). Die λ_i sind hierbei die EWe von T. Zu jedem $\lambda \in \sigma(T)$ gibt es seiner geometrischen Vielfachheit entsprechend viele Jordanblöcke und die Gesamtdimension der Blöcke zum EW λ ist gleich der algebraischen Vielfachheit von λ. Im Bild oben ist also $\lambda_i = \lambda_j$ für $i \neq j$ durchaus möglich. Ist T durch Multiplikation mit der Matrix A gegeben, so gibt es eine invertierbare Matrix S mit $M_{\mathcal{B}}(T) = SAS^{-1}$.

KAPITEL III

Stetige Funktionen

III.1. Wiederholung

Es seien D und W Mengen. Eine *Funktion*, oder *Abbildung*, von D nach W ordnet jedem $x \in D$ genau ein $y \in W$ zu. Wir schreiben

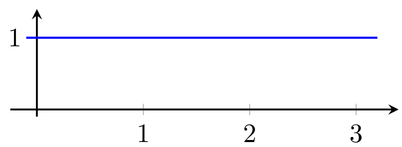
$$f\colon D \to W,\ x \mapsto f(x)$$

und nennen D den *Definitionsbereich* und W den *Wertebereich* der Funktion f. Im Folgenden sind stets D und W Teilmengen von $\mathbb{R}$ oder $\mathbb{C}$, und wir nehmen immer $D \neq \emptyset$ und $W \neq \emptyset$ an, ohne dies explizit zu notieren. Definiert man Funktionen mithilfe von Relationen, so wie wir das in Mathematik 1 gemacht haben, dann sieht man, dass es genau eine Abbildung $\emptyset \to W$ gibt bei beliebigem W, und keine Abbildung $D \to \emptyset$ für $D \neq \emptyset$.

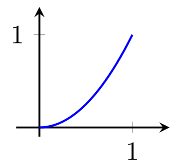
Schließlich kürzen wir „Es sei $D \subset \mathbb{R}$ und $f\colon D \to \mathbb{R}$“ ab durch „Es sei $f\colon \mathbb{R} \supset D \to \mathbb{R}$“ oder $f\colon D \to \mathbb{R}$.

Beispiele III.1.1.

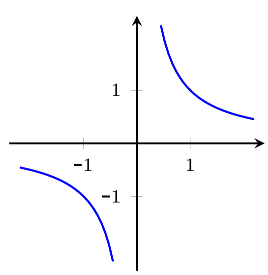
(1) $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = 1$:



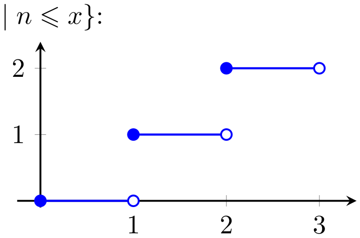
(2) $g\colon [0,1] \to \mathbb{R}$, $g(x) = x^2$:



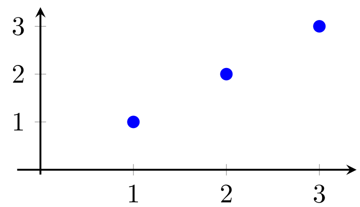
(3) $h\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $h(x) = 1/x$:



(4) $i\colon [0,\infty) \to \mathbb{R}$, $i(x) = \max\{n \in \mathbb{Z} \mid n \leqslant x\}$:



(5) $j\colon \mathbb{N} \to \mathbb{R}$, $j(x) = x$:



(6) $k\colon [0,1] \to \mathbb{R}$, $k(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & \text{sonst.} \end{cases}$

Die Funktion k heißt *Dirichletfunktion* und wird oft als 'unendlich feiner Kamm' beschrieben.

(7) $\ell\colon \{z \in \mathbb{C} \mid |z - \frac{2}{3}| < \frac{1}{3}\} \to \mathbb{C}$, $\ell(z) = \sum\limits_{k=0}^{\infty} 3^k \big(z - \frac{2}{3}\big)^k$.

(8) $\exp\colon \mathbb{C} \to \mathbb{C}$.

III.2. Stetigkeit

Definition III.2.1. Es sei $f\colon \mathbb{R} \supset D \to \mathbb{R}$.

(i) f ist *stetig an der Stelle* $x_0 \in D$ (oder *im Punkt* $x_0 \in D$), falls gilt

$$\forall\, \varepsilon > 0\; \exists\, \delta > 0\; \forall\, x \in D\colon |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

(ii) f ist *stetig* (*auf* D), falls f stetig an jeder Stelle $x_0 \in D$ ist, d.h., falls gilt

$$\forall\, x_0 \in D,\ \varepsilon > 0\; \exists\, \delta > 0\; \forall\, x \in D\colon |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

Bemerkungen III.2.2.

(i) In III.2.1 bezeichnet $|\cdot|$ den Betrag in $\mathbb{R}$.

(ii) Ersetzen wir in III.2.1 den Wertebereich $\mathbb{R}$ durch W mit $\operatorname{Bild}(f) = \{f(x) \mid x \in D\} \subset W \subset \mathbb{R}$, so ist die Definition unabhängig von W.

(iii) Ist eine komplexe Funktion $f\colon \mathbb{C} \supset D \to \mathbb{C}$ gegeben, dann können wir $|\cdot|$ als Betrag von $\mathbb{C}$ lesen und genau wie in III.2.1 definieren, was Stetigkeit in $x_0 \in D$ bzw. auf ganz D bedeutet.

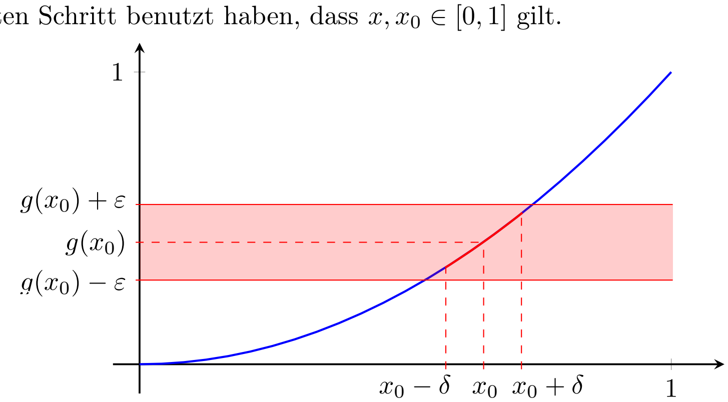
Beispiele III.2.3. Für die Beispiele von oben ergibt sich:

(1) $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = 1$, ist stetig auf $\mathbb{R}$, denn $|f(x) - f(x_0)| = 0$ für alle $x, x_0 \in \mathbb{R}$.

(2) $g\colon [0,1] \to \mathbb{R}$, $g(x) = x^2$, ist stetig auf $[0,1]$: Es seien $x_0 \in [0,1]$ und $\varepsilon > 0$ gegeben. Wir wählen $\delta = \frac{\varepsilon}{2}$. Es sei $x \in [0,1]$ mit $|x - x_0| < \delta$ gegeben. Dann gilt

$$|g(x) - g(x_0)| = |x^2 - x_0^2| = |x - x_0||x + x_0| < \delta(|x| + |x_0|) \leqslant 2\delta = \varepsilon,$$

wobei wir im letzten Schritt benutzt haben, dass $x, x_0 \in [0,1]$ gilt.

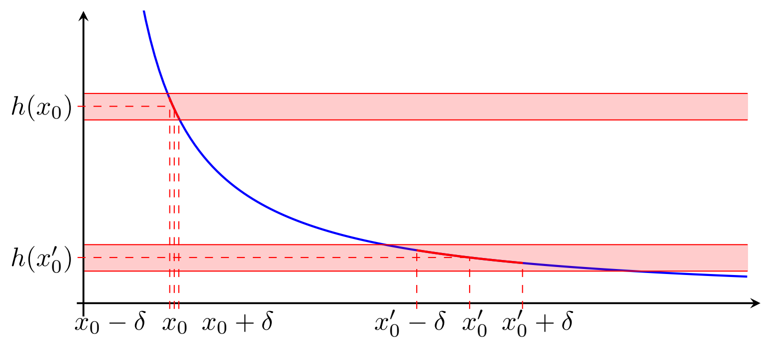


(3) $h\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $h(x) = \frac{1}{x}$, ist stetig auf $\mathbb{R}\backslash\{0\}$: Es seien $x_0 \in \mathbb{R}\backslash\{0\}$ und $\varepsilon > 0$ gegeben. Wir wählen $\delta = \min\{\frac{|x_0|}{2}, \frac{\varepsilon|x_0|^2}{2}\}$. Dann gilt

$$|h(x) - h(x_0)| = \Big|\frac{1}{x} - \frac{1}{x_0}\Big| = \Big|\frac{x_0 - x}{xx_0}\Big| = \frac{|x_0 - x|}{|x||x_0|} < \frac{\delta}{\frac{|x_0|}{2}\cdot|x_0|} = \frac{2\delta}{|x_0|^2} \leqslant \varepsilon,$$

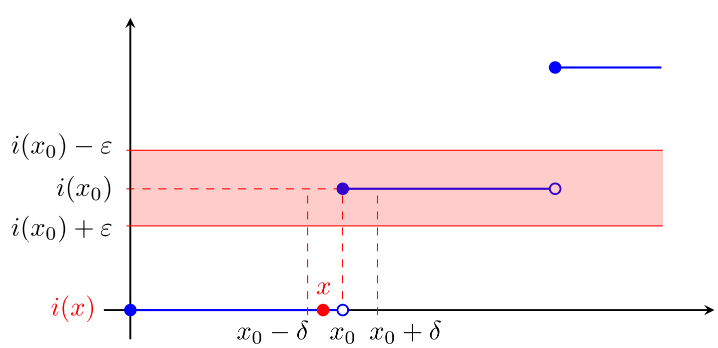
wobei wir für die erste Abschätzung $|x| = |x_0+x-x_0| \geqslant |x_0|-|x-x_0| > |x_0|-\delta \geqslant |x_0|-\frac{|x_0|}{2} = \frac{|x_0|}{2}$ und für die zweite $\delta \leqslant \frac{\varepsilon}{2}|x_0|^2$ benutzt haben.

In diesem Beispiel hängt δ nicht nur von ε sondern auch von x_0 ab. Im Bild unten heißt ein zweites Vorkommen von x_0 hier x_0', um beide Werte zu unterscheiden:



(4) $i\colon [0,\infty) \to \mathbb{R}$, $i(x) = \max\{n \in \mathbb{Z} \mid n \leqslant x\}$, ist genau dann stetig an der Stelle $x_0 \in [0,\infty)$, wenn $x_0 \notin \mathbb{N}$ gilt: Wir zeigen zuerst, dass i in $x_0 \in \mathbb{N}$ nicht stetig ist. Dafür behaupten wir:

$$\exists\, \varepsilon > 0\, \forall\, \delta > 0\, \exists\, x \in [0,\infty)\colon |x - x_0| < \delta \;\wedge\; |i(x) - i(x_0)| \geqslant \varepsilon.$$

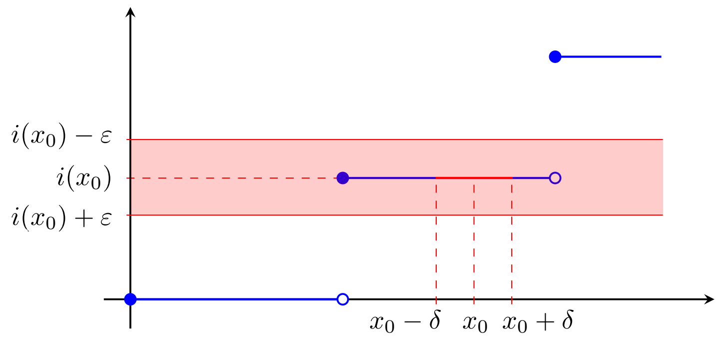


Wir wählen $\varepsilon = \frac{1}{2}$. Es sei $\delta > 0$ beliebig. Wähle $x \in (x_0 - \min\{\delta, 1\}, x_0)$. Dann gilt $|x - x_0| < \delta$ und $i(x) - i(x_0)| = |(x_0 - 1) - x_0| = |-1| = 1 \geqslant \frac{1}{2} = \varepsilon$.

Als nächstes zeigen wir, dass i in $x_0 \in [0,\infty)\backslash\mathbb{N}$ stetig ist. Es seien $\varepsilon > 0$ und $x_0 \in [0,\infty)\backslash\mathbb{N}$ dafür beliebig gegeben.

Fall 1: Wenn $x_0 = 0$ ist, dann wählen wir $\delta = \frac{1}{2}$. Es sei $x \in [0,\infty)$ mit $|x - x_0| = |x| < \delta = \frac{1}{2}$. Dann gilt $|i(x) - i(x_0)| = |i(x) - i(0)| = |0 - 0| < \varepsilon$.

Fall 2: Wenn $x_0 \neq 0$ ist, dann existiert genau ein $n \in \mathbb{N}_0$ mit $n < x_0 < n + 1$.



Wir setzen $\delta = \min\{\frac{|x_0-n|}{2}, \frac{|x_0-(n+1)|}{2}\} > 0$. Es sei $x \in [0,\infty)$ mit $|x-x_0| < \delta$. Dann ist $i(x) = i(x_0)$, also $|i(x) - i(x_0)| = 0 < \varepsilon$.

(5) $j\colon \mathbb{N} \to \mathbb{R}$, $j(x) = x$, ist stetig. Es seien $x_0 \in \mathbb{N}$ und $\varepsilon > 0$ gegeben. Wähle $\delta = \frac{1}{2}$. Es sei $x \in \mathbb{N}$ mit $|x - x_0| < \delta$ gegeben. Das geht aber nur wenn $x = x_0$ ist. Also folgt $|j(x) - j(x_0)| = 0 < \varepsilon$.

(6) $k\colon [0,1] \to \mathbb{R}$, $k(x) = \left\{\begin{matrix} 1, & x \in \mathbb{Q} \\ 0, & x \in \mathbb{R}\backslash\mathbb{Q} \end{matrix}\right\}$ ist nirgends stetig: Es sei $x_0 \in [0,1]$.

Fall 1: Ist $x_0 \in \mathbb{Q}$, so betrachten wir $\varepsilon = \frac{1}{2}$. Es sei $\delta > 0$ beliebig. Aus der Mathematik 1/2 wissen wir, dass es ein $x \in \mathbb{R}\backslash\mathbb{Q}$ gibt mit $|x - x_0| < \delta$. Dann gilt aber $|k(x) - k(x_0)| = |1 - 0| = 1 \geqslant \frac{1}{2} = \varepsilon$.

Fall 2: Ist $x_0 \in \mathbb{R}\backslash\mathbb{Q}$, so geht man analog vor.

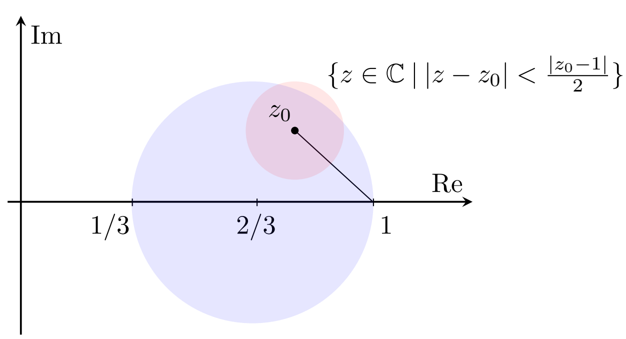
(7) Wir können die Funktion ℓ als $\ell\colon D := \{z \in \mathbb{C} \mid |z - \frac{2}{3}| < \frac{1}{3}\} \to \mathbb{C}$, $\ell(z) = \frac{1}{3-3z}$ schreiben (s. Mathematik 2). Damit sehen wir, dass ℓ sehr ähnlich zu der Funktion h in (iii) ist. In der Tat ist ℓ auf dem gesamten Definitionsbereich stetig: Es seien $z_0 \in D$ und $\varepsilon > 0$ beliebig gegeben. Wir setzen $\delta := \min\{\frac{|1-z_0|}{2}, \frac{3\varepsilon|1-z_0|^2}{2}\}$. Es sei $z \in D$ mit $|z - z_0| < \delta$. Dann gilt

$$\begin{aligned}|\ell(z) - \ell(z_0)| &= \left|\frac{1}{3-3z} - \frac{1}{3-3z_0}\right| = \frac{1}{3}\left|\frac{1-z_0-(1-z)}{(1-z)(1-z_0)}\right| = \frac{1}{3}\frac{|z-z_0|}{|1-z||1-z_0|} \\ &< \frac{1}{3}\frac{\delta}{\frac{|1-z_0|}{2}\cdot|1-z_0|} = \frac{2}{3}\frac{\delta}{|1-z_0|^2} < \varepsilon,\end{aligned}$$

wobei wir für die erste Abschätzung

$$|1-z| = |1-z+z_0-z_0| \geqslant |1-z_0| - |z_0-z| > |1-z_0| - \delta \geqslant |1-z_0| - \frac{|1-z_0|}{2} = \frac{|1-z_0|}{2}$$

und für die zweite $\delta \leqslant \frac{3\varepsilon|1-z_0|^2}{2}$ benutzt haben.



(8) $\exp\colon \mathbb{C} \to \mathbb{C}$ ist stetig auf ganz $\mathbb{C}$. Es seien $\varepsilon > 0$ und $z_0 \in \mathbb{C}$ gegeben. Wir wählen $\delta = \frac{\varepsilon}{|e^{z_0}|+\varepsilon}$. Dann gilt für $|z - z_0| < \delta$

$$|e^z - e^{z_0}| = |e^{z-z_0+z_0} - e^{z_0}| = |e^{z-z_0} - 1||e^{z_0}| = \left|e^{z-z_0} - \sum_{k=0}^{0} \frac{1}{k!}(z-z_0)^k\right||e^{z_0}|$$
$$\leqslant \frac{|z-z_0|^{0+1}}{(0+1)!} e^{|z-z_0|}|e^{z_0}| < \delta e^{\delta}|e^{z_0}| \leqslant \delta \frac{1}{1-\delta}|e^{z_0}| = \varepsilon,$$

wobei wir erst die Fehlerabschätzung der Exponentialfunktion benutzt haben, dann, dass $\exp\colon \mathbb{R} \to \mathbb{R}$ strikt wächst, und schließlich die fundamentale Abschätzung. Letzteres ist möglich, weil $\delta < 1$ gilt. Die Gleichung am Ende folgt durch Auflösen von $\delta = \frac{\varepsilon}{|e^{z_0}|+\varepsilon}$ nach ε.

Bemerkungen III.2.4.

- In der Definition der Stetigkeit kommt es auf kleine $\varepsilon > 0$ an; genauer: Gilt die Bedingung $\forall\, \varepsilon \in (0, \varepsilon_0) \ldots$ mit einem festen $\varepsilon_0 > 0$, dann gilt auch für alle $\forall\, \varepsilon > 0 \ldots$. Beim Nachweis der Stetigkeit kann man also ohne Einschränkung $\varepsilon < 1$ oder etwas Ähnliches annehmen, wenn dies beim Abschätzen hilft.
- In der am Ende zu zeigenden Abschätzung genügt es auch $|f(x) - f(x_0)| \leqslant \varepsilon$ oder $\leqslant 2\varepsilon$ oder $\leqslant \varepsilon^2$ zu zeigen (statt $< \varepsilon$).
- Stetigkeit (in einem Punkt) ist eine lokale Eigenschaft: Es seien $f, g\colon D \subset \mathbb{R} \to \mathbb{R}$ und $x_0 \in D$. Wenn es ein $\varepsilon > 0$ gibt mit $f(x) = g(x)$ für alle $x \in D$ mit $|x - x_0| < \varepsilon$, dann ist g genau dann stetig in x_0, wenn f stetig in x_0 ist.

Definition III.2.5. Es seien D, W Mengen und $f\colon D \to W$ eine Abbildung.

(i) Für $D' \subset D$ bezeichnen wir mit $f|_{D'}\colon D' \to W$, $f|_{D'}(x) = f(x)$ die *Einschränkung* von f auf D'.

(ii) Für $W \supset W' \supset \mathrm{Bild}(f)$ bezeichnen wir mit $f|^{W'}\colon D \to W'$, $f|^{W'}(x) = f(x)$, die *Ko-Einschränkung* von f nach W'.

(iii) Beides kann kombiniert werden, z.B. kann $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, (ko-)eingeschränkt werden zu $f|_{[0,2]}^{[0,5]}\colon [0, 2] \to [0, 5]$. Wenn keine Verwechselungsgefahr besteht, schreiben wir allerdings einfach f für die (Ko-)Einschränkung.

Proposition III.2.6. Es sei $f\colon \mathbb{R} \supset D \to \mathbb{R}$ stetig. Dann gilt:

(i) Jede Einschränkung von f ist stetig.

(ii) Jede wohldefinierte Ko-Einschränkung von f ist stetig.

Beweis. Da f stetig ist, gilt nach Definition

$$\forall\, x_0 \in D,\ \varepsilon > 0\ \exists\, \delta > 0\ \forall\, x \in D\colon |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon.$$

Für (i) sieht man nun, dass die Bedingung auch gilt, wenn man D durch eine kleinere Menge $D' \subset D$ ersetzt. Für (ii) sieht man, dass es sich um exakt die gleiche Bedingung handelt, weil der Wertebereich gar nicht vorkommt. □

Satz III.2.7. *Es seien $D, F, G, H \subset \mathbb{R}$ und es seien $f\colon D \to F$, $g\colon H \to G$ Abbildungen mit $f(D) \subset H$ (d.h. $g \circ f\colon D \to G$, $x \mapsto g(f(x))$ ist wohldefiniert). Falls f in $x_0 \in D$ stetig ist und $g \in f(x_0) \in H$ stetig ist, dann ist $g \circ f$ stetig in x_0.*

Beweis. Es sei $\varepsilon > 0$ beliebig. Wir wählen $\tilde{\delta} > 0$, sodass für $y \in H$ mit $|y - f(x_0)| < \tilde{\delta}$ die Abschätzung $|g(y) - g(f(x_0))| < \varepsilon$ gilt. Nun wählen wir $\delta > 0$, sodass für $x \in D$ mit $|x - x_0| < \delta$ die Abschätzung $|f(x) - f(x_0)| < \tilde{\delta}$ gilt. Setzen wir nun $y = f(x)$, dann gilt für $x \in D$ mit $|x - x_0| < \delta$ stets $|g(f(x)) - g(f(x_0))| < \varepsilon$. □

Korollar III.2.8. *Es seien $D, F, G, H \subset \mathbb{R}$. Sind $f\colon D \to F$ und $g\colon G \to H$ komponierbar und stetig, so ist ihre Komposition $g \circ f\colon D \to H$ ebenfalls stetig.* □

III.3. Folgenkriterium

Definition III.3.1. Es sei $D \subset \mathbb{R}$ und $f\colon D \to \mathbb{R}$ eine Abbildung.

(i) Wir nennen $a \in \mathbb{R}$ einen *Berührpunkt* von D, falls eine Folge $(x_n)_{n\in\mathbb{N}}$ in D existiert, sodass $\lim_{n\to\infty} x_n = a$ gilt.

(ii) Es sei $a \in \mathbb{R}$ ein Berührpunkt von D und $c \in \mathbb{R}$. Wir definieren

$$c = \lim_{x\to a} f(x) \;:\Longleftrightarrow \text{Für jede Folge } (x_n)_{n\in\mathbb{N}} \text{ in } D \text{ mit } \lim_{n\to\infty} x_n = a \text{ gilt } \lim_{n\to\infty} f(x_n) = c.$$

(iii) Es sei $a \in \mathbb{R}$ ein Berührpunkt von $D \cap (a,\infty)$ und $c \in \mathbb{R}$. Wir definieren

$$c = \lim_{x\searrow a} f(x) \;:\Longleftrightarrow \text{Für jede Folge } (x_n)_{n\in\mathbb{N}} \text{ in } D\cap(a,\infty) \text{ mit } \lim_{n\to\infty} x_n = a \text{ gilt } \lim_{n\to\infty} f(x_n) = c.$$

(iv) Es sei D nicht von oben beschränkt und $c \in \mathbb{R}$. Wir definieren

$$c = \lim_{x\to\infty} f(x) \;:\Longleftrightarrow \text{Für jede Folge } (x_n)_{n\in\mathbb{N}} \text{ in } D \text{ mit } \lim_{n\to\infty} x_n = \infty \text{ gilt } \lim_{n\to\infty} f(x_n) = c.$$

(v) Wir definieren

$$\lim_{x\nearrow a} f(x) = c,\ \lim_{x\to-\infty} f(x) = c,\ \lim_{x\searrow a} f(x) = \pm\infty,\ \lim_{x\to\pm\infty} f(x) = \pm\infty, \text{ etc.}$$

analog und sprechen in allen Fällen vom *Grenzwert der Funktion f* im Punkt a oder in $\pm\infty$, bzw. vom linksseitigen oder rechtsseitigen Grenzwert, vgl. III.3.3(iv).

Bemerkungen III.3.2.

(a) Ist $a \in \mathbb{R}$ kein Berührpunkt, so ist die rechte Seite in III.3.1(ii) wahr für jedes $c \in \mathbb{R}$. In einem solchen Fall können wir '$\lim_{x\to a} f(x)$' nicht in sinnvoller Weise erklären. Ist a Berührpunkt, aber die rechte Seite in III.3.1(ii) gilt nicht, dann sagen wir *der Grenzwert von f in a existiert nicht.* Analoges gilt für Grenzwerte in $\pm\infty$.

(b) Endliche Grenzwerte von Funktionen in Punkten können genauso für $f\colon D \subset \mathbb{C} \to \mathbb{C}$ definiert werden. Bei rechts- oder linksseitigen Grenzwerten, bei $\pm\infty$ als Grenzwert und bei Grenzwerten im Unendlichen geht das im Allgemeinen nicht, aber in Spezialfällen, z.B. kann $\lim_{x\searrow a} f(z)$ für $f\colon D \subset \mathbb{R} \subset \mathbb{C} \to \mathbb{C}$ definiert werden, wenn a ein Berührpunkt von $D \cap (a,\infty)$ ist. Bei komplexen Funktionen $f\colon \mathbb{C} \to \mathbb{C}$ betrachtet man oft auch $\lim_{|z|\to\infty} f(z)$.

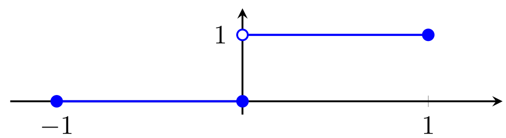
(c) Ist $a \in D$, so ist a automatisch Berührpunkt: Die konstante Folge $x_n \equiv a$ konvergiert gegen a und gehört zu D. In diesen Fall muss $\lim_{x\to a} f(x) = f(a)$ sein, falls nur der Grenzwert von f im Punkt a existiert. Es können auch Punkte $a \notin D$ Berührpunkte sein (z.B. $a = 1$, $D = (0,1)$, $x_n = 1 - \frac{1}{n}$). Wenn in diesem Fall für jede Folge $(x_n)_{n\in\mathbb{N}}$ in D mit $x_n \to a$ die Folge $(f(x_n))_{n\in\mathbb{N}}$ konvergiert, dann ist deren Grenzwert unabhängig von $(x_n)_{n\in\mathbb{N}}$: Angenommen, es gibt $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ mit $x_n, y_n \to a$ aber $f(x_n) \to c \neq b \leftarrow f(y_n)$. Definiere die 'Zick-Zack-Folge' $(z_n)_{n\in\mathbb{N}} := (x_1, y_1, x_2, y_2, \dots)$. Dann ist $(f(z_n))_{n\in\mathbb{N}}$ divergent, weil es zwei konvergente Teilfolgen mit verschiedenen Grenzwerten gibt.

(d) Für den einseitigen Grenzwert betrachten wir mit Absicht nur Folgen, die den Wert a nicht annehmen.

Beispiele III.3.3.

(1) $f\colon [0,1) \to \mathbb{R}$, $f(x) = x^2$. Dann gilt $\lim_{x\to 1} f(x) = 1$: Gilt $[0,1) \ni x_n \to 1$, so folgt $f(x_n) = x_n^2 \to 1$.
(2) $g\colon (0,1] \to \mathbb{R}$, $f(x) = \frac{1}{x}$. Dann gilt $\lim_{x\to 0} g(x) = \infty$: Gilt $(0,1] \ni x_n \to 0$, so folgt $g(x_n) = \frac{1}{x_n} \to \infty$.
(3) $h\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $h(x) = \frac{1}{x}$. Dann existiert $\lim_{x\to 0} h(x)$ nicht, denn z.B. gilt $\lim_{n\to\infty} h(1/n) = \lim_{n\to\infty} \frac{1}{1/n} = \infty$, aber $\lim_{n\to\infty} h(-1/n) = -\infty$.

(4) $i\colon [-1,1] \to \mathbb{R}$, $i(x) = \begin{cases} 0, & x \leqslant 0, \\ 1, & x > 0. \end{cases}$



Dann gilt $\lim_{x \nearrow 0} i(x) = 0$, denn für $(x_n)_{n\in\mathbb{N}}$ in $[-1,0)$ mit $x_n \to 0$ folgt $0 \equiv i(x_n) \to 0$. Analog gilt $\lim_{x \searrow 0} i(x) = 1$. Beachten Sie hier, dass wir bei einseitigen Limiten nur Folgen betrachten, die den Punkt a nicht annehmen. Der Grenzwert $\lim_{x\to 0} i(x)$ existiert nicht: Hier kann man entweder wie in (3) argumentieren und Folgen $(1/n)_{n\in\mathbb{N}}$ sowie $(-1/n)_{n\in\mathbb{N}}$ einsetzen, sodass die Bildfolgen gegen verschieden Werte konvergieren. Da $i(0) = 0$ ist, könnte man hier auch die Folge konstant Null statt $(-1/n)_{n\in\mathbb{N}}$ verwenden. Schließlich könnte man auch z.B. $(x_n)_{n\in\mathbb{N}} = (1, -1/2, 1/3, -1/4, \dots)$ betrachten und argumentieren, dass dies auf die divergente Bildfolge $(1,0,1,0,\dots)$ führt.

(5) Es gilt
$$\lim_{x\to 1} \frac{x^2+x-2}{x-1} = \lim_{x\to 1} \frac{(x-1)(x+2)}{x-1} = \lim_{x\to 1}(x+2) = 3.$$
Beachten Sie, dass hier keine Definitionsbereiche angegeben sind, aber dass es natürlich ist, zunächst $f\colon \mathbb{R}\backslash\{1\} \to \mathbb{R}$, $f(x) = \frac{x^2+x-2}{x-1}$ zu betrachten, sodass 1 Berührpunkt des Definitionsbereichs ist. Auf dem Definitionsbereich gilt dann aber $f(x) = x-2$ wie man durch Faktorisieren und Kürzen sieht. Schließlich sieht man die letzte Gleichung sofort; formal muss man aber hier auch wieder $(x_n)_{n\in\mathbb{N}}$ mit $x_n \to 1$ und $x_n \neq 1$ einsetzen und dann sehen, dass $x_n + 2 \to 3$ für jede solche Folge gilt.

(6) Ist $j\colon [0,1] \to \mathbb{R}$ eine beliebige Funktion, so ist $\lim_{x\to 2} j(x)$ nicht definiert, da 2 kein Berührpunkt von $[0,1]$ ist.

Satz III.3.4. *(Folgenkriterium für Stetigkeit) Die Funktion $f\colon D \subset \mathbb{R} \to \mathbb{R}$ ist genau dann stetig in $x_0 \in D$, wenn für jede Folge $(x_n)_{n\in\mathbb{N}}$ in D mit $x_n \to x_0$ gilt $f(x_n) \to f(x_0)$, also $\lim_{x\to x_0} f(x) = f(x_0)$ gilt.*

Beweis. '$\Longrightarrow$' Es sei $(x_n)_{n\in\mathbb{N}}$ in D mit $x_n \to x_0$ gegeben. Es sei $\varepsilon > 0$. Wähle $\delta > 0$, sodass $|f(x) - f(x_0)| < \varepsilon$ gilt für $|x - x_0| < \delta$. Das ist möglich, weil f in x_0 stetig ist. Wir wählen nun $n_0 \in \mathbb{N}$ derart, dass $|x_n - x_0| < \delta$ gilt für $n \geqslant n_0$. Das ist möglich, weil $x_n \to x_0$ gilt. Dann folgt, dass die Abschätzung $|f(x_n) - f(x_0)| < \varepsilon$ für jedes $n \geqslant n_0$ gilt. Damit haben wir aber gerade $f(x_n) \to f(x_0)$ gezeigt.

'$\Longleftarrow$' Angenommen, f ist nicht stetig in x_0. Dann gilt
$$\exists\, \varepsilon > 0\, \forall\, \delta > 0\, \exists\, x \in [0,\infty)\colon |x - x_0| < \delta \;\wedge\; |f(x) - f(x_0)| \geqslant \varepsilon.$$
Wir wählen ε wie in der Bedingung. Für $n \in \mathbb{N}$ setzen wir $\delta := \frac{1}{n}$ in die Bedingung ein und erhalten $x_n \in D$ mit $|x_n - x_0| < \frac{1}{n}$ und $|f(x_n) - f(x_0)| \geqslant \varepsilon$. Letztere Abschätzungen für jedes $n \in \mathbb{N}$ implizieren, dass die Folge $(x_n)_{n\in\mathbb{N}}$ gegen x_0 konvergiert und dass die Folge $(f(x_n))_{n\in\mathbb{N}}$ nicht gegen $f(x_0)$ konvergieren kann. Widerspruch. □

Bemerkung III.3.5. Ist $D \subset \mathbb{R}$, a ein Berührpunkt von D und $f\colon D \to \mathbb{R}$, dann zeigt man wie in III.3.4, dass gilt
$$\lim_{x\to a} f(x) = c \;:\Longleftrightarrow\; \forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x \in D\colon |x - a| < \delta \Rightarrow |f(x) - c| < \varepsilon.$$
Analog können $\lim_{x\nearrow a} f(x) = c$, $\lim_{x\to a} f(x) = \pm\infty$, $\lim_{x\to\pm\infty} f(x) = c$ usw. charakterisiert werden (Schreiben Sie's auf!).

Proposition III.3.6. Es seien $f, g\colon D \subset \mathbb{R} \to \mathbb{R}$ stetig in $x_0 \in D$ und es sei $\lambda \in \mathbb{R}$. Dann sind $f+g\colon D \to \mathbb{R}$, $\lambda f\colon D \to \mathbb{R}$ und $f \cdot g\colon D \to \mathbb{R}$ stetig in x_0. Wenn $g(x_0) \neq 0$ ist, dann ist auch $\frac{f}{g}\colon \{x \in D \mid g(x) \neq 0\} \to \mathbb{R}$ stetig in x_0.

Beweis. Mit III.3.4 genügt es zu zeigen, dass
$$\lim_{n\to\infty} (f+g)(x_n) = (f+g)(x_n),\; \lim_{n\to\infty} (f\cdot g)(x_n) = (f \cdot g)(x_0),\; \text{usw.}$$

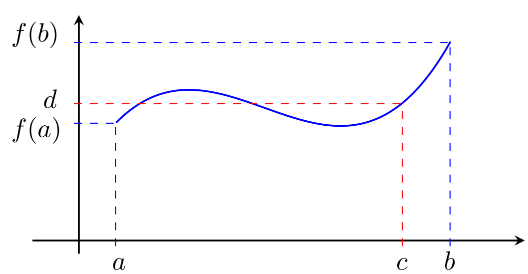
gelten für $(x_n)_{n\in\mathbb{N}}$ in D mit $x_n \to x_0$. Da nach Annahme für eine solche Folge $f(x_n) \to f(x_0)$ und $g(x_n) \to g(x_0)$ gelten, folgt obiges aus den Rechenregeln für Grenzwerte konvergenter Folgen aus der Mathematik 2. $\square$

Korollar III.3.7. *Polynomfunktionen sind auf jedem Definitionsbereich stetig, rationale Funktion* $\left(\frac{p(x)}{q(x)}\right.$ *mit Polynomen* p *und* $\left.q\right)$ *sind stetig auf jedem Definitionsbereich, auf dem sie wohldefiniert sind.*

Bemerkung III.3.8. III.2.4, III.2.6, III.2.7, III.2.8, III.3.4, III.3.5, III.3.6, III.3.7 gelten auch für komplexe Funktionen $f\colon \mathbb{C} \supset D \to \mathbb{C}$ mit der Einschränkung, dass man bei III.3.5 mit einseitigen Grenzwerten und mit $\pm\infty$ vorsichtig sein muss. Im nächsten Satz brauchen wir schon in der Formulierung Abschätzungen '$<$' bzw. '$>$', d.h. hier können wir nicht ohne weiteres $\mathbb{R}$ durch $\mathbb{C}$ ersetzen.

III.4. Zwischenwertsatz

Satz III.4.1. *(Zwischenwertsatz) Es sei* $-\infty < a < b < +\infty$, $f\colon [a,b] \to \mathbb{R}$ *stetig mit* $f(a) < f(b)$ *[bzw.* $f(a) > f(b)$*]. Es sei* $f(a) < d < f(b)$ *[bzw.* $f(a) > d > f(b)$*]. Dann existiert ein* $c \in [a,b]$ *mit* $f(c) = d$.



Beweis.

- Wir betrachten zunächst den Spezialfall $f(a) < 0 = d < f(b)$ und benutzen ein sogenanntes Intervallhalbierungsverfahren: Wir definieren induktiv eine Folge von Intervallen $[a_n, b_n] \subset [a,b]$ für $n \in \mathbb{N}_0$ mit

  (1) $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$,

  (2) $f(a_n) \leqslant 0$ und $f(b_n) \geqslant 0$,

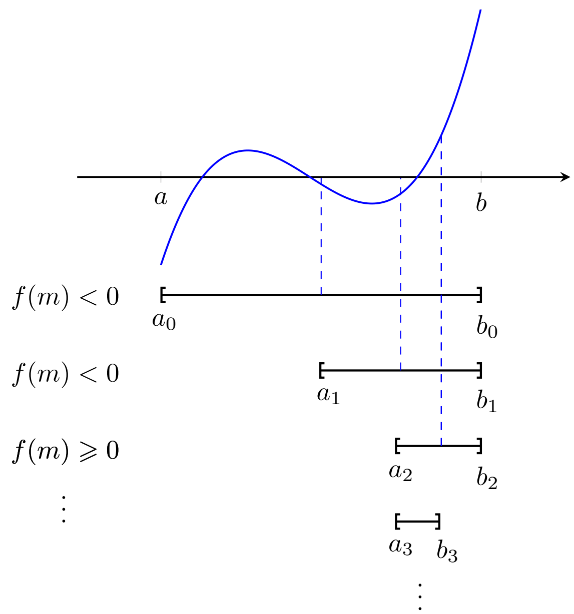
  (3) $b_n - a_n = 2^{-n}(b-a)$,

  für jedes $n \in \mathbb{N}_0$. Wir starten mit $[a_0, b_0] := [a,b]$. Ist $[a_n, b_n]$ bereits definiert, so setzen wir zunächst $m := \frac{a_n+b_n}{2}$ und unterscheiden dann zwei Fälle:

  Fall 1: Wenn $f(m) \geqslant 0$ ist, dann definieren wir $[a_{n+1}, b_{n+1}] := [a_n, m]$.

  Fall 2: Wenn $f(m) < 0$ ist, dann definieren wir $[a_{n+1}, b_{n+1}] := [m, b_n]$.

  Dann gelten (1) und (2) nach Konstruktion und (3) gilt, da wir $b_{n+1} - a_{n+1} = m - a_n = \frac{a_n+b_n}{2} - a_n = \frac{1}{2}(b_n - a_n)$ im Fall 1 haben und etwas Analoges im Fall 2 herauskommt.



Nach Konstruktion ist $(a_n)_{n\in\mathbb{N}}$ wachsend und beschränkt und $(b_n)_{n\in\mathbb{N}}$ ist fallend und beschränkt. Damit konvergieren beide Folgen und wegen (3) sind ihre Grenzwerte gleich. Wir definieren

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n =: c \in [a,b].$$

Da f stetig ist, gilt $f(a_n) \to f(c) \leqslant 0$ aber auch $f(b_n) \to f(c) \geqslant 0$. Also muss $f(c) = 0$ gelten, was zu zeigen war.

- Es sei jetzt $f(a) < d < f(b)$. Wir definieren $g\colon [a,b] \to \mathbb{R}$, $g(x) := f(x) - d$. Dann ist g stetig und $g(a) < 0 < g(b)$. Nach dem ersten Teil existiert also $c \in [a,b]$ mit $g(c) = 0$, d.h. $f(c) = g(c) + d = d$.
- Es sei schließlich $f(a) > d > f(b)$. Dann setzen wir $g\colon [a,b] \to \mathbb{R}$, $g(x) := -f(x)$. Es gilt dann $g(a) < -d < g(b)$ und es existiert $c \in [a,b]$ mit $g(c) = -d$, also $f(c) = d$.

□

Korollar III.4.2. *Es sei $I \subset \mathbb{R}$ ein beliebiges Intervall*[1] *und $f\colon I \to \mathbb{R}$ sei stetig. Dann ist $f(I)$ wieder ein Intervall.*

Beweis. Setze $A := \inf f(I) \in \mathbb{R} \cup \{-\infty\}$ und $B := \sup f(I) \in \mathbb{R} \cup \{\infty\}$. Es genügt zu zeigen, dass $(A,B) \subset f(I)$ gilt denn dann folgt $f(I) \in \{(A,B), [A,B], (A,B], [A,B)\}$. Es sei also $y \in (A,B)$. Dann existieren $a, b \in I$ mit $f(a) < y < f(b)$. Nach dem Zwischenwertsatz existiert also ein $x \in [a,b] \subset I$ mit $f(x) = y$, also $y \in f(I)$. □

III.5. Umkehrsatz

Satz III.5.1. *(Umkehrsatz) Es sei $f\colon I \subset \mathbb{R} \to \mathbb{R}$ stetig auf einem beliebigen Intervall und strikt wachsend (d.h. für $x < x'$ in I gilt $f(x) < f(x')$). Dann ist $I' := f(I) \subset \mathbb{R}$ nach III.4.2 ein Intervall. Ferner ist $f\colon I \to I'$ bijektiv und $f^{-1}\colon I' \to I$ ist strikt wachsend und stetig.*

Beweis. Wir zeigen, in dieser Reihenfolge, dass f bijektiv ist, dass f^{-1} strikt wachsend, und dass f^{-1} stetig ist.

- Da wir $I' = \mathrm{Bild}(f)$ definiert haben, ist klar, dass $f\colon I \to I'$ surjektiv ist. Ferner ist f nach Voraussetzung strikt wachsend und damit insbesondere injektiv. Folglich ist f bijektiv.

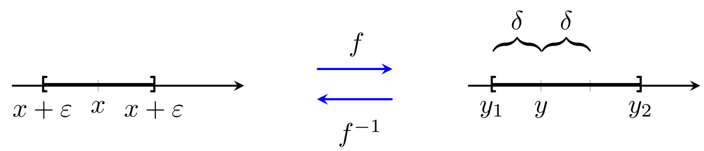
---

[1] Das Intervall I kann offen, abgeschlossen, halboffen, beschränkt, oder unbeschränkt sein. Wir können hier sogar $I = [a,a] = \{a\}$ oder $I = (a,a) = \emptyset$ erlauben, da in diesen pathologischen Fällen die Aussage der Folgerung offenbar richtig bleibt. Wir setzen aber dennoch im Folgenden stets voraus, dass Intervalle 'mindestens zwei Punkte enthalten' (und damit dann natürlich unendlich viele).

- Angenommen, f^{-1} ist nicht strikt wachsend. Dann gibt es $y, y' \in I'$ mit $y < y'$ aber $f^{-1}(y) \geqslant f^{-1}(y')$ wobei aber '=' nicht gelten kann, weil f^{-1} bijektiv ist. Mit $x := f^{-1}(y)$, $x' := f^{-1}(y')$ erhalten wir zwei Punkte $x, x' \in I$ mit $x > x'$ und $f(x) = y < y' = f(x')$. Dann wäre f aber nicht strikt wachsend im Widerspruch zur Voraussetzung.
- Wir fixieren $y' \in I'$ und setzen $x := f^{-1}(y)$, d.h. $f(x) = y$.

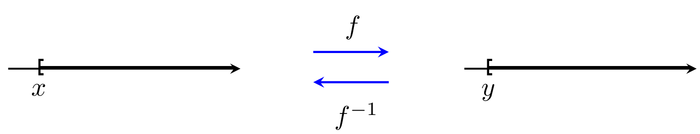
  Fall 1: Wir nehmen zunächst an, dass y kein Randpunkt von I' ist, d.h. wenn $I' = [a, b]$ mit $-\infty < a < b < \infty$ ist, dann ist $y \in (a, b)$ und entsprechend bei Intervallen der Form $(a, b]$ mit $-\infty \leqslant a < b < \infty$ bzw. $[a, b)$ mit $-\infty < a < b \leqslant \infty$.

  Wenn aber $y \in I'$ kein Randpunkt ist, dann kann $x \in I$ auch kein Randpunkt sein, weil wir bereits wissen, dass f streng wachsend ist. Wir finden also $\varepsilon_0 > 0$, sodass $[x - \varepsilon, x + \varepsilon] \subset I$ für beliebiges $0 < \varepsilon < \varepsilon_0$ gilt und können definieren $y_1 := f(x - \varepsilon)$, $y_2 := f(x + \varepsilon)$. Dann ist $f\colon [x - \varepsilon, x + \varepsilon] \to [y_1, y_2]$ bijektiv. Es sei $\delta := \min(y - y_1, y_2 - y)$, d.h. der Abstand von y zu dem Randpunkt von $[y_1, y_2]$ der näher an y liegt:



  Dann gilt $f^{-1}((y - \delta, y + \delta)) \subset (x - \varepsilon, x + \varepsilon)$ oder, anders aufgeschrieben,
$$\forall\, \tilde{y} \in I'\colon |\tilde{y} - y| < \delta \;\Rightarrow\; |f^{-1}(\tilde{y}) - x| < \varepsilon.$$
  Da $x = f(y)$ nach Definition gilt, haben wir gerade gezeigt, dass f^{-1} in y stetig ist.

  Fall 2: Wenn y ein Randpunkt ist, dann ersetzen wir in obigem Beweis das Intervall $[x-\varepsilon, x+\varepsilon]$ durch $[x, x + \varepsilon]$ im Fall, dass x gleich dem linken Randpunkt ist bzw. durch $[x - \varepsilon, x]$ im Fall, dass x gleich dem rechten Randpunkt ist. Da f streng wachsend ist, gilt dann das Entsprechende auch für y.



  (Das Bild zeigt nur den Fall, dass x gleich dem linken Randpunkt ist.)

$\square$

Bemerkung III.5.2. Ein analoger Satz gilt für 'strikt fallend'.

Beispiele III.5.3.

(1) $\exp\colon \mathbb{R} \to (0, \infty)$ ist stetig als Einschränkung der komplexen Exponentialfunktion $\exp\colon \mathbb{C} \to \mathbb{C}$, bijektiv und strikt wachsend. Letzteres wissen wir aus der Mathematik 2. Nach III.4.2 ist daher $\log\colon (0, \infty) \to \mathbb{R}$ stetig und strikt wachsend.

(2) $f\colon [0, \infty) \to \mathbb{R}$, $x \mapsto x^\alpha$ für $\alpha > 0$ ist stetig. Hierfür sei zunächst $x > 0$. Dann gilt $x^\alpha = e^{\log x^\alpha} = e^{\alpha \log x}$. Nach (1) ist $x \mapsto \log x$ stetig. Dann ist aber nach III.3.6 auch $x \mapsto \alpha \log x$ stetig und schließlich $x \mapsto e^{\alpha \log x} = x^\alpha$ nach III.2.8. Für $x = 0$ klappt die obige Argumentation nicht, hier wenden wir daher das Folgenkriterium an und sehen $f(0) = 0^\alpha = 0 = \lim_{x \searrow 0} e^{\alpha \log x}$ (denn nach der fundamentalen Abschätzung gilt für $y < 1$ stets $0 < e^y < \frac{1}{1-y} \to 0$ für $y \to -\infty$ und für $x > 0$ gilt $\log x \leqslant x - 1 \to -\infty$ für $x \searrow 0$).

(3) Mit (1) folgt auch, dass $f\colon \mathbb{R} \to (0, \infty)$, $x \mapsto \exp(x \log a) = a^x$ stetig ist für $a > 0$ und ebenso die Umkehrfunktion $\log_a\colon (0, \infty) \to \mathbb{R}$.

(4) Aus Blatt 3, Aufgabe 1 wissen wir, dass $\sin, \cos\colon \mathbb{R} \to \mathbb{R}$ beide stetig sind. Wir zeigen jetzt, dass $\sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ bijektiv ist und strikt wächst. Daraus folgt, dass $\cos\colon [0, \pi] \to [-1, 1]$ bijektiv ist und strikt fällt, da für jedes $x \in \mathbb{R}$ die Beziehung $\cos(x) = \sin\left(\frac{\pi}{2} - x\right)$ gilt.

(a) Zunächst halten wir fest, dass aus $\sin(-\frac{\pi}{2}) = -1$, $\sin(\frac{\pi}{2}) = 1$ und dem Zwischenwertsatz folgt, dass $\sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ surjektiv ist.

(b) Da ferner aus der Mathematik 2 bekannt ist, dass $\sin|_{[-\frac{\pi}{2}, \frac{\pi}{2}]}(x) = 0$ genau dann gilt, wenn $x = 0$ ist, folgt durch nochmalige Anwendung des Zwischenwertsatzes, dass

$$\sin|_{(0,\frac{\pi}{2}]} > 0 \ \text{ und } \ \sin|_{[-\frac{\pi}{2},0)} < 0$$

gelten. Analog wissen wir, dass $\cos\frac{\pi}{2} = 0$ ist und sehen, dass

$$\cos|_{[0,\frac{\pi}{2})} > 0$$

gilt. Es folgt, dass für $x \neq y$ und $-\frac{\pi}{2} \leqslant x \leqslant 0 \leqslant y \leqslant \frac{\pi}{2}$ stets $\sin x < \sin y$ gilt. In der Tat ist für $x = 0 < y$ dann $\sin x = 0 < \sin y$, für $x < 0 = y$ ist $\sin x < 0 = \sin y$ und für $x < 0 < y$ haben wir $\sin x < 0 < \sin y$.

(c) Es seien nun $0 < x < y \leqslant \frac{\pi}{2}$. Wir setzen $u := \frac{x+y}{2} \in (0, \frac{\pi}{2})$, $v := \frac{x-y}{2} \in (-\frac{\pi}{2}, 0)$, d.h. es gilt $x = u + v$ und $y = u - v$ und nach obigem $\cos u > 0$ und $\sin v < 0$. Es folgt

$$\begin{aligned}\sin y - \sin x &= \sin(u - v) - \sin(u + v)\\ &= \sin u \cos v - \cos u \sin v - (\sin u \cos v + \cos u \sin v)\\ &= -2\cos u \sin v > 0\end{aligned}$$

und damit $\sin y > \sin x$.

(d) Mit der aus der Mathematik 2 bekannten Symmetrieeigenschaft $\sin(-x) = -\sin(x)$ folgt für $-\frac{\pi}{2} \leqslant x < y < 0$, also $0 < -y < -x \leqslant -\frac{\pi}{2}$, dann mit (c)

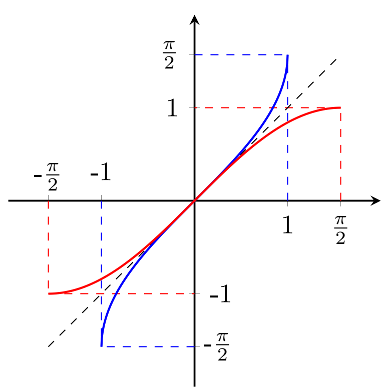
$$\sin y = -\sin(-y) > -\sin(-x) = \sin x$$

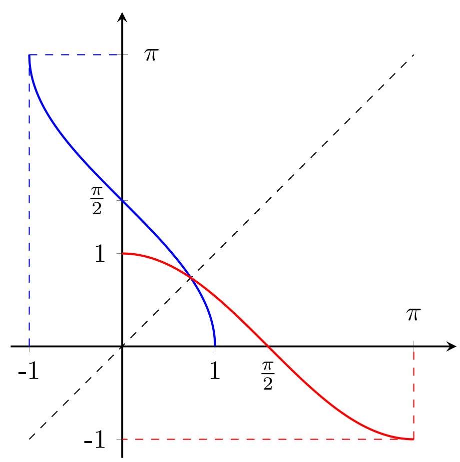
und wir erhalten, dass der Sinus auch auf $(-\frac{\pi}{2}, 0]$ strikt wächst.

Wir bezeichnen mit

$$\arcsin := \sin^{-1}\colon [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}] \ \text{und} \ \arccos := \cos^{-1}\colon [-1, 1] \to [0, \frac{\pi}{2}]$$

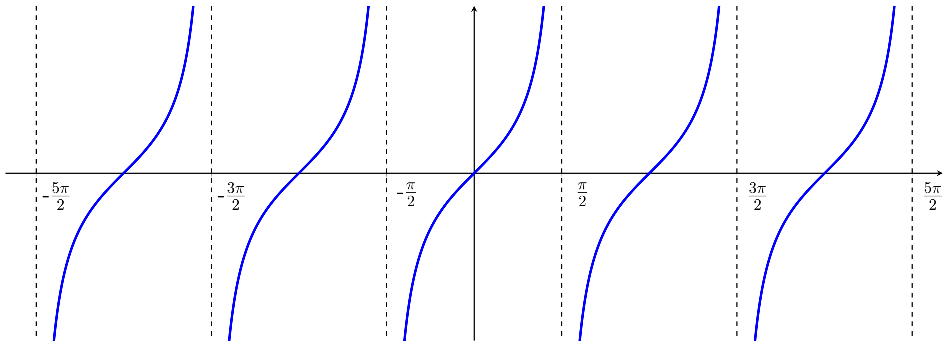
die nach obigem wohldefinierten und nach III.5.1 stetigen Umkehrfunktionen *Arkussinus* bzw. *Arkuskosinus*. Die arccos-Funktion hatten wir schon früher kennengelernt.





Wir beachten, dass wir den Definitionsbereich von Sinus und Kosinus auch anders hätten wählen können, z.B. ist $\sin\colon [\frac{\pi}{2}, \frac{3}{2}\pi] \to [-1, 1]$ ebenfalls bijektiv.

(5) Nach III.5.3 (4) und III.3.6 ist der Tangens $\tan\colon \mathbb{R}\backslash\{\frac{\pi}{2} + k\pi \mid k \in \mathbb{Z}\} \to \mathbb{R}$, $\tan x = \frac{\sin x}{\cos x}$ stetig.



Wir zeigen jetzt, dass $\tan\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ bijektiv und strikt wachsend ist.

(a) Es seien zunächst $0 \leqslant x < y < \frac{\pi}{2}$. Dann gilt

$$\tan x = \frac{\sin x}{\cos x} < \frac{\sin y}{\cos x} < \frac{\sin y}{\cos y} = \tan y,$$

wobei wir für die erste Ungleichung benutzt haben, dass sin wächst, und für die zweite, dass cos fällt, siehe (3).

(b) Als nächstes seien $-\frac{\pi}{2} < x < y \leqslant 0$. Dann gilt

$$\tan x = \frac{\sin x}{\cos x} = \frac{-\sin(-x)}{\cos(-x)} = -\tan(-x) < -\tan(-y) = \tan y,$$

wobei wir für die Abschätzung den vorherigen Teil benutzt haben und davor und danach Symmetrieeigenschaften von Sinus und Kosinus, die wir in der Mathematik 2 gezeigt haben.

(c) Ist $-\frac{\pi}{2} < x < 0 < y < \frac{\pi}{2}$, so folgt $\tan x < 0 < \tan y$ durch Betrachtung der Vorzeichen von sin und cos die wir oben notiert haben.

(d) Als letztes müssen wir noch zeigen, dass $\tan\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ surjektiv ist. Dazu berechnen wir zuerst

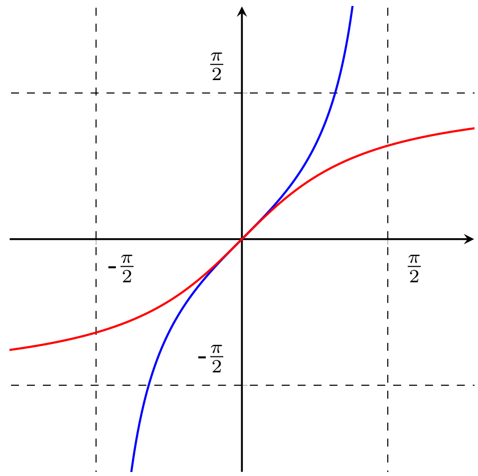
$$\lim_{x \nearrow \frac{\pi}{2}} \tan x = \lim_{x \nearrow \frac{\pi}{2}} \frac{\sin x}{\cos x} = +\infty \quad \text{und} \quad \lim_{x \searrow -\frac{\pi}{2}} \tan x = \lim_{x \searrow -\frac{\pi}{2}} \frac{\sin x}{\cos x} = -\infty,$$

weil $\sin x \to \pm 1$, $\cos x \to 0$ für $x \to \pm\frac{\pi}{2}$ und $\cos x > 0$ für $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ gelten. Da der Tangens stetig ist, folgt aus III.4.2, dass $\operatorname{Bild}(\tan|_{(-\frac{\pi}{2}, \frac{\pi}{2})}) \subset \mathbb{R}$ ein Intervall sein muss, was nach dem Vorherigen aber nur geht, wenn das Intervall $(-\infty, +\infty) = \mathbb{R}$ ist.

Die nach III.5.1 stetige Inverse von $\tan\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ bezeichen wir mit

$$\arctan := \tan^{-1}\colon \mathbb{R} \to (\frac{\pi}{2}, \frac{\pi}{2})$$

und sprechen vom *Arkustangens*.



Wir beachten wieder, dass man auch hier einen anderen Zweig des Tangens hätte betrachten können.

III.6. Extremalsatz und gleichmäßige Stetigkeit

Definition III.6.1. Es sei $D \subset \mathbb{R}$ beliebig. Eine Funktion $f\colon D \to \mathbb{R}$ heißt *beschränkt*, falls $f(D) \subset \mathbb{R}$ eine beschränkte Menge ist. Explizit heißt dies

$$\exists\, m,\, M \in \mathbb{R}\, \forall\, x \in D\colon m \leqslant f(x) \leqslant M.$$

Entsprechend mit nur einer Abschätzung definiert man, was es heißt, dass f *von oben* bzw. *von unten beschränkt* ist.

Satz III.6.2. *(Extremalsatz) Es sei $-\infty < a < b < \infty$. Jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$ ist beschränkt und nimmt Minimum und Maximum an, d.h. es gibt $p, q \in [a,b]$ mit*

$$f(p) = \sup\{f(x) \mid x \in [a,b]\} \text{ und } f(q) = \inf\{f(x) \mid x \in [a,b]\}.$$

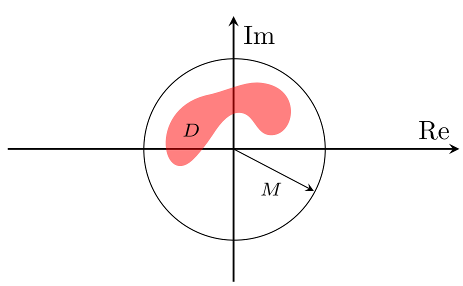
Insbesondere ist das obige Supremum also ein Maximum und das Infimum ein Minimum.

Beweis. Wir setzen $A := \sup\{f(x)|x \in [a,b]\} \in \mathbb{R}\cup\{\infty\}$ und wählen $(x_n)_{n\in\mathbb{N}} \subset [a,b]$ mit $\lim_{n\to\infty} f(x_n) = A$. Da $(x_n)_{n\in\mathbb{N}}$ beschränkt ist, besagt der Satz von Bolzano-Weierstraß, dass es eine konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$ gibt und wir können $p := \lim_{k\to\infty} x_{n_k} \in [a,b]$ definieren. Da f stetig ist, gilt nach dem Folgenkriterium $A = \lim_{k\to\infty} f(x_{n_k}) = f(p)$. D.h. aber $A \in \mathbb{R}$, also ist f von oben beschränkt und nimmt sein Supremum an, welches daher ein Maximum ist. Der andere Teil geht genauso. $\square$

Bemerkung III.6.3. Intervalle der Form $[a,b]$ mit $-\infty < a < b < \infty$ heißen *kompakte Intervalle*. III.6.2 wird falsch wenn der Definitionsbereich nicht kompakt ist oder wenn f nicht stetig ist.

Definition III.6.4. Wir nennen $D \subset \mathbb{C}$ *beschränkt*, wenn gilt

$$\exists\, M \geqslant 0\, \forall\, z \in D\colon |z| \leqslant M.$$



Eine Funktion $f\colon D \to \mathbb{C}$ mit $D \subset \mathbb{C}$ beliebig (also vielleicht auch unbeschränkt) heißt *beschränkt*, wenn $f(D) \subset \mathbb{C}$ beschränkt ist. Explizit heißt dies

$$\exists\, M \geqslant 0\, \forall\, x \in D\colon |f(x)| \leqslant M.$$

Für Funktionen $f\colon D \subset \mathbb{R} \to \mathbb{R}$ ist obiges äquivalent zur Beschränktheit, wie wir sie in III.6.1 definiert haben.

Bemerkung III.6.5. III.6.2 gilt komplex zumindest in folgendem Sinne: Ist $M \geqslant 0$ und $f\colon \{z \in \mathbb{C} \,\big|\, |z| \leqslant M\} \to \mathbb{R}$ stetig, so ist f beschränkt und nimmt Minimum und Maximum an. Hat man eine stetige Funktion $f\colon \{z \in \mathbb{C} \,\big|\, |z| \leqslant M\} \to \mathbb{C}$, so zeigt die Betrachtung von $|f|\colon \{z \in \mathbb{C} \,\big|\, |z| \leqslant M\} \to \mathbb{R}$, dass f beschränkt ist.

Definition III.6.6. Eine Funktion $f\colon D \subset \mathbb{R} \to \mathbb{R}$ ist *gleichmäßig stetig*, falls gilt

$$\forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x, y \in D\colon |x - y| < \delta \;\Rightarrow\; |f(x) - f(y)| < \varepsilon.$$

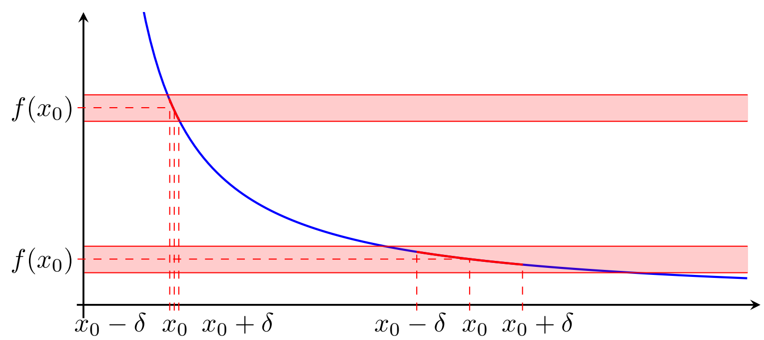
Bemerkungen III.6.7.

(i) Jede gleichmäßig stetige Funktion ist stetig, denn die Implikation

$$\forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x, y\colon\; \dots \;\Longrightarrow\; \forall\, x\, \forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, y\colon\; \dots$$

gilt unabhängig von der Aussage, die nach den Quantoren folgt.

(ii) Die Umkehrung von (i) ist falsch: Betrachten Sie dazu $f\colon (0,1) \to \mathbb{R}$, $f(x) = \frac{1}{x}$. Nach III.2.3(iii) und III.2.6 ist f stetig. Wir hatten aber bereits in III.2.3(iii) beobachtet, dass wir bei einem fixierten $\varepsilon > 0$ den Wert von $\delta > 0$ immer weiter verkleinern müssen, wenn wir x_0 näher an Null wählen.



Für einen formalen Beweis nehmen wir an, dass f gleichmäßig stetig ist. Dann existiert für $\varepsilon = 1$ ein $\delta > 0$ mit

$$\forall\, x, y \in (0,1)\colon |x - y| < \delta \;\Rightarrow\; \big|f(x) - f(y)\big| < 1.$$

Wir setzen nun $x = \frac{1}{n}$ und $y = \frac{1}{2n}$. Dann gilt $|x - y| = |\frac{2-1}{2n}| = \frac{1}{2n} \to 0$ und $|\frac{1}{x} - \frac{1}{y}| = |n - 2n| = n \geqslant 1$. D.h. wir finden $n \in \mathbb{N}$ mit $|x - y| < \delta$ aber $|f(x) - f(y)| \geqslant 1$ im Widerspruch zur Annahme.

(iii) Auf kompakten Intervallen sind die zwei Begriffe äquivalent wie der nächste Satz zeigt.

Satz III.6.8. *Es sei $-\infty < a < b < \infty$ und $f\colon [a,b] \to \mathbb{R}$ sei stetig. Dann ist f gleichmäßig stetig.*

Beweis. Angenommen, f ist nicht gleichmäßig stetig. D.h. es gilt

$$\exists\, \varepsilon > 0\, \forall\, \delta > 0\, \exists\, x, y \in D\colon |x - y| < \delta \wedge |f(x) - f(y)| \geqslant \varepsilon.$$

Wir wählen ein $\varepsilon > 0$ wie oben. Für $n \in \mathbb{N}$ setzen wir $\delta = \frac{1}{n}$ und finden dazu x_n, $y_n \in D$ mit $|x_n - y_n| < \frac{1}{n}$ und $|f(x_n) - f(y_n)| \geqslant \varepsilon$. Nach dem Satz von Bolzano-Weierstraß existiert dann eine konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$ von $(x_n)_{n\in\mathbb{N}}$. Wir definieren $c := \lim_{k\to\infty} x_{n_k}$. Da $|x_{n_k} - y_{n_k}| < \frac{1}{n_k} \to 0$ für $k \to \infty$ gilt, folgt mit der Dreiecksungleichung

$$|y_{n_k} - c| = |y_{n_k} - x_{n_k} + x_{n_k} - c| \leqslant |y_{n_k} - x_{n_k}| + |x_{n_k} - c| \to 0$$

und daher $\lim_{k\to\infty} y_{n_k} = c$. Da f stetig ist, folgt weiter

$$\lim_{k\to\infty} [f(x_{n_k}) - f(y_{n_k})] = f(c) - f(c) = 0$$

im Widerspruch zu $|f(x_{n_k}) - f(y_{n_k})| \geqslant \varepsilon > 0$ für alle $k \in \mathbb{N}$. $\square$

Bemerkung III.6.9. III.6.8 gilt auch im Komplexen, z.B. für $f\colon \{z \in \mathbb{C} \mid |z - z_0| \leqslant R\} \to \mathbb{C}$ mit $z_0 \in \mathbb{C}$ und $R > 0$ beliebig.

KAPITEL IV

Der Fundamentalsatz der Algebra

In diesem Kapitel behandeln wir einen Beweis des Fundamentalsatzes der Algebra, den wir schon in der Mathematik 2 benutzt haben. Dieser besagt, dass jedes nichtkonstante Polynom mit komplexen Koeffizienten mindestens eine komplexe Nullstelle besitzt.

IV.1. Beweis

Lemma IV.1.1. *Es sei $p\colon \mathbb{C} \to \mathbb{C}$, $p(z) = \sum_{k=0}^{n} c_k z^k$ eine komplexe Polynomfunktion vom Grad $n \geqslant 1$ und es sei $a \in \mathbb{C}$ mit $p(a) \neq 0$. Es sei weiter $R > 0$ und $D := \{z \in \mathbb{C} \,\big|\, |a - z| < R\}$. Dann existiert $b \in D$ mit $|p(b)| < |p(a)|$.*

Beweis. Wir bemerken zunächst, dass die Punkte in D von der Form $a + w$ mit $|w| < R$ sind. Nun behaupten wir, dass

$$(\star) \quad p(a + w) = p(a) + cw^m(1 + r(w))$$

gilt mit $c \in \mathbb{C}\backslash\{0\}$, $1 \leqslant m \leqslant n$ und einer Polynomfunktion $r\colon \mathbb{C} \to \mathbb{C}$ vom Grad $n - m$, die $r(0) = 0$ erfüllt.

In der Tat gilt nach dem binomischen Lehrsatz

$$p(a + w) = \sum_{k=0}^{n} c_k (a + w)^k = \sum_{k=0}^{n} c_k \left(\sum_{i=0}^{k} \tbinom{k}{i} a^{k-i} \right) w^i = \sum_{i=0}^{n} \left(\sum_{k=i}^{n} \tbinom{k}{i} c_k a^{k-i} \right) w^i,$$

wobei wir $\binom{k}{i} = 0$ für $i > k$ benutzt haben. Also ist $w \mapsto p(a + w)$ eine komplexe Polynomfunktion vom Grad n, und damit ist auch

$$\tilde{p}\colon \mathbb{C} \to \mathbb{C}, \quad \tilde{p}(w) := p(a + w) - p(a)$$

eine Polynomfunktion vom Grad n, die wegen $n \geqslant 1$ nicht konstant ist, und an der Stelle $w_0 = 0$ eine Nullstelle besitzt.

Ist $m \in \mathbb{N}_0$ die Ordnung der Nullstelle $w_0 = 0$, so gilt $1 \leqslant m \leqslant n$ und wir können nach dem Satz über Polynomdivision aus der Mathematik 1 für jedes $w \in \mathbb{C}$

$$\tilde{p}(w) = w^m q(w)$$

schreiben, mit einer Polynomfunktion $q\colon \mathbb{C} \to \mathbb{C}$ vom Grad $n - m$, sodass $q(0) \neq 0$. Wir betrachten nun die Polynomfunktion $r\colon \mathbb{C} \to \mathbb{C}$ gegeben durch $r(w) := \frac{q(w)-q(0)}{q(0)}$. Dann hat r Grad n und es gilt $r(0) = 0$, sowie $q(w) = q(0)(1 + r(w))$. Damit haben wir $(\star)$ für $c := q(0)$ gezeigt.

Als nächstes schätzen wir $|cw^m|$ und $|r(w)|$ nach oben ab für Werte w, die nahe bei Null liegen:

(1) Für $|w| < \rho_1 := \sqrt[m]{|p(a)/c|}$ (wohldefiniert und strikt positiv!) gilt:

$$|cw^m| = |c||w|^m < |c||p(a)/c| = |p(a)|.$$

(2) Da $r\colon D \to \mathbb{C}$ stetig ist und $r(0) = 0$ gilt, folgt für $\varepsilon = 1$, dass ein $\rho_2 > 0$ existiert, sodass für alle $w \in \mathbb{C}$ gilt

$$|w| < \rho_2 \;\Rightarrow\; |r(w)| < 1.$$

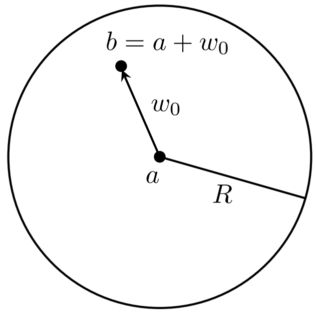
Kombinieren wir (1) und (2), so erhalten wir

$$(\circ) \quad \forall\, w \in \mathbb{C} \;\text{ mit } |w| < \rho := \min(\rho_1, \rho_2)\colon |cw^m| < |p(a)| \text{ und } |r(w)| < 1.$$

Es sei nun ζ eine m-te Wurzel von $-\frac{p(a)/c}{|p(a)/c|}$, d.h. es gilt

$$\zeta \in \mathbb{C},\ |\zeta| = 1 \ \text{ und } \ \zeta^m = -\frac{p(a)/c}{|p(a)/c|}.$$

Es sei weiter $0 < \varepsilon < \min(\rho, R)$. Wir setzen $w_0 := \varepsilon\zeta$ und behaupten, dass $b := a + w_0$ die im Lemma gewünschten Eigenschaften hat. In der Tat gilt zunächst $|w_0| = \varepsilon < R$ und daher ist $b \in D$:



Jetzt muss noch $|p(b)| < |p(a)|$ gezeigt werden. Mit ($\star$) folgt

$$|p(b)| = |p(a + w_0)| = |p(a) + cw_0^m(1 + r(w_0))|$$

und weiter mit $\delta := \frac{\varepsilon^m}{|p(a)/c|}$

$$(+) \quad cw_0^m = c\varepsilon^m\zeta^m = -\frac{\varepsilon^m}{|p(a)/c|}p(a) = -\delta p(a),$$

wobei wir erst die Definition von w_0, dann die Definition von ζ und schließlich die Definition von δ benutzt haben. Da $\varepsilon > 0$ ist, sehen wir, dass $\delta > 0$ gilt. Wegen $\varepsilon < \rho$ folgt mit ($\circ$), dass $\delta < 1$ gilt. Damit schätzen wir nun $|p(b)|$ weiter ab. Mit (+) erhalten wir

$$\begin{aligned}
|p(b)| &= |p(a) - \delta p(a)(1 + r(w_0))| \\
&= |(1-\delta)p(a) - \delta p(a) r(w_0)| \\
&\leqslant (1-\delta)|p(a)| + \delta|p(a)||r(w_0)| && \text{(Dreiecksungleichung)} \\
&< (1-\delta)|p(a)| + \delta|p(a)| && \text{(weil } |r(w_0)| < 1 \text{ nach } (\circ)) \\
&= |p(a)|.
\end{aligned}$$

□

Satz IV.1.2. *(Fundamentalsatz der Algebra) Jedes nichtkonstante Polynom mit komplexen Koeffizienten besitzt mindestens eine komplexe Nullstelle.*

Beweis. Es sei $p\colon \mathbb{C} \to \mathbb{C}$, $p(z) = \sum_{k=0}^n c_k z^k$ eine Polynomfunktion vom Grad $n \geqslant 1$. Wir betrachten

$$\lim_{|z|\to\infty} \frac{p(z)}{z^n} = \lim_{|z|\to\infty} \big[c_n + \frac{c_{n-1}}{z} + \frac{c_{n-2}}{z^2} + \cdots + \frac{c_0}{z^n}\big] = c_n,$$

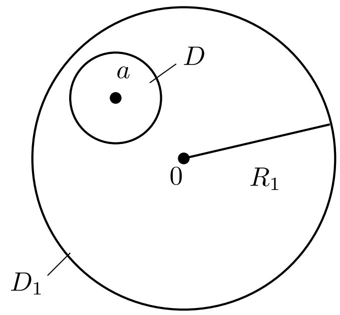
woraus folgt

$$\lim_{|z|\to\infty} |p(z)| = \lim_{|z|\to\infty} \Big|\frac{p(z)}{z^n}\Big| \cdot |z^n| = +\infty,$$

weil $c_n \neq 0$ nach Voraussetzung gilt. D.h. es existiert ein $R_1 > 0$, sodass $|p(z)| > |p(0)|$ für alle z auf dem Kreis $\{z \in \mathbb{C} \,\big|\, |z| = R_1\}$ gilt. Nun ist

$$|p|\colon \overline{D}_1 := \{z \in \mathbb{C} \,\big|\, |z| \leqslant R_1\} \to \mathbb{R}$$

stetig und nimmt gemäß III.6.5 sein Minimum in einem Punkt $a \in \overline{D}_1$ an. Wegen obigem kann a aber nicht auf dem Rand liegen, also $a \in D_1 = \{z \in \mathbb{C} \,\big|\, |z| < R_1\}$. Jetzt nehmen wir $R > 0$, sodass $D := \{z \in \mathbb{C} \,\big|\, |z - a| < R\} \subset D_1$ gilt:



Angenommen, $p(a) \neq 0$, dann gäbe es nach IV.1.1 ein $b \in D \subset \overline{D}_1$ mit $|p(b)| < |p(a)|$ und a wäre keine Minimalstelle von $|p|$ auf $\overline{D}_1$. Widerspruch. $\square$

Korollar IV.1.3. *Über $\mathbb{C}$ zerfällt jedes Polynom p vom Grad $n \geqslant 1$ in Linearfaktoren. Genauer: Es gibt $z_1, \ldots, z_n \in \mathbb{C}$ und $c \in \mathbb{C}$ mit*

$$p(z) = c \prod_{i=1}^{n} (z - z_i).$$

Beweis. Wir wenden IV.1.2 und den Satz über Polynomdivision aus der Mathematik 2 n Mal an. $\square$

KAPITEL V

Differenzierbarkeit

V.1. Definition

Definition V.1.1. Es sei $f\colon \mathbb{R} \supset D \to \mathbb{R}$ eine Funktion und es sei $a \in D$ ein Berührpunkt von $D\backslash\{a\}$. Dann ist f *differenzierbar in* a, falls der Grenzwert

$$f'(a) := \lim_{\substack{x\to a\\ x\in D\backslash\{a\}}} \frac{f(x)-f(a)}{x-a}$$

existiert. Der Wert $f'(a)$ heißt die *Ableitung von* f *an der Stelle* a. Falls jeder Punkt $a \in D$ Berührpunkt von $D\backslash\{a\}$ ist und f differenzierbar in jedem $a \in D$ ist, dann sagen wir, dass f (*überall*) *differenzierbar* ist. Wir bekommen in diesem Fall eine Abbildung

$$f'\colon D \to \mathbb{R},\ a \mapsto f'(a),$$

die wir als *Ableitung von* f bezeichnen.

Bemerkungen V.1.2.

(i) Dass $a \in D$ ein Berührpunkt von $D\backslash\{a\}$ ist, ist nötig damit der Grenzwert in V.1.1 wohldefiniert ist: Da wir $x = a$ nicht einsetzen können, brauchen wir Punkte in D, die nah bei a liegen, aber nicht gleich a sind. Oft lässt man '$x \in D\backslash\{a\}$' unter dem Limes weg, weil man am nachfolgenden Bruch sieht, dass $x = a$ nicht eingesetzt werden darf.

(ii) In V.1.1 bedeutet 'der Grenzwert existiert', dass dieser in $\mathbb{R}$ existiert; $\pm\infty$ sind nicht erlaubt.

(iii) Substituiert man $h := x - a$, so sieht man, dass

$$f'(a) := \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$$

gilt, falls f differenzierbar ist. Man kann auch die Definition mit obigem Ausdruck formulieren. Wenn man ganz genau sein will, muss man noch $a + h \in D\backslash\{a\}$ oder $h \neq 0$ unter den Limes schreiben, aber auch hier lassen wir dies meistens weg.

(iv) Andere Notationen für die Ableitung sind

$$f'(a) = \frac{\mathrm{d}f}{\mathrm{d}x}(a) = \frac{\mathrm{d}f(a)}{\mathrm{d}x} = \frac{\mathrm{d}f(x)}{\mathrm{d}x}\Big|_{x=a}$$

oder $f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x}(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x}$ bzw. kurz $f' = \frac{\mathrm{d}f}{\mathrm{d}x}$. Hierbei ist zu beachten, dass $\frac{\mathrm{d}f}{\mathrm{d}x}$ lediglich eine Abkürzung ist und kein Bruch reeller Zahlen. Wir werden allerdings später an manchen Stellen so tun, als wäre dies doch ein Bruch mit dem wir ganz normal rechnen könnten! Solche 'Rechnungen' sind dann aber nur als Heuristiken zu werten und haben keinen Beweiswert. Wir werden sehen, dass dies einerseits in der Praxis durchaus hilfreich sein kann und manchmal nur eine etwas unsaubere Schreibweise eines fundierten mathematischen Satzes ist. Andererseits werden wir auch sehen, dass es zu schweren Fehlern führen kann, wenn man allzu sorglos damit umgeht.

Es gilt z.B.

$$f'(0) = \frac{\mathrm{d}f}{\mathrm{d}x}(0) = \frac{\mathrm{d}f(0)}{\mathrm{d}x} = \frac{\mathrm{d}f(x)}{\mathrm{d}x}\Big|_{x=0}$$

aber nicht $f'(0) = \frac{\mathrm{d}f}{\mathrm{d}0}(0)$.

Beispiele V.1.3.

(i) Es sei $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = c$ für ein $c \in \mathbb{R}$. Dann gilt

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \frac{c - c}{x - a} = 0.$$

(ii) Für $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = cx$ mit $c \in \mathbb{R}$ gilt

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{c(x+h) - cx}{h} = \lim_{h\to 0} \frac{cx + ch - cx}{h} = \lim_{h\to 0} \frac{ch}{h} = c.$$

(iii) Es sei $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Dann gilt

$$f'(x) = \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} = \lim_{h\to 0} (2x + h) = 2x.$$

(iv) Für die Funktion $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $f(x) = \frac{1}{x}$ gilt

$$f'(x) = \lim_{h\to 0} \frac{1}{h}\Big(\frac{1}{x+h} - \frac{1}{x}\Big) = \lim_{h\to 0} \frac{x - (x+h)}{h(x+h)x} = \lim_{h\to 0} \frac{-1}{(x+h)x} = -\frac{1}{x^2}.$$

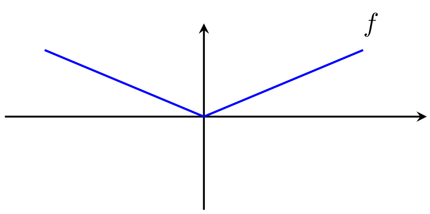
(v) Es sei $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ die Betragsfunktion. Dann gilt für $x \neq 0$

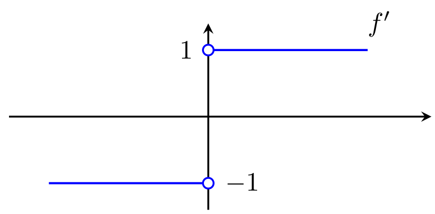
$$f'(x) = \lim_{h\to 0} \frac{|x+h| - |x|}{h} = \lim_{h\to 0} \left\{ \begin{matrix} \frac{x+h-x}{h}, & x > 0 \\ \frac{-(x+h)-(-x)}{h}, & x < 0 \end{matrix} \right\} = \left\{ \begin{matrix} 1, & x > 0 \\ -1, & x < 0 \end{matrix} \right\},$$

wobei wir in der ersten geschweiften Klammer benutzt haben, dass für $x > 0$ und h geeignet klein ebenfalls $x + h > 0$ gilt und analog für $x < 0$ und h geeignet klein $x + h < 0$ gilt. Für $x = 0$ existiert

$$\lim_{h\to 0} \frac{|0+h| - |0|}{h} = \lim_{h\to 0} \frac{|h|}{h}$$

also nicht und die Betragsfunktion ist in Null nicht differenzierbar.





V.2. Ableiten und lineare Approximation

Satz V.2.1. *(Lineare Approximation) Es sei $D \subset \mathbb{R}$ und es sei $x_0 \in D$ Berührpunkt von $D\backslash\{x_0\}$. Dann sind folgende Aussagen äquivalent.*

(i) *f ist in x_0 differenzierbar.*

(ii) *Es existieren $\alpha \in \mathbb{R}$ und $\varphi\colon D \to \mathbb{R}$, sodass gilt:*
   - $\forall\, x \in D\colon f(x) = f(x_0) + \alpha(x - x_0) + \varphi(x)$,
   - $\lim\limits_{\substack{x\to x_0\\ x\neq x_0}} \dfrac{\varphi(x)}{x - x_0} = 0.$

Falls obiges der Fall ist, gilt $\alpha = f'(x_0)$.

Beweis. (i) $\Rightarrow$ (ii): Es sei f differenzierbar in x_0. Wir setzen $\alpha := f'(x_0)$ und $\varphi(x) := f(x) - f(x_0) - \alpha(x - x_0)$. Dann gilt die Formel aus (ii) und weiter

$$\frac{\varphi(x)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \xrightarrow{x\to x_0} 0$$

(ii) $\Rightarrow$ (i): Es seien α und φ wie in (ii). Dann gilt

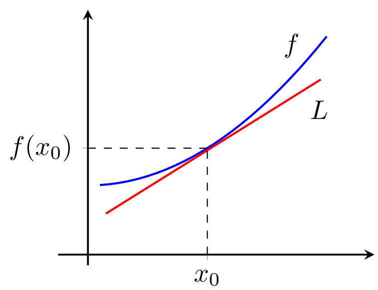
$$\frac{f(x) - f(x_0)}{x - x_0} - \alpha = \frac{f(x) - f(x_0) - \alpha(x - x_0)}{x - x_0} = \frac{\varphi(x)}{x - x_0} \xrightarrow{x\to x_0} 0$$

und damit existiert der Grenzwert $\lim_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0}$ und ist gleich α. $\square$

Bemerkung V.2.2. Es sei $f\colon D \subset \mathbb{R} \to \mathbb{R}$ differenzierbar in $x_0 \in \mathbb{R}$. Dann ist

$$L\colon D \to \mathbb{R},\ L(x) := f(x_0) + f'(x_0)(x - x_0)$$

eine affin–lineare Approximation von f. Den Graphen von L kann man sich als Tangente an den Graphen von f im Punkt $(x_0, f(x_0))$ vorstellen. In der Schule wird die Ableitung vielleicht manchmal als Steigung der Tagente in einem Punkt 'definiert'. Hierfür müsste man allerdings zuerst definieren, was eine Tangente an einen Graphen in einem Punkt sein soll. Dennoch passt das nachfolgende Bild gut zu der üblichen Erklärung, bei der man Sekanten einzeichnet, die durch $(x_0, f(x_0))$ und $(x, f(x))$ gehen, und sich dann vorstellt, dass x immer näher an x_0 heranrückt und die Sekante im Grenzfall mit dem Graphen von L übereinstimmt.



Die Abbildung φ aus V.2.1 kann als $\varphi = f - L$ geschrieben und als Fehler interpretiert werden, den wir machen, wenn wir f durch L ersetzen. V.2.1(ii) sagt dann, dass dieser Fehler schneller als $x - x_0$ gegen Null geht, wenn wir uns mit x an x_0 annähern, denn wenn wir $\varphi(x)$ durch $x - x_0$ teilen, was selbst gegen Null geht, dann geht der Quotient immer noch gegen Null.

Korollar V.2.3. *Es sei $f\colon D \to \mathbb{R}$ eine Abbildung. Es sei $x_0 \in D$. Dann gilt: Ist f differenzierbar in x_0, so ist f ist auch stetig in x_0. Die Umkehrung gilt nicht.*

Beweis. '$\Longrightarrow$' Wir wählen α und φ wie in V.2.1(ii). Dann gilt

$$\begin{aligned}\lim_{x\to x_0} f(x) &= \lim_{x\to x_0}\big(f(x_0) + f'(x_0)(x - x_0) + \varphi(x)\big)\\ &= f(x_0) + \lim_{x\to x_0} f'(x_0)(x - x_0) + \lim_{x\to x_0}\varphi(x) = f(x_0).\end{aligned}$$

'$\not\Longleftarrow$' Die Betragsfunktion $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ ist im Punkt $x_0 = 0$ zwar stetig, aber nicht differenzierbar. $\square$

V.3. Rechenregeln für Ableitungen

Satz V.3.1. *(Ableitungsregeln) Es seien $f, g\colon D \subset \mathbb{R} \to \mathbb{R}$ differenzierbar in $x \in D$ und sei $\lambda \in \mathbb{R}$. Dann sind auch die Abbildungen $f + g$, λf und $f \cdot g\colon D \to \mathbb{R}$ differenzierbar in x und es gilt:*

$$\begin{aligned}&(f+g)'(x) = f'(x) + g'(x)\\ &(\lambda f)' = \lambda f'(x)\\ &(f\cdot g)'(x) = f'(x)g(x) + f(x)g'(x) \quad \textit{Produkt- oder Leibnizregel.}\end{aligned}$$

Die ersten beiden Regeln zusammen besagen, dass Ableiten eine $\mathbb{R}$-lineare Abbildung ist. Ist zusätzlich zu den Voraussetzungen oben noch $g(x) \neq 0$ für alle $x \in D$, dann ist auch $\frac{f}{g}\colon D \to \mathbb{R}$ in x differenzierbar und es gilt die Quotientenregel

$$\Big(\frac{f}{g}\Big)'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}.$$

Beweis. Die ersten zwei Aussagen folgen direkt aus den Rechenregeln für Grenzwerte, die wir in der Mathematik 2 behandelt haben. Für die Produktregel berechnen wir

$$\begin{aligned}(fg)'(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\Big(f(x+h)\big[g(x+h) - g(x)\big] + \big[f(x+h) - f(x)\big]g(x)\Big) \\ &= \lim_{h\to 0} \Big(f(x+h)\frac{g(x+h)-g(x)}{h} + \frac{f(x+h)-f(x)}{h}g(x)\Big) \\ &= f(x)g'(x) + f'(x)g(x).\end{aligned}$$

Für die Quotientenregel betrachten wir erst den Spezialfall, dass der Zähler konstant Eins ist:

$$\begin{aligned}\big(\frac{1}{g}\big)'(x) &= \lim_{h\to 0} \frac{1}{h}\Big(\frac{1}{g(x+h)} - \frac{1}{g(x)}\Big) \\ &= \lim_{h\to 0} \frac{1}{h}\frac{g(x) - g(x+h)}{g(x+h)g(x)} \\ &= \lim_{h\to 0} \frac{1}{g(x+h)g(x)}\frac{g(x)-g(x+h)}{h} \\ &= \frac{1}{g(x)^2}\big(-g'(x)\big).\end{aligned}$$

Hierbei haben wir im letzten Schritt benutzt, dass g in x wegen V.2.3 stetig ist. Die allgemeine Version der Quotientenregel können wir nun mithilfe der bereits bewiesenen Produktregel zeigen:

$$\big(\frac{f}{g}\big)'(x) = \big(f\cdot\frac{1}{g}\big)'(x) = f'(x)\frac{1}{g(x)} + \frac{-g'(x)}{g(x)^2}f(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}.$$ □

Satz V.3.2. *(Kettenregel) Es seien $f\colon D\subset\mathbb{R}\to\mathbb{R}$ und $g\colon F\subset\mathbb{R}\to\mathbb{R}$ mit $f(D)\subset F$ gegeben. Es sei f differenzierbar in $x\in D$ und g differenzierbar in $y := f(x)\in F$. Dann ist $g\circ f\colon D\to\mathbb{R}$ differenzierbar in x und es gilt*

$$(g\circ f)'(x) = g'(f(x))f'(x).$$

Beweis. Wir definieren zunächst die Hilfsfunktion

$$g^*\colon F\to\mathbb{R},\ g^*(\tilde{y}) = \begin{cases} \frac{g(\tilde{y})-g(y)}{\tilde{y}-y}, & \tilde{y}\neq y, \\ g'(y), & \tilde{y} = y.\end{cases}$$

Da g in y differenzierbar ist, gilt dann $\lim_{\tilde{y}\to y} g^*(\tilde{y}) = g'(y) = g^*(y)$. Daraus folgt

$$(g\circ f)'(x) = \lim_{\tilde{x}\to x}\frac{g(f(\tilde{x})) - g(f(x))}{\tilde{x}-x} = \lim_{\tilde{x}\to x} g^*(f(\tilde{x}))\cdot\frac{f(\tilde{x}) - f(x)}{\tilde{x}-x} = g'(f(x))\cdot f'(x),$$

wobei die letzte Gleichheit aus V.2.3 zusammen mit dem in der Zeile davor ausgerechneten Grenzwert folgt. Die mittlere Gleichung sieht man am besten rückwärts durch Fallunterscheidung: Ist $f(\tilde{x})\neq f(x)$ so kann der erste Faktor gemäß dem ersten Fall in der Definition von g^* ersetzt werden. Dann kürzen sich aber die Terme $f(\tilde{x}) - f(x)$ weg und auf beiden Seiten bleibt das gleiche übrig. Ist $f(\tilde{x}) = f(x)$, so sind beide Seiten gleich Null. □

Beispiele V.3.3.

(i) Es sei $f_n\colon\mathbb{R}\to\mathbb{R}$, $f_n(x) = x^n$ für $n\in\mathbb{N}_0$. Für $n\in\mathbb{N}$ gilt dann $f_n' = nf_{n-1}$. Wir zeigen dies mit Induktion.

Die Fälle $n = 1, 2$ haben wir schon in V.1.2 gesehen.

Für den Induktionsschritt von n auf $n+1$ erhalten wir

$$\begin{aligned}(f_{n+1})'(x) = (f_1 f_n)'(x) &= f_1'(x)f_n(x) + f_1(x)f_n'(x) \\ &= 1\cdot x^n + x\cdot nx^{n-1} = (1+n)x^n = (n+1)f_n(x),\end{aligned}$$

wobei wir für die zweite Gleichung die Produktregel und für die dritte die Induktionsannahme sowie die Formel für $n = 1$ aus dem Induktionsanfang benutzt haben.

(ii) Es sei $g_n \colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $g_n(x) = \frac{1}{x^n}$ für $n \in \mathbb{N}$. Für $n \in \mathbb{N}$ gilt dann $g_n' = (-n) \cdot g_{n+1}$. In der Tat folgt mit der Quotientenregel und (i):

$$\begin{aligned} g_n'(x) = \big(\frac{1}{f_n}\big)'(x) &= \frac{0 \cdot f_n(x) - 1 \cdot f_n'(x)}{f_n(x)^2} \\ &= \frac{-nx^{n-1}}{(x^n)^2} = -\frac{n}{x^{2n-n+1}} = -\frac{n}{x^{n+1}} = (-n)\, g_{n+1}(x). \end{aligned}$$

(iii) Beispiele (i) und (ii) schreibt man oft zusammen als $(x^n)' = nx^{n-1}$ für $n \in \mathbb{Z}$. Dabei muss man aber beachten, dass diese Funktionen für unterschiedliche n verschiedene Definitionsbereiche haben.

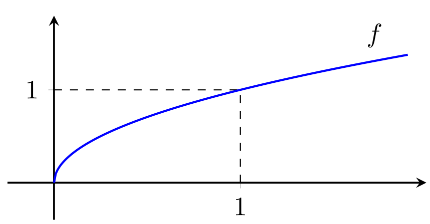
(iv) $f\colon [0,\infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$. Dann gilt $f'(x) = \frac{1}{2\sqrt{x}}$ für $x > 0$:

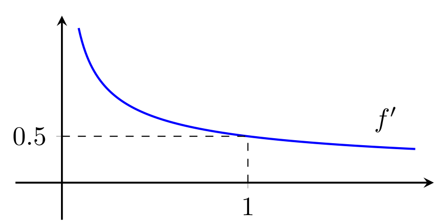
$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\sqrt{x+h} - \sqrt{x}}{h} \\ &= \lim_{h\to 0} \frac{(\sqrt{x+h} - \sqrt{x})(\sqrt{x+h} + \sqrt{x})}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h\to 0} \frac{(x+h) - x}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \lim_{h\to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} \\ &= \frac{1}{2\sqrt{x}}. \end{aligned}$$

Beachten Sie, dass hier nur strikt positive Werte von h betrachtet werden, da der Grenzwert über $h \in [0,\infty)\backslash\{0\}$ gebildet wird, vgl. V.1.1 und V.1.2(i). In $x = 0$ ist f nicht differenzierbar, weil der Grenzwert

$$\lim_{h\to 0} \frac{\sqrt{0+h} - \sqrt{0}}{h} = \lim_{h\to 0} \frac{\sqrt{h}}{h}$$

nicht existiert. Im folgenden Bild sieht man, dass die Funktion immer steiler wird, je näher man der Null kommt; für die Ableitung(sfunktion) $f'\colon (0,\infty) \to \mathbb{R}$ gilt $\lim_{x\to 0} f'(x) = +\infty$. In Null existiert die Ableitung nicht.





Satz V.3.4. *Die Exponentialfunktion* $\exp\colon \mathbb{R} \to \mathbb{R}$ *ist differenzierbar und es gilt* $\exp' = \exp$.

Beweis. Es sei $x \in \mathbb{R}$ und $0 < |h| < 1$. Dann gilt (mit demselben Trick wie in einer Präsenzaufgabe):

$$\begin{aligned} \big|\frac{e^{x+h} - e^x}{h} - e^x\big| = e^x\big|\frac{e^h - 1 - h}{h}\big| &= \frac{e^x}{|h|}|e^h - (1+h)| \\ &= \frac{e^x}{|h|}\big|e^h - \sum_{k=0}^{1} \frac{h^k}{k!}\big| \leqslant \frac{e^x}{|h|}\frac{|h|^{1+1}}{(1+1)!}e^{|h|} \\ &= e^x\frac{|h|}{2}e^{|h|} \to 0, \end{aligned}$$

wobei wir zuerst die Fehlerabschätzung für die Exponentialfunktion angewendet haben und dann, dass nach III.5.3(i) die Exponentialfunktion (insbesondere in Null) stetig ist. Es folgt also

$$\lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = e^x$$

und damit die Behauptung. □

Satz V.3.5. *Es sei $I \subset \mathbb{R}$ ein beliebiges Intervall. Es sei $f\colon I \to \mathbb{R}$ stetig, strikt wachsend und $g := f^{-1}\colon I' \to \mathbb{R}$ die Inverse, wobei wir $I' := \operatorname{Bild}(f)$ setzen. Es sei nun zusätzlich f differenzierbar in $x \in I$ und gelte $f'(x) \neq 0$. Dann ist g differenzierbar in $y := f(x)$ und es gilt*

$$g'(y) = \frac{1}{f'(x)} \Big(= \frac{1}{f'(g(y))} \Big).$$

Der Satz gilt analog für strikt fallende Funktionen.

Beweis. Es sei $(y_n)_{n\in\mathbb{N}} \subset I'\backslash\{y\}$ gegeben mit $\lim_{n\to\infty} y_n = y$. Wir setzen $x_n := g(y_n)$. Da g stetig ist, folgt $\lim_{n\to\infty} x_n = g(y) = x$. Da g bijektiv ist, gilt $x_n \neq x$ für alle $n \in \mathbb{N}$. Nun folgt

$$\lim_{n\to\infty} \frac{g(y_n) - g(y)}{y_n - y} = \lim_{n\to\infty} \frac{x_n - x}{f(x_n) - f(x)} = \frac{1}{f'(x)}$$

und da $(y_n)_{n\in\mathbb{N}}$ beliebig war, haben wir $g'(y) = \frac{1}{f'(x)}$ gezeigt. □

Korollar V.3.6. *Es gilt* $\log'(x) = \dfrac{1}{\exp'(\log x)} = \dfrac{1}{\exp(\log x)} = \dfrac{1}{x}$ *für $x > 0$ und unter Benutzung von V.3.5 und V.3.4.*

Satz V.3.7. *Die Funktionen* $\sin, \cos\colon \mathbb{R} \to \mathbb{R}$ *sind differenzierbar und es gilt* $\sin' = \cos$ *und* $\cos' = -\sin$.

Beweis. Wir notieren zunächst

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\sin(x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h\to 0} \Big(\sin x \cdot \frac{\cos h - 1}{h} + \cos x \cdot \frac{\sin h}{h} \Big) \\ &= \sin x \cdot \lim_{h\to 0} \frac{\cos h - 1}{h} + \cos x \cdot \lim_{h\to 0} \frac{\sin h}{h} =: (\star)\end{aligned}$$

und sehen, dass es genügt,

$$\lim_{h\to 0} \frac{\cos h - 1}{h} = 0 \quad \text{und} \quad \lim_{h\to 0} \frac{\sin h}{h} = 1$$

zu zeigen. Wir behandeln den ersten Grenzwert hier und den zweiten in einer Präsenzaufgabe. Da wir im Beweis von V.3.4 die Exponentialfunktion mithilfe der Fehlerabschätzung ihrer Reihendarstellung behandelt haben, und wir für den Kosinus ebenfalls die Reihendarstellung $\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$ kennen, liegt es nahe auch hier dieselbe zu verwenden. Dies führt zunächst auf

$$\Big|\frac{\cos h - 1}{h}\Big| = \frac{1}{|h|}|\cos h - 1| = \frac{1}{|h|}\Big|\cos h - \sum_{k=0}^{0} (-1)^k \frac{h^{2k}}{(2k)!}\Big|$$

und wir benötigen nun eine Fehlerabschätzung. Hierfür berechnen wir für $n \geqslant 0$

$$\operatorname{Re}\Big(\sum_{k=0}^{2n} \frac{(ix)^k}{k!}\Big) = \sum_{k=0}^{2n} \operatorname{Re}(i^k)\frac{x^k}{k!} \quad \text{und} \quad i^k = \begin{cases} 1, & k = 0, 4, 8, \dots \\ i, & k = 1, 5, 9, \dots \\ -1, & k = 2, 6, 10, \dots \\ -i, & k = 3, 7, 11, \dots, \end{cases}$$

woraus für $|x| < 1$ folgt

$$\begin{aligned}\Big|\cos x - \sum_{k=0}^{n}(-1)^k \frac{x^{2k}}{(2k)!}\Big| &= \Big|\mathrm{Re}(e^{ix}) - \mathrm{Re}\Big(\sum_{k=0}^{2n+1}\frac{(ix)^k}{k!}\Big)\Big| \\ &= \Big|\mathrm{Re}\Big(e^{ix} - \sum_{k=0}^{2n+1}\frac{(ix)^k}{k!}\Big)\Big| \\ &\leqslant \Big|e^{ix} - \sum_{k=0}^{2n+1}\frac{(ix)^k}{k!}\Big| \\ &\leqslant \frac{|ix|^{2n+2}}{(2n+2)!}e^{|ix|} \\ &\leqslant 3\frac{|x|^{2n+2}}{(2n+2)!},\end{aligned}$$

wobei wir für die erste Ungleichung $|z| = \sqrt{(\mathrm{Re}z)^2 + (\mathrm{Im}z)^2} \geqslant \sqrt{(\mathrm{Re}z)^2} = |\mathrm{Re}z|$ benutzt haben, für die zweite die Fehlerabschätzung der Exponentialfunktion und für die dritte $|i| = 1$ und dass die reelle Exponentialfunktion wachsend ist und daher $e^{|x|} \leqslant e^1 \leqslant 3$ gilt.

Wir erhalten also insgesamt

$$\Big|\cos x - \sum_{k=0}^{n}(-1)^k \frac{x^{2k}}{(2k)!}\Big| \leqslant 3\frac{|x|^{2n+2}}{(2n+2)!}$$

Um zu der Abschätzung von $|\frac{\cos h-1}{h}|$ zurückzukommen, setzen wir oben $x = h$ und $n = 0$ und teilen dann nochmal durch $|h| = |x|$ und erhalten nun die gewünschte Abschätzung

$$\Big|\frac{\cos h - 1}{h}\Big| = \frac{1}{|h|}\Big|\cos(h) - \sum_{k=0}^{0}(-1)^k\frac{h^{2k}}{(2k)!}\Big| \leqslant \frac{3}{|h|}\frac{|h|^2}{2!} = \frac{3}{2}|h| \to 0$$

für $h \to 0$.

Wir zeigen $\lim_{h\to 0}\frac{\sin h}{h} = 1$ in einer Präsenzaufgabe und bekommen damit

$$(\star) = \sin x \cdot \lim_{h\to 0}\frac{\cos h - 1}{h} + \cos x \cdot \lim_{h\to 0}\frac{\sin h}{h} = \cos x.$$

Analog zeigt man $\cos' = -\sin$. □

Bemerkungen V.3.8.

(i) Mit V.3.7 und der Quotientenregel kann man sehen, dass auch der Tangens differenzierbar ist und $\tan'$ bestimmen.

(ii) Mit (i) und V.3.5 folgt, dass die inversen trigonometrischen Funktionen differenzierbar sind; ihre Ableitungen werden wir in einer Übungsaufgabe bestimmen.

V.4. Höhere Ableitungen

Ist eine Funktion $f\colon D \subset \mathbb{R} \to \mathbb{R}$ auf ganz D differenzierbar, so betrachten wir die Ableitung als neue Funktion $f'\colon D \to \mathbb{R}$. Es liegt dann nahe, diese Funktion wieder auf Differenzierbarkeit oder zumindest auf Stetigkeit zu untersuchen und, im Fall, dass f' wieder überall differenzierbar sein sollte, für die Ableitung der Ableitung $(f')' =: f''$ wieder genauso zu verfahren. Im Folgenden beschränken wir uns auf Definitionsbereiche die nichtleere offene Intervalle sind und definieren induktiv, was es heißt, k-mal differenzierbar zu sein, wobei wir den Anfang in (i) für $k = 1$ schon in V.1.1 erledigt haben und nur noch bemerken müssen, dass wir statt f' unten $f^{(1)}$ schreiben:

Definition V.4.1. Es sei I ein nichtleeres offenes Intervall und $f\colon I \to \mathbb{R}$.

(i) Für $k \geqslant 2$ sagen wir, dass f *k-mal differenzierbar* ist, falls f $(k-1)$–mal differenzierbar ist und $f^{(k-1)}\colon I \to \mathbb{R}$ differenzierbar ist. Wir bezeichnen in diesem Fall $f^{(k)} := (f^{(k-1)})'$ als die *k–te Ableitung*. (Die wird manchmal auch als $\frac{\mathrm{d}^k f}{\mathrm{d}x^k}$ notiert.)

(ii) Ist für $k \geqslant 1$ die k–te Ableitung zusätzlich stetig, so sagen wir f ist *k–mal stetig differenzierbar*. Im Fall einer 1–mal stetig differenzierbaren Abbildung lassen wir '1–' weg und sprechen nur von einer *stetig differenzierbaren Abbildung.*

(iii) Wir führen die folgende Notation ein

$$\mathrm{C}^0(I) := \mathrm{C}(I) := \{f\colon I \to \mathbb{R} \mid f \text{ ist stetig}\}$$
$$\mathrm{C}^k(I) := \{f\colon I \to \mathbb{R} \mid f \text{ ist } k\text{–mal stetig differenzierbar}\} \text{ für } k \in \mathbb{N}$$
$$\mathrm{C}^\infty(I) := \bigcap_{k\in\mathbb{N}} \mathrm{C}^k(I)$$

Für $k \in \mathbb{N}_0 \cup \{\infty\}$ nennen wir die Elemente von $\mathrm{C}^k(I)$ auch C^k–Funktionen. Wir bezeichnen manchmal mit $f^{(0)} := f$ die Funktion selbst ('0–te Ableitung'). Wir beachten, dass C^∞–Funktionen *beliebig oft* (stetig) differenzierbar sind.

Beispiele V.4.2.

(1) $x \mapsto x^2$, sin, cos, exp $\in \mathrm{C}^\infty(\mathbb{R})$, $\log \in \mathrm{C}^\infty((0,\infty))$.

(2) $|\cdot| \in \mathrm{C}^0(\mathbb{R})\setminus \mathrm{C}^1(\mathbb{R})$; in der Tat ist $|\cdot|$ in $x = 0$ nicht differenzierbar also insbesondere nicht stetig differenzierbar auf $\mathbb{R}$.

(3) $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = \left\{\begin{matrix} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{matrix}\right\}$ liegt ebenfalls in $\mathrm{C}^0(\mathbb{R})\setminus \mathrm{C}^1(\mathbb{R})$, ist aber überall differenzierbar jedoch mit unstetiger Ableitung.

(4) $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = \left\{\begin{matrix} x^3 \sin\left(\frac{1}{x}\right), & x \neq 0 \\ 0, & x = 0 \end{matrix}\right\}$ gehört zu $\mathrm{C}^1(\mathbb{R})\setminus \mathrm{C}^2(\mathbb{R})$. Hier ist f' in Null nicht differenzierbar.

KAPITEL VI

Der Mittelwertsatz und seine Konsequenzen

VI.1. Extrema und der Mittelwertsatz

Definition VI.1.1. Es sei $f\colon (a,b) \to \mathbb{R}$ mit $-\infty \leqslant a < b \leqslant +\infty$ gegeben. Dann hat f ein *lokales Maximum* [bzw. *Minimum*] in $x \in (a,b)$, falls ein $\varepsilon > 0$ existiert mit $(x-\varepsilon, x+\varepsilon) \subset (a,b)$ und $f(x) \geqslant f(\xi)$ [bzw. $f(x) \leqslant f(\xi)$] für alle $\xi \in (x-\varepsilon, x+\varepsilon)$. Das lokale Maximum [bzw. Minimum] ist *strikt* falls '=' nur für $\xi = x$ gilt. Wir verwenden das Wort *Extremum* für beides, Maximum und Minimum. Lokale Extrema nennt man auch *relative Extrema*.

Satz VI.1.2. *(notwendige Bedingung für lokale Extrema) Es sei $f\colon (a,b) \to \mathbb{R}$ mit $-\infty \leqslant a < b \leqslant +\infty$ gegeben. Wenn f im Punkt $x \in (a,b)$ ein lokales Extremum hat und f in x differenzierbar ist, dann gilt $f'(x) = 0$.*

Beweis. Wir behandeln nur den Fall, dass f in x ein lokales Maximum hat. Dann existiert ein $\varepsilon > 0$, sodass $(x-\varepsilon, x+\varepsilon) \subset (a,b)$ und $f(\xi) \leqslant f(x)$ gilt für alle $\xi \in (x-\varepsilon, x+\varepsilon)$. Weil f differenzierbar ist, folgt

$$f'(x) = \lim_{\xi \to x} \frac{f(\xi) - f(x)}{\xi - x} = \lim_{\xi \searrow x} \frac{f(\xi) - f(x)}{\xi - x} = \lim_{\xi \nearrow x} \frac{f(\xi) - f(x)}{\xi - x}.$$

In den beiden letzten Brüchen sind für alle ξ die hinreichend nah bei x sind, die Zähler negativ oder Null. Der Nenner ist aber im letzten Bruch stets negativ und im vorletzten stets positiv. D.h. der Grenzwert ist gleichzeitig kleiner gleich Null und größer gleich Null und kann damit nur gleich Null sein. □

Bemerkungen VI.1.3.

(i) $f'(x) = 0$ ist nur notwendig, aber nicht hinreichend für ein Extremum: $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$ erfüllt $f'(0) = 0$, aber in Null liegt kein Extremum vor.

(ii) Ein Extremum kann vorliegen, ohne dass die Funktion differenzierbar ist, z.B. hat $|\cdot|\colon \mathbb{R} \to \mathbb{R}$ ein striktes lokales Minimum in Null.

(iii) Betrachten wir $f\colon [a,b] \to \mathbb{R}$ mit $-\infty < a < b < +\infty$, so kann $f(a) = \max_{x \in [a,b]} f(x)$ oder $f(a) = \min_{x \in [a,b]} f(x)$ gelten und f in a differenzierbar sein, ohne dass die Ableitung in diesem Punkt verschwindet, so zum Beispiel für $f\colon [0,1] \to \mathbb{R}$, $f(x) = x$.

Satz VI.1.4. *(Satz von Rolle; ausgesprochen 'Roll') Es sei $-\infty < a < b < +\infty$ und es sei $f\colon [a,b] \to \mathbb{R}$ stetig mit $f(a) = f(b)$. Weiterhin sei die Einschränkung von f auf (a,b), $f\colon (a,b) \to \mathbb{R}$, differenzierbar. Dann existiert ein $\xi \in (a,b)$ mit $f'(\xi) = 0$.*

Beweis. Falls f konstant ist, dann ist die Aussage trivial. Es sei f also nicht konstant.

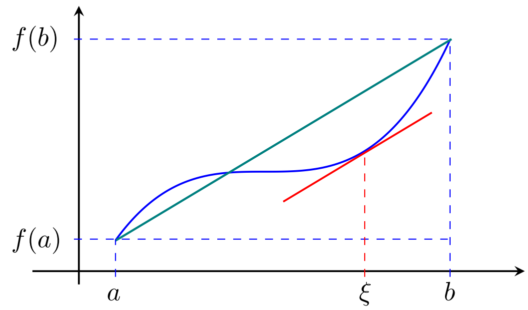
Fall 1: Es gibt ein $x_0 \in (a,b)$ mit $f(x_0) > f(a)$. Nach dem Extremalsatz (III.6.2) existiert $x \in [a,b]$ mit $f(x) = \max_{y \in [a,b]} f(y)$. Dieses x muss aber ungleich a und auch ungleich b sein, da wir ja wissen, dass $f(x_0) > f(a)\big(= f(b)\big)$ gilt und damit das Maximum nicht in a bzw. b angenommen werden kann. D.h. es gilt $x \in (a,b)$ und daher folgt mit VI.1.2, dass $f'(x) = 0$ ist.

Fall 2: Wenn es kein $x_0 \in (a,b)$ mit $f(x_0) > f(a)$ gibt, dann muss es ein x_0 mit $f(x_0) < f(a)$ geben und man kann völlig analog zum ersten Fall argumentieren. □

Satz VI.1.5. *(Mittelwertsatz) Es sei $-\infty < a < b < +\infty$ und $f\colon [a,b] \to \mathbb{R}$ sei stetig und $f\colon (a,b) \to \mathbb{R}$ differenzierbar. Dann existiert $\xi \in (a,b)$ mit*

$$\frac{f(b) - f(a)}{b - a} = f'(\xi).$$

Im Bild hat die rote Gerade die Steigung $f'(\xi)$ und ist parallel zur grünen Verbindungsstrecke mit Steigung $\frac{f(b)-f(a)}{b-a}$:



Beweis. Wir definieren die Hilfsfunktion

$$F\colon [a,b] \to \mathbb{R}, \quad F(x) := f(x) - \frac{f(b)-f(a)}{b-a}(x-a).$$

Dann ist F stetig auf $[a,b]$ und differenzierbar auf (a,b) nach Ergebnissen aus Kapitel III und Kapitel V. Außerdem gilt $F(a) = f(a) = F(b)$ wie man durch Einsetzen direkt überprüft. Nach dem Satz von Rolle existiert daher $\xi \in (a,b)$ mit $F'(\xi) = 0$. Damit gilt aber

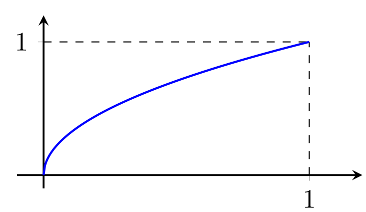
$$0 = F'(\xi) = f'(\xi) - \frac{f(b)-f(a)}{b-a}$$

und dies zeigt die Behauptung, wenn man nach $f'(\xi)$ auflöst. $\square$

Bemerkungen VI.1.6.

(i) Der Spezialfall $f(a) = f(b)$ in VI.1.5 liefert den Satz von Rolle.

(ii) Man könnte in VI.1.4 und VI.1.5 auch $\xi = a, b$ erlauben; Differenzierbarkeit und Ableitung in Randpunkten sind mit V.1.1 definiert. Der Satz ist so aber besser, da wir die Differenzierbarkeit in a und b nicht voraussetzen müssen und die Aussage $\xi \in (a,b)$ stärker ist als $\xi \in [a,b]$. Ein konkretes Beispiel dafür ist die Wurzelfunktion $f\colon [0,1] \to \mathbb{R}$, $f(x) = \sqrt{x}$, die in Null nicht differenzierbar ist, aber die Voraussetzungen des Mittelwertsatzes erfüllt:



(iii) Es gibt einen 'verallgemeinerten Mittelwertsatz': Es seien $f, g\colon [a,b] \to \mathbb{R}$ stetig und auf (a,b) differenzierbar und es sei außerdem $g'(x) \neq 0$ für alle $x \in (a,b)$. Dann ist $g(a) \neq g(b)$ und es existiert $\xi \in (a,b)$ mit

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(\xi)}{g'(\xi)}.$$

Der Beweis wird in einer Präsenzaufgabe behandelt. Setzt man $g(x) = x$, so erhält man den Mittelwertsatz in der Fassung von VI.1.5.

Korollar VI.1.7. *Es sei $f\colon (a,b) \to \mathbb{R}$ differenzierbar mit $-\infty \leqslant a < b \leqslant +\infty$. Falls f' auf (a,b) beschränkt ist (d.h. es gilt $\exists\, m,\, M \in \mathbb{R}\, \forall\, \xi \in (a,b)\colon m \leqslant f'(\xi) \leqslant M$) so folgt*

$$\forall\, x_1,\, x_2 \in (a,b) \text{ mit } x_1 < x_2\colon m(x_2 - x_1) \leqslant f(x_2) - f(x_1) \leqslant M(x_2 - x_1)$$

Beweis. Für $x_1 < x_2$ in (a,b) existiert $\xi \in (x_1, x_2)$ mit $\frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(\xi)$ und daher nach Voraussetzung $m \leqslant \frac{f(x_2)-f(x_1)}{x_2-x_1} \leqslant M$. $\square$

Korollar VI.1.8. *Es seien $-\infty < a < b < +\infty$ und $f\colon [a,b] \to \mathbb{R}$ sei stetig und auf (a,b) differenzierbar mit $f'(x) = 0$ für alle $x \in (a,b)$. Dann ist f konstant.*

Beweis. Wir wenden VI.1.7 mit $m = M = 0$ an. □

Satz VI.1.9. *Es sei $c \in \mathbb{R}$ und $f\colon \mathbb{R} \to \mathbb{R}$ sei eine differenzierbare Funktion mit $f'(x) = cf(x)$ für alle $x \in \mathbb{R}$. Es sei weiter $a := f(0)$. Dann gilt $f(x) = ae^{cx}$ für alle $x \in \mathbb{R}$.*

Beweis. Wir definieren die Hilfsfunktion $F\colon \mathbb{R} \to \mathbb{R}$, $F(x) = f(x)e^{-cx}$. Dann ist F differenzierbar mit

$$F'(x) = f'(x)e^{-cx} - cf(x)e^{-cx} = (f'(x) - cf(x))e^{-cx} = 0.$$

Nach VI.1.8 ist F also konstant und wegen $F(0) = f(0)e^0 = f(0) = a$ folgt, dass $F(x) = a$ für jedes $x \in \mathbb{R}$ gelten muss. Umstellen von $a = F(x) = f(x)e^{-cx}$ liefert $f(x) = ae^{cx}$ für alle $x \in \mathbb{R}$. □

Bemerkung VI.1.10. VI.1.8 bedeutet, dass $\exp\colon \mathbb{R} \to \mathbb{R}$ durch die Eigenschaften $\exp' = \exp$ und $\exp(0) = 1$ eindeutig bestimmt ist.

Satz VI.1.11. *Es seien $-\infty < a < b < +\infty$ und $f\colon [a,b] \to \mathbb{R}$ sei stetig und auf (a,b) differenzierbar.*

(i) *Falls $f'(x) \geqslant 0$ [bzw. > 0, $\leqslant 0$, < 0] für alle $x \in (a,b)$ gilt, dann ist f auf $[a,b]$ wachsend [bzw. strikt wachsend, fallend, strikt fallend].*

(ii) *Ist f auf $[a,b]$ wachsend [bzw. fallend], dann ist $f'(x) \geqslant 0$ [bzw. $\leqslant 0$] für alle $x \in (a,b)$.*

Beweis. (i) Es sei $f'(x) \geqslant 0$ für alle $x \in (a,b)$. Angenommen, f ist nicht wachsend. Dann existieren x_1, $x_2 \in [a,b]$ mit $x_1 < x_2$ aber $f(x_1) > f(x_2)$. Nach dem Mittelwertsatz existiert dann $\xi \in (x_1, x_2)$, sodass $f'(\xi) = \frac{f(x_2)-f(x_1)}{x_2-x_1} < 0$ gilt. Widerspruch.

Es sei jetzt $f'(x) > 0$ für alle $x \in (a,b)$. Angenommen, f ist nicht strikt wachsend. Dann existieren x_1, $x_2 \in [a,b]$ mit $x_1 < x_2$ aber $f(x_1) \geqslant f(x_2)$. Nach dem Mittelwertsatz existiert dann $\xi \in (x_1, x_2)$, sodass $f'(\xi) = \frac{f(x_2)-f(x_1)}{x_2-x_1} \leqslant 0$ gilt. Widerspruch.

Die Aussagen mit $f'(x) \leqslant 0$ und < 0 zeigt man analog.

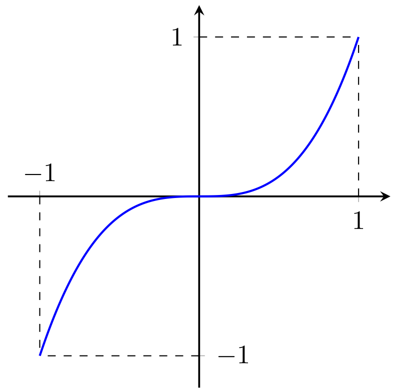
(ii) Es sei nun f wachsend. Dann gilt für x, $\xi \in (a,b)$ mit $x \neq \xi$ stets $\frac{f(\xi)-f(x)}{\xi-x} \geqslant 0$. Wir fixieren $x \in (a,b)$ und nehmen den Grenzwert $\xi \to x$. Dann folgt $f'(x) \geqslant 0$.

Die zweite Aussagen mit f fallend behandelt man analog. □

Bemerkungen VI.1.12.

(i) Hätten wir im Mittelwertsatz $\xi = a$, b erlaubt (siehe die Kommentare in VI.1.6(ii)), dann hätten wir oben in (i) Differenzierbarkeit in a, b fordern und auch in diesen Punkten $f'(x) \geqslant 0$ [bzw. > 0, $\leqslant 0$, < 0] fordern müssen. Das hätte unseren Satz schwächer gemacht, z.B. kann man so mit VI.1.11(i) schließen, dass $f\colon [0,1] \to \mathbb{R}$, $f(x) = x^2$ und $g\colon [0,1] \to \mathbb{R}$, $g(x) = \sqrt{x}$ auf $[0,1]$ strikt wachsen, obwohl $f'(0) = 0$ ist und $g'(0)$ nicht existiert. In der oben skizzierten Variante des Satzes hätte man 'strikt wachsend' auf $(0,1]$ schließen können und hätte dann nochmal genauer untersuchen müssen, was passiert, wenn man den Punkt 0 noch hinzufügt.

(ii) In VI.1.11(ii) wurde die Implikation 'f strikt wachsend $\Rightarrow f' > 0$' nicht vergessen, sondern sie ist falsch: Z.B. ist $f\colon [-1,1] \to \mathbb{R}$, $f(x) = x^3$ strikt wachsend aber es gilt $f'(0) = 0$.



Satz VI.1.13. *(Hinreichende Bedingung für lokale Extrema) Es seien $-\infty \leqslant a < b \leqslant +\infty$ und sei $f\colon (a,b) \to \mathbb{R}$ überall differenzierbar und an der Stelle $x \in (a,b)$ zweimal differenzierbar*[1]*. Es gelte $f'(x) = 0$ und $f''(x) > 0$ [bzw. < 0]. Dann hat f an der Stelle x ein striktes lokales Minimum [bzw. Maximum].*

Beweis. Es sei $f''(x) > 0$. Wegen

$$f''(x) = \lim_{\xi \to x} \frac{f'(\xi) - f'(x)}{\xi - x} > 0$$

existiert ein $\varepsilon > 0$, sodass $\frac{f'(\xi)-f'(x)}{\xi - x} > 0$ für alle $\xi \in (x-\varepsilon, x+\varepsilon)\backslash\{x\}$ gilt. (Angenommen nicht, dann existiert eine Folge $(\xi_n)_{n\in\mathbb{N}} \subset (a,b)$ mit $\xi_n \to x$ und $\frac{f'(\xi_n)-f(x)}{\xi_n - x} \leqslant 0$ für alle $n \in \mathbb{N}$, also $\lim_{n\to\infty} \frac{f'(\xi_n)-f(x)}{\xi_n - x} \leqslant 0$ im Widerspruch dazu, dass letzterer Grenzwert für jede Folge $(\xi_n)_{n\in\mathbb{N}}$ echt größer Null ist.) Da $f'(x) = 0$ gilt, erhalten wir also

$$\forall\, \xi \in (x-\varepsilon, x+\varepsilon)\backslash\{x\} \colon \frac{f'(\xi)}{\xi - x} > 0$$

und daher $f'(\xi) < 0$ für $\xi \in (x-\varepsilon, x)$ und $f'(\xi) > 0$ für $\xi \in (x, x+\varepsilon)$. Mit VI.1.11 folgt, dass $f|_{[x-\varepsilon,x]}$ strikt fallend und $f|_{[x,x+\varepsilon]}$ strikt wachsend ist. Also muss in x ein striktes Minimum vorliegen.

Wenn $f''(x) < 0$ ist, argumentiert man analog. □

Bemerkung VI.1.14. Der letzte Teil des obigen Beweises zeigt das *Vorzeichenwechselkriterium*: Ist $f\colon (a,b) \to \mathbb{R}$ differenzierbar und gilt $f'(x) = 0$ für ein $x \in (a,b)$ und gibt es ein $\varepsilon > 0$, so dass $f'(\xi) < 0$ für $\xi \in (x-\varepsilon, x)$ sowie $f'(\xi) > 0$ für $\xi(x, x+\varepsilon)$ [bzw. $f'(\xi) > 0$ für $\xi \in (x-\varepsilon, x)$ sowie $f'(\xi) > 0$ für $\xi(x, x+\varepsilon)$], dann hat f in x ein striktes lokales Minimum [bzw. Maximum].

VI.2. Konvexe Funktionen

Definition VI.2.1. Es sei $I \subset \mathbb{R}$ ein beliebiges Intervall. $f\colon I \to \mathbb{R}$ heißt

(i) *konvex*, falls $\forall\, a, b \in I,\ \lambda \in [0,1] \colon f((1-\lambda)a + \lambda b) \leqslant (1-\lambda)f(a) + \lambda f(b)$,

(ii) *strikt konvex*, falls $\forall\, a \neq b \in I,\ \lambda \in (0,1) \colon f((1-\lambda)a + \lambda b) < (1-\lambda)f(a) + \lambda f(b)$,

(iii) [*strikt*] *konkav*, falls $-f$ [strikt] konvex ist.

Bemerkungen VI.2.2.

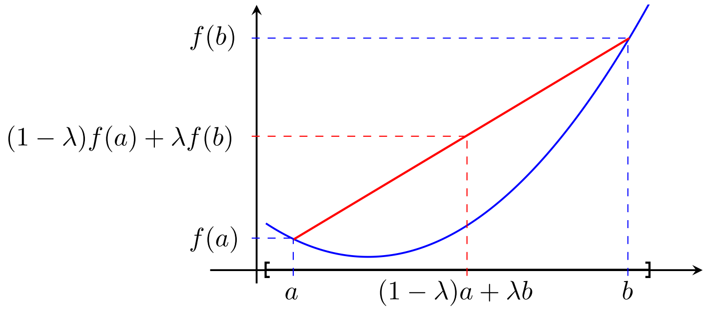
(i) In VI.2.1(i) könnte man auch $\lambda \in (0,1)$ schreiben, da für $\lambda \in \{0,1\}$ beide Seiten der geforderten Ungleichung gleich sind. Aus demselben Grund kann man in VI.2.1(ii) nicht $\lambda \in [0,1]$ schreiben, weil dann die strikte Ungleichung für $\lambda \in \{0,1\}$ immer verletzt wäre.

(ii) [Strikte] Konkavität wird durch die Bedingungen in VI.2.1(i) [bzw. VI.2.1(ii)] mit umgedrehter Abschätzung charakterisiert.

---

[1] Streng genommen haben wir in Kapitel V.4 nur definiert, was 2–mal stetig differenzierbar auf einem offenen Intervall I heißt, aber es ist natürlich klar, dass wir hier meinen, dass $f'\colon (a,b) \to \mathbb{R}$ im Punkt x differenzierbar ist, und dass wir mit $f''(x)$ die Ableitung von f' in x bezeichnen.

(iii) Die Bedingung in VI.2.1(i) hat die folgende anschauliche Interpretation:

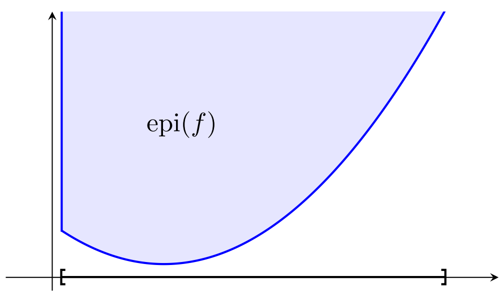


Für alle a, $b \in I$ ist die gerade Verbindungsline von $(a, f(a))$ nach $(b, f(b))$, oder genauer der Graph der Abbildung

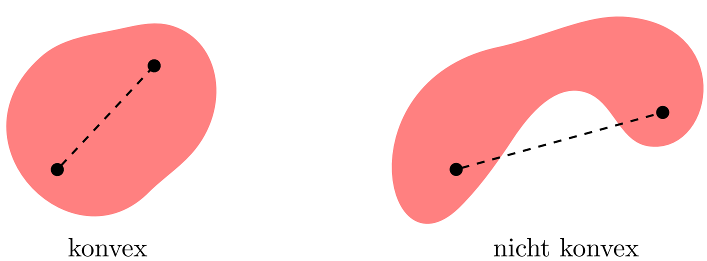
$$\phi\colon [0,1] \to \mathbb{R}^2,\ \ \phi(\lambda) = (1-\lambda)\left(\begin{smallmatrix} a \\ f(a) \end{smallmatrix}\right) + \lambda\left(\begin{smallmatrix} b \\ f(b) \end{smallmatrix}\right) = \left(\begin{smallmatrix} a \\ f(a) \end{smallmatrix}\right) + \lambda\left(\left(\begin{smallmatrix} b \\ f(b) \end{smallmatrix}\right) - \left(\begin{smallmatrix} a \\ f(a) \end{smallmatrix}\right)\right),$$

oberhalb des Graphen von f (wobei '=' erlaubt ist). Zeichnet man den sogenannten *Epigraph* von f,

$$\operatorname{epi}(f) := \{(x,y) \in \mathbb{R}^2 \mid x \in I,\, y \geqslant f(x)\},$$



so sieht man, wo der Name 'konvex' herkommt: f ist konvex genau dann, wenn $\operatorname{epi}(f) \subset \mathbb{R}^2$ konvex ist (d.h. die Verbindungsline von je zwei Punkten bleibt in der Menge):



Lemma VI.2.3. *Es sei $I \subset \mathbb{R}$ ein beliebiges Intervall und $f\colon I \to \mathbb{R}$. Dann sind folgende Aussagen äquivalent.*

(i) *f ist konvex.*

(ii) $\forall\, a < x < b\colon f(x) \leqslant f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$.

(iii) $\forall\, a < x < b\colon \frac{f(x)-f(a)}{x-a} \leqslant \frac{f(b)-f(a)}{b-a} \leqslant \frac{f(b)-f(x)}{b-x}$.

(iv) $\forall\, a < x < b\colon \frac{f(x)-f(a)}{x-a} \leqslant \frac{f(b)-f(x)}{b-x}$.

Ersetzt man in (ii)–(iv) '$\leqslant$' durch '$<$', so sind diese Aussagen äquivalent dazu, dass f strikt konvex ist. Durch Umdrehen der Abschätzungen erhält man Charakterisierungen [strikt] konkaver Funktionen.

Beweis. Wir zeigen nur die erste Aussage; der Rest geht genauso bzw. folgt durch Multiplikation mit -1.

(i) $\Longrightarrow$ (ii): Es seien $a < x < b$ in I. Setze $\lambda := \frac{x-a}{b-a}$. Dann ist $\lambda \in (0,1)$ und es gilt

$$\begin{aligned}(1-\lambda)a+\lambda b &= \Big(1-\frac{x-a}{b-a}\Big)a+\frac{x-a}{b-a}\,b\\ &= \frac{(b-a-x+a)a+(x-a)b}{b-a}\\ &= \frac{ba-xa+xb-ab}{b-a}\\ &= \frac{(b-a)x}{b-a}=x.\end{aligned}$$

Daher können wir benutzen, dass f konvex ist, und mit VI.2.1(i) wie folgt abschätzen

$$\begin{aligned}f(x)=f((1-\lambda)a+\lambda b) &\leqslant (1-\lambda)f(a)+\lambda f(b)\\ &= \Big(1-\frac{x-a}{b-a}\Big)f(a)+\frac{x-a}{b-a}\,f(b)\\ &= f(a)+\frac{-f(a)+f(b)}{b-a}(x-a)\\ &= f(a)+\frac{f(b)-f(a)}{b-a}(x-a).\end{aligned}$$

(ii) $\Longrightarrow$ (iii): Es seien wieder $a < x < b$ in I. Wir starten mit der Ungleichung in (ii) und subtrahieren erst $f(a)$ und teilen dann durch $(x-a)$. Das liefert die erste Ungleichung in (iii). Jetzt multiplizieren wir die Ungleichung aus (ii) mit -1 und addieren $f(b)$ auf beiden Seiten. Dies liefert

$$\begin{aligned}f(b)-f(x) &\geqslant f(b)-f(a)-\frac{f(b)-f(a)}{b-a}(x-a)\\ &= \frac{(f(b)-f(a))(b-a)-(f(b)-f(a))(x-a)}{b-a}\\ &= \frac{f(b)-f(a)}{b-a}(b-a-x+a)\\ &= \frac{f(b)-f(a)}{b-a}(b-x).\end{aligned}$$

Nun teilen wir beide Seiten der obigen Ungleichung durch $(b-x)$ und beachten, dass letzteres positiv ist. Dann folgt

$$\frac{f(b)-f(x)}{b-x}\geqslant\frac{f(b)-f(a)}{b-a}$$

und das ist gerade die zweite Ungleichung in (iii).

(iii) $\Longrightarrow$ (iv): Dies ist eine Spezialisierung, weil man lediglich den mittleren Term weglässt.

(iv) $\Longrightarrow$ (i): Es seien $a, b \in I$, $\lambda \in (0,1)$ und ohne Einschränkung sei $a < b$. Wir setzen $x := (1-\lambda)a+\lambda b$. Dann gilt $x = (1-\lambda)a+\lambda b < (1-\lambda)b+\lambda b = b$ und analog $x > (1-\lambda)a+\lambda a = a$, d.h. wir haben $a < x < b$

und können mit der Ungleichung aus (iv) beginnen:

$$\begin{aligned}
&& \frac{f(x)-f(a)}{x-a} &\leqslant \frac{f(b)-f(x)}{b-x} \\
&\Longrightarrow & (b-x)\big(f(x)-f(a)\big) &\leqslant (x-a)\big(f(b)-f(x)\big) \\
&\Longrightarrow & (b-x)f(x)+(x-a)f(x) &\leqslant (x-a)f(b)+(b-x)f(a) \\
&\Longrightarrow & \big(b-x+x-a\big)f(x) &\leqslant (x-a)f(b)+(b-x)f(a) \\
&\Longrightarrow & f(x) &\leqslant \frac{x-a}{b-a}f(b)+\frac{b-x}{b-a}f(a).
\end{aligned}$$

Jetzt müssen wir noch in den zwei Brüchen der letzten Zeile wieder $x=(1-\lambda)a+\lambda b$ einsetzen. Dies liefert

$$\frac{x-a}{b-a}=\frac{(1-\lambda)a+\lambda b-a}{b-a}=\frac{a-\lambda a+\lambda b-a}{b-a}=\lambda\frac{b-a}{b-a}=\lambda$$

und

$$\frac{b-x}{b-a}=\frac{b-(1-\lambda)a-\lambda b}{b-a}=\frac{(1-\lambda)b-(1-\lambda)a}{b-a}=(1-\lambda)\frac{b-a}{b-a}=1-\lambda,$$

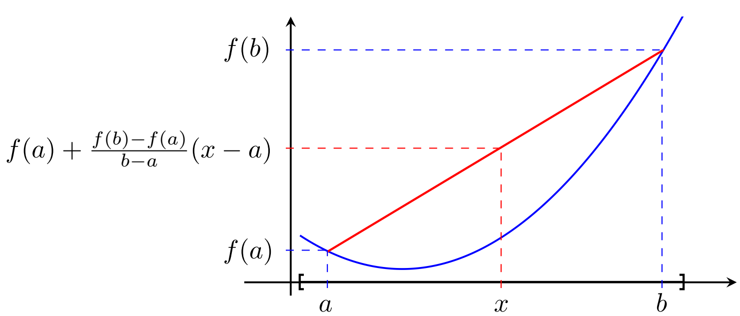
zusammen also

$$f((1-\lambda)a+\lambda b)=f(x)\leqslant\lambda f(b)+(1-\lambda)f(b)$$

was gerade die für (i) benötigte Ungleichung ist. □

Bemerkungen VI.2.4.

(i) VI.2.3(i) zeigt nochmal (und vielleicht etwas besser als unsere Erklärung in VI.2.2(iii)), dass f genau dann konvex ist, wenn für alle $a<b$ der Graph von $f|_{(a,b)}$ unterhalb der Verbindungsstrecke von $(a,f(a))$ nach $(b,f(b))$ liegt.

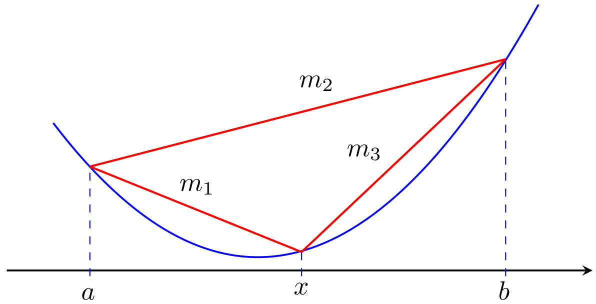


Diese Verbindungsstrecke kann mithilfe von VI.2.3(i) jetzt als Graph der (Einschränkung der) affin–linearen Funktion

$$(a,b)\to\mathbb{R},\quad x\mapsto f(a)+\frac{f(b)-f(a)}{b-a}(x-a)$$

in aus der Schule bekannter ‘Punkt-Steigungs-Form’ beschrieben werden.

(ii) VI.2.3(iii) bedeutet anschaulich, dass die Steigung der Verbindungsstrecke von $(a,f(a))$ nach $(x,f(x))$ kleiner ist als die der Verbindungsstrecke von $(a,f(a))$ nach $(b,f(b))$ und dass diese wiederum kleiner ist als die Steigung der Verbindungsstrecke von $(x,f(x))$ nach $(b,f(b))$.



Im Bild bezeichnen m_1, m_2, m_3 die Steigungen und dann besagt VI.2.3(iii), dass $m_1 \leqslant m_2 \leqslant m_3$ und VI.2.3(iv), dass $m_1 \leqslant m_3$ gilt.

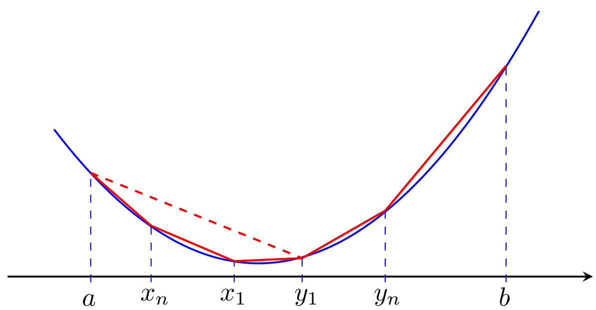
(iii) Da VI.2.3 also mit der Änderung von Steigungen zu tun hat, suggeriert dies, dass Konvexität etwas mit der (zweiten) Ableitung zu tun hat. Dies präzisieren wir im folgenden Satz.

Satz VI.2.5. *Es sei $I \subset \mathbb{R}$ ein Intervall und sei $f\colon I \to \mathbb{R}$ differenzierbar. Dann gilt:*

(i) *f ist genau dann [strikt] konvex, wenn f' [strikt] wächst.*

(ii) *f ist genau dann [strikt] konkav, wenn f' [strikt] fällt.*

Beweis. Für (i) '$\Longrightarrow$': Es sei f strikt konvex und $a < b$ in I. Wähle $(x_n)_{n\in\mathbb{N}} \subset I$ strikt fallend mit Grenzwert a, $(y_n)_{n\in\mathbb{N}} \subset I$ strikt wachsend mit Grenzwert b und derart, dass $x_1 < y_1$ gilt. Dann folgt mit VI.2.3(iii) bzw. VI.2.3(iv)



$$\frac{f(x_n)-f(a)}{x_n-a} < \frac{f(x_1)-f(a)}{x_1-a} < \frac{f(y_1)-f(x_1)}{y_1-x_1} < \frac{f(b)-f(y_n)}{b-y_n}$$

für jedes $n \in \mathbb{N}$. Da f in a und b differenzierbar ist, folgt durch Grenzübergang $n \to \infty$

$$f'(a) \leqslant \frac{f(x_1)-f(a)}{x_1-a} < \frac{f(y_1)-f(x_1)}{y_1-x_1} \leqslant f'(b).$$

Im Fall, dass f nur konvex (und nicht notwendig strikt konvex) ist, haben wir oben überall '$\leqslant$', d.h. aus $a < b$ folgt $f'(a) \leqslant f'(b)$.

Für (i) '$\Longleftarrow$': Es seien $a < x < b$ und I. Zweimalige Anwendung des Mittelwertsatzes liefert, dass es $\xi \in (a,x)$ und $\eta \in (x,b)$ gibt mit

$$\frac{f(x)-f(a)}{x-a} = f'(\xi) \quad \text{und} \quad \frac{f(b)-f(x)}{b-x} = f'(\eta).$$

Wenn nun f' strikt wächst, dann folgt $\frac{f(x)-f(a)}{x-a} < \frac{f(b)-f(x)}{b-x}$ also nach VI.2.3 strikte Konvexität. Wächst f lediglich (eventuell nicht strikt) so gilt die Ungleichung mit '$\leqslant$' und VI.2.3 zeigt, dass f immerhin konvex ist.

(ii) Da f' genau dann [strikt] fällt, wenn $-f' = (-f)'$ [strikt] wächst, folgt (ii) aus (i). □

Korollar VI.2.6. *Es sei $I \subset \mathbb{R}$ ein Intervall und $f\colon I \to \mathbb{R}$ sei zweimal differenzierbar. Dann gilt:*

(i) *f ist genau dann konvex, wenn $f''(x) \geqslant 0$ für alle $x \in I$ gilt.*

(ii) *Ist $f''(x) > 0$ für alle $x \in I$, dann ist f strikt konvex.*

Für Konkavität gelten analoge Aussagen.

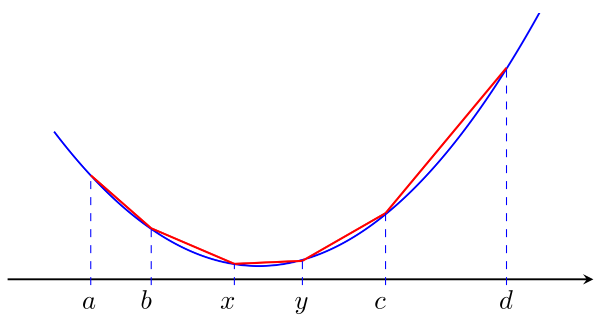
Beweis. Beide Aussagen folgen durch Kombination von VI.2.5 mit VI.1.11. □

Bemerkung VI.2.7. Die Umkehrung von VI.2.6(ii) gilt nicht, z.B. $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$.

Korollar VI.2.8. *(aus Lemma VI.2.3) Es sei $I \subset \mathbb{R}$ ein Intervall und sei $f\colon I \to \mathbb{R}$ konvex. Dann ist f in jedem Punkt $x \in I$, der kein Randpunkt ist, stetig. Für konkaves f gilt das gleiche.*

Beweis. Es sei $x \in I$ kein Randpunkt. Dann gibt es $a < b < x < c < d$ mit $[a, d] \subset I$. Wir behaupten, dass $f|_{[b,c]}$ gleichmäßig stetig ist. Das sieht formal stärker aus als das, was in der Folgerung steht, aber tatsächlich ist es das doch nicht, da auf kompakten Intervallen Stetigkeit und gleichmäßige Stetigkeit äquivalent sind, vgl. III.6.8.

Es seien also jetzt $x, y \in [b, c]$ mit $x \neq y$ beliebig gegeben. Ohne Einschränkung dürfen wir $x < y$ annehmen. Mit VI.2.3 erhalten wir



$$\frac{f(b) - f(a)}{b - a} \leqslant \frac{f(y) - f(x)}{y - x} \leqslant \frac{f(d) - f(c)}{d - c},$$

woraus $\left|\frac{f(y)-f(x)}{y-x}\right| \leqslant \max\left\{\left|\frac{f(c)-f(a)}{c-a}\right|, \left|\frac{f(d)-f(c)}{d-c}\right|\right\} =: K \geqslant 0$ folgt. Ist jetzt $\varepsilon > 0$ gegeben, so können wir $\delta := \frac{\varepsilon}{K+1}$ setzen. Für $x \neq y$ in $[b, c]$ mit $|x - y| < \delta$ gilt dann

$$|f(y) - f(x)| \leqslant K|x - y| < K\delta = \frac{K\,\varepsilon}{K + 1} \leqslant \varepsilon$$

und folglich ist $f|_{[b,c]}$ gleichmäßig stetig. □

Bemerkung VI.2.9. In Randpunkten können konvexe Funktionen durchaus unstetig sein, z.B. ist

$$f\colon [0, 1] \to \mathbb{R},\quad f(x) = \begin{cases} 0, & x < 1, \\ 1, & x = 1, \end{cases}$$

konvex!

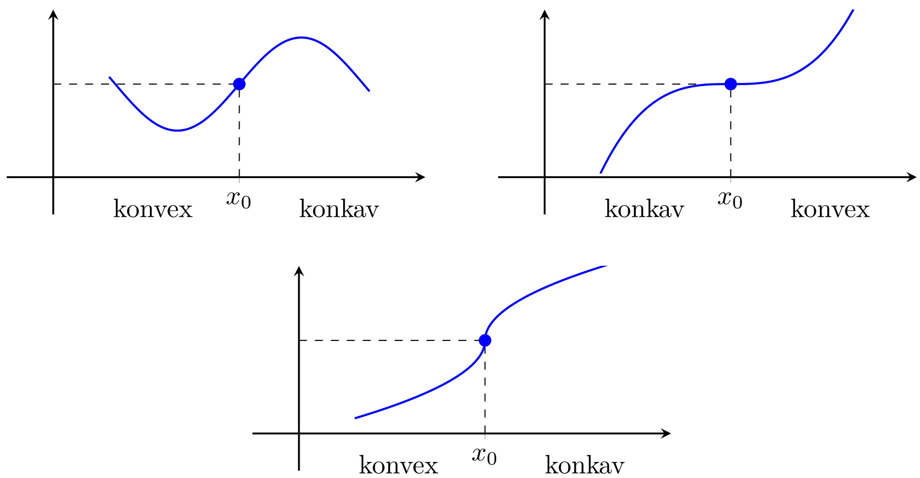
VI.3. Wendepunkte

Definition VI.3.1. Es sei $I \subset \mathbb{R}$ ein Intervall und $f\colon I \to \mathbb{R}$ sei stetig.

(i) Ein Punkt $x_0 \in I$ (manchmal auch $(x_0, f(x_0)) \in I \times \mathbb{R}$) heißt *Wendepunkt* von f, falls $\varepsilon > 0$ existiert mit $(x_0 - \varepsilon, x_0 + \varepsilon) \subset I$, sodass $f|_{(x_0-\varepsilon,x_0)}$ konvex und $f|_{(x_0,x_0+\varepsilon)}$ konkav ist, oder umgekehrt.

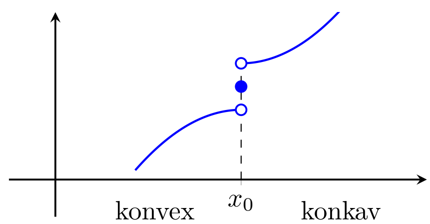
(ii) Ist x_0 ein Wendepunkt, f differenzierbar in x_0 und $f'(x_0) = 0$, so nennt man x_0 [oder $(x_0, f(x_0))$] einen *Sattelpunkt*.

Die Bilder unten zeigen einen Wendepunkt, der kein Sattelpunkt ist, dann einen Sattelpunkt und als letztes einen Wendepunkt, an dem die Abbildung nicht differenzierbar ist.



Bemerkungen VI.3.2.

(i) In VI.3.1 hätte es genügt vorauszusetzen, dass f in x_0 stetig ist, denn die Stetigkeit von $f|_{(x_0-\varepsilon,x_0)}$ und $f|_{(x_0,x_0+\varepsilon)}$ folgt aus der Konvexität bzw. Konkavität mit VI.2.8. Durch die Forderung der Stetigkeit in x_0 schließen wir Beispiele des folgenden Typs als Wendepunkt aus:



(ii) Der Begriff (wie auch z.B. 'lokales Maximum/Minimum') hat gewisse Pathologien, z.B. ist für konstantes $f\colon \mathbb{R} \to \mathbb{R}$ jeder Punkt $x_0 \in \mathbb{R}$ Wendepunkt (wie auch lokales Maximum und lokales Minimum). Dies könnte man beheben, indem man in VI.3.1 strikte Konvexität bzw. Konkavität fordert. In einer Übungsaufgabe werden wir allerdings eine Funktion sehen, die nicht konstant ist, aber trotzdem an einer Stelle x_0 ein Minimum und gleichzeitig einen Wendepunkt hat.

(iii) Es gibt einige weitere Begriffe (z.B. 'Flachpunkte') und in der Literatur sind Definitionen nicht einheitlich. Z.B. wird oft definiert

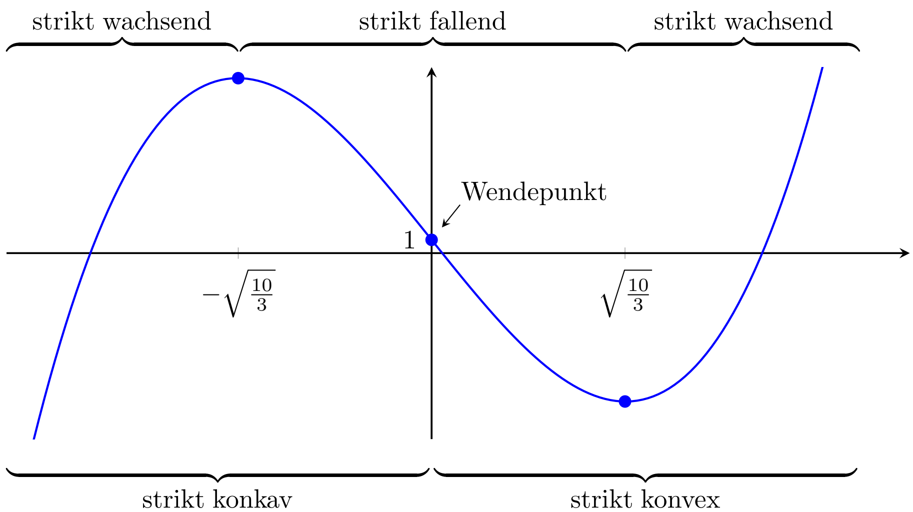
$$\text{‘}x_0 \text{ Wendepunkt} :\Longleftrightarrow f''(x_0) = 0 \text{ und } f'' \text{ wechselt das Vorzeichen’}.$$

Dies setzt erstens voraus, dass f zweimal differenzierbar ist, was wir nicht vorausgesetzt haben, und 'Vorzeichenwechsel' kann man so verstehen, dass $f''|_{(x_0-\varepsilon,x_0)} > 0$ und $f''|_{(x_0,x_0+\varepsilon)} < 0$ oder umgekehrt gelten soll, was strikter Konvexität bzw. strikter Konkavität entspricht und daher stärker ist als was in VI.3.1 definiert wurde.

Beispiel VI.3.3. ('Kurvendiskussion') Wir untersuchen $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3 - 10x + 1$.

- Da f eine Polynomfunktion ist, ist klar, dass $f \in \mathrm{C}^\infty(\mathbb{R})$ liegt. Die erste und zweite Ableitung können wir mit den üblichen Regeln bestimmen: $f'(x) = 3x^2 - 10$, $f''(x) = 6x$.
- Als nächstes stellen wir fest, dass $f'(x) = 0$ genau dann gilt, wenn $x = \pm\sqrt{10/3}$ ist und dass $f''(+\sqrt{10/3}) > 0$ sowie $f''(-\sqrt{10/3}) < 0$ gilt. Daraus können wir mit VI.1.13 schließen, dass f in $x_1 := -\sqrt{10/3}$ ein striktes lokales Maximum und in $x_2 := \sqrt{10/3}$ ein striktes lokales Minimum hat.

- Weiter sehen wir, dass $f'(x) < 0$ genau dann gilt, wenn $x_1 < x < x_2$. Damit ist f mit VI.1.11(i) auf $(-\infty, x_1]$ strikt wachsend[2], auf $[x_1, x_2]$ strikt fallend und auf $[x_2, \infty)$ strikt wachsend.
- Schließlich ist $f''(x) < 0$ für $x < 0$ und $f''(x) > 0$ für $x > 0$. Damit ist f auf $(-\infty, 0]$ strikt konkav und auf $[0, \infty)$ strikt konvex. Insbesondere liegt in $x_3 = 0$ ein Wendepunkt vor der wegen $f'(x_3) = -10 \neq 0$ kein Sattelpunkt ist.
- Anhand der Funktion sieht man, dass $\lim_{x\to\infty} f(x) = +\infty$ und $\lim_{x\to-\infty} f(x) = -\infty$ gilt; f ist also nach oben und unten unbeschränkt. Wegen des Zwischenwertsatzes ist $\operatorname{Bild} f = \mathbb{R}$ und die bestimmten lokalen Extrema sind nicht global.
- Um ein einigermaßen vernünftiges Bild von f zu malen, müssen nun noch einige Punkte des Graphen explizit ausgerechnet werden. Wir sehen, dass offenbar $f(x_3) = f(0) = 1$ gilt, aber dass sowohl $f(x_1)$ als auch $f(x_2)$ ohne ein Computeralgebrasystem (CAS) eher schwierig auszurechnen sind. Was man ohne CAS sehen kann, ist allerdings, dass $f(x_1) > 1$ sein muss, denn wenn dem nicht so wäre, dann könnte $f|_{(x_1,0)}$ nicht strikt fallend sein. Mit einem ähnlichen Argument kann man $f(x_2)$ abschätzen: Wir berechnen $f(1) = 1^3 - 10 \cdot 1 + 1 = -8$. Da $f|_{[x_1,x_2]}$ strikt fallend ist und $1 \in (x_1, x_2)$ liegt, muss $f(x_2) < f(1) = -8$ sein.
- Die Nullstellen von f sind ebenfalls schwierig explizit auszurechnen. Aus dem bisherigen können wir aber mithilfe des Zwischenwertsatzes schließen, dass es eine Nullstelle echt kleiner als x_1 geben muss, eine zwischen 0 und x_2 und eine weitere echt größer als x_2 geben muss. Da f wegen des Fundamentalsatzes der Algebra aber nicht mehr als drei komplexe Nullstellen (also auch nicht mehr als drei reelle) haben kann, folgt, dass dies alle Nullstellen von f sind.
- Daraus ergibt sich qualitativ folgendes Bild:



[2] In VI.1.11(i) haben wir nur beschränkte Intervalle betrachtet, aber wir können hier, wenn $y_1 < y_2$ in $(-\infty, x_1]$ gegeben sind, die Einschränkung $f|_{[y_1,y_2]}$ betrachten. Diese ist dann mit VI.1.11(i) strikt wachsend, also gilt $f(y_1) < f(y_2)$ wie gewünscht.

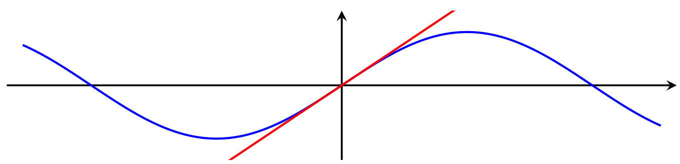
KAPITEL VII

Die Regel von L'Hôpital

Das Ziel dieses Kapitels ist es, eine Methode zu finden, mit der zum Beispiel die Grenzwerte

$$\lim_{x\to 0}\frac{\sin x}{x},\quad \lim_{x\to 0}\frac{\cos x-1}{x}\quad \text{oder}\quad \lim_{x\to\infty}\frac{\sqrt{x}}{\log x},$$

einfach berechnet werden können. Betrachtet man im Fall des ersten Grenzwerts oben die Funktionen $x\mapsto \sin x$ und $x\mapsto x$ in der Nähe von Null,



so kommt man auf die Idee, dass der Grenzwert 1 ist (wie wir aus dem Beweis von V.3.7 wissen), weil beide Funktionen in Null die gleiche Steigung haben, und vermutet, dass es möglich sein könnte Grenzwerte wie oben durch Differenzieren von Zähler und Nenner zu bestimmen.

Satz VII.0.1. *(Regel von L'Hôpital) Es seien $-\infty\leqslant a<b\leqslant+\infty$ und $f,g\colon(a,b)\to\mathbb{R}$ differenzierbar. Es gelte*

(i) $\lim\limits_{x\to b}f(x)=\lim\limits_{x\to b}g(x)=0$ *oder* $\lim\limits_{x\to b}f(x)=\lim\limits_{x\to b}g(x)=\infty$,

(ii) $g'(x)\neq 0$ *für alle* $x\in(a,b)$,

(iii) $\lim\limits_{x\to b}\dfrac{f'(x)}{g'(x)}=c\in\overline{\mathbb{R}}\ \big(=\mathbb{R}\cup\{+\infty,\,-\infty\}\big)$ *existiert.*

Dann existiert $x_0\in[a,b)$, sodass $g(x)\neq 0$ ist für $x\in(x_0,b)$ und es gilt

$$\lim_{x\to b}\frac{f(x)}{g(x)}=\lim_{x\to b}\frac{f'(x)}{g'(x)}.$$

Bemerkung VII.0.2. Eine analoge Aussage gilt für $x\to a$. Im Fall $\lim_{x\to 0}\frac{\sin x}{x}$ müssen wir L'Hôpital dann erst für $\frac{\sin x}{x}|_{(-\pi/2,0)}$ und dann für $\frac{\sin x}{x}|_{(0,\pi/2)}$ anwenden. Insbesondere muss (ii) nur in der Nähe von b gelten, weil wir das Intervall anpassen können.

Beweis. (von VII.0.1)

(1) Wir behandeln zuerst den Fall, bei dem $\lim\limits_{x\to b}f(x)=\lim\limits_{x\to b}g(x)=0$ ist. Wir nehmen an, dass $c:=\lim_{x\to b}f'(x)/g'(x)\in\mathbb{R}$ gilt. Wegen (ii) und dem Satz von Rolle muss $g\colon(a,b)\to\mathbb{R}$ injektiv sein: Gäbe es $x_1\neq x_2$ mit $g(x_1)=g(x_2)$, so hätte g' eine Nullstelle zwischen x_1 und x_2.

Damit ist g dann strikt wachsend oder strikt fallend. Ohne Einschränkung nehmen wir an, dass g fällt. Sonst ersetzen wir g durch $-g$. Wegen $\lim_{x\to b}g(x)=0$ muss dann $g(x)>0$ gelten für alle $x\in(a,b)$. Es folgt, dass

$$\frac{f}{g}\colon(a,b)\to\mathbb{R}$$

wohldefiniert ist und wir können $\lim_{x\to b}\frac{f(x)}{g(x)}$ untersuchen.

Wir setzen nun $A := \lim_{x\to a} g(x) \in \mathbb{R}_{>0}\cup\{\infty\}$. Dann ist $\mathrm{Bild}(g) = (0, A)$ und daher $g\colon (a, b) \to (0, A)$ bijektiv. Wir definieren $\psi := g^{-1}\colon (0, A) \to (a, b)$ und betrachten

$$F := f \circ \psi\colon (0, A) \to \mathbb{R}.$$

Dann gilt mit $y \in (0, A)$:

$$\text{(a)} \qquad \lim_{y\to 0} F(y) = \lim_{y\to 0} f(\psi(y)) = \lim_{x\to b} f(x) = 0,$$

wobei die letzte Gleichung nach Voraussetzung (i) gilt, und wir für die vorletzte benutzt haben, dass die Implikation '$y_n \to 0 \Rightarrow \psi(y_n) \to b$' für jede Folge $(y_n)_{n\in\mathbb{N}} \subset (0, A)$ gilt.

Wir berechnen

$$\text{(b)} \qquad \lim_{y\to 0} F'(y) = \lim_{y\to 0} f'(\psi(y)) \cdot \psi'(y) = \lim_{y\to 0} f'(\psi(y)) \frac{1}{g'(\psi(y))} = \lim_{x\to b} \frac{f'(x)}{g'(x)},$$

wobei wir erst die Kettenregel, dann die Umkehrregel, dann wieder die Implikation '$y_n \to 0 \Rightarrow \psi(y_n) \to b$' und im letzten Schritt Voraussetzung (iii) benutzt haben. Die Umkehrregel ist wegen Voraussetzung (ii) anwendbar.

Falls die in der folgenden Rechnung vorkommenden Grenzwerte existieren, erhalten wir

$$\text{(c)} \qquad \lim_{x\to b} \frac{f(x)}{g(x)} = \lim_{x\to b} \frac{f(\psi(g(x)))}{g(x)} = \lim_{y\to 0} \frac{f(\psi(y))}{y} = \lim_{y\to 0} \frac{F(y)}{y},$$

wobei nun die Implikation '$x_n \to b \Rightarrow g(x_n) \to 0$' für jede Folge $(x_n)_{n\in\mathbb{N}} \subset (a, b)$ benutzt wurde.

Wir zeigen jetzt aber, dass (a) und (b) implizieren, dass $\lim_{y\to 0} F(y)/y = c$ gilt. Dann zeigt (c) tatsächlich, dass $\lim_{x\to b} f(x)/g(x)$ existiert und gleich c ist, und wir sind fertig.

Wegen (a) können wir $F\colon [0, A) \to \mathbb{R}$ durch $F(0) := 0$ stetig fortsetzen.

Fall 1: Ist $c = 0$, dann impliziert (b), dass $\lim_{y\to 0} F'(y) = 0$ gilt. Daraus folgt, indem wir die Definition des Grenzwertes benutzen, und dann die Folgerung VI.1.7 aus dem Mittelwertsatz mit $a := 0$, $b := y_0$, $m := -\varepsilon$, $M := +\varepsilon$, $x_1 := 0$ und $x_2 := y$ anwenden und dabei beachten, dass für $a, b \in \mathbb{R}$ die Schlussfolgerung auch für $x_1, x_2 \in [a, b]$ gilt:

$$\begin{aligned}
&\forall\, \varepsilon > 0\ \exists\, y_0 \in (0, A)\ \forall\, y \in (0, y_0]\colon |F'(y)| < \varepsilon \\
\Longrightarrow\ &\forall\, \varepsilon > 0\ \exists\, y_0 \in (0, A)\ \forall\, y \in (0, y_0]\colon -\varepsilon(y - 0) \leqslant F(y) - F(0) \leqslant \varepsilon(y - 0) \\
\Longrightarrow\ &\forall\, \varepsilon > 0\ \exists\, y_0 \in (0, A)\ \forall\, y \in (0, y_0]\colon |F(y)| \leqslant \varepsilon y \\
\Longrightarrow\ &\forall\, \varepsilon > 0\ \exists\, y_0 \in (0, A)\ \forall\, y \in (0, y_0]\colon \left|\tfrac{F(y)}{y}\right| \leqslant \varepsilon
\end{aligned}$$

Letzteres bedeutet aber gerade $\lim_{y\to\infty} F(y)/y = 0 = c$ wie gewünscht.

Fall 2: Den Fall $c \neq 0$ führen wir auf den ersten Fall zurück. Dazu definieren wir $\tilde{F}\colon (0, A) \to \mathbb{R}$, $\tilde{F}(y) := F(y) - cy$. Dann gelten

$$\lim_{y\to 0} \tilde{F}(y) = 0 \quad \text{und} \quad \lim_{y\to 0} \tilde{F}'(y) = \lim_{y\to 0} F'(y) - c = c - c = 0$$

wegen (a) bzw. wegen (b). Daher können wir Fall 1 auf $\tilde{F}$ anwenden und erhalten $\lim_{y\to 0} \tilde{F}(y)/y = 0$ und damit

$$\lim_{y\to 0} \frac{F(y)}{y} = \lim_{y\to 0} \frac{\tilde{F}(y) + cy}{y} = 0 + c$$

was den Fall $\lim\limits_{x\to b} f(x) = \lim\limits_{x\to b} g(x) = 0$ und $c = \lim_{x\to b} f'(x)/g'(x) \in \mathbb{R}$ abschließt.

(2) Wir nehmen nun an, dass $\lim\limits_{x\to b} f(x) = \lim\limits_{x\to b} g(x) = 0$ ist, aber nun sei $c = \infty$. Es gilt also

$$\lim_{x\to b} \frac{f'(x)}{g'(x)} = \infty.$$

Die natürliche Idee ist, dass wir diesen Fall durch Kehrwertbildung auf den erstem Fall zurückführen. Dazu bemerken wir zuerst, dass ein $x_0 \in (a, b)$ existieren muss, sodass $f'(x) \neq 0$ für alle $x \geqslant x_0$ gilt (andernfalls gäbe es $(x_n)_{n\in\mathbb{N}} \subset (a,b)$ mit $x_n \to b$ sowie $\frac{f'(x_n)}{g'(x_n)} = 0$ für alle $n \in \mathbb{N}$ und der Quotient könnte nicht gegen $+\infty$ konvergieren). Wir beachten, dass wir nach Voraussetzung auch wissen, dass $g'(x) \neq 0$ für alle $x \in (x_0, b)$ gilt, dass also f' und g' in (x_0, b) keine Nullstellen haben.

Obiges bedeutet insbesondere, dass $\frac{g'}{f'}\colon (x_0, b) \to \mathbb{R}$ wohldefiniert ist und dass

$$\lim_{x\to b} \frac{g'(x)}{f'(x)} = 0$$

gilt. Damit sind alle drei Voraussetzungen (i)–(iii), für g/f statt f/g, und mit endlichem Grenzwert von g'/f', erfüllt. Es folgt also aus (1), eventuell nach Vergrößern von x_0, dass $f(x) \neq 0$ für $x \in (x_0, b)$ und dass

$$\lim_{x\to b} \frac{g(x)}{f(x)} = 0$$

gilt.

Um jetzt hieraus zu schließen, dass der Kehrwert für $x \to b$ gegen $+\infty$ konvergiert, behaupten wir, dass $\frac{g(x)}{f(x)} > 0$ für alle $x \in (x_0, b)$ gilt. In der Tat folgt aus $f'(x) \neq 0$ für $x \in (x_0, b)$ analog unserer Argumentation am Beginn von (1) für g, dass $f\colon (x_0, b) \to \mathbb{R}$ entweder strikt wächst oder strikt fällt, und deswegen, da außerdem $\lim_{x\to b} f(x) = 0$ gilt, entweder auf ganz (x_0, b) strikt positiv oder auf ganz (x_0, b) strikt negativ sein muss. Da das gleiche für g gilt, haben also f und g entweder auf ganz (x_0, b) gleiche Vorzeichen oder auf ganz (x_0, b) verschiedene Vorzeichen. Letzteres geht aber nicht, denn dann wäre eine der Funktionen strikt wachsend (also Ableitung positiv nach VI.1.11 und dann strikt positiv, da wir bereits wissen, dass f' und g' keine Nullstellen in (x_0, b) haben) und die andere strikt fallend (also Ableitung strikt negativ), im Widerspruch dazu, dass der Quotient der Ableitungen ja gegen $+\infty$ strebt. Es folgt also, dass $\frac{g(x)}{f(x)} > 0$ für alle $x \in (x_0, b)$ ist und wir bekommen schließlich

$$\lim_{x\to b} \frac{f(x)}{g(x)} = \lim_{x\to b} \left(\frac{g(x)}{f(x)}\right)^{-1} = \infty.$$

Den Fall, dass $c = -\infty$ ist, behandelt man analog.

(3) Jetzt sei wieder $c \in \mathbb{R}$, aber wir behandeln den Fall $\lim_{x\to b} f(x) = \infty = \lim_{x\to b} g(x)$. Wir beginnen damit, wie in (1) festzustellen, dass $g' \neq 0$ auf (a, b) sein muss, und daher g entweder strikt wächst oder strikt fällt. Letzteres geht aber diesmal nicht, da wir voraussetzen, dass $\lim_{x\to b} g(x) = \infty$ ist. Also muss g strikt wachsen, wir können $B := \lim_{x\to a} g(x) \in \{-\infty\} \cup \mathbb{R}$ definieren und erhalten, dass $g\colon (a, b) \to (B, \infty)$ bijektiv ist. Wir definieren $\psi := g^{-1}\colon (B, \infty) \to (a, b)$, betrachten

$$F := f \circ \psi\colon (B, \infty) \to \mathbb{R}$$

und erhalten für $y \in (B, \infty)$ mit Kettenregel, Umkehrregel und Voraussetzung (ii), dass gilt

$$\text{(b)} \qquad \lim_{y\to\infty} F'(y) = \lim_{y\to\infty} \frac{f'(\psi(y))}{g'(\psi(y))} = \lim_{x\to b} \frac{f'(x)}{g'(x)} = c.$$

Nach der Definition von ψ und F gilt

$$\text{(c)} \qquad \lim_{x\to b} \frac{f(x)}{g(x)} = \lim_{y\to\infty} \frac{f(\psi(y))}{y} = \lim_{y\to\infty} \frac{F(y)}{y}.$$

Es genügt nun aus (b) zu folgern, dass $\lim_{y\to\infty} F(y)/y = c$ ist. Hierzu nehmen wir wieder erstmal an, dass $c = 0$ ist. Dann gilt

$$\begin{aligned} &\forall\, \varepsilon > 0\ \exists\, y_0 \in (B, \infty)\ \forall\, y \in [y_0, \infty)\colon |F'(y)| < \varepsilon \\ \implies\ &\forall\, \varepsilon > 0\ \exists\, y_0 \in (B, \infty)\ \forall\, y \in [y_0, \infty)\colon |F(y) - F(y_0)| < \varepsilon |y - y_0| \end{aligned}$$

wieder unter Benutzung von VI.1.7. Für $\varepsilon > 0$ wähle nun y_0 wie oben. Für $y \geqslant \max\{y_0, |F(y_0)|/\varepsilon\} > 0$ folgt dann mit $\frac{|y-y_0|}{y} = \frac{y-y_0}{y} \leqslant 1$ und $|F(y_0)| \leqslant \varepsilon y$, dass gilt

$$\left|\frac{F(y)}{y}\right| \leqslant \frac{|F(y)-F(y_0)|}{y} + \frac{|F(y_0)|}{y} \leqslant \varepsilon\frac{|y-y_0|}{y} + \frac{|F(y_0)|}{y} \leqslant 2\varepsilon$$

also

$$\lim_{y\to\infty} \frac{F(y)}{y} = 0 = c.$$

Im Fall, dass $c \neq 0$ ist, betrachtet man wie in (1) die Hilfsfunktion $\tilde{F}(y) = F(y) - cy$.

(4) Für den Fall, dass $\lim_{x\to b} f(x) = \infty = \lim_{x\to b} g(x)$ ist und $c = \pm\infty$, geht man wie in (2) vor und führt dies auf (3) zurück durch Kehrwertbildung.

□

Wir schreiben im Folgenden manchmal $\frac{0}{0}$ beziehungsweise $\frac{\infty}{\infty}$, um anzudeuten, welchen der L'Hôpital'schen Fälle wir benutzen.

Beispiele VII.0.3.

(1) Es gilt

$$\lim_{x\to 0} \frac{e^x - 1}{x} \overset{\text{‘}\frac{0}{0}\text{’}}{=} \lim_{x\to 0} \frac{(e^x-1)'}{x'} = \lim_{x\to 0} \frac{e^x}{1} = e^0 = 1,$$

wobei wir den ‘$\frac{0}{0}$’–Fall der Regel von L'Hôpital anwenden können, weil der Grenzwert des Quotienten rechts existiert und die Ableitung des Nenners ungleich Null ist.

(2) Es gilt

$$\lim_{x\to\infty} \frac{\log x}{x} \overset{\text{‘}\frac{\infty}{\infty}\text{’}}{=} \lim_{x\to\infty} \frac{(\log x)'}{x'} = \lim_{x\to\infty} \frac{1/x}{1} = 0,$$

wobei wir wieder den ‘$\frac{\infty}{\infty}$’–Fall der Regel von L'Hôpital anwenden können, da der Grenzwert des Quotienten rechts existiert und die Ableitung des Nenners ungleich Null ist.

(3) Es gilt

$$\lim_{x\to\infty} \frac{x^2}{e^x} \overset{\text{‘}\frac{\infty}{\infty}\text{’}}{=} \lim_{x\to\infty} \frac{(x^2)'}{(e^x)'} = \lim_{x\to\infty} \frac{2x}{e^x} \overset{\text{‘}\frac{\infty}{\infty}\text{’}}{=} \lim_{x\to\infty} \frac{(2x)'}{(e^x)'} = \lim_{x\to\infty} \frac{2}{e^x}.$$

Wir beachten hierbei, dass die Rechnung nun am besten von rechts nach links gelesen werden sollte, denn die erste Gleichung ist ja erst gültig, wenn alle drei Voraussetzungen von L'Hôpital erfüllt sind — da wir aber wieder auf einen Grenzwert des Typs ‘$\frac{\infty}{\infty}$’ kommen, können wir dies nicht direkt sehen. Erst nachdem wir nochmal abgeleitet haben, kommt ein Bruch heraus, bei dem wir sehen, dass der Grenzwert existiert. Demnach ist die zweite Anwendung von L'Hôpital gerechtfertigt, der Grenzwert in der Mitte existiert und dann ist auch die erste Anwendung von L'Hôpital möglich, da auch hier nun alle drei Voraussetzungen erfüllt sind.

(4) Es gilt

$$\lim_{x\searrow 0} x^x = \lim_{x\searrow 0} e^{x\log x} = e^0 = 1,$$

weil exp stetig ist und weil wir für den Exponenten den ‘$\frac{\infty}{\infty}$’–Fall von L'Hôpital anwenden können:

$$\lim_{x\searrow 0} x\log x = \lim_{x\searrow 0} \frac{\log x}{1/x} = \lim_{x\searrow 0} \frac{(\log x)'}{(1/x)'} = \lim_{x\searrow 0} \frac{1/x}{-1/x^2} = \lim_{x\searrow 0}(-x) = 0.$$

Bemerkung VII.0.4. VII.0.1 Man kann andere Fälle durch Umformungen auf die Fälle ‘$\frac{0}{0}$’ und ‘$\frac{\infty}{\infty}$’ zurückführen. Konvergiert $f(x)$ zum Beispiel gegen 0 und $g(x)$ gegen ∞, so kommen wir mit

$$\underbrace{f(x)\cdot g(x)}_{\text{‘}0\cdot\infty\text{’}} = \underbrace{\frac{f(x)}{\frac{1}{g(x)}}}_{\text{‘}\frac{0}{0}\text{’}}$$

wieder in einen Standardfall zurück.

Seien Sie bitte vorsichtig, dass Sie L'Hôpital nur anwenden, wenn die Voraussetzungen des Satzes erfüllt sind. So ist

$$\lim_{x \to 1} \frac{x^2}{4x} = \frac{1}{4},$$

aber

$$\lim_{x \to 1} \frac{2x}{4} = \frac{2}{4} = \frac{1}{2}.$$

KAPITEL VIII

Das Riemann-Integral

Im gesamten Kapitel seien a und b reelle Zahlen mit $a < b$.

VIII.1. Treppenfunktionen

Definition VIII.1.1. Eine Abbildung $\varphi\colon [a,b] \to \mathbb{R}$ heißt eine *Treppenfunktion*, falls eine *Partition* oder *Zerlegung* $a = x_0 < x_1 < \cdots < x_n = b$ des Intervalls $[a,b]$ existiert, sodass $\varphi|_{(x_{k-1},x_k)}$ für jedes $k = 1,\ldots,n$ konstant ist. Wir setzen

$$\mathrm{T}[a,b] := \big\{\varphi\colon [a,b] \to \mathbb{R} \;\big|\; \varphi \text{ ist Treppenfunktion}\big\}$$

und nennen $\mathrm{T}[a,b]$ den *Raum der Treppenfunktionen.*

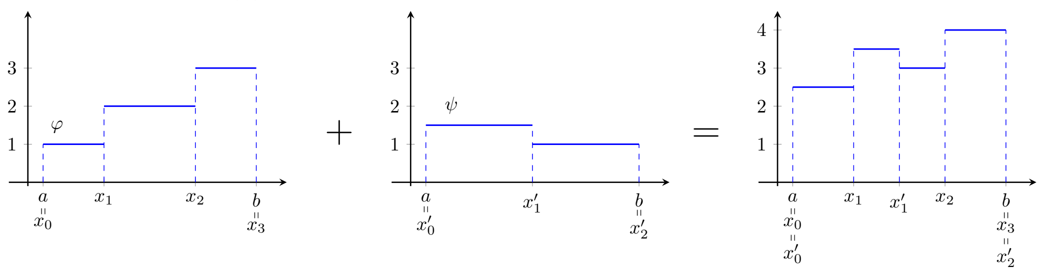
Proposition VIII.1.2. Die Menge $\mathrm{T}[a,b] \subset \{f\colon [a,b] \to \mathbb{R}\}$ ist ein Untervektorraum des $\mathbb{R}$-Vektorraums aller Funktionen von $[a,b]$ nach $\mathbb{R}$. Insbesondere gilt:

(i) $0 \in \mathrm{T}[a,b]$.

(ii) $\varphi, \psi \in \mathrm{T}[a,b] \implies \varphi + \psi \in \mathrm{T}[a,b]$,

(iii) $\varphi \in \mathrm{T}[a,b]$, $\lambda \in \mathbb{R} \implies \lambda\varphi \in \mathrm{T}[a,b]$,

Ein Beispiel zum Punkt (ii) ist:



Beweis. (i) und (iii) sind trivial.

(ii) Was wir brauchen ist eine *gemeinsame Verfeinerung der Partitionen*: Es sei $a = x_0 < x_1 < \cdots < x_n = b$ eine Partition für φ und $a = x'_0 < x'_1 < \cdots < x'_m = b$ eine Partition für ψ. Wir wählen nun $a = t_0 < t_1 < \cdots t_\ell = b$, sodass

$$\{t_0, t_1, \ldots, t_\ell\} = \{x_0, x_1, \ldots, x_n\} \cup \{x'_0, x'_1, \ldots, x'_m\}$$

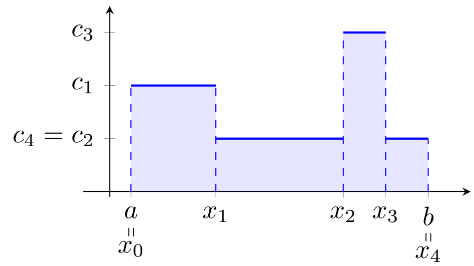
gilt. Dann sind φ und ψ auf allen Intervallen (t_{j-1}, t_j) für $j = 1, \ldots, \ell$ jeweils konstant und daher ist $\varphi + \psi$ dort ebenfalls konstant. D.h. $\varphi + \psi \in \mathrm{T}[a,b]$. □

Definition VIII.1.3. (Integral einer Treppenfunktion) Es sei $\varphi \in \mathrm{T}[a,b]$ gegeben durch die Partition $a = x_0 < x_1 < \cdots < x_n = b$ mit $\varphi|_{(x_{k-1},x_k)} = c_k$ für $k = 1,\ldots,n$. Dann definieren wir

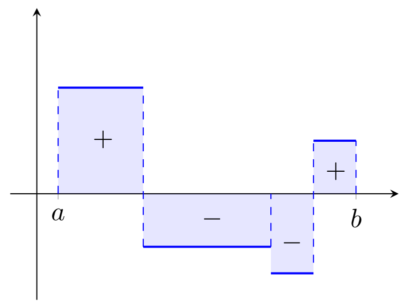
$$\int_a^b \varphi(x)\,\mathrm{d}x := \sum_{k=1}^{n} c_k(x_k - x_{k-1}).$$

Bemerkungen VIII.1.4.

- Wir haben die folgende Interpretation:

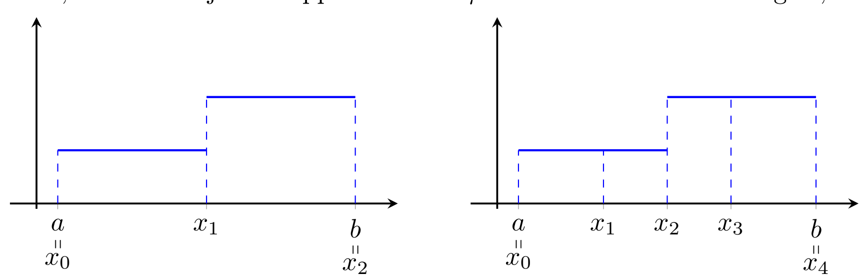


Falls $\varphi(x) \geqslant 0$ für alle $x \in [a,b]$ (oder zumindest $c_k \geqslant 0$ in der Notation von VIII.1.3) gilt, dann kann $\int_a^b \varphi(x)\,\mathrm{d}x$ als Fläche zwischen der x–Achse und dem Graphen von φ interpretiert werden.



Falls φ auf Teilintervallen der Partition negativ ist, dann zählen wir die entsprechenden Flächen mit negativem Vorzeichen.

- In VIII.1.3 definieren wir $\int_a^b \varphi(x)\,\mathrm{d}x$ für $\varphi \in \mathrm{T}[a,b]$, indem wir eine Partition auswählen — es ist aber natürlich so, dass es für jede Treppenfunktion φ stets mehre Partitionen gibt, z.B.



Es muss also gezeigt werden, dass $\int_a^b \varphi(x)\,\mathrm{d}x$ unabhängig von der Wahl der Partition ist, um zu garantieren, dass $\int_a^b \varphi(x)\,\mathrm{d}x$ wohldefiniert ist. Es seien dazu

$$Z\colon a = x_0 < x_1 < \cdots < x_n = b \quad \text{und} \quad Z'\colon a = t_0 < t_1 < \cdots < t_m = b$$

Partitionen für $\varphi \in \mathrm{T}[a,b]$ mit $\varphi|_{(x_{i-1},x_i)} = c_i$ und $\varphi|_{(t_{j-1},t_j)} = c'_j$ für $i = 1,\dots,n$ und $j = 1,\dots,m$. Nur für den Moment definieren wir

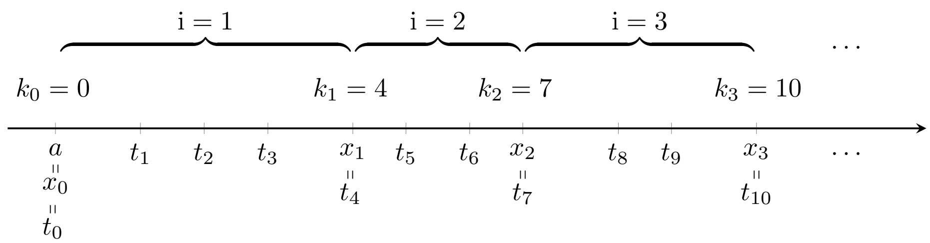
$$\int_Z \varphi := \sum_{i=0}^{n} c_i(x_i - x_{i-1}), \quad \int_{Z'} \varphi := \sum_{j=0}^{m} c'_j(t_j - t_{j-1})$$

und behaupten $\int_Z \varphi = \int_{Z'} \varphi$.

(1) Wir nehmen zunächst an, dass jeder Punkt von Z auch in Z' vorkommt. D.h. für jedes i gibt es ein k_i mit $x_i = t_{k_i}$, und wir haben also

$$x_{i-1} = t_{k_{i-1}} < t_{k_{i-1}+1} < \cdots < t_{k_i} = x_i,$$

woraus folgt $c'_j = c_i$ für $k_{i-1} < j \leqslant k_i$ und $i = 1,\dots,n$.



Nun erhalten wir mithilfe einer Teleskopsumme für die vierte Gleichung:

$$\begin{aligned}\int_{Z'} \varphi &= \sum_{j=1}^{m} c'_j(t_j - t_{j-1}) = \sum_{i=1}^{n} \sum_{j=k_{i-1}+1}^{k_i} c_i(t_j - t_{j-1}) \\ &= \sum_{i=1}^{n} c_i(t_{k_i} - t_{k_{i-1}-1+1}) = \sum_{i=1}^{n} c_i(x_i - x_{i-1}) = \int_Z \varphi.\end{aligned}$$

(2) Falls Z und Z' beliebige Partitionen sind, wähle Z'', sodass alle Punkte von Z in Z'' vorkommen und auch alle Punkte von Z' in Z'' vorkommen (gemeinsame Verfeinerung wie im Beweis von VIII.1.2(ii)). Dann kann zweimal (1) benutzt werden, um zu sehen, dass

$$\int_{Z'} \varphi = \int_{Z''} \varphi = \int_Z \varphi$$

gilt.

Satz VIII.1.5. *Es seien* φ, $\psi \in \mathrm{T}[a, b]$ *und* $\lambda \in \mathbb{R}$. *Dann gilt*

(i) $\displaystyle\int_a^b (\varphi + \psi)(x)\,\mathrm{d}x = \int_a^b \varphi(x)\,\mathrm{d}x + \int_a^b \psi(x)\,\mathrm{d}x$,

(ii) $\displaystyle\int_a^b (\lambda\varphi)(x)\,\mathrm{d}x = \lambda \int_a^b \varphi(x)\,\mathrm{d}x$,

(iii) $\displaystyle\varphi \leqslant \psi \implies \int_a^b \varphi(x)\,\mathrm{d}x \leqslant \int_a^b \psi(x)\,\mathrm{d}x$.

Beweis. VIII.1.4 hat gezeigt, dass wir annehmen können, dass φ und ψ durch dieselbe Partition definiert sind. Hat man das einmal, so sind alle drei Aussagen mit einer einfachen Rechnung zu prüfen. □

VIII.2. Riemann-Integrierbarkeit

Definition VIII.2.1. (Ober- und Unterintegral) Es sei $f\colon [a, b] \to \mathbb{R}$ eine beschränkte Funktion. Dann definieren wir:

$$\int_a^{b\,*} f(x)\,\mathrm{d}x := \inf\left\{\textstyle\int_a^b \varphi(x)\,\mathrm{d}x \;\middle|\; \varphi \in \mathrm{T}[a, b],\ \varphi \geqslant f\right\} \quad \textit{Oberintegral}$$

$$\int_{a\,*}^{b} f(x)\,\mathrm{d}x := \sup\left\{\textstyle\int_a^b \varphi(x)\,\mathrm{d}x \;\middle|\; \varphi \in \mathrm{T}[a, b],\ \varphi \leqslant f\right\} \quad \textit{Unterintegral}$$

Nach Definition gilt dann $\int_{a\,*}^{b} f(x)\,\mathrm{d}x \leqslant \int_a^{b\,*} f(x)\,\mathrm{d}x$.

Beispiele VIII.2.2.

(1) Für $\varphi \in \mathrm{T}[a, b]$ gilt $\int_{a\,*}^{b} \varphi(x)\,\mathrm{d}x = \int_a^{b\,*} \varphi(x)\,\mathrm{d}x$.

(2) Für die Dirichletfunktion $k\colon [0, 1] \to \mathbb{R}$ (siehe III.2.3 (6)) gilt $\int_{a\,*}^{b} k(x)\,\mathrm{d}x = 0$ und $\int_a^{b\,*} k(x)\,\mathrm{d}x = 1$.

Definition VIII.2.3. Eine beschränkte Funktion $f\colon [a, b] \to \mathbb{R}$ ist *Riemann–integrierbar*, oder kurz *R–integrierbar*, falls $\int_{a\,*}^{b} f(x)\,\mathrm{d}x = \int_a^{b\,*} f(x)\,\mathrm{d}x$ gilt. Ist dies der Fall, so definieren wir das *Riemann–Integral*, oder kurz *R–Integral*, von f per

$$\int_a^b f(x)\,\mathrm{d}x := \int_a^{b\,*} f(x)\,\mathrm{d}x.$$

Beispiel VIII.2.4. Alle Treppenfunktionen sind R–integrierbar. Die Dirichletfunktion ist nicht R–integrierbar.

Bemerkungen VIII.2.5.

(i) Es gibt noch andere Integralbegriffe neben dem Riemann–Integral und mit einigen davon wird dann z.B. die Dirichletfunktion integrierbar. Wenn klar ist, dass wir über Riemannintegrierbarkeit sprechen (und in dieser Vorlesung tun wir das stets), dann lassen wir 'Riemann–' bzw. 'R–' weg.

(ii) Für die 'Integrationsvariable' können wir auch andere Buchstaben benutzen. Es ist üblich diese, und dann auch das $\mathrm{d}x$, ganz wegzulassen:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(t)\,\mathrm{d}t = \int_a^b f(\varepsilon)\,\mathrm{d}\varepsilon = \int_a^b f.$$

Insbesondere ist '$\mathrm{d}x$' hier erst einmal nur ein Symbol, welches anzeigt, welcher Buchstable die Variable ist, z.B. ist ja $\int_0^1 te^x\,\mathrm{d}x = t(e-1)$ aber $\int_0^1 te^x\,\mathrm{d}t = \frac{1}{2}e^x$. Die Symbole '$\mathrm{d}x$' oder '$\mathrm{d}t$' alleine sind hier zunächst nicht definiert.

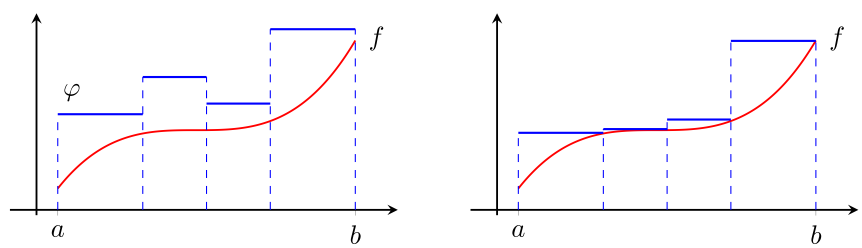
(iii) Man kann zeigen, dass gilt

$$\int_a^{b\,*} f(x)\,\mathrm{d}x = \inf\big\{\mathcal{O}_{f,Z}\;\big|\; Z : a = x_0 < x_1 < \cdots < x_n = b\big\},$$

wobei

$$\mathcal{O}_{f,Z} = \sum_{i=1}^{n}\big(\sup_{x\in(x_{i-1},x_i)} f(x)\big)\cdot(x_i - x_{i-1})$$

die *Obersumme* von f zur Zerlegung Z von $[a,b]$ ist. Eine analoge Aussage gilt für $\int_{a\,*}^{b} k(x)\,\mathrm{d}x$ mit den *Untersummen* $\mathcal{U}_{f,Z}$. Der Unterschied z.B. zwischen dem Infimum der Obersummen und dem Oberintegral ist, dass wir für das Oberintegral das Infimum über die Integrale *aller* Treppenfunktionen $\varphi \geqslant f$ nehmen — und daher auch solche vorkommen, die 'weit weg' von f sind. Nach Definition können wir jede Obersumme auch als Integral einer Treppenfunktion (mit Zerlegung Z und Stufenhöhen $c_k := \sup_{x\in(x_{k-1},x_k)} f(x)$) lesen. Wir nehmen dann ebenfalls ein Infimum über Integrale von Treppenfunktionen die größer gleich f sind — aber eben nur solche die schon 'nah' an f dran liegen:



Das Ergebnis stimmt überein, weil wir eben nur solche Treppenfunktionen weglassen, die ohnehin keinen Einfluss auf das Infimum haben.

Proposition VIII.2.6. ('Einquetschungskriterium') Es sei $f\colon [a,b] \to \mathbb{R}$ beschränkt. Dann ist f genau dann R–integrierbar, wenn gilt

$$\forall\,\varepsilon > 0\;\exists\,\varphi,\,\psi \in \mathrm{T}[a,b]\colon \varphi \leqslant f \leqslant \psi \quad\text{und}\quad \int_a^b \psi(x)\,\mathrm{d}x - \int_a^b \varphi(x)\,\mathrm{d}x < \varepsilon.$$

Beweis. '$\Longrightarrow$' Wenn f Riemann–integrierbar ist, dann existiert ein $R_0 \in \mathbb{R}$ mit

$$\sup\big\{\textstyle\int_a^b \varphi \;\big|\; \varphi \leqslant f\big\} = R_0 = \inf\big\{\textstyle\int_a^b \psi \;\big|\; f \leqslant \psi\big\}.$$

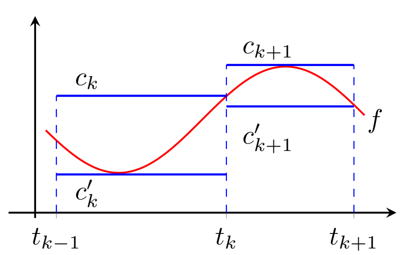
Nach Definition von Supremum und Infimum existieren daher $\varphi \leqslant f$ und $\psi \geqslant f$ mit

$$R_0 - \frac{\varepsilon}{2} < \int_a^b \varphi < R_0 \quad\text{und}\quad R_0 < \int_a^b \psi < R_0 + \frac{\varepsilon}{2}.$$

Daraus folgt
$$\int_a^b \psi < R_0 + \frac{\varepsilon}{2} < \Big(\int_a^b \varphi + \frac{\varepsilon}{2}\Big) + \frac{\varepsilon}{2}$$
und daher
$$\int_a^b \psi - \int_a^b \varphi < \varepsilon.$$
'$\Longleftarrow$' Es sei $\varepsilon > 0$. Wähle $\varphi \leqslant f \leqslant \psi$ in $\mathrm{T}[a,b]$ mit $\int_a^b \psi - \int_a^b \varphi < \varepsilon$. Aus der ersten Ungleichung folgt
$$\int_a^b \varphi \leqslant \sup\{\textstyle\int_a^b \tilde{\varphi} \mid \tilde{\varphi} \leqslant f\} = \displaystyle\int_{a*}^b f \leqslant \int_a^{b*} f = \inf\{\textstyle\int_a^b \tilde{\psi} \mid f \leqslant \tilde{\psi}\} \leqslant \displaystyle\int_a^b \psi,$$
und dann mit der zweiten
$$\Big|\int_a^{b*} f - \int_{a*}^b f\Big| \leqslant \int_a^b \psi - \int_a^b \varphi < \varepsilon.$$
Gilt obiges aber für jedes $\varepsilon > 0$, dann kann die linke Seite, welche unabhängig von ε ist, nur gleich Null sein. Damit sind dann Ober- und Unterintegral gleich und f ist R–integrierbar. $\square$

Satz VIII.2.7. *Jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$ ist integrierbar.*

Beweis. Wir wollen die Bedingung im Einquetschungskriterium prüfen. Dazu behaupten wir
$$(*) \quad \forall\, \varepsilon > 0\, \exists\, \varphi,\, \psi \in \mathrm{T}[a,b]\colon \varphi \leqslant f \leqslant \psi \text{ und } \forall\, x \in [a,b]\colon |\varphi(x) - \psi(x)| < \varepsilon.$$
Nach III.6.8 ist $f\colon [a,b] \to \mathbb{R}$ gleichmäßig stetig, d.h. wir wissen, dass
$$(\circ) \quad \forall\, \varepsilon > 0\, \exists\, \delta > 0\, \forall\, x, y \in [a,b]\colon |x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$
gilt. Wir zeigen damit nun $(*)$. Es sei $\varepsilon > 0$ gegeben. Wähle $\delta > 0$ wie in $(\circ)$. Wähle dann $n \in \mathbb{N}$, sodass $\frac{b-a}{n} < \delta$ gilt und definiere $t_k := a + k \cdot \frac{b-a}{n}$ für $k = 0, \dots, n$. Das liefert eine äquidistante Zerlegung $a = t_0 < t_1 < \cdots < t_n = b$. Setze nun
$$c_k := \sup_{x \in [t_{k-1}, t_k]} f(x) \quad \text{und} \quad c'_k := \inf_{x \in [t_{k-1}, t_k]} f(x)$$
für $k = 1, \dots, n$. Da f stetig ist, können wir den Extremalsatz anwenden und finden $\xi_k, \xi'_k \in [t_{k-1}, t_k]$, sodass $c = f(\xi_k)$ und $c'_k = f(\xi'_k)$ gilt. Da $|\xi_k - \xi'_k| \leqslant \frac{b-a}{n} < \delta$ gilt, folgt mit $(\circ)$, dass
$$|f(\xi_k) - f(\xi'_k)| = |c_k - c'_k| < \varepsilon$$
für jedes $k = 1, \dots, n$ gilt. Jetzt definieren wir $\varphi(t_k) := \psi(t_k) := f(t_k)$ für $k = 0, \dots, n$ und $\varphi(x) := c'_k$, $\psi(x) = c_k$ für $x \in (t_{k-1}, t_k)$, $k = 1, \dots, n$.



Es gelten also beide Bedingungen in $(*)$ und wir folgern nun daraus die Bedingung in VIII.2.6: Es sei $\varepsilon > 0$ gegeben. Wähle nach $(*)$ Treppenfunktionen φ, $\psi \in \mathrm{T}[a,b]$ mit $\varphi \leqslant f \leqslant \psi$ und $\psi(x) - \varphi(x) \leqslant \frac{\varepsilon}{b-a}$ für alle $x \in [a,b]$. Dann gilt
$$\int_a^b \psi(x)\,\mathrm{d}x - \int_a^b \varphi(x)\,\mathrm{d}x = \int_a^b (\psi(x) - \varphi(x))\,\mathrm{d}x \leqslant \int_a^b \theta(x)\,\mathrm{d}x = \frac{\varepsilon}{b-a} \cdot (b-a) = \varepsilon,$$
wobei wir erst VIII.1.5(i) und dann VIII.1.5(iii) mit der konstanten Funktion $\theta\colon [a,b] \to \mathbb{R}$ mit $\theta(x) := \frac{\varepsilon}{b-a}$ benutzt haben. Beachten Sie, dass VIII.1.5 anwendbar ist, da wir oben ausschließlich Treppenfunktionen integriert haben. $\square$

Satz VIII.2.8. *Jede monotone Funktion $f\colon [a,b] \to \mathbb{R}$ ist integrierbar.*

Beweis. Es sei f wachsend; für fallende f geht das Argument analog. Wir prüfen wieder das Einquetschungskriterium aus VIII.2.6: Es sei $\varepsilon > 0$ gegeben. Wähle $n \in \mathbb{N}$, sodass $\frac{b-a}{n}(f(b) - f(a)) < \varepsilon$. Setze $t_k = a + k \cdot \frac{b-a}{n}$ für $k = 0, \dots, n$. Jetzt definieren wir

$$\begin{aligned}
\varphi(x) &:= f(t_{k-1}) \text{ für } x \in [t_{k-1}, t_k) \text{ und } k = 1, \dots, n,\\
\psi(x) &:= f(t_k) \text{ für } x \in [t_{k-1}, t_k) \text{ und } k = 1, \dots, n,\\
\varphi(b) &:= \psi(b) := f(b).
\end{aligned}$$

Dann sind φ, $\psi \in \mathrm{T}[a,b]$. Da f wachsend ist, gilt $\varphi \leqslant f \leqslant \psi$. Wir erhalten

$$\begin{aligned}
\int_a^b \psi - \int_a^b \varphi &= \sum_{k=1}^n f(t_k)(t_k - t_{k-1}) - \sum_{k=1}^n f(t_{k-1})(t_k - t_{k-1})\\
&= \frac{b-a}{n} \sum_{k=1}^n f(t_k) - f(t_{k-1})\\
&= \frac{b-a}{n}(f(t_n) - f(t_0))\\
&= \frac{b-a}{n}(f(b) - f(a)) < \varepsilon
\end{aligned}$$

und VIII.2.6 impliziert nun, dass f integrierbar ist. □

VIII.3. Eigenschaften des Integrals

Proposition VIII.3.1. (Linearität und Monotonie des Integrals) Es seien f, $g\colon [a,b] \to \mathbb{R}$ integrierbar und sei $\lambda \in \mathbb{R}$. Dann sind $f + g$ und λf integrierbar und es gilt:

(i) $\displaystyle\int_a^b (f+g)(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x,$

(ii) $\displaystyle\int_a^b (\lambda f)(x)\,\mathrm{d}x = \lambda \int_a^b f(x)\,\mathrm{d}x,$

(iii) $\displaystyle f \leqslant g \implies \int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b g(x)\,\mathrm{d}x.$

Beweis. (iii) Es gilt

$$\begin{aligned}
\int_a^b f(x)\,\mathrm{d}x &= \sup\left\{ \textstyle\int_a^b f \;\middle|\; \varphi \in \mathrm{T}[a,b],\, \varphi \leqslant f \right\}\\
&\leqslant \sup\left\{ \textstyle\int_a^b f \;\middle|\; \varphi \in \mathrm{T}[a,b],\, \varphi \leqslant g \right\} = \textstyle\int_a^b g(x)\,\mathrm{d}x
\end{aligned}$$

Für (i) und (ii) benutzen wir wieder VIII.2.6:

(i) Es sei $\varepsilon > 0$. Wähle $\varphi_1, \varphi_2, \psi_1, \psi_2 \in \mathrm{T}[a,b]$ mit $\varphi_1 \leqslant f \leqslant \psi_1$, $\varphi_2 \leqslant g \leqslant \psi_2$ und $\int_a^b \psi_j - \int_a^b \varphi_j < \frac{\varepsilon}{2}$ für $j = 1, 2$. Dann gilt

$$\varphi_1 + \varphi_2 \leqslant f + g \leqslant \psi_1 + \psi_2 \quad \text{und} \quad \int_a^b \psi_1 + \psi_2 - \int_a^b \varphi_1 + \varphi_2 < \varepsilon,$$

wobei wir für die zweite Ungleichung VIII.1.5 benutzen können, weil wir hier nur Treppenfunktionen betrachten. Da $\varphi_1 + \varphi_2, \psi_1 + \psi_2 \in \mathrm{T}[a,b]$ gilt, haben wir gezeigt, dass $f + g$ integrierbar ist. Jetzt zeigen wir die Gleichung in (ii) durch Abschätzen, nämlich zuerst nach oben

$$\begin{aligned}\int_a^b f+g-\int_a^b f-\int_a^b g &\leqslant \int_a^b \psi_1+\psi_2-\int_a^b \varphi_1-\int_a^b \varphi_2\\ &=\int_a^b \psi_1+\int_a^b \psi_2-\int_a^b \varphi_1-\int_a^b \varphi_2\\ &=\int_a^b \psi_1-\varphi_1+\int_a^b \psi_2-\varphi_2\\ &<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon\end{aligned}$$

und dann nach unten in analoger Weise

$$\int_a^b f+g-\int_a^b f-\int_a^b g>-\varepsilon,$$

wobei wir jeweils im ersten Schritt den Teil (iii) des aktuellen Satzes benutzt haben, den wir ja bereits gezeigt haben, und danach die Rechenregeln aus VIII.1.5 für Treppenfunktionen. Zusammen erhalten wir also

$$\forall\,\varepsilon>0\colon\; -\varepsilon<\int_a^b f+g-\Big(\int_a^b f+\int_a^b g\Big)<\varepsilon$$

und das geht nur, wenn der mittlere Term verschwindet, woraus die Gleichung

$$\int_a^b f+g=\int_a^b f+\int_a^b g$$

folgt.

(ii) Wir teilen den Beweis in vier Teile, in denen wir verschiedene Werte von λ betrachten.

(1) Für $\lambda=0$ ist die Aussage trivialerweise erfüllt.

(2) Für $\lambda=-1$ haben wir

$$\begin{aligned}\int_{a\,*}^{b} -f &=\sup\{\textstyle\int_a^b\varphi \;\big|\; \varphi\in \mathrm{T}[a,b],\, \varphi\leqslant -f\}\\ &=\sup\big\{\textstyle\int_a^b\varphi \;\big|\; \varphi\in \mathrm{T}[a,b],\, -\varphi\geqslant f\big\}\\ &=\sup\big\{-\textstyle\int_a^b\psi \;\big|\; \psi\in \mathrm{T}[a,b],\, \psi\geqslant f\big\}\\ &=-\inf\big\{\textstyle\int_a^b\psi \;\big|\; \psi\in \mathrm{T}[a,b],\, \psi\geqslant f\big\}\\ &=-\int_a^{b\,*} f(x)\,\mathrm{d}x\\ &=-\int_a^b f,\end{aligned}$$

wobei wir VIII.1.5 benutzt haben, sowie dass $\psi\mapsto-\psi$ eine Bijektion von $\mathrm{T}[a,b]$ auf sich selbst ist. Analog zeigt man, dass $\int_a^{b\,*}(-f)=-\int_a^b f$ gilt, woraus folgt, dass $\int_a^{b\,*}(-f)=\int_{a\,*}^{b}(-f)$ gilt, also $-f$ integrierbar ist mit $\int_a^b(-f)=-\int_a^b f$.

(3) Es sei $\lambda>0$: Es sei $\varepsilon>0$, wähle $\varphi,\,\psi\in\mathrm{T}[a,b]$ mit $\varphi\leqslant f\leqslant\psi$ und $\int_a^b\psi-\int_a^b\varphi<\frac{\varepsilon}{\lambda}$. Dann folgt

$$\lambda\varphi\leqslant\lambda f\leqslant\lambda\psi \quad\text{und}\quad \int_a^b\lambda\psi-\int_a^b\lambda\varphi<\varepsilon,$$

wobei wir für die zweite Ungleichung VIII.1.5 verwenden können, da $\lambda\psi$ und $\lambda\varphi$ Treppenfunktionen sind. D.h. VIII.2.6 zeigt, dass λf integrierbar ist. Nun können wir ausrechnen (mit der Substitution

$\psi := \frac{1}{\lambda}\varphi$ bzw. $\varphi = \lambda\psi$)

$$\begin{aligned}\int_a^b \lambda f &= \sup\left\{\textstyle\int_a^b \varphi \;\middle|\; \varphi \in \mathrm{T}[a,b],\ \varphi \leqslant \lambda f\right\}\\ &= \sup\left\{\textstyle\int_a^b \lambda\psi \;\middle|\; \lambda\psi \in \mathrm{T}[a,b],\ \psi \leqslant f\right\}\\ &= \lambda \sup\left\{\textstyle\int_a^b \psi \;\middle|\; \psi \in \mathrm{T}[a,b],\ \psi \leqslant f\right\},\end{aligned}$$

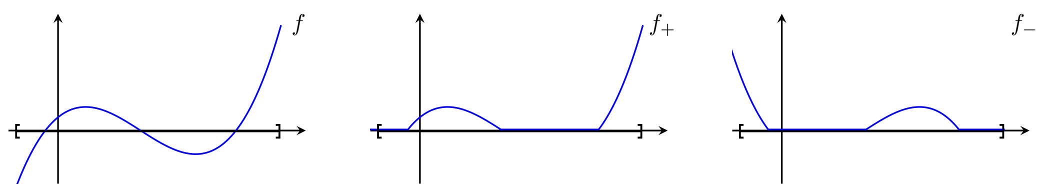
wobei wir benutzt haben, dass $\psi \mapsto \lambda\psi$ eine Bijektion $\mathrm{T}[a,b] \to \mathrm{T}[a,b]$ ist.

(4) Den Fall $\lambda < 0$ können wir in Kombination von (2) und (3) beweisen.

□

Definition VIII.3.2. Es sei $f\colon D \subset \mathbb{R} \to \mathbb{R}$ eine Funktion. Dann definieren wir den *Positivteil* $f_+\colon D \to \mathbb{R}$ und den *Negativteil* $f_-\colon D \to \mathbb{R}$ von f als

$$f_+(x) := \begin{Bmatrix} f(x), & \text{falls } f(x) > 0 \\ 0, & \text{sonst} \end{Bmatrix}, \quad f_-(x) := \begin{Bmatrix} -f(x), & \text{falls } f(x) < 0 \\ 0, & \text{sonst} \end{Bmatrix}.$$



Wir beachten, dass $f_+,\ f_- \geqslant 0$, $f = f_+ - f_-$ und $|f| = f_+ + f_-$. Insbesondere ist der Negativteil f_- eine nicht-negative Funktion.

Proposition VIII.3.3. Es seien $f,\ g\colon [a,b] \to \mathbb{R}$ integrierbar. Dann gilt:

(i) $f_+,\ f_-$ sind integrierbar und es gilt $\left|\int_a^b f(x)\,\mathrm{d}x\right| \leqslant \int_a^b |f(x)|\,\mathrm{d}x$,

(ii) Für $p \in [1,\infty)$ ist $|f|^p$ integrierbar,

(iii) $f \cdot g\colon [a,b] \to \mathbb{R}$ ist integrierbar.

Beweis. (i) Wir fangen an mit f_+: Nach Voraussetzung finden wir für $\varepsilon > 0$ Treppenfunktionen $\varphi \leqslant f \leqslant \psi$ mit $\int_a^b \psi - \int_a^b \varphi < \varepsilon$. Dann folgt $\varphi_+ \leqslant f_+ \leqslant \psi_+$ und $\varphi_- \geqslant f_- \geqslant \psi_-$. Nun haben wir

$$\varepsilon > \int_a^b \psi - \varphi = \int_a^b \psi_+ - \psi_- - (\varphi_+ - \varphi_-) = \int_a^b (\psi_+ - \varphi_+) + (\varphi_- - \psi_-) \geqslant \int_a^b \psi_+ - \varphi_+,$$

also ist f_+ nach VIII.2.6 integrierbar. Analog zeigt man, dass f_- integrierbar ist. Da $|f| = f_+ - f_-$ ist, folgt mit VIII.3.1, dass $|f|$ integrierbar ist. Wegen $f \leqslant |f|$, $-f \leqslant |f|$ folgt auch mit VIII.3.1

$$\int_a^b f \leqslant \int_a^b |f| \quad \text{sowie} \quad -\int_a^b f = \int_a^b -f \leqslant \int_a^b |f|$$

und damit

$$\left|\int_a^b f\right| \leqslant \int_a^b |f|.$$

Zu (ii):

(1) Wir betrachten zuerst den Fall, dass $0 \leqslant f \leqslant 1$ gilt, und prüfen wieder das Einquetschungskriterium VIII.2.6. Es sei dazu $\varepsilon > 0$. Wir wählen $\varphi,\ \psi \in \mathrm{T}[a,b]$ mit $0 \leqslant \varphi \leqslant f \leqslant \psi \leqslant 1$ und $\int_a^b \psi - \varphi < \frac{\varepsilon}{p}$. Dann gilt $\varphi^p,\ \psi^p \in \mathrm{T}[a,b]$ und $\varphi^p \leqslant f^p \leqslant \psi^p$. Es sei $x \in [a,b]$. Falls $\varphi(x) < \psi(x)$, betrachten wir

$$g\colon [\varphi(x), \psi(x)] \to \mathbb{R},\ g(y) = y^p.$$

Die Funktion g ist stetig und auf dem offenen Intervall $(\varphi(x), \psi(x))$ differenzierbar, d.h. wir können den Mittelwertsatz anwenden und finden $\xi \in (\varphi(x), \psi(x))$ mit

$$\frac{\psi(x)^p - \varphi(x)^p}{\psi(x) - \varphi(x)} = \frac{g(\psi(x)) - g(\varphi(x))}{\psi(x) - \varphi(x)} = g'(\xi) = p\xi^{p-1} \leqslant p,$$

wobei wir im letzten Schritt $0 < \xi < 1$ benutzt haben. Es gilt also $\psi(x)^p - \varphi(x)^p \leqslant p(\psi(x) - \varphi(x))$. Falls $\varphi(x) = \psi(x)$ ist, gilt letztere Gleichung trivialerweise. D.h. wir haben

$$\psi(x)^p - \varphi(x)^p \leqslant p(\psi(x) - \varphi(x))$$

für alle $x \in [a, b]$. Es folgt mit VIII.3.1

$$\int_a^b \psi^p - \varphi^p \leqslant \int_a^b p(\psi - \varphi) < p \cdot \frac{\varepsilon}{p} = \varepsilon$$

und mit VIII.2.6 ist f^p integrierbar.

(2) Es sei jetzt f beliebig aber nicht $f \equiv 0$[1]. Dann gilt $0 < s := \sup_{x\in[a,b]} |f(x)| < \infty$, da f als R–integrierbare Funktion nach Definition beschränkt ist. Wir setzen

$$\tilde{f} \colon [a, b] \to \mathbb{R},\ \tilde{f}(x) := \frac{|f(x)|}{s}.$$

Dann gilt $0 \leqslant \tilde{f} \leqslant 1$ und nach (1) ist $\tilde{f}^p$ integrierbar. Dann ist aber auch $|f|^p = s^p \tilde{f}^p$ nach VIII.2.6 integrierbar.

Teil (iii) folgt aus (ii) mit $f \cdot g = \frac{1}{4}\big((f + g)^2 - (f - g)^2\big)$. □

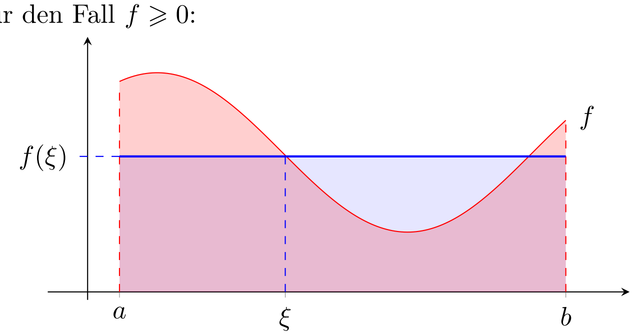
Satz VIII.3.4. $\big($*Verallgemeinerter Mittelwertsatz der Integralrechnung*$\big)$ *Es seien f, $g\colon\ [a, b] \to \mathbb{R}$ gegeben und sei f stetig und $g \geqslant 0$ sei integrierbar. Dann existiert $\xi \in [a, b]$ mit*

$$\int_a^b f(x)g(x)\,\mathrm{d}x = f(\xi)\int_a^b g(x)\,\mathrm{d}x.$$

Ist $g \equiv 1$, so erhalten wir den (einfachen) Mittelwertsatz der Integralrechnung: Für eine stetige Funktion $f\colon [a, b] \to \mathbb{R}$ existiert $\xi \in [a, b]$ mit

$$\int_a^b f(x)\,\mathrm{d}x = f(\xi)(b - a).$$

Eine Veranschaulichung für den Fall $f \geqslant 0$:



Die Fläche unter dem Graphen ist gleich der Fläche des Rechtecks, wobei dessen Höhe als Funktionswert auftritt.

Beweis. Wir setzen $m := \inf_{x\in[a,b]} f(x)$ und $M := \sup_{x\in[a,b]} f(x)$. Dann gilt $mg \leqslant fg \leqslant Mg$ und daher $m\int_a^b g \leqslant \int_a^b fg \leqslant M\int_a^b g$. Mit dem Zwischenwertsatz angewandt auf die Funktion $[m, M] \to \mathbb{R}$, $x \mapsto x \cdot \int_a^b g$ folgt, dass ein $\mu \in [m, M]$ existiert mit

$$\int_a^b fg = \mu \cdot \int_a^b g.$$

Nochmalige Anwendung des Zwischenwertsatzes (nun auf f) zeigt, dass ein ξ existiert mit $f(\xi) = \mu$. □

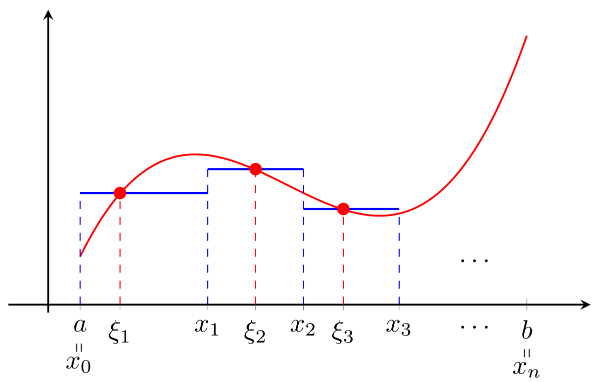
[1]Für f konstant Null ist die Aussage erstens trivialerweise erfüllt. Zweitens ist dieser Fall bereits in Teil (1) enthalten.

Bemerkung VIII.3.5. Um Integrale auszurechnen, ist die reine Definition sehr ungünstig, denn wir müssen stets Infima/Suprema über beliebige Zerlegungen nehmen. Das Konzept der *Riemannsummen* wird dieses Problem lösen.

Definition VIII.3.6. Es sei $f\colon [a,b] \to \mathbb{R}$ eine Funktion und sei $Z : a = x_0 < x_1 < \cdots < x_n = b$ eine Zerlegung.

(i) $\xi = (\xi_i)_{i=1,\dots,n}$ heißt der Zerlegung zugehöriger *Zwischenvektor*, falls $\xi_i \in [x_{i-1}, x_i]$ gilt für $i = 1, \dots, n$.

(ii) $\mathcal{R}(Z, \xi, f) := \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$ heißt *Riemannsumme von f bezüglich Zerlegung Z und Zwischenvektor ξ.*



(iii) $\mu(Z) := \max_{i=1,\dots,n}(x_i - x_{i-1})$ heißt *Feinheit der Zerlegung Z.*

(iv) Eine Folge von Zerlegungen $(Z^{(k)})_{k\in\mathbb{N}}$, d.h.

$$Z^{(k)} : a = x_0^{(k)} < x_1^{(k)} < x_2^{(k)} < \cdots < x_{n_k}^{(k)} = b$$

heißt *Zerlegungsnullfolge*, falls $\mu(Z^{(k)}) \to 0$ gilt für $k \to \infty$.

Satz VIII.3.7. *Es sei $f\colon [a,b] \to \mathbb{R}$ integrierbar. Dann gilt*

$$\lim_{k\to\infty} \mathcal{R}(Z^{(k)}, \xi^{(k)}, f) = \int_a^b f(x)\,\mathrm{d}x$$

für jede Zerlegungsnullfolge $(Z^{(k)})_{k\in\mathbb{N}}$ und jede Folge von zugehörigen Zwischenvektoren $(\xi^{(k)})_{k\in\mathbb{N}}$.

Beispiel VIII.3.8. $\int_0^1 x\,\mathrm{d}x = \frac{1}{2}$.

Da $f\colon [0,1] \to \mathbb{R}$, $f(x) = x$ stetig (und auch monoton) ist, folgt, dass f integrierbar ist. D.h. VIII.3.7 kann benutzt werden.

Wir definieren äquidistante Zerlegungen

$$Z^{(k)} : 0 = x_0^{(k)} < x_1^{(k)} < \cdots < x_k^{(k)} = 1$$

mit Feinheit $\frac{1}{k}$ für $k \geqslant 1$, d.h.

$$x_i^{(k)} := \frac{i}{k}$$

für $i = 0, \dots, k$. Dann wählen wir Zwischenvektoren $\xi^{(k)}$ mit

$$\xi_i^{(k)} := x_i^{(k)},$$

d.h. jeweils den rechten Randpunkt des jeweiligen Intervalls. Dies liefert eine Zerlegungsnullfolge und eine Folge von zugehörigen Zwischenvektoren. Nun berechnen wir die zugehörigen Riemannsummen und dann deren Grenzwert

$$\begin{aligned}\mathcal{R}(Z^{(k)},\xi^{(k)},f) &= \sum_{i=1}^{n} f(\xi_i^{(k)})(x_i^{(k)} - x_{i-1}^{(k)}) \\ &= \sum_{i=1}^{n} \frac{i}{k}\Big(\frac{i}{k} - \frac{i-1}{k}\Big) \\ &= \frac{1}{k^2}\sum_{i=1}^{k} i \\ &= \frac{1}{k^2}\cdot\frac{k(k+1)}{2} \\ &= \frac{1}{2}\Big(\frac{k+1}{k}\Big) \xrightarrow{k\to\infty} \frac{1}{2},\end{aligned}$$

welcher nach VIII.3.7 gerade der Wert des gesuchten Integrals ist.

Beweis. (von VIII.3.7) Es sei $f\colon [a,b]\to\mathbb{R}$ integrierbar. Wir behaupten

$$(*)\qquad \forall\,\varepsilon>0\;\exists\,\delta>0\;\forall\,Z \text{ mit } \mu(Z)<\delta\;\forall\,\xi\colon \Big|\mathcal{R}(Z,\xi,f) - \int_a^b f\,\Big| < \varepsilon.$$

Hieraus folgt sofort VIII.3.7, denn für eine Zerlegungsnullfolge $(Z^{(k)})_{k\in\mathbb{N}}$ und $\varepsilon>0$ kann $\delta>0$ wie in $(*)$ gewählt werden und dann kann $k_0\in\mathbb{N}$ mit $\mu(Z^{(k)})<\delta$ für $k\geqslant k_0$ gefunden werden, sodass also für diese $k\geqslant k_0$ dann gerade $\big|\mathcal{R}(Z,\xi,f)-\int_a^b f\,\big|<\varepsilon$ gilt, was die in VIII.3.7 behauptete Konvergenz zeigt.

Es sei jetzt $f\not\equiv 0$[2] und $\varepsilon>0$ gegeben. Wir wählen gemäß des Einquetschungskriteriums $\varphi,\,\psi\in\mathrm{T}[a,b]$ mit $\varphi\leqslant f\leqslant\psi$ und $\int_a^b\psi-\int_a^b\varphi<\frac{\varepsilon}{2}$. Für φ und ψ finden wir dann eine gemeinsame Partition $a=t_0<t_1<\cdots<t_m=b$. Da f als R–integrierbare Funktion beschränkt ist, gilt $0<M:=\sup_{x\in[a,b]}|f(x)|<\infty$ und damit $-M\leqslant\pm f\leqslant M$. Wir setzen $\delta:=\frac{\varepsilon}{4Mm}>0$. Es sei nun $Z: a=x_0<x_1<\cdots<x_n=b$ eine beliebige Partition mit $\mu(Z)<\delta$ und sei $\xi=(\xi_k)_{k=1,\dots,n}$ ein zugehöriger Zwischenvektor.

Als nächstes schreiben wir die Riemannsumme als Integral einer Treppenfunktion F (vgl. das Bild in VIII.3.6(ii)). Definiere hierzu

$$F\colon[a,b]\to\mathbb{R},\; F(x):=\begin{cases} f(a), & \text{für } x=a\\ f(\xi_i), & \text{für } x\in(x_{i-1},x_i],\ i=1,\dots,n.\end{cases}$$

Dann gilt $F\in\mathrm{T}[a,b]$ und

$$\int_a^b F=\sum_{i=1}^{n} f(\xi_i)(x_i-x_{i-1})=\mathcal{R}(Z,\xi,f)$$

und außerdem $-M\leqslant\pm F\leqslant M$. Um $(*)$ zu zeigen, müssen wir also

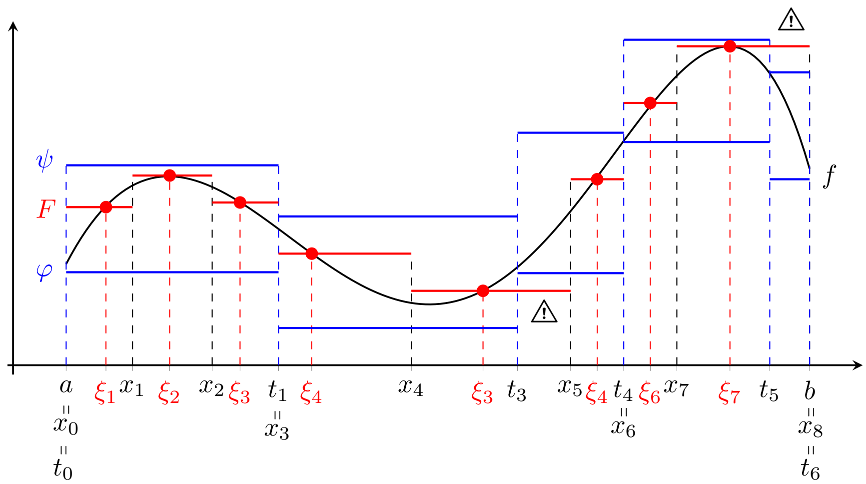
$$(+)\qquad \Big|\int_a^b F-\int_a^b f\Big|<\varepsilon \quad\text{bzw.}\quad -\varepsilon<\int_a^b F-\int_a^b f<\varepsilon$$

zeigen und dafür muss $\int_a^b F$ abgeschätzt werden.

Wir unterscheiden zwei Sorten von Intervallen:

(1) Ist $[x_{i-1},x_i]\subset[t_{j-1},t_j]$ für ein $1\leqslant i\leqslant n$ und ein $1\leqslant j\leqslant m$, so gilt die Abschätzung $\varphi\leqslant F\leqslant\psi$ auf $[x_{i-1},x_i]$.
(2) Den zweiten Intervalltyp veranschaulichen wir im folgenden Bild.

[2]Ist $f\equiv 0$, dann ist sind alle Riemannsummen, unabhähngig von Zerlegung und Zwischenvektor, Null, und folglich steht auf beiden Seiten der behaupteten Gleichung Null.



An den mit einem Ausrufezeichen markierten Stellen (genauer für $x \in [t_3, x_5]$ und $x \in [t_5, x_8]$) ist die Ungleichung $\varphi(x) \leqslant F(x) \leqslant \psi(x)$ *nicht* erfüllt und wir können daher nicht $\int_a^b \varphi \leqslant \int_a^b F \leqslant \int_a^b \psi$ schließen.

Überlappen die Intervalle wie bei den Ausrufezeichen, so kann wie erwähnt $F(x) \leqslant \varphi(x)$ oder $F(x) \geqslant \psi(x)$ an manchen Stellen x vorkommen. Hier gilt dann aber

$$-M \leqslant F \leqslant M, \quad -M \leqslant -f \leqslant M \quad \text{und} \quad \varphi \leqslant f \leqslant \psi,$$

also durch Addieren der drei Ungleichungen

$$\varphi - 2M \leqslant F \leqslant \psi + 2M.$$

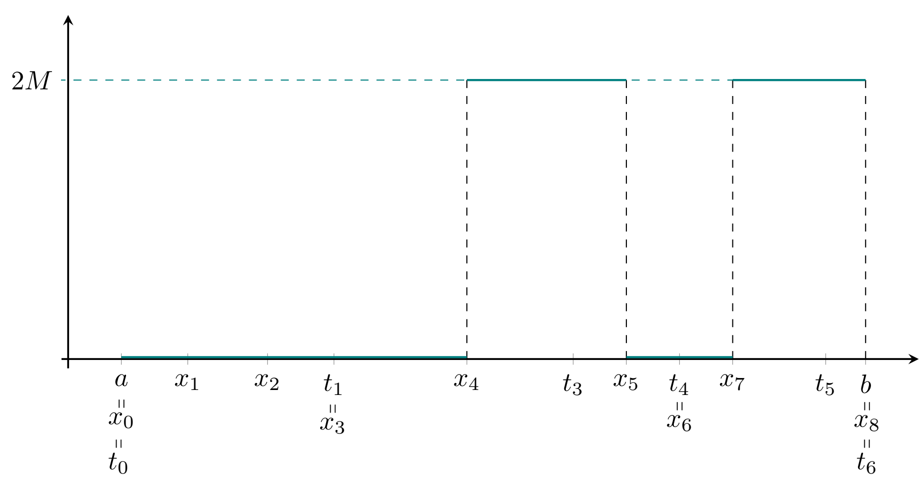
Wir definieren nun

$$A := \bigcup_{\substack{i=1,\ldots,n \text{ s.d. } \exists\, 1 \leqslant j \leqslant m:\\ [x_{i-1},x_i] \subset [t_{j-1},t_j]}} [x_{i-1}, x_i]$$

d.h. wir vereinigen alle Intervalle des ersten Typs der Zerlegung Z. Weiter definieren wir

$$S\colon [a,b] \to \mathbb{R}, \quad S(x) := \begin{cases} 0, & \text{falls } x \in A, \\ 2M, & \text{falls } x \in [a,b] \backslash A \end{cases}$$

und bemerken $S \in \mathrm{T}[a,b]$, sowie $\varphi - S \leqslant F \leqslant \psi + S$ auf $[a,b]$: Auf den Intervallen vom Typ (1) gilt $\varphi \leqslant F \leqslant \psi$ und S ist Null; auf Intervallen vom Typ (2) erledigt S für uns die Korrektur mit $\pm 2M$. D.h. wir brauchen eine Abschätzung für $\int_a^b S$:



In der Tat gilt

$$\int_a^b S(x)\,\mathrm{d}x = \text{ Anzahl der Intervalle vom Typ (2) } \cdot \text{ deren Länge } \cdot 2M,$$

wobei wir die Länge der Intervalle durch die Feinheit $\mu(Z)$ und daher durch δ abschätzen können. Um die Anzahl der Intervalle vom Typ (2) festzustellen, müssen wir die Frage stellen, für wieviele $i = 1, \dots, n$ es kein $j \in \{1, \dots, m\}$ gibt mit $[x_{i-1}, x_i] \subset [t_{j-1}, t_j]$. Dies sind aber genau diejenigen Indizes i, für die mindestens ein $t_j \in (x_{i-1}, x_i)$ liegt und da es nur m–viele Werte t_j gibt, kann letzteres höchstens m–mal passieren, und wir können abschätzen

$$\int_a^b S(x)\,\mathrm{d}x \leqslant m \cdot \delta \cdot 2M \leqslant m \cdot \frac{\varepsilon}{4Mm} \cdot 2M = \frac{\varepsilon}{2}.$$

Damit folgt nun

$$\int_a^b F - \int_a^b f \leqslant \int_a^b \psi + S - \int_a^b f \leqslant \int_a^b \psi - \int_a^b \varphi + \frac{\varepsilon}{2} \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

und analog

$$\int_a^b F - \int_a^b f \geqslant -\varepsilon$$

also die Aussage in $(+)$. □

Bemerkungen VIII.3.9.

(1) Bei allen technischen Details behalten Sie die Idee im Kopf, dass man durch Treppenfunktionen von oben und unten approximiert und dann die Z–Partitionen so fein macht, dass der Beitrag der Intervalle vom Typ (2) beliebig klein wird.

(2) Es gilt auch die Umkehrung von VIII.3.7, d.h. wenn $R_0 \in \mathbb{R}$ existiert, sodass für jede Zerlegungsnullfolge $(Z^{(k)})_{k\in\mathbb{N}}$ und jede Folge von zugehörigen Zwischenvektoren $\lim_{k\to\infty} \mathcal{R}(Z^{(k)}, \xi^{(k)}, f) = R_0$ gilt, dann ist f R–integrierbar und es ist $\int_a^b f = R_0$.

Proposition VIII.3.10. Es sei $f\colon [a,b] \to \mathbb{R}$ und $a < c < b$. Dann ist f genau dann integrierbar, wenn $f|_{[a,c]}$ und $f|_{[c,b]}$ beide integrierbar sind. In diesem Fall gilt

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x + \int_c^b f(x)\,\mathrm{d}x.$$

Beweis.[3] Wir überlegen uns zuerst die Äquivalenz. Ist f integrierbar, so existieren für vorgegebenes $\varepsilon > 0$ nach dem Einqueschungskriterium VIII.2.6 $\varphi, \psi \in \mathrm{T}[a,b]$ mit $\varphi \leqslant f \leqslant \psi$ und

$$\int_a^b \psi - \int_a^b \varphi < \varepsilon.$$

Dann sind aber $\varphi|_{[a,c]}, \psi|_{[a,c]} \in T[a,c]$, es gilt $\varphi|_{[a,c]} \leqslant f|_{[a,c]} \leqslant \psi|_{[a,c]}$ und aus

$$\int_a^c \psi|_{[a,c]} - \int_a^c \varphi|_{[a,c]} = \int_a^c \psi - \int_a^c \varphi \leqslant \int_a^c \psi - \int_a^c \varphi + \underbrace{\int_c^b \psi - \int_c^b \varphi}_{\geqslant 0} = \int_a^b \psi - \int_a^b \varphi < \varepsilon$$

folgt die Integrierbarkeit von $f|_{[a,c]}$, wieder aus VIII.2.6. Für $f|_{[c,b]}$ argumentiert man analog.

Sind für die Umgekehrung $f|_{[a,c]}$ und $f|_{[c,b]}$ beide integrierbar, so können wir VIII.2.6 zweimal anwenden und für gegebenes $\varepsilon > 0$ Treppenfunktionen $\varphi_1, \psi_1 \in \mathrm{T}[a,c]$, $\varphi_2, \psi_2 \in \mathrm{T}[c,b]$ mit $\varphi_1 \leqslant f|_{[a,c]} \leqslant \psi_1$, $\varphi_2 \leqslant f|_{[c,b]} \leqslant \psi_2$ und

$$\int_a^c \psi_1 - \int_a^c \varphi_1 < \varepsilon/2 \quad \text{sowie} \quad \int_c^b \psi_1 - \int_c^b \varphi_1 < \varepsilon/2$$

[3]Wir geben einen elementaren Beweis, der zeigt, dass obiges praktisch aus der Definition, bzw. dem sehr nah an der Definition liegenden Einquetschungskriterium, folgt. Ein sehr eleganter alternativer Beweis kann mithilfe von VIII.3.3(iii) und durch Benutzung von charakteristischen Funktionen, z.B. $1_{[a,c]}$, $1_{[c,b]}$ und $1_{(c,b]}$ geführt werden.

finden. Nun definiert man $\varphi, \psi \in \mathrm{T}[a,b]$ als

$$\varphi(x) = \begin{cases} \varphi_1(x) & \text{für } x \in [a,c], \\ \varphi_2(x) & \text{sonst,} \end{cases} \quad \text{und} \quad \psi(x) = \begin{cases} \psi_1(x) & \text{für } x \in [a,c], \\ \psi_2(x) & \text{sonst.} \end{cases}$$

Dann folgt mit VIII.2.6, dass f integrierbar ist, da $\varphi \leqslant f \leqslant \psi$ und

$$\int_a^b \psi - \int_a^b \varphi = \int_a^c \psi_1 + \int_c^b \psi_2 - \int_a^c \varphi_1 - \int_c^b \varphi_2 < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

In der folgenden Rechnung benutzen wir für die dritte Gleichung, dass jede Treppenfunktion in $\mathrm{T}[a,b]$ eine Darstellung hat, in der c als Punkt in der Zerlegung vorkommt:

$$\begin{aligned} \int_a^b f &= \sup\left\{ \textstyle\int_a^b \varphi \;\middle|\; \varphi \in \mathrm{T}[a,b],\, \varphi \leqslant f \right\} \\ &= \sup\left\{ \textstyle\int_a^c \varphi + \int_c^b \varphi \;\middle|\; \varphi \in \mathrm{T}[a,b],\, \varphi \leqslant f \right\} \\ &= \sup\left\{ \textstyle\int_a^c \varphi_1 + \int_c^b \varphi_2 \;\middle|\; \varphi_1 \in \mathrm{T}[a,c],\, \varphi_2 \in \mathrm{T}[c,b],\, \varphi_1 \leqslant f|_{[a,c]},\, \varphi_2 \leqslant f|_{[c,b]} \right\} \\ &= \sup\left\{ \textstyle\int_a^c \varphi_1 \;\middle|\; \varphi_1 \in \mathrm{T}[a,c],\, \varphi_1 \leqslant f|_{[a,c]} \right\} + \sup\left\{ \textstyle\int_c^b \varphi_2 \;\middle|\; \varphi_2 \in \mathrm{T}[c,b],\, \varphi_2 \leqslant f|_{[c,b]} \right\} \\ &= \int_a^c f + \int_c^b f. \end{aligned}$$

Damit ist auch die Gleichung gezeigt. Da obiges für das Oberintegral genau so funktioniert, hat man auf diese Weise auch nochmal die Rückrichtung der Äquivalenz gezeigt. □

Definition VIII.3.11. Wir definieren $\int_a^a f(x)\,\mathrm{d}x := 0$ und für $a > b$ setzen wir

$$\int_a^b f(x)\,\mathrm{d}x := -\int_b^a f(x)\,\mathrm{d}x.$$

VIII.4. Differentiation und Integration

Abschnitt ist $I \subset \mathbb{R}$ stets ein Intervall und wir nehmen an, dass I mindestens zwei Punkte enthält.

Proposition VIII.4.1. Es sei $f\colon I \to \mathbb{R}$ stetig und $a \in I$. Dann ist $F\colon I \to \mathbb{R}$, $F(x) := \int_a^x f(t)\,\mathrm{d}t$ differenzierbar und es gilt $F' = f$.

Beweis. Für $x \in I$ und $h \neq 0$, sodass $x + h \in I$ ist, gilt

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\Big[\int_a^{x+h} f(t)\,\mathrm{d}t - \int_a^x f(t)\,\mathrm{d}t\Big] = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t,$$

wobei wir für die Zusammenfassung der Integrale VIII.3.10 benutzt haben. Mit dem Mittelwertsatz der Integralrechnung VIII.3.4 folgt nun

$$\forall\, h\ \exists\, \xi_h \in [x, x+h]\ \big(\text{bzw. } \xi_h \in [x+h, x]\big)\colon \int_x^{x+h} f(t)\,\mathrm{d}t = h \cdot f(\xi_h)$$

und damit

$$F'(x) = \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t = \lim_{h\to 0} f(\xi_h) = f(x),$$

wobei wir im letzten Schritt benutzt haben, dass f stetig ist. □

Definition VIII.4.2. Eine differenzierbare Funktion $F\colon I \to \mathbb{R}$ heißt *Stammfunktion* von $f\colon I \to \mathbb{R}$, falls $F' = f$ gilt.

Proposition VIII.4.3. Es sei $F\colon I \to \mathbb{R}$ eine Stammfunktion von $f\colon I \to \mathbb{R}$. Dann ist $G\colon I \to \mathbb{R}$ genau dann eine Stammfunktion von f, wenn $F - G$ konstant ist.

Beweis. '$\Longleftarrow$' Es sei $F - G \equiv c$ mit $c \in \mathbb{R}$ fest. Dann gilt $G' = (F-c)' = F' = f$.

'$\Longrightarrow$' Es sei G Stammfunktion. Dann gilt $(G-F)' = G' - F' = f - f = 0$ und mit VI.1.8 folgt, dass $F - G$ konstant ist. □

Satz VIII.4.4. *(Hauptsatz der Differential- und Integralrechnung) Es sei $f\colon I \to \mathbb{R}$ integrierbar und F sei eine Stammfunktion von f. Dann gilt für alle $a, b \in I$*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Beweis. Ohne Einschränkung gelte $a < b$. Es sei $\varepsilon > 0$. Nach dem Einquetschungskriterium VIII.2.6 existieren φ, $\psi \in \mathrm{T}[a,b]$ mit $\varphi \leqslant f \leqslant \psi$ und $\int_a^b \psi - \int_a^b \varphi < \varepsilon$. Es sei $Z : a = x_0 < x_1 < \cdots < x_n = b$ eine gemeinsame Zerlegung für φ und ψ. Für jedes $k = 1, \ldots, n$ wenden wir nun den Mittelwertsatz VI.1.5 auf die Funktion $F\colon [x_{k-1}, x_k] \to \mathbb{R}$ an und finden diesem zufolge $\xi_k \in (x_{k-1}, x_k)$ mit

$$F'(\xi_k) = \frac{F(x_k) - F(x_{k-1})}{x_k - x_{k-1}}.$$

Dann folgt mit einem Teleskopsummenargument

$$F(b) - F(a) = \sum_{k=1}^n F(x_k) - F(x_{k-1}) = \sum_{k=1}^n f(\xi_k)(x_k - x_{k-1}).$$

Weil φ, ψ Treppenfunktionen sind und $\varphi \leqslant f \leqslant \psi$ gilt, erhalten wir als nächstes

$$\int_a^b \varphi = \sum_{k=1}^n \varphi(\xi_k)(x_k - x_{k-1}) \leqslant \underbrace{\sum_{k=1}^n f(\xi_k)(x_k - x_{k-1})}_{=F(b)-F(a)} \leqslant \sum_{k=1}^n \psi(\xi_k)(x_k - x_{k-1}) = \int_a^b \psi.$$

Schließlich folgt daraus

$$-\varepsilon \leqslant \int_a^b \varphi - \int_a^b \psi \leqslant F(b) - F(a) - \int_a^b f \leqslant \int_a^b \psi - \int_a^b \varphi \leqslant \varepsilon.$$

Da obiges für beliebiges $\varepsilon > 0$ gilt, muss der mittlere Term verschwinden, also $\int_a^b f = F(b) - F(a)$ sein. □

Bemerkungen VIII.4.5.

(1) Häufig schreibt man $F(b) - F(a) =: F(x)\big|_a^b =: \big[F(x)\big]_a^b$.

(2) Manchmal schreibt man $\int f(x)\,\mathrm{d}x = F(x)$ als Abkürzung für 'F ist Stammfunktion von f'. Beachten Sie, dass dann auch gilt $\int f(x)\,\mathrm{d}x = F(x) + c$ für jedes $c \in \mathbb{R}$. Dies so zu schreiben ist daher formal nicht korrekt (aber üblich). Außerdem ist bei dieser Schreibweise der Definitionsbereich von f nicht angegeben, d.h. man kann leicht Fehler machen, z.B. ist

$$\int \frac{1}{x}\,\mathrm{d}x = \log x + c \;\text{ auf }\; (0,\infty) \;\text{ aber }\; \int \frac{1}{x}\,\mathrm{d}x = \log(-x) + c \;\text{ auf }\; (-\infty, 0).$$

Überzeugen Sie sich durch Ableiten von diesen Gleichungen.

(3) Wenn $f\colon I \to \mathbb{R}$ stetig ist, dann impliziert VIII.4.1, dass f eine Stammfunktion hat, die dann in $\mathrm{C}^1(I)$ liegt. In vielen Büchern wird der Hauptsatz so formuliert und durch die Formel

$$\frac{\mathrm{d}}{\mathrm{d}x} \int_a^x f(t)\,\mathrm{d}t = f(x)$$

ausgedrückt.

(4) Vorsicht: Für $f\colon [a,b] \to \mathbb{R}$ impliziert die Integrierbarkeit von f *nicht*, dass f eine Stammfunktion hat. Falls f eine Stammfunktion besitzt, dann impliziert dies *nicht*, dass f integrierbar ist.

Beispiele VIII.4.6.

- $f(x) = x^n$, $n \neq -1$.[4] Dann ist $F(x) = \frac{x^{n+1}}{n+1}$ eine Stammfunktion und wir erhalten zum Beispiel sofort: $\int_0^1 x^2 \,\mathrm{d}x = \frac{x^3}{3}\big|_0^1 = \frac{1}{3}$.
- $f(x) = \frac{1}{x}$, $x > 0$. $F(x) = \log x$ ist Stammfunktion und das impliziert sofort $\int_1^x \frac{1}{t} \,\mathrm{d}t = \log x$, was jetzt sogar für alle $x > 0$ gezeigt ist.
- $\int_1^5 \frac{1}{10-x} \,\mathrm{d}x = (-1)\log(10-x)\big|_1^5 = -\big(\log(10-5) - \log(10-1)\big) = \log(\frac{9}{5})$, wobei man hier z.B. so vorgehen kann, dass man versucht die Stammfunktion zu erraten (und dabei vielleicht den Faktor -1 erstmal nicht errät), und dann durch Ableiten aber sieht, dass man ein Minus zu viel bekommt, was man durch den Faktor -1 korrigiert.

Bemerkungen VIII.4.7.

(1) Mit dem was wir in vorhergehenden Kapiteln über Ableitungen gelernt haben, können wir jetzt eine Liste elementarer Stammfunktionen für Polynome, sin, cos, exp etc aufstellen.

(2) Linearkombination können als $\int_a^b \lambda f + \mu g = \lambda \int_a^b f + \mu \int_a^b g$ integriert werden.

(3) Bei Kompositionen oder Produkten kann manchmal eine Stammfunktion geraten werden (wie in VIII.4.6(iii)) oder es können die zwei folgenden Integrationsregeln verwendet werden. Aber auch hier muss man oft raten oder Verschiedenes ausprobieren. Es gibt keinen Algorithmus, der Sie immer zum Erfolg führt.

Satz VIII.4.8. *(Substitutionsregel) Es sei $f\colon I \to \mathbb{R}$ stetig und $\varphi\colon [a,b] \to \mathbb{R}$ sei eine C^1–Abbildung mit $\varphi([a,b]) \subset I$. Dann gilt*

$$\int_a^b f(\varphi(t))\varphi'(t) \,\mathrm{d}t = \int_{\varphi(a)}^{\varphi(b)} f(x) \,\mathrm{d}x.$$

Beweis. Es sei $F\colon I \to \mathbb{R}$ eine Stammfunktion von f. Dann gilt mit der Kettenregel

$$(F \circ \varphi)'(t) = F'(\varphi(t))\varphi'(t) = f(\varphi(t))\varphi'(t)$$

und weiter durch zweimalige Anwendung des Hauptsatzes, nämlich erst auf $(F \circ \varphi)'$ und dann auf f, folgt

$$\int_a^b f(\varphi(t))\varphi(t) \,\mathrm{d}t = (F \circ \varphi)(t)\big|_a^b = F(\varphi(b)) - F(\varphi(a)) = \int_{\varphi(a)}^{\varphi(b)} f(x) \,\mathrm{d}x$$

wie gewünscht. □

Beispiele VIII.4.9.

- Es gilt

$$\begin{aligned}
\int_0^{\sqrt{\frac{\pi}{2}}} 2t\cos(t^2) \,\mathrm{d}t &= \int_0^{\sqrt{\frac{\pi}{2}}} \varphi'(t) f(\varphi(t)) \,\mathrm{d}t \quad (f(x) := \cos(x),\ \varphi(t) := t^2) \\
&= \int_{\varphi(0)}^{\varphi(\sqrt{\frac{\pi}{2}})} f(x) \,\mathrm{d}x \quad \text{(Substitutionsregel)} \\
&= \int_0^{\frac{\pi}{2}} \cos x \,\mathrm{d}x = \sin x\big|_0^{\frac{\pi}{2}} = \sin\frac{\pi}{2} - \sin 0 = 1.
\end{aligned}$$

[4] Für $n \in \mathbb{N}$ kann man als Definitionsbereich jedes beliebige Intervall nehmen. Für negative ganzzahlige n darf das Intervall die Null nicht enthalten. Ist n keine ganze Zahl, so ist man sicher unterwegs mit einem Intervall das Teilmenge von $\mathbb{R}_{>0}$ ist, kann aber in Spezialfällen auch negative x erlauben (z.B. wenn man $x^{2/3} = \sqrt[3]{x^2}$ liest) muss dann aber bei der Formel für die Stammfunktion eventuell nochmal nachbessern.

- Es gilt
$$\begin{aligned}\int_1^2 (2t+1)^2\,\mathrm{d}t &= \frac{1}{2}\int_1^2 2(2t+1)^2\,\mathrm{d}t \quad (\text{nahrhafte Eins } 1=\frac{1}{2}\cdot 2)\\ &= \frac{1}{2}\int_1^2 \varphi'(t)f(\varphi(t))\,\mathrm{d}t \quad (f(x):=x^2,\ \varphi(t):=2t+1)\\ &= \frac{1}{2}\int_{\varphi(1)}^{\varphi(2)} f(x)\,\mathrm{d}x \quad (\text{Substitutionsregel})\\ &= \frac{1}{2}\int_3^5 x^2\,\mathrm{d}x = \frac{1}{2}\cdot\frac{x^3}{3}\Big|_3^5 = \frac{49}{3}.\end{aligned}$$
Man muss, bzw. kann, manche Integrale durch geschicktes Umschreiben erst auf die Form bringen, die für die Anwendung der Substitutionsregel nötig ist. Dies gilt nicht nur im Sinne, dass Konstanten ergänzt werden, sondern kann auch von x abhängige Terme umfassen.

Bemerkung VIII.4.10. Ein sehr effizienter Weg die Beispiele aus VIII.4.9 aufzuschreiben, ist die folgende, oft von Physiker:innen benutzte, Methode
$$\int_0^{\sqrt{\frac{\pi}{2}}} 2t\underset{\substack{\uparrow\\ x:=t^2\\ \Rightarrow \frac{\mathrm{d}x}{\mathrm{d}t}=2t\\ \Rightarrow \mathrm{d}x=2t\,\mathrm{d}t,}}{\cos(t^2)}\,\mathrm{d}t = \int_{x(0)}^{x(\sqrt{\frac{\pi}{2}})} \cos x\,\mathrm{d}x = \sin x\Big|_0^{\frac{\pi}{2}} = \sin\frac{\pi}{2} - \sin 0 = 1.$$
wobei wir von links nach rechts die $2t\,\mathrm{d}t$ durch $\mathrm{d}x$ ersetzen und
$$\int_1^2 (2t\underset{\substack{\uparrow\\ x:=2t+1\\ \Rightarrow \frac{\mathrm{d}x}{\mathrm{d}t}=2\\ \Rightarrow \mathrm{d}t=\frac{\mathrm{d}x}{2},}}{+}1)^2\,\mathrm{d}t = \int_{x(1)}^{x(2)} x^2\frac{\mathrm{d}x}{2} = \frac{1}{2}\cdot\frac{x^3}{3}\Big|_3^5 = \frac{49}{3}.$$
wo wir $\mathrm{d}t = \frac{\mathrm{d}x}{2}$ einsetzen. Beachten Sie hierbei aber:
- Links ist x eine Funktion, rechts eine Variable; dies nennt man im Englischen ‘abuse of notation’.
- Auflösen von $\frac{\mathrm{d}x}{\mathrm{d}t} = \cdots$ nach $\mathrm{d}x$ oder $\mathrm{d}t$ ist eine formale Rechnung; was $\mathrm{d}x$ und $\mathrm{d}t$ einzeln bedeuten, haben wir gar nicht definiert und wir wissen daher schon gar nicht, ob wir mit diesen Ausdrücken Bruchrechnen dürfen.
- Insbesondere das zweite Beispiel suggeriert, dass wir so jedes Integral einer Verkettung behandeln können, da wir nicht auf eine bestimmte Form des Integranden angewiesen sind, sondern durch die formale Rechnung mit $\mathrm{d}t$ und $\mathrm{d}x$ immer das, was fehlt, automatisch herausbekommen. Dies ist aber nicht so.
- Aufgrund der letzten beiden Punkte ist die obige Methode sehr mit Vorsicht zu genießen.

Bemerkung VIII.4.11. Wir können VIII.4.8 auch dazu benutzen, um Stammfunktionen zu bestimmen, weil $F(x) := \int_a^x f(t)\,\mathrm{d}t$ eine ist, wenn f stetig ist:
$$\int x\cos x^2\,\mathrm{d}x = \frac{1}{2}\int 2x\cos x^2\,\mathrm{d}x = \frac{1}{2}\int \cos u\,\mathrm{d}u = \frac{1}{2}\sin u + C = \frac{1}{2}\sin x^2 + C,$$
wobei wir hier mit $u := x^2$ substituiert haben und dann am Ende wieder re-substituieren müssen. Vorsicht! In der Rechnung oben ist verschleiert, dass u eine Funktion von x ist.

Beim Bestimmen von Stammfunktionen kann man am Ende durch Ableiten sehr leicht prüfen, ob tatsächlich eine Stammfunktion vorliegt. Tut man dies, so ist dies Beweis genug, und es ist nebensächlich, wie man die Stammfunktion gefunden hat.

Satz VIII.4.12. *(Partielle Integration) Es seien f, $g\colon I \to \mathbb{R}$ beides C^1–Abbildungen. Dann gilt für a, $b \in I$*

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = f(x)g(x)\big|_a^b - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

Beweis. Wir setzen $F := fg$. Dann gilt mit der Produktregel $F' = f'g + fg'$ und daher

$$\int_a^b f'g + \int_a^b fg' = \int_a^b F' = F\big|_a^b = fg\big|_a^b,$$

wobei die vorletzte Gleichung aus dem Hauptsatz folgt. Umstellen liefert die behauptete Gleichung. □

Beispiel VIII.4.13. Für $f(x) = x^2$ und $g(x) = e^x$ gilt

$$\begin{aligned}\int_0^1 e^x\,x^2\,\mathrm{d}x &= e^x\,x^2\,\big|_0^1 - \int_0^1 e^x\,2x\,\mathrm{d}x\\ &= e^1 1^2 - e^0 0^2 - 2\int_0^1 e^x\,x\,\mathrm{d}x\\ &= e - 2\Big(e^x\,x\,\big|_0^1 - \int_0^1 e^x\,1\,\mathrm{d}x\Big)\\ &= e - 2\big(e^1 1 - e^0 0 - e^x\big|_0^1\big) = e - 2\end{aligned}$$

Bemerkungen VIII.4.14.

(1) Es ist egal ob man mit $g'f$ oder $f'g$ anfängt. Wichtig ist, dass wir mit partieller Integration die Ableitung von einem Faktor auf den anderen hinüberschieben können.

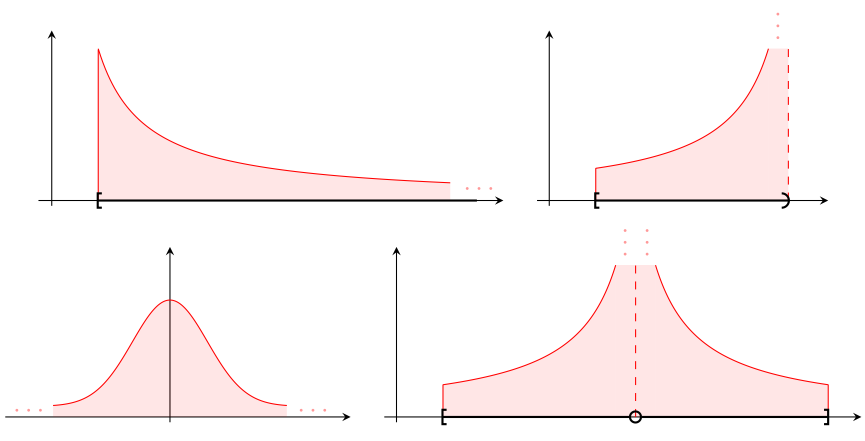
(2) Die Wahl am Anfang sollte schon so gemacht werden, dass das neue Integral einfacher zu berechnen ist als das alte. Wählt man in VIII.4.13 $f(x) = x^3/3$ und $g(x) = e^x$, dann erhält man

$$\int_0^1 e^x\,x^2\,\mathrm{d}x = e^x\,x^3/3\,\big|_0^1 - \int_0^1 e^x\,x^3/3\,\mathrm{d}x.$$

Dies ist zwar korrekt, aber es verschlimmert die Situation.

VIII.5. Uneigentliche Integrale

Bisher haben wir Integrale beschränkter Funktionen über beschränkten Intervallen definiert und berechnet. Jetzt erweitern wir dies und behandeln zum Beispiel Integrale von Funktionen wie unten.



Definition VIII.5.1. Es sei $a \in \mathbb{R}$ und $f\colon [a,\infty) \to \mathbb{R}$ eine Abbildung, sodass $f|_{[a,R]}$ für jedes $a < R < \infty$ R–integrierbar ist. Wir definieren

$$\int_a^\infty f(x)\,\mathrm{d}x := \lim_{R\to\infty} \int_a^R f(x)\,\mathrm{d}x,$$

falls dieser Grenzwert in $\mathbb{R}$ existiert. Wir nennen dann $\int_a^\infty f(x)\,\mathrm{d}x$ ein *uneigentliches Integral*. Manchmal sagen wir, dass das *Integral konvergiert*, wenn der obige Grenzwert existiert.

Beispiele VIII.5.2.

(1) Für $s > 1$ gilt

$$\int_1^\infty \frac{1}{x^s}\,\mathrm{d}x = \lim_{R\to\infty}\int_1^R x^{-s}\,\mathrm{d}x = \lim_{R\to\infty}\frac{x^{-s+1}}{-s+1}\Big|_1^R = \lim_{R\to\infty}\frac{1}{1-s}\big(\frac{1}{R^{s-1}}-1\big) = \frac{1}{s-1}.$$

(2) Dagegen konvergiert $\displaystyle\int_1^\infty \frac{1}{x}\,\mathrm{d}x$ nicht: $\displaystyle\int_1^R \frac{1}{x}\,\mathrm{d}x = \log R - \log 1 = \log R \xrightarrow{R\to\infty} \infty$

Definition VIII.5.3. Es sei $f\colon (a,b] \to \mathbb{R}$ mit $a < b$ in $\mathbb{R}$ und sei $f|_{[a+\varepsilon,b]}$ R–integrierbar für alle $0 < \varepsilon < b-a$. Wir definieren

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{\varepsilon\searrow 0}\int_{a+\varepsilon}^b f(x)\,\mathrm{d}x,$$

falls dieser Grenzwert existiert. Funktionen $f\colon [a,b) \to \mathbb{R}$ behandeln wir analog.

Beispiele VIII.5.4.

(1) Für $s < 1$ gilt

$$\int_0^1 \frac{1}{x^s}\,\mathrm{d}x = \lim_{\varepsilon\searrow 0}\int_{0+\varepsilon}^1 x^{-s}\,\mathrm{d}x = \lim_{\varepsilon\searrow 0}\frac{x^{-s+1}}{-s+1}\Big|_\varepsilon^1 = \lim_{\varepsilon\searrow 0}\frac{1}{1-s}\big(1-\varepsilon^{1-s}\big) = \frac{1}{1-s}.$$

(2) $\displaystyle\int_0^1 \frac{1}{x}\,\mathrm{d}x$ konvergiert nicht: $\displaystyle\int_\varepsilon^1 \frac{1}{x}\,\mathrm{d}x = \log 1 - \log\varepsilon \xrightarrow{\varepsilon\searrow 0} \infty.$

Definition VIII.5.5. Es sei $f\colon (a,b) \to \mathbb{R}$ mit $-\infty \leqslant a < b \leqslant +\infty$ und sei $f|_{[\alpha,\beta]}$ R–integrierbar für alle $[\alpha,\beta] \subset (a,b)$. Es sei $c \in (a,b)$ beliebig. Wir definieren

$$\int_a^b f(x)\,\mathrm{d}x := \lim_{\alpha\searrow a}\int_\alpha^c f(x)\,\mathrm{d}x + \lim_{\beta\nearrow b}\int_c^\beta f(x)\,\mathrm{d}x,$$

falls beide Grenzwert (in $\mathbb{R}$) existieren. Beachten Sie, dass VIII.3.11 in diesem Fall sicherstellt, dass $\int_a^b f(x)\,\mathrm{d}x$ unabhängig von c ist.

Beispiel VIII.5.6. Es gilt

$$\begin{aligned}\int_{-\infty}^\infty \frac{1}{1+x^2}\,\mathrm{d}x &= \lim_{R\to\infty}\int_{-R}^0 \frac{1}{1+x^2}\,\mathrm{d}x + \lim_{R\to\infty}\int_0^R \frac{1}{1+x^2}\,\mathrm{d}x\\ &= \lim_{R\to\infty}\Big(\arctan x\Big|_{-R}^0\Big) + \lim_{R\to\infty}\Big(\arctan x\Big|_0^R\Big)\\ &= \lim_{R\to\infty} -\arctan R + \lim_{R\to\infty}\arctan R = \pi.\end{aligned}$$

Die letzte verbleibende Definition (‘Singularität in der Mitte’) und weitere Beispiele behandeln wir auf dem Übungsblatt. Dort sehen wir auch, dass im Gegensatz zu VIII.5.2(2) und VIII.5.4(2), wo die Integrale immerhin *uneigentlich* gegen $+\infty$ konvergieren, auch Divergenz eintreten kann.

KAPITEL IX

Euklidische und unitäre Räume

IX.1. Skalarprodukte und Normen

Im Folgenden sei V stets ein Vektorraum über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Wir setzen am Anfang nicht voraus, dass V endlichdimensional ist.

Definition IX.1.1. Ein *Skalarprodukt* auf V ist eine Abbildung $\langle -, - \rangle \colon V \times V \to \mathbb{K}$, $(u, v) \mapsto \langle u, v \rangle$ mit folgenden Eigenschaften:

(SP1) $\forall\, \alpha_1, \alpha_2 \in \mathbb{K},\, u, v_1, v_2 \in V \colon \langle u, \alpha_1 v_1 + \alpha_2 v_2 \rangle = \alpha_1 \langle u, v_1 \rangle + \alpha_2 \langle u, v_2 \rangle$ (Linearität im 2. Argument),

(SP2) $\forall\, u, v \in V \colon \langle u, v \rangle = \overline{\langle v, u \rangle}$ (Symmetrie ($\mathbb{K} = \mathbb{R}$) bzw. Hermitezität ($\mathbb{K} = \mathbb{C}$)),

(SP3) $\forall\, v \in V \colon \langle v, v \rangle \geqslant 0$ und $(\langle v, v \rangle = 0 \iff v = 0)$ (Positive Definitheit).

Ist $\langle -, - \rangle$ ein Skalarprodukt auf V, so heißt das Paar $(V, \langle -, - \rangle)$ *Skalarproduktraum.* Im Fall $\mathbb{K} = \mathbb{R}$ spricht man von *euklidischen Räumen*, im Fall $\mathbb{K} = \mathbb{C}$ von *unitären Räumen.* Häufig schreibt man auch nur 'Es sei V ein Skalarproduktraum', ohne das Skalarprodukt explizit anzugeben.

Beispiele IX.1.2.

(1) $V = \mathbb{R}^n$ über $\mathbb{K} = \mathbb{R}$ mit

$$\langle u, v \rangle = \left\langle \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \right\rangle := \sum_{i=1}^{n} u_i v_i. \qquad \text{(Standardskalarprodukt des } \mathbb{R}^n\text{)}$$

(2) $V = \mathbb{C}^n$ über $\mathbb{K} = \mathbb{C}$ mit

$$\langle u, v \rangle = \left\langle \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \right\rangle := \sum_{i=1}^{n} \overline{u}_i v_i. \qquad \text{(Standardskalarprodukt des } \mathbb{C}^n\text{)}$$

(3) $V = \ell^2 = \big\{ (x_k)_{k \in \mathbb{N}} \subset \mathbb{K} \mid \sum_{k=1}^{\infty} |x_k|^2 < \infty \big\}$ über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ mit

$$\langle (x_k)_{k \in \mathbb{N}}, (y_k)_{k \in \mathbb{N}} \rangle := \sum_{i=1}^{\infty} \overline{x}_i y_i$$

heißt *Hilbert'scher Folgenraum.* Im Gegensatz zu (1) und (2), wo man außer den Skalarprodukteigenschaften nichts zeigen muß und diese durch 'scharfes Hingucken' sofort einsieht, muss man hier prüfen, dass ℓ^2 ein Vektorraum ist und zeigen, dass das Skalarprodukt wohldefiniert ist, dass also die Reihe $\sum_{i=1}^{\infty} \overline{x}_i y_i$ in $\mathbb{C}$ konvergiert, wenn die Folgen $(x_k)_{k \in \mathbb{N}}$, $(y_k)_{k \in \mathbb{N}}$ in ℓ^2 liegen. Am geschicktesten ist es hier, zunächst letzteres zu erledigen und zwar mithilfe der allgemeinen Form der Cauchy–Schwarz–Ungleichung, die wir gleich in IX.1.6(ii) beweisen werden. Dann kann man IX.1.6(i) benutzen und damit das Unterraumkriterium für $\ell^2 \subset \mathrm{Abb}(\mathbb{N}, \mathbb{K})$ nachweisen.

(4) $V = \mathrm{C}([a, b], \mathbb{R})$ über $\mathbb{K} = \mathbb{R}$ mit

$$\langle f, g \rangle = \int_a^b f(x) g(x) \,\mathrm{d}x$$

ist ebenfalls ein Skalarproduktraum. Hier zeigt man $\langle f, f \rangle = 0 \implies f = 0$ mithilfe der Stetigkeit von f. Alle anderen Eigenschaften folgen direkt aus den Rechenregeln für das Riemann–Integral.

Bemerkungen IX.1.3.

- Für $\mathbb{K} = \mathbb{R}$ ist ein Skalarprodukt auch im 1. Argument linear, weil
$$\begin{aligned} \langle \alpha_1 u_1 + \alpha_2 u_2, v \rangle &\overset{\text{(SP2)}}{=} \langle v, \alpha_1 u_1 + \alpha_2 u_2 \rangle \\ &\overset{\text{(SP1)}}{=} \alpha_1 \langle v, u_1 \rangle + \alpha_2 \langle v, u_2 \rangle \\ &\overset{\text{(SP2)}}{=} \alpha_1 \langle u_1, v \rangle + \alpha_2 \langle u_2, v \rangle. \end{aligned}$$
Für $\mathbb{K} = \mathbb{C}$ ist ein Skalarprodukt im 1. Argument *konjugiert–linear*, d.h. Skalare können heraus- bzw. hineingezogen werden, müssen aber dabei komplex konjugiert werden:
$$\begin{aligned} \langle \alpha_1 u_1 + \alpha_2 u_2, v \rangle &\overset{\text{(SP2)}}{=} \overline{\langle v, \alpha_1 u_1 + \alpha_2 u_2 \rangle} \\ &\overset{\text{(SP1)}}{=} \overline{\alpha_1 \langle v, u_1 \rangle + \alpha_2 \langle v, u_2 \rangle} \\ &= \overline{\alpha}_1 \overline{\langle v, u_1 \rangle} + \overline{\alpha}_2 \overline{\langle v, u_2 \rangle} \\ &\overset{\text{(SP2)}}{=} \overline{\alpha}_1 \langle u_1, v \rangle + \overline{\alpha}_2 \langle u_2, v \rangle. \end{aligned}$$
- Für $\mathbb{K} = \mathbb{R}$ ist ein Skalarprodukt eine *Bilinearform*, für $\mathbb{K} = \mathbb{C}$ spricht man von einer Sesquilinearform (sesqui = eineinhalb).
- Je nach Literatur wird Linearität im 1. Argument gefordert und es folgt dann die (konjugierte) Linearität im 2. Argument.

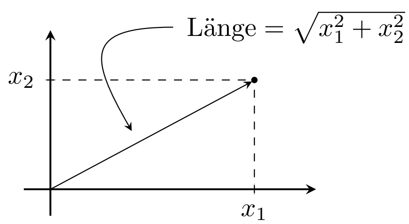
Definition IX.1.4. Eine Abbildung $\|\cdot\| \colon V \to \mathbb{R}$ heißt *Norm*, falls gilt:

(N1) $\forall\, v \in V \colon \|v\| \geqslant 0$ und $\|v\| = 0 \iff v = 0$ (Positive Definitheit),

(N2) $\forall\, \alpha \in \mathbb{K},\ v \in V \colon \|\alpha v\| = |\alpha| \|v\|$ (Homogenität),

(N3) $\forall\, u,\ v \in V \colon \|u + v\| \leqslant \|u\| + \|v\|$ (Dreiecksungleichung).

Beispiel IX.1.5. In $\mathbb{R}^n$ ist
$$\left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\| := \sqrt{x_1^2 + \cdots + x_n^2}$$
eine Norm; in $\mathbb{R}^2$ (oder $\mathbb{R}^3$) kann die Norm von $x = \binom{x_1}{x_2}$ als Länge des Vektors x interpretiert werden:



Das hatten wir uns in I.2 schon angesehen.

Dass obiges tatsächlich eine Norm definiert, kann man direkt nachrechnen. Wir machen dies aber hier nicht, weil es auch sofort aus dem nächsten Satz folgt.

Satz IX.1.6. *Es sei $(V, \langle\text{-}, \text{-}\rangle)$ ein Skalarproduktraum. Dann gilt*

(i) $\|v\| := \sqrt{\langle v, v \rangle}$ *für $v \in V$ definiert eine Norm auf V.*

(ii) $\forall\, u, v \in V \colon |\langle u, v \rangle| \leqslant \|u\| \|v\|$ *(Cauchy–Schwarz–Ungleichung).*

Im Spezialfall des $\mathbb{R}^3$ hatten wir die Cauchy–Schwarz–Ungleichung in I.2.6 behandelt.

Beweis. Wir zeigen zunächst die Cauchy–Schwarz–Ungleichung:
Ist $v = 0$, dann sind beide Seiten Null.
Es sei $v \neq 0$. Setze $z := u - \frac{\langle v, u \rangle}{\langle v, v \rangle} v$, d.h. $u = z + \frac{\langle v, u \rangle}{\langle v, v \rangle} v$. Wir benutzen, dass das SP im 1. Argument konjugiert linear ist, und erhalten
$$(*) \quad \langle z, v \rangle = \left\langle u - \frac{\langle v, u \rangle}{\langle v, v \rangle} v, v \right\rangle = \langle u, v \rangle - \frac{\overline{\langle v, u \rangle}}{\langle v, v \rangle} \langle v, v \rangle = \langle u, v \rangle - \langle u, v \rangle = 0.$$

Mit der Linearität im 2. Argument gilt außerdem

$$\begin{aligned}
\|u\|^2 &= \langle u,u\rangle = \Big\langle z+\frac{\langle v,u\rangle}{\langle v,v\rangle}v, z+\frac{\langle v,u\rangle}{\langle v,v\rangle}v\Big\rangle \\
&= \langle z,z\rangle + \Big\langle z, \frac{\langle v,u\rangle}{\langle v,v\rangle}v\Big\rangle + \Big\langle \frac{\langle v,u\rangle}{\langle v,v\rangle}v, z\Big\rangle + \Big\langle \frac{\langle v,u\rangle}{\langle v,v\rangle}v, \frac{\langle v,u\rangle}{\langle v,v\rangle}v\Big\rangle \\
&= \langle z,z\rangle + \underbrace{\frac{\langle v,u\rangle}{\langle v,v\rangle}\langle z,v\rangle + \overline{\frac{\langle v,u\rangle}{\langle v,v\rangle}}\langle v,z\rangle}_{\overset{(*)}{=}0} + \overline{\frac{\langle v,u\rangle}{\langle v,v\rangle}}\frac{\langle v,u\rangle}{\langle v,v\rangle}\langle v,v\rangle \\
&= \|z\|^2 + \frac{|\langle v,u\rangle|^2}{\langle v,v\rangle^2}\langle v,v\rangle \geqslant \frac{|\langle v,u\rangle|^2}{\|v\|^2}.
\end{aligned}$$

Multiplikation mit $\|v\|^2$ und Wurzelziehen zeigt $|\langle u,v\rangle| \leqslant \|u\|\|v\|$ wie gewünscht, und beendet damit den Beweis von (ii).

Wir zeigen nun (i):

Da nach (SP3) $\langle v,v\rangle \geqslant 0$ ist, folgt zunächst, dass $\|\cdot\|: V \to \mathbb{R}$ wohldefiniert ist und nach Definition nur nichtnegative Werte annimmt. Gilt $0 = \|v\| = \sqrt{\langle v,v\rangle}$, so muss $\langle v,v\rangle = 0$ sein und dann nach (SP3) also $v=0$. Das gibt (N1).

Für (N2): Es gilt $\|\alpha v\|^2 = \langle \alpha v, \alpha v\rangle = \overline{\alpha}\alpha\langle v,v\rangle = |\alpha|^2\|v\|^2$ was durch Wurzelziehen die Behauptung zeigt.

Für (N3) benutzen wir die Cauchy–Schwarz–Ungleichung, die wir mit CSU abkürzen.

(N3) Es gilt

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle = |\langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle| \\
&\leqslant \|u\|^2 + |\langle u,v\rangle| + |\langle v,u\rangle| + \|v\|^2 \\
&\leqslant \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\
&= (\|u\|+\|v\|)^2,
\end{aligned}$$

wobei die erste Abschätzung die Dreiechsungleichung für den Betrag komplexer Zahlen ist und die zweite die CSU. Wurzelziehen zeigt nun die gewünsche Ungleichung $\|u+v\| \leqslant \|u\| + \|v\|$. Dies beendet auch den Beweis von (i). □

Bemerkungen IX.1.7.

- In jedem Skalarproduktraum gilt die *Parallelogrammgleichung*

$$\forall\, u,v \in V\colon \|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2.$$

- In euklidischen/unitären Räumen gelten die *Polarisationsidentitäten*

$$\begin{aligned}
\langle u,v\rangle &= \frac{1}{4}\big(\|u+v\|^2 - \|u-v\|^2\big) \quad (\mathbb{K}=\mathbb{R}) \\
\langle u,v\rangle &= \frac{1}{4}\big(\|u+v\|^2 - \|u-v\|^2 + i\|u-iv\|^2 - i\|u+iv\|^2\big) \quad (\mathbb{K}=\mathbb{C}).
\end{aligned}$$

  Dies zeigen wir in einer Aufgabe.

IX.2. Orthogonalität

Definition IX.2.1. Es sei V ein Skalarproduktraum.

(i) Zwei Vektoren $u,v \in V$ heißen *orthogonal*, falls $\langle u,v\rangle = 0$ gilt.

(ii) Für $M \subset V$ heißt $M^\perp := \big\{u \in V \mid \forall\, m \in M\colon \langle m,u\rangle = 0\big\}$ das *orthogonale Komplement von M*.

(iii) Eine endliche oder abzählbare Menge $(v_i)_{i\in I} \subset V$ heißt *Orthonormalsystem (ONS)*, falls gilt

$$\forall\, i,k \in I\colon \langle v_i, v_k\rangle = \delta_{ik} = \begin{cases} 1, & \text{falls } i=k, \\ 0, & \text{sonst.}\end{cases}$$

Beispiele IX.2.2.

(1) In $\mathbb{R}^n$ und $\mathbb{C}^n$ bilden die Einheitsvektoren $(e_1, \dots, e_n)$ jeweils ein ONS.
(2) In ℓ^2 ist $(e_i)_{i\in\mathbb{N}}$ mit $e_i = (0, \dots, 0, 1, 0, \dots)$, wobei die 1 an der i–ten Stelle steht, ein ONS.
(3) In $\mathrm{C}[-\pi, \pi]$ mit Skalarprodukt $\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x$ bilden die Funktionen

$$f_1(x) = \tfrac{1}{\sqrt{2\pi}},\ f_2(x) = \tfrac{1}{\sqrt{2\pi}}\sin x,\ f_3(x) = \tfrac{1}{\sqrt{2\pi}}\cos x,$$
$$f_4(x) = \tfrac{1}{\sqrt{2\pi}}\sin 2x,\ f_5(x) = \tfrac{1}{\sqrt{2\pi}}\cos 2x,\ f_6(x) = \tfrac{1}{\sqrt{2\pi}}\sin 3x,\ \dots$$

ein ONS.

Lemma IX.2.3. *Es sei V ein Skalarproduktraum. Dann gilt:*

(i) *Für alle Teilmengen $M \subset V$ ist $M^\perp$ ein Untervektorraum von V.*
(ii) *Jedes ONS von V ist eine linear unabhängige Menge*[1].
(iii) *Es sei V ein n–dimensionaler Skalarproduktraum und $(v_1, \dots, v_n)$ ein ONS. Dann ist $(v_1, \dots, v_n)$ eine Basis von V.*

Den Beweis behandeln wir in einer Lernwerkstatt. □

Definition IX.2.4. Basen, die gleichzeitig ein Orthonormalsystem bilden, nennen wir *Orthonormalbasen, (ONB).*

Satz IX.2.5. *Es sei V ein n–dimensionaler Skalarproduktraum und es sei $(v_1, \dots, v_n)$ eine ONB.*

(i) *Jedes $v \in V$ hat die eindeutige Darstellung*
$$v = \sum_{k=1}^{n} \langle v_k, v\rangle v_k$$
bzgl. der ONB. D.h. die Koeffizienten sind $\langle v_k, v\rangle$.
(ii) *Für $u, v \in V$ gilt*
$$\langle u, v\rangle = \sum_{k=1}^{n} \overline{\langle v_k, u\rangle}\langle v_k, v\rangle \Big(= \sum_{k=1}^{n} \langle u, v_k\rangle\langle v_k, v\rangle \Big).$$
(iii) *Für $v \in V$ gilt die* Parseval'sche Gleichung
$$\|v\|^2 = \sum_{k=1}^{n} |\langle v_k, v\rangle|^2.$$

Beweis. (i) Dass wir v nach jeder Basis und damit auch nach der gegebenen ONB eindeutig entwickeln können, wissen wir. Es sei also $v = c_1v_1 + \dots + c_nv_n$ mit geeigneten $c_i \in \mathbb{K}$. Dann folgt für $1 \leqslant i \leqslant n$
$$\langle v_i, v\rangle = \Big\langle v_i, \sum_{k=1}^{n} c_k v_k \Big\rangle = \sum_{k=1}^{n} c_k \underbrace{\langle v_i, v_k\rangle}_{=\delta_{ik}} = c_i.$$
(ii) Es gilt (mit der Darstellung aus (i) für die erste Gleichung):
$$\langle u, v\rangle = \Big\langle \sum_{k=1}^{n}\langle v_k, u\rangle v_k, \sum_{j=1}^{n}\langle v_j, v\rangle v_j \Big\rangle = \sum_{k,j=1}^{n} \overline{\langle v_k, u\rangle}\langle v_j, v\rangle \underbrace{\langle v_k, v_j\rangle}_{=\delta_{kj}} = \sum_{k=1}^{n} \overline{\langle v_k, u\rangle}\langle v_k, v\rangle.$$
(iii) Hier gilt schließlich mit (ii):
$$\|v\|^2 = \langle v, v\rangle = \sum_{k=1}^{n} \overline{\langle v_k, v\rangle}\langle v_k, v\rangle = \sum_{k=1}^{n} |\langle v_k, v\rangle|^2.$$
□

[1] Erinnern Sie sich bitte daran, dass eine unendliche Menge $X \subset V$ linear unabhängig ist, wenn jede endliche Teilmenge von X linear unabhängig ist.

Satz IX.2.6. *(Satz über die Orthogonalprojektion) Es sei V ein Skalarproduktraum, es sei $(v_1, \dots, v_\ell)$ ein ONS und W sei der ℓ-dimensionale Untervektorraum von V, der von den v_1 bis v_ℓ erzeugt wird: $W := \mathsf{Span}\{v_1, \dots, v_\ell\}$. Dann gilt:*

(i) *Jedes $v \in V$ lässt sich eindeutig schreiben als $v = w + w^\perp$ mit $w \in W$ und $w^\perp \in W^\perp$ und zwar als*

$$w := \sum_{k=1}^{\ell} \langle v_k, v \rangle v_k, \quad w^\perp := v - w.$$

(ii) *Die Abbildung*

$$P\colon V \to V,\ Pv := w$$

ist linear, es gilt $P^2 = P$ und $\operatorname{Bild} P = W$*. Man nennt P die* orthogonale Projektion auf W.

(iii) *Es gilt $\langle Pu, v \rangle = \langle u, Pv \rangle$ für alle $u, v \in V$.*

(iv) *Es gilt $\|v - Pv\| = \min\big\{\|v - w\| \;\big|\; w \in W\big\}$ für jedes $v \in V$.*

Beweis. (i) Da $w \in W$ nach Definition gilt, muss für die Existenz der Zerlegung nur gezeigt werden, dass $w^\perp \in W^\perp$ liegt, was wiederum nach Definition genau dann der Fall ist, wenn $\langle u, w^\perp \rangle = 0$ für alle $u \in W$ gilt. Wir nehmen also einen beliebigen Vektor $u \in W$ und entwickeln diesen mit IX.2.5(i) nach der ONB $(v_1, \dots, v_\ell)$ von W, d.h.

$$u = \sum_{k=1}^{\ell} \langle v_k, u \rangle v_k.$$

Außerdem gilt

$$w^\perp = v - w = v - \sum_{k=1}^{\ell} \langle v_k, v \rangle v_k.$$

Daraus folgt

$$\begin{aligned}
\langle u, w^\perp \rangle &= \Big\langle \sum_{k=1}^{\ell} \langle v_k, u \rangle v_k, v - \sum_{j=1}^{\ell} \langle v_j, v \rangle v_j \Big\rangle \\
&= \sum_{k=1}^{\ell} \overline{\langle v_k, u \rangle} \langle v_k, v \rangle - \sum_{k,j=1}^{n} \overline{\langle v_k, u \rangle} \langle v_j, v \rangle \underbrace{\langle v_k, v_j \rangle}_{=\delta_{kj}} \\
&= \sum_{k=1}^{\ell} \overline{\langle v_k, u \rangle} \langle v_k, v \rangle - \sum_{k=1}^{n} \overline{\langle v_k, u \rangle} \langle v_k, v \rangle = 0,
\end{aligned}$$

womit die Existenz gezeigt ist.

Um die Eindeutigkeit der Zerlegung zu zeigen, nehmen wir an, dass $v = w + w^\perp$ und $v = u + u^\perp$ mit einem $u \in W$ und $u^\perp \in W^\perp$. Dann folgt

$$(*) \qquad 0 = w + w^\perp - (u + u^\perp) = \underbrace{w - u}_{\in W} + \underbrace{w^\perp - u^\perp}_{\in W^\perp}$$

und mit der Definition von $W^\perp$ gilt $\langle w - u, w^\perp - u^\perp \rangle = 0$. Auflösen von $(*)$ nach $w^\perp - u^\perp$ zeigt weiter $w^\perp - u^\perp = -(w - u)$. Damit folgt

$$0 = \langle w - u, w^\perp - u^\perp \rangle = \langle w - u, -(w - u) \rangle = -\|w - u\|^2$$

und (N1) impliziert dann, dass $w = u$ sein muss. Dann liefert $(*)$ aber sofort auch $w^\perp = u^\perp$.

(ii) P ist linear, da $\langle$-,-$\rangle$ im 2. Argument linear ist. Dass $P^2 = P$ gilt, rechnen wir für $v \in V$ nach:

$$\begin{aligned}
P^2 v &= P(Pv) = P\Big(\sum_{k=1}^{\ell} \langle v_k, v \rangle v_k\Big) = \sum_{k=1}^{\ell} \langle v_k, v \rangle P v_k \\
&= \sum_{k=1}^{\ell} \langle v_k, v \rangle \sum_{j=1}^{\ell} \langle v_j, v_k \rangle v_j = \sum_{k=1}^{\ell} \langle v_k, v \rangle v_k = Pv.
\end{aligned}$$

Dass Bild $P = W$ gilt, folgt daraus, dass für $w \in W$ die Entwicklung von w nach der ONB $(v_1, \dots, v_\ell)$ von W mit Pw übereinstimmt.

(iii) Es seien $u, v \in V$ beliebig. Entsprechend (i) zerlegen wir $v = w + w^\perp$ und $u = x + x^\perp$ mit $w, x \in W$ und $w^\perp, x^\perp \in W^\perp$. Dann gilt

$$\langle Pu, v\rangle = \langle x, w + w^\perp\rangle = \langle x, w\rangle + \underbrace{\langle x, w^\perp\rangle}_{=0} = \langle x, w\rangle + \underbrace{\langle x^\perp, w\rangle}_{=0} = \langle x + x^\perp, w\rangle = \langle u, Pv\rangle,$$

wobei wir benutzt haben, dass $Pv = w$ und $Pu = x$ gilt und dass $\langle y, z\rangle = 0$ falls $y \in W$ und $z \in W^\perp$ sind (oder umgekehrt).

(iv) $Pv \in W$ impliziert $\|v - Pv\| \geqslant \inf\big\{\|v - u\| \,\big|\, u \in W\big\}$ und es muss

$$(*) \quad \forall\, u \in W\colon \|v - Pv\| \leqslant \|v - u\|$$

gezeigt werden. Es sei $w \in W$ beliebig. Wegen (i) haben wir $\langle v - Pv, Pv - w\rangle = 0$, da $v - Pv \in W^\perp$ und $Pv - w \in W$ gelten. Daraus folgt

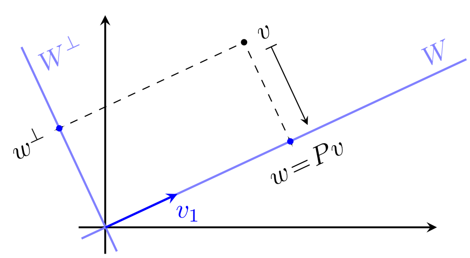
$$\begin{aligned} \|v - w\|^2 &= \|\underbrace{(v - Pv)}_{=:x} + \underbrace{(Pv - w)}_{=:y}\|^2 = \langle x + y, x + y\rangle \\ &= \langle x, x\rangle + \underbrace{\langle x, y\rangle + \langle y, x\rangle}_{=0} + \langle y, y\rangle = \|x\|^2 + \|y\|^2 \\ &= \|v - Pv\|^2 + \|Pv - w\|^2 \geqslant \|v - Pv\|^2. \end{aligned}$$

Dies liefert nach Wurzelziehen genau die Ungleichung in $(*)$. Damit haben wir

$$\|v - Pv\| = \inf\big\{\|v - u\| \,\big|\, u \in W\big\}$$

und insbesondere ist das Infimum ein Minimum, weil $Pv \in W$ gilt. □

Bemerkung IX.2.7. In $V = \mathbb{R}^2$ und für $\ell = 1$ haben wir die folgende Veranschaulichung:



D.h. $W = \mathsf{Span}\{v_1\}$ ist eine Gerade durch den Ursprung und $W^\perp$ ist diejenige Gerade durch den Ursprung, die senkrecht auf W steht. Die Abbildung P schickt dann einen Punkt $v \in \mathbb{R}^2$ auf genau den Punkt $w \in W$ der senkrecht unter v liegt. Zeichnen Sie die Vektoraddition $w + (v - w) = v$ ein. Dann kann $v - w$ als Pfeil vom Punkt w zum Punkt v interpretiert werden und dieser steht senkrecht auf W. Formal gilt das, weil $\langle w, v - w\rangle = \langle w, w^\perp\rangle = 0$. IX.2.6 sagt nun insbesondere noch, dass sich nichts mehr ändert, wenn man $w = Pv$ ein zweites Mal projiziert und dass $w = Pv$ genau der Punkt in W ist, der minimalen Abstand zu v hat.

Beachten Sie: Wären wir im Raum $\mathbb{R}^3$ und $W = \mathsf{Span}\{v_1\}$ ist eindimensional, so ist $W^\perp$ genau die Ebene durch den Ursprung, die von der Geraden W senkrecht durchstoßen wird. Wäre in $\mathbb{R}^3$ der Raum $W = \mathsf{Span}\{v_1, v_2\}$ zweidimensional, d.h. eine Ebene durch den Ursprung, dann ist umgekehrt $W^\perp$ genau diejenige Gerade welche senkrecht auf W steht und durch den Ursprung geht.

Satz IX.2.8. *(Gram–Schmidt–Orthonormalisierungsverfahren) Es sei V ein n–dimensionaler Skalarproduktraum. Dann gilt:*

(i) *V besitzt eine ONB.*

(ii) *Jedes ONS* $(v_1,\dots,v_\ell)$ *lässt sich zu einer ONB* $(v_1,\dots,v_n)$ *ergänzen*[2].

Beweis. Wir zeigen $(*)$: Zu einem System linear unabhängiger Vektoren $u_1,\dots,u_m$ existiert stets ein ONS $(v_1,\dots,v_m)$ mit $\mathsf{Span}\{u_1,\dots,u_m\} = \mathsf{Span}\{v_1,\dots,v_m\}$.

Die Aussage (i) des Satzes folgt dann durch Betrachtung einer beliebigen Basis $(u_1,\dots,u_n)$ aus dem Spezialfall $m=n$ von $(*)$. Aussage (ii) folgt, indem man die Vektoren $v_1\dots,v_\ell$ zu einer Basis ergänzt und dann das folgende Verfahren anwendet, von dem wir sehen werden, dass es die $v_1,\dots,v_\ell$ unverändert lässt:

(1) Da $(u_1,\dots,u_m)$ linear unabhängig sind, ist $u_1\neq 0$ und
$$v_1 := \frac{u_1}{\|u_1\|} = \frac{1}{\|u_1\|}\cdot u_1$$
ist wohldefiniert und hat Norm 1. Außerdem gilt $\mathsf{Span}\{u_1\} = \mathsf{Span}\{v_1\}$.

(2) Es sei P_1 die orthogonale Projektion auf $W_1 := \mathsf{Span}\{u_1\} = \mathsf{Span}\{v_1\}$. Dann gilt $u_2 - P_1u_2 \in W^\perp$ nach IX.2.6(i) und $u_2 - P_1u_2 \neq 0$, da u_1,u_2 linear unabhängig sind. Wir setzen also
$$v_2 := \frac{u_2 - P_1u_2}{\|u_2 - P_1u_2\|}$$
und erhalten einen Vektor mit Norm 1, $\langle v_1,v_2\rangle = 0$ und $\mathsf{Span}\{u_1,u_2\} = \mathsf{Span}\{v_1,v_2\}$.

$\vdots$

(m) Angenommen $v_1,\dots,v_{m-1}$ sind bereits konstruiert. Es sei dann P_{m-1} die orthogonale Projektion auf $W_{m-1} := \mathsf{Span}\{u_1,\dots,u_{m-1}\} = \mathsf{Span}\{v_1,\dots,v_{m-1}\}$. Dann ist $u_m - P_{m-1}u_m \in W_{m-1}^\perp$ und $u_m - P_{m-1}u_m \neq 0$. Wir setzen
$$v_m := \frac{u_m - P_{m-1}u_m}{\|u_m - P_{m-1}u_m\|}$$
und erhalten einen Vektor mit Norm 1, $\langle v_m,v_k\rangle = 0$ für $k=1,\dots,m-1$ und $\mathsf{Span}\{u_1,\dots,u_m\} = \mathsf{Span}\{v_1,\dots,v_m\}$.

□

IX.3. Adjungierte Abbildungen

Bemerkungen IX.3.1. Jetzt und im Folgenden sei $(V,\langle\text{-},\text{-}\rangle)$ stets ein n–dimensionaler Skalarproduktraum. Wir benutzen die folgenden Bezeichnungen:

(i) $\mathrm{End}(V) = \{T\colon V\to V \mid T \text{ linear}\}$.

(ii) $V^* = \{f\colon V\to\mathbb{K} \mid f \text{ linear}\}$ heißt *Dualraum*; seine Elemente heißen *Funktionale* oder Linearformen) auf V.

(iii) Da $\langle\text{-},\text{-}\rangle$ im 2. Argument linear ist, ist für festes $u\in V$ die Abbildung $f_u\colon V\to\mathbb{K}$, $f_u(v) := \langle u,v\rangle$ ein Funktional. Der nächste Satz zeigt, dass in der Tat alle Funktionale von dieser Form sind.

Satz IX.3.2. *(Satz von Riesz; endlichdimensionale Version) Es sei V ein n–dimensionaler Skalarproduktraum und $f\colon V\to\mathbb{K}$ sei linear. Dann existiert genau ein $u\in V$ mit $f(v)=\langle u,v\rangle$ für alle $v\in V$.*

Beweis. Es sei $(v_1,\dots,v_n)$ eine ONB in V. Wir wissen aus der Mathematik 2, dass f eindeutig bestimmt ist durch die Werte auf dieser ONB. Es sei $\alpha_i := f(v_i)\in\mathbb{K}$. Wenn wir ein beliebiges $v\in V$ nach der ONB entwickeln als $v = \beta_1v_1+\cdots+\beta_nv_n$ mit $\beta_i\in\mathbb{K}$, so folgt
$$f(v) = f\Big(\sum_{i=1}^n \beta_iv_i\Big) = \sum_{i=1}^n \beta_i f(v_i) = \sum_{i=1}^n \beta_i\alpha_i.$$

[2] Hieraus folgt insbesondere, dass wir in der Situation von Satz IX.2.6 unter der Zusatzannahme $n := \dim V < \infty$ die $v_1,\dots,v_\ell$ zu einer ONB $(v_1,\dots,v_n)$ ergänzen können und dass dann das orthogonale Komplement $W^\perp$ gerade von den $v_{\ell+1},\dots,v_n$ aufgespannt wird. Dies klärt die Frage, wie man $W^\perp$ ausrechnen oder sich vorstellen kann (vgl. auch das Bild in IX.2.7).

Wir setzen $u := \sum_{i=1}^n \overline{\alpha}_i v_i$. Dann gilt für v wie oben

$$\langle u, v\rangle = \Big\langle \sum_{i=1}^n \overline{\alpha}_i v_i, \sum_{j=1}^n \beta_j v_j \Big\rangle = \sum_{i,j=1}^n \alpha_i \beta_j \langle v_i, v_j\rangle = \sum_{i=1}^n \alpha_i \beta_i.$$

Daraus folgt, dass $u := \sum_{i=1}^n \overline{\alpha}_i v_i$ die gewünschte Eigenschaft hat und außerdem eindeutig ist: Es sei $\tilde{u} \in V$ ein weiterer Vektor mit $f_{\tilde{u}} = f$. Dann gilt

$$\forall\, v \in V\colon \langle \tilde{u}, v\rangle = f(v) = \langle u, v\rangle,$$

woraus folgt, dass $\langle \tilde{u} - u, v\rangle = 0$ gilt für alle $v \in V$, also insbesondere $\langle \tilde{u} - u, \tilde{u} - u\rangle = 0$. Wegen (SP3) bzw. (N1) erzwingt dies $\tilde{u} - u = 0$. □

Satz IX.3.3. *(Adjungierte lineare Abbildung) Es sei V ein n–dimensionaler Skalarproduktraum. Zu jedem $T \in \mathrm{End}(V)$ existiert genau ein $T^{ad} \in \mathrm{End}(V)$ mit der Eigenschaft*

$$\forall\, u, v \in V\colon \langle u, Tv\rangle = \langle T^{ad} u, v\rangle.$$

Die Abbildung $T \mapsto T^{ad}$ von $\mathrm{End}(V)$ in sich hat die folgenden Eigenschaften:

(i) $\forall\, \alpha \in \mathbb{K},\ T, S \in \mathrm{End}(V)\colon (\alpha T + S)^{ad} = \overline{\alpha} T^{ad} + S^{ad}$,

(ii) $\forall\, T, S \in \mathrm{End}(V)\colon (T \circ S)^{ad} = S^{ad} \circ T^{ad}$,

(iii) $\forall\, T \in \mathrm{End}(V)\colon T^{ad^{ad}} = T$.

Definition IX.3.4. Die Abbildung T^{ad} wie in IX.3.5 heißt die zu T *adjungierte Abbildung.*

Beweis. Es sei $u \in V$ fest. Wir betrachten

$$g\colon V \to \mathbb{K},\ g(v) := \langle u, Tv\rangle.$$

Dann ist $g \in V^*$ und nach dem Satz von Riesz (IX.3.2) existiert genau ein $u' \in V$ mit $g = \langle u', \text{-}\rangle$, also

$$\forall\, v \in V\colon \langle u, Tv\rangle = \langle u', v\rangle.$$

Wir können also $T^{ad}\colon V \to V$ definieren als $T^{ad} u := u'$ und müssen jetzt prüfen, dass T^{ad} linear ist. Dies machen wir wieder mit der (einfachen aber wichtigen) Äquivalenz

$$x = y \iff \forall\, v \in V\colon \langle x, v\rangle = \langle y, v\rangle.$$

Es seien $\alpha \in \mathbb{K},\ u, w \in V$ fest. Für $v \in V$ gilt dann

$$\begin{aligned} \langle T^{ad}(\alpha u + w), v\rangle &= \langle \alpha u + w, Tv\rangle \\ &= \overline{\alpha}\langle u, Tv\rangle + \langle w, Tv\rangle \\ &= \overline{\alpha}\langle T^{ad} u, v\rangle + \langle T^{ad} w, v\rangle \\ &= \langle \alpha T^{ad} u + T^{ad} w, v\rangle \end{aligned}$$

also $T^{ad}(\alpha u + w) = \alpha T^{ad} u + T^{ad} w$.

(i)–(iii) beweist man ähnlich; dies behandeln wir in der Lernwerkstatt. □

Bemerkung IX.3.5. Es sei V ein n–dimensionaler Skalarproduktraum und sei $\mathcal{B} = (b_1, \dots, b_n)$ eine feste ONB. Es sei $T \in \mathrm{End}(V)$. Dann ist

$$M_{\mathcal{B}}(T) = \big[(Tb_1)_{\mathcal{B}} \ \cdots \ (Tb_n)_{\mathcal{B}}\big] =: \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} =: A \ \text{ und } \ Tb_k = \sum_{i=1}^n a_{ik} b_i.$$

Daraus folgt

$$\langle b_j, Tb_k\rangle = \Big\langle b_j, \sum_{i=1}^n a_{ik} b_i \Big\rangle = \sum_{i=1}^n a_{ik} \langle b_j, b_i\rangle = a_{jk}.$$

D.h. die Einträge der Darstellungsmatrix von T können durch obige Formel berechnet werden, solange wir eine ONB benutzen, und das machen wir ab jetzt in Skalarprodukträumen immer.

Es sei nun $B := M_{\mathcal{B}}(T^{ad})$. Dann gilt

$$b_{jk} = \langle b_j, T^{ad} b_k\rangle = \overline{\langle T^{ad} b_k, b_j\rangle} = \overline{\langle b_k, Tb_j\rangle} = \overline{a}_{kj},$$

d.h. $B = \overline{A}^t$ über $\mathbb{K} = \mathbb{C}$ bzw. $B = A^t$ über $\mathbb{K} = \mathbb{R}$. Ist also $T \in \mathrm{End}(V)$ durch eine Matrix A dargestellt, so kann die Darstellungsmatrix der adjungierten Abbildung T^{ad} durch Transponieren und Konjugieren bestimmt werden, bzw. nur durch transponieren, wenn $\mathbb{K} = \mathbb{R}$ ist.

Definition IX.3.6. Es sei $A \in M(n \times n, \mathbb{C})$. Dann bezeichnet $\overline{A}^t$ die *hermitesch konjugierte Matrix* oder auch die *adjungierte Matrix*.

Definition IX.3.7. Es sei V ein n–dimensionaler Skalarproduktraum. $T \in \mathrm{End}(V)$ heißt

(i) *normal*, falls $T \circ T^{ad} = T^{ad} \circ T$ gilt,

(ii) *selbstadjungiert*, falls $T = T^{ad}$ gilt,

(iii) *unitär* ($\mathbb{K} = \mathbb{C}$) bzw. *orthogonal* ($\mathbb{K} = \mathbb{R}$) falls $T^{ad} = T^{-1}$ gilt,

(iv) *isometrisch*, falls $\|Tv\| = \|v\|$ für alle $v \in V$ gilt.

Bemerkung IX.3.8. Mit IX.3.5 ergeben sich für die oben definierten Eigenschaften für $A = M_{\mathcal{B}}(T)$ mit einer ONB $\mathcal{B}$:

T normal ($\mathbb{K} = \mathbb{C}$)	$\iff$	$\overline{A}^t A = A\overline{A}^t$	*normale*
T normal ($\mathbb{K} = \mathbb{R}$)	$\iff$	$A^t A = A A^t$	*Matrix*
T selbstadjungiert ($\mathbb{K} = \mathbb{C}$)	$\iff$	$A = \overline{A}^t$	*selbstadjungierte* oder *hermitesche Matrix*
T selbstadjungiert ($\mathbb{K} = \mathbb{R}$)	$\iff$	$A = A^t$	*symmetrische Matrix*
T unitär ($\mathbb{K} = \mathbb{C}$)	$\iff$	$\overline{A}^t = A^{-1}$	*unitäre Matrix*
T orthogonal ($\mathbb{K} = \mathbb{R}$)	$\iff$	$A^t = A^{-1}$	*orthogonale Matrix*

Satz IX.3.9. *Es sei V ein n–dimensionaler Skalarproduktraum.*

(i) *Selbstadjungierte und unitäre/orthogonale Abbildungen sind stets normal.*

(ii) *$T \in \mathrm{End}(V)$ ist isometrisch genau dann, wenn $\langle Tu, Tv\rangle = \langle u, v\rangle$ für alle $u, v \in V$ gilt.*

(iii) *Unitäre/orthogonale Abbildungen sind stets isometrisch.*

Beweis. (i) folgt direkt.

(ii) '$\Longleftarrow$' Es gilt $\|v\|^2 = \langle v, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2$ woraus durch Wurzelziehen die Behauptung folgt.

'$\Longrightarrow$' Wir benutzen die Polarisationsidentität und führen den Fall $\mathbb{K} = \mathbb{C}$ aus. Im Fall $\mathbb{K} = \mathbb{R}$ müssen unten jeweils die letzten zwei Terme in der Klammer weggelassen werden:

$$\begin{aligned}\langle Tu, Tv\rangle &= \frac{1}{4}\big(\|T(u+v)\|^2 - \|T(u-v)\|^2 + i\|T(u-iv)\|^2 - i\|T(u+iv)\|^2\big)\\ &= \frac{1}{4}\big(\|u+v\|^2 - \|u-v\|^2 + i\|u-iv\|^2 - i\|u+iv\|^2\big)\\ &= \langle u, v\rangle.\end{aligned}$$

(iii) Es gilt $\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^{ad}Tv, v\rangle = \langle T^{-1}Tv, v\rangle = \langle v, v\rangle = \|v\|^2$. □

Satz IX.3.10. *Es sei $A \in M(n \times n, \mathbb{K})$.*

(i) *Die Matrix A ist genau dann unitär/orthogonal, wenn die Spalten (äquivalent die Zeilen) eine ONB von $\mathbb{K}^n$ bzgl. des Standardskalarprodukts bilden.*

(ii) *Wenn A unitär/orthogonal ist, gilt $|\det A| = 1$.*

(iii) *Wenn A unitär/orthogonal ist, dann haben alle Eigenwerte von A den Betrag 1; ist $\mathbb{K} = \mathbb{R}$, so heißt dies, dass die Eigenwerte nur -1 oder 1 sein können.*

Beweis. (i) Es sei $\mathbb{K} = \mathbb{C}$ und

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nk} & \cdots & a_{nn} \end{pmatrix}, \quad \text{d.h. } \overline{A}^t = \begin{pmatrix} \overline{a}_{11} & \cdots & \overline{a}_{n1} \\ \vdots & & \vdots \\ \overline{a}_{1k} & \cdots & \overline{a}_{nk} \\ \vdots & & \vdots \\ \overline{a}_{1n} & \cdots & \overline{a}_{nn} \end{pmatrix}.$$

Weil $\overline{A}^t = A^{-1}$ ist, also $\overline{A}^t A = E_n$ gilt, folgt

$$\left\langle \begin{pmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix}, \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix} \right\rangle = \sum_{j=1}^n \overline{a}_{jk} a_{ji} = \underbrace{\begin{pmatrix} \overline{a}_{1k} & \cdots & \overline{a}_{nk} \end{pmatrix}}_{\substack{k\text{–te Zeile} \\ \text{von } \overline{A}^t}} \underbrace{\begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}}_{\substack{i\text{–te Spalte} \\ \text{von } A}} = (\overline{A}^t A)_{ki} = (E_n)_{ki} = \delta_{ki}.$$

Dies zeigt, dass die Spalten von A eine ONB bilden. Für die Zeilen nutzt man $A\overline{A}^t = E_n$ aus und geht analog vor. Im Fall $\mathbb{K} = \mathbb{R}$ ersetzt man $\overline{A}^t$ durch A^t.

(ii) Es gilt

$$1 = \det E_n = \det \overline{A}^t A = \det \overline{A}^t \det A = \overline{\det A^t} \det A = \overline{\det A} \det A = |\det A|^2$$

für $\mathbb{K} = \mathbb{C}$; für $\mathbb{K} = \mathbb{R}$ lassen Sie die komplexe Konjugation weg.

(iii) Wenn A unitär bzw. orthogonal ist, dann folgt mit IX.3.9(iii), dass $A\colon \mathbb{K}^n \to \mathbb{K}^n$, $x \mapsto Ax$ isometrisch ist. Es sei nun $\lambda \in \sigma(A)$ und u ein zugehöriger Eigenvektor. Dann gilt $\|u\| = \|Au\| = \|\lambda u\| = |\lambda|\|u\|$ und folglich $|\lambda| = 1$. □

KAPITEL X

Eigenwerte und Normalformen in Spezialfällen

Wir hatten im vorigen Abschnitt spezielle Arten von Endomorphismen kennengelernt. In diesen Fällen kann man Aussagen über deren Eigenwerte treffen, sowie in günstigen Fällen etwas zur Diagnonalisierbarkeit sagen.

X.1. Der selbstadjungierte Fall

Im Folgenden sei stets $(V, \langle\text{-},\text{-}\rangle)$ ein endlich-dimensionaler Skalarproduktraum und $T\colon V \to V$ linear und selbstadjungiert, das heißt es gilt $T = T^{ad}$, was äquivalent ist zu $\langle u, Tv\rangle = \langle Tu, v\rangle$ für alle $u, v \in V$.

Satz X.1.1. *Es seien T und V wie oben. Dann gilt:*

(i) *Jeder Eigenwert von T ist reell*[1].

(ii) *Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal zueinander.*

Beweis. (i) Es sei $Tu = \lambda u$ mit $u \neq 0$. Dann gilt einerseits

$$\langle u, Tu\rangle = \langle u, \lambda u\rangle = \lambda\langle u, u\rangle$$

und andererseits, weil T selbstadjungiert ist,

$$\langle u, Tu\rangle = \langle Tu, u\rangle = \langle \lambda u, u\rangle = \overline{\lambda}\langle u, u\rangle.$$

Da $\langle u, u\rangle \neq 0$ ist, folgt $\lambda = \overline{\lambda}$.

(ii) Es gelte $Tu = \lambda u$ und $Tv = \mu v$ mit $\lambda \neq \mu$ und $u \neq 0 \neq v$. Dann gilt wieder einerseits

$$\langle u, Tv\rangle = \langle u, \mu v\rangle = \mu\langle u, v\rangle$$

und andererseits, weil T selbstadjungiert ist,

$$\langle u, Tv\rangle = \langle Tu, v\rangle = \langle \lambda u, v\rangle = \overline{\lambda}\langle u, v\rangle \overset{\text{(i)}}{=} \lambda\langle u, v\rangle.$$

Daraus folgt $(\mu - \lambda)\langle u, v\rangle = 0$. Da nach Voraussetzung $\mu - \lambda \neq 0$ ist, muss $\langle u, v\rangle = 0$ sein. □

Lemma X.1.2. *Es sei $V \neq \{0\}$ ein endlich-dimensionaler Skalarproduktraum. Dann hat jede selbstadjungierte lineare Abbildung einen Eigenvektor zu einem reellen Eigenwert.*

Beweis.

(1) Für $\mathbb{K} = \mathbb{C}$ zerfällt das charakteristische Polynom in Linearfaktoren, hat also mindestens eine Nullstelle $\lambda \in \mathbb{C}$. Diese ist ein EW, also nach X.1.1 reell und es gibt nach Definition einen Eigenvektor zum Eigenwert λ.

(2) Für $\mathbb{K} = \mathbb{R}$ wählen wir eine ONB $\mathcal{B}$ in V. Dann ist $A := M_{\mathcal{B}}(T) \in M(n \times n, \mathbb{R})$ symmetrisch. Fassen wir A als komplexe Matrix in $M(n \times n, \mathbb{C})$ auf, dann ist sie hermitesch; es gilt also $A = \overline{A}^t$. Damit ist $A\colon \mathbb{C}^n \to \mathbb{C}^n$, $x \mapsto Ax$ selbstadjungiert und hat nach (1) einen reellen Eigenwert. D.h. P_A hat eine reelle Nullstelle und damit hat T einen reellen Eigenwert.

□

Korollar X.1.3. *Jede hermitsche komplexe Matrix und jede symmetrische reelle Matrix hat einen reellen Eigenwert.*

[1] Über $\mathbb{K} = \mathbb{R}$ ist diese Aussage nach Definition wahr, aber über $\mathbb{K} = \mathbb{C}$ natürlich nicht!

Bemerkung X.1.4. Vorsicht: Komplexe symmetrische Matrizen haben nicht unbedingt einen reellen Eigenwert. Zum Beispiel hat $\left(\begin{smallmatrix} i & 0 \\ 0 & i \end{smallmatrix}\right)$ offenbar keinen.

Lemma X.1.5. *Es sei V ein endlichdimensionaler Skalarproduktraum und $T \in \mathrm{End}(V)$ sei selbstadjungiert. Es sei weiter $U \subset V$ ein Unterraum, der bzgl. T invariant ist, d.h. es gilt $TU \subset U$. Dann gilt auch $TU^\perp \subset U^\perp$ (d.h. $U^\perp$ ist auch T–invariant) und $T|_U^U$ sowie $T|_{U^\perp}^{U^\perp}$ sind wieder selbstadjungiert.*

Beweis. Es sei $v \in U^\perp$. Es gilt also $\langle u, v\rangle = 0$ für alle $u \in U$. Wir müssen zeigen, dass $Tv \in U^\perp$ ist, also dass $\langle u, Tv\rangle = 0$ ist für alle $u \in U$. Wir rechnen nach, dass für $u \in U$ gilt

$$\langle u, Tv\rangle = \langle \underbrace{Tu}_{\in U}, v\rangle = 0,$$

weil U unter T invariant ist. Dass die Einschränkungen wieder selbstadjungiert sind, ist klar. □

Satz X.1.6. *(Hauptsatz über selbstadjungierte lineare Abbildungen) Jede selbstadjungierte lineare Abbildung T auf einem endlich-dimensionalen Skalarproduktraum V ist diagonalisierbar.*

Beweis. Wir müssen zeigen, dass V eine ONB aus EVen besitzt. Das machen wir durch Induktion nach $n := \dim V$.

$n = 1$: Hier ist die Aussage trivial, denn alle linearen Abbildungen $V \to V$ sind von der Form $x \mapsto \lambda x$ mit einem $\lambda \in \mathbb{K}$.

$n-1$ auf n: Es sei $n \geqslant 2$ und gelte die Aussage für $(n-1)$–dimensionale Vektorräume. V sei ein Vektorraum mit Dimension n. Nach X.1.2 existiert ein Eigenvektor v und wir können ohne Einschränkung annehmen, dass $\|v\| = 1$ gilt. Wir setzen $U := \mathsf{Span}\{v\}$, also $\dim U = 1$. Für $u \in U$ gilt $Tu = \lambda u \in U$, wobei λ den zu v gehörigen Eigenwert bezeichne. Insbesondere ist also $TU \subset U$. Gemäß X.1.5 ist dann ebenfalls $TU^\perp \subset U^\perp$ und $T|_{U^\perp}^{U^\perp}$ ist selbstadjungiert. Außerdem ist $\dim U^\perp = n-1$, d.h. nach Induktionsvorraussetzung existiert für $U^\perp$ eine ONB $(v_1, \ldots, v_{n-1})$ aus Eigenvektoren zu $T|_{U^\perp}^{U^\perp}$, also auch zu T. Es folgt, dass $(v_1, \ldots, v_{n-1}, v)$ in V eine ONB aus Eigenvektoren ist. □

Bemerkung X.1.7. Es seien $\mu_1, \ldots, \mu_r$ die paarweise verschiedenen Eigenwerte von T und $\mathrm{E}_{\mu_1}, \ldots, \mathrm{E}_{\mu_r}$ die zugehörigen Eigenräume. Nach einer Übungsaufgabe gilt

$$V = \mathrm{E}_{\mu_1} \oplus \ldots \oplus \mathrm{E}_{\mu_r}.$$

Also hat jedes $v \in V$ eine eindeutige Darstellung $v = w_1 + w_2 + \cdots + w_r$ mit $w_i \in \mathrm{E}_{\mu_i}$. Nach X.1.1(ii) wissen wir jetzt noch, dass die Eigenräume orthogonal aufeinander stehen, d.h. $w_i = P_i v$, wenn P_i die orthogonale Projektion auf E_{μ_i} bezeichnet (vgl. IX.2.6). Also gilt

$$v = P_1 v + \ldots + P_r v.$$

Beachten Sie, dass $P_i v \in \mathrm{E}_{\mu_i}$ gilt, und damit $P_i v$ entweder Null ist oder ein Eigenvektor zum Eigenwert μ_i ist. Anwenden von T liefert also

$$Tv = TP_i v + \ldots + TP_r v = \mu_1 P_1 v + \ldots + \mu_r P_r v.$$

Da dies für jedes $v \in V$ gilt, können wir obiges auch auch schreiben als

$$T = \mu_1 P_1 + \ldots + \mu_r P_r.$$

Letzteres wird die *Spektraldarstellung* von T genannt, denn $\{\mu_1, \ldots, \mu_r\} = \sigma(T)$ ist gerade das Spektrum von T.

X.2. Normalform symmetrischer und hermitescher Matrizen

Satz X.2.1. *(Satz über die Diagonalisierbarkeit hermitescher/symmetrischer Matrizen) Jede komplexe hermitesche bzw. reelle symmetrische Matrix ist diagonalisierbar. Genauer:*

(i) *Für jede hermitesche Matrix $A \in M(n \times n, \mathbb{C})$ existiert eine unitäre Matrix $S \in M(n \times n, \mathbb{C})$, sodass $SAS^{-1} = SA\overline{S}^t$ Diagonalgestalt hat.*
(ii) *Für jede symmetrische Matrix $A \in M(n \times n, \mathbb{R})$ existiert eine orthogonale Matrix $S \in M(n \times n, \mathbb{R})$, sodass $SAS^{-1} = SAS^t$ Diagonalgestalt hat.*

Beweis. Nach X.1.7 existiert eine ONB $(b_1, \ldots, b_n)$ aus Eigenvektoren (in $\mathbb{K}^n$). Aus Kapitel II wissen wir, dass dann SAS^{-1} diagonal ist und dass die Spalten von S^{-1} gerade die Vektoren $b_1, \ldots b_n$ sind. Nach IX.3.10(i) ist S^{-1} dann unitär bzw. orthogonal, also $\overline{S^{-1}}^t = (S^{-1})^{-1} = S$. Weil zweimaliges Konjugieren die Identität ist und ebenso zweimaliges Transponieren, ist dann auch S unitär bzw. orthogonal. $\square$

Bemerkung X.2.2. Die Matrix S war weder in Kapitel II durch die Forderung, dass SAS^{-1} diagonal ist, eindeutig bestimmt, noch ist sie das in der Situation von X.2.1.

Beispiele X.2.3.

(i) Die Matrizen $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right) \in M(2 \times 2, \mathbb{R})$ und $\left(\begin{smallmatrix} 1 & i \\ -i & -1 \end{smallmatrix}\right) \in M(2 \times 2, \mathbb{C})$ sind diagonalisierbar nach X.2.1 und zwar ohne weitere Rechnung.

(ii) Vorsicht: X.2.1 ist keine Äquivalenz. Zum Beispiel ist $\left(\begin{smallmatrix} i & 0 \\ 0 & i \end{smallmatrix}\right) \in M(2 \times 2, \mathbb{C})$ bereits eine Diagonalmatrix, ohne hermitesch zu sein.

X.3. Normalformen normaler, unitärer und orthogonaler Endomorphismen

Für normale lineare Abbildungen in *komplexen* endlichdimensionalen Skalarprodukträumen gilt ein zu X.1.6 analoges Resultat:

Satz X.3.1. *Es sei V ein komplexer, endlichdimensionaler Skalarproduktraum und $T \in \mathrm{End}(V)$ sei normal, also $T^{ad} \circ T = T \circ T^{ad}$. Dann ist T diagonalisierbar.*

Beweis. Da wir über $\mathbb{C}$ arbeiten, hat P_T mindestens eine Nullstelle, also existiert mindestens ein Eigenwert λ, der jetzt aber nicht unbedingt reell ist. Wie vorher ist $U := \mathrm{E}_\lambda$ unter T invariant. Da T normal ist, gilt $T^{ad}U \subset U$, denn für $v \in U$ haben wir

$$Tv = \lambda v \overset{T^{ad}(\cdot)}{\Longrightarrow} T^{ad}(Tv) = \lambda(T^{ad}v) \overset{T^{ad}\circ T = T\circ T^{ad}}{\Longrightarrow} T(T^{ad}v) = \lambda(T^{ad}v),$$

also ist $T^{ad}v$ Eigenvektor zum Eigenwert λ und damit ist $T^{ad}v \in \mathrm{E}_\lambda = U$.

Wir zeigen jetzt, dass $TU^\perp \subset U^\perp$ und $T^{ad}U^\perp \subset U^\perp$ gelten:

Es sei $v \in U^\perp$, d.h. $\langle u, v\rangle = 0$ für alle $u \in U$. Dann gilt $\langle u, Tv\rangle = \langle T^{ad}u, v\rangle = 0$ für $u \in U$, also $Tv \in U^\perp$.

Analog haben wir $\langle u, T^{ad}v\rangle = \langle Tu, v\rangle = 0$ für beliebiges $u \in U$, also $T^{ad}v \in U^\perp$.

Da $T|_{U^\perp}^{U^\perp} \in \mathrm{End}(U^\perp)$ normal ist, kann nun der Induktionsbeweis von X.1.6 kopiert werden. $\square$

Bemerkung X.3.2. Da wir den Beweis übernommen haben, erhalten wir auch im Fall normaler komplexer Matrizen $A \in M(n \times n, \mathbb{C})$, dass es eine unitäre Matrix S gibt, so dass SAS^{-1} Diagonalgestalt hat.

Beispiel X.3.3. Die Matrix $\left(\begin{smallmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{smallmatrix}\right)$ ist normal, d.h. über $\mathbb{C}$ diagonalisierbar. Sie hat komplexe Eigenwerte und kann daher über $\mathbb{R}$ nicht diagonalisierbar sein.

Korollar X.3.4.

(i) *Unitäre lineare Abbildungen sind stets diagonalisierbar.*

(ii) *Unitäre Matrizen $A \in M(n \times n, \mathbb{C})$ sind stets diagonalisierbar.*

Bemerkung X.3.5. Orthogonale Matrizen $A \in M(n \times n, \mathbb{R})$ sind im Allgemeinen nicht diagonalisierbar. So hat zum Beispiel die Drehmatrix $A = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$ für $\theta \in (0, 2\pi) \setminus \{\pi\}$ keine reellen Eigenwerte, ist aber orthogonal:

$$P_A(X) = (\cos\theta - X)^2 + \sin^2\theta = X^2 - 2\cos\theta X + \cos^2\theta + \sin^2\theta = X^2 - 2\cos\theta X + 1$$

und dies hat die Nullstellen $\cos\theta \pm i \sin\theta$.

Diese sind also für $\theta \in (0, 2\pi) \setminus \{\pi\}$ nicht reell.

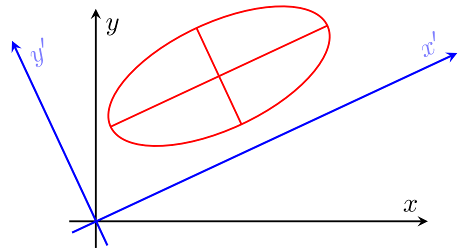
Die Normalform für orthogonale Matrizen ist daher etwas komplizierter:

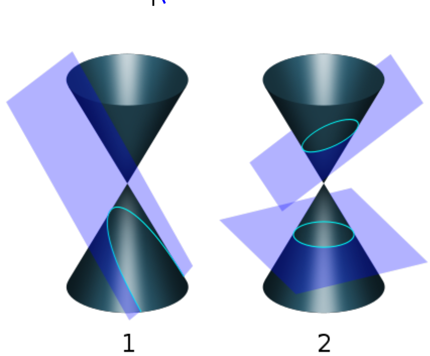
Satz X.3.6. *Es sei $T \in \mathrm{End}_{\mathbb{R}}(V)$ orthogonal, wobei V ein endlich-dimensionaler euklidischer Vektorraum ist. Dann gibt es eine ONB $\mathcal{B}$ von V, so dass*

$$M_{\mathcal{B}}(T) = \begin{pmatrix} E_r & 0 & \dots & \dots & 0 \\ 0 & -E_s & \ddots & & \vdots \\ \vdots & \ddots & D_1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & D_k \end{pmatrix}.$$

Hierbei ist $D_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$ mit $\theta_i \in (0, 2\pi) \setminus \{\pi\}$ und $s, r, k \geqslant 0$. □

Bemerkung X.3.7. Orthogonale Diagonalisierbarkeit ist die zentrale Zutat bei der sogenannten *Hauptachsentransformation*, bei der man z.B. für eine Ellipse und allgemeinere Kegelschnitte dasjenige Koordinatensystem sucht, in welchem man die Halbachsen ablesen kann:





Quelle: wiki commons; `https://commons.wikimedia.org/wiki/File:Conic_sections_with_plane.svg`

Die Ellipse wird hierbei durch eine Gleichung beschrieben, in der eine symmetrische Matrix vorkommt. Die orthogonale Matrix enspricht einem Koordinatenwechsel, der die Ellipse nicht deformiert, weil Längen und Winkel unter orthogonalen Abbildungen erhalten bleiben, vgl. IX.3.9.

X.4. Positive und negative Definitheit

Wichtig für das nächste Semester, und zwar bei der Extremwertberechnung in mehreren Veränderlichen, sind die folgenden Begriffe:

Definition X.4.1. Ein symmetrische Matrix $A \in M(n \times n, \mathbb{R})$ heißt

(i) *positiv semidefinit* ($A \geqslant 0$), falls $\forall\, u \in \mathbb{R}^n \colon \langle u, Au\rangle \geqslant 0$,

(ii) *positiv definit* ($A > 0$), falls $\forall\, u \in \mathbb{R}^n \backslash \{0\} \colon \langle u, Au\rangle > 0$,

(iii) *negativ semidefinit* ($A \leqslant 0$), falls $\forall\, u \in \mathbb{R}^n \colon \langle u, Au\rangle \leqslant 0$,

(iv) *negativ definit* ($A < 0$), falls $\forall\, u \in \mathbb{R}^n \backslash \{0\} \colon \langle u, Au\rangle < 0$,

(v) *indefinit* ($A < 0$), falls $\exists\, u, v \in \mathbb{R}^n \colon \langle u, Au\rangle < 0 \wedge \langle v, Av\rangle > 0$.

Beispiele X.4.2.

(1) Ist $A = E_n$, so ist $\langle u, Au\rangle = \langle u, u\rangle \geqslant 0$ und $\langle u, Au\rangle = 0$ gilt genau dann wenn $u = 0$ ist, d.h. die Einheitsmatrix E_n ist positiv definit und $A = -E_n$ ist entsprechend negativ definit.

(2) Für $A = \left(\begin{smallmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{smallmatrix}\right)$ erhalten wir

$$\left\langle \left(\begin{smallmatrix} u_1 \\ u_2 \\ u_3 \end{smallmatrix}\right), \left(\begin{smallmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{smallmatrix}\right) \left(\begin{smallmatrix} u_1 \\ u_2 \\ u_3 \end{smallmatrix}\right) \right\rangle = \lambda_1 u_1^2 + \lambda_2 u_2^2 + \lambda_3 u_3^2$$

und es folgt:

- A ist positiv semidefinit $\iff \forall\, i\colon \lambda_i \geqslant 0$.
- A ist positiv definit $\iff \forall\, i\colon \lambda_i > 0$.
- A ist negativ semidefinit $\iff \forall\, i\colon \lambda_i \leqslant 0$.
- A ist negativ definit $\iff \forall\, i\colon \lambda_i < 0$.
- A ist indefinit $\iff \exists\, \lambda_i < 0,\ \lambda_j > 0$.

Schließlich gilt der folgende

Satz X.4.3. *Es sei $A \in M(n \times n, \mathbb{R})$ symmetrisch; insbesondere ist $\sigma(A) \subset \mathbb{R}$. Dann gilt*

(1) $A \geqslant 0 \iff$ *Alle Eigenwerte von A sind größer gleich Null.*

(2) $A > 0 \iff$ *Alle Eigenwerte von A sind echt größer Null.*

Entsprechende Aussagen gelten für $A < 0$, $A \leqslant 0$ und A indefinit.

Beweis. Nach X.1.6 existiert eine ONB $(v_1, \dots, v_n)$ aus Eigenvektoren mit Eigenwerten $\lambda_1, \dots, \lambda_n$. Wir stellen $u \in \mathbb{R}^n$ bezüglich dieser Basis dar, d.h. $u = \alpha_1 v_1 + \dots + \alpha_n v_n$. Dann gilt

$$\langle u, Au\rangle = \Big\langle \sum_{i=1}^n \alpha_i v_i, \sum_{j=1}^n \alpha_j A v_j \Big\rangle = \Big\langle \sum_{i=1}^n \alpha_i v_i, \sum_{j=1}^n \alpha_j \lambda_j v_j \Big\rangle = \sum_{i=1}^n \alpha_i^2 \lambda_i,$$

woraus beide Aussagen folgen. □

KAPITEL XI

Gleichmäßige Konvergenz

Wir wiederholen den Begriff der gleichmäßigen Konvergenz und betrachten dann Eigenschaften wie Stetigkeit, Integrierbarkeit und Ableitbarkeit.

XI.1. Wiederholung

Wiederholung XI.1.1. Es sei $D \subset \mathbb{R}$ und $(f_n)_{n\in\mathbb{N}}$ mit $f_n\colon D \to \mathbb{R}$ sei eine *Folge von Funktionen* (oder auch *Funktionenfolge*). Es sei $f\colon D \to \mathbb{R}$.

(i) $(f_n)_{n\in\mathbb{N}}$ *konvergiert punktweise* gegen f, falls $\lim_{n\to\infty} f_n(x) = f(x)$ für alle $x \in D$ gilt, also

$$\forall\, x \in D,\, \varepsilon > 0\; \exists\, n_0\; \forall\, n \geqslant n_0\colon |f_n(x) - f(x)| < \varepsilon.$$

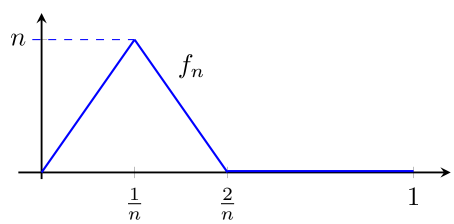
(ii) $(f_n)_{n\in\mathbb{N}}$ *konvergiert gleichmäßig* gegen f, falls gilt:

$$\forall\, \varepsilon > 0\; \exists\, n_0\; \forall\, n \geqslant n_0,\, x \in D\colon |f_n(x) - f(x)| < \varepsilon.$$

Bemerkung XI.1.2. Konvergiert f_n gleichmäßig gegen f, so auch punktweise, aber die Umkehrung gilt nicht. Wir nennen f die *Grenzfunktion*.

Beispiel XI.1.3. Für $n \geqslant 2$ definieren wir

$$f_n\colon [0,1] \to \mathbb{R},\quad f_n(x) := \begin{cases} n^2 x, & \text{falls } 0 \leqslant x \leqslant \frac{1}{n} \\ 2n - n^2 x, & \text{falls } \frac{1}{n} < x \leqslant \frac{2}{n} \\ 0, & \text{falls } \frac{2}{n} < x \leqslant 1 \end{cases}$$



- Die Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ konvergiert punktweise gegen $f \equiv 0$: Ist $x = 0$, dann ist $f_n(x) = 0$ für alle $n \geqslant 2$, also gilt $f_n(x) \to 0$. Für $x > 0$ wählen wir $n_0 \geqslant \max\{2, \frac{2}{x}\}$. Für $n \geqslant n_0$ gilt dann $\frac{2}{n} \leqslant x$ und damit gilt $f_n(x) = 0$. Also $f_n(x) \to 0$.
- Die Funktionenfolge $(f_n)_{n\in\mathbb{N}}$ konvergiert *nicht* gleichmäßig gegen $f \equiv 0$: Wir zeigen

  $$\exists\, \varepsilon > 0\; \forall\, n_0\; \exists\, n \geqslant n_0,\, x \in [0,1]\colon |f_n(x) - f(x)| \geqslant \varepsilon.$$

  Es sei $\varepsilon = 1$. Für n_0 beliebig setzen wir $n := n_0$ und $x := \frac{1}{n_0}$. Dann gilt $|f_n(x) - f(x)| = |f_{n_0}(\frac{1}{n_0}) - 0| = n_0 \geqslant 1 = \varepsilon$.
- Die Folge $(f_n)_{n\in\mathbb{N}}$ konvergiert in der Tat gegen keine Funktion $f\colon [0,1] \to \mathbb{R}$ gleichmäßig: Gäbe es eine solche Funktion, so würde $(f_n)_{n\in\mathbb{N}}$ insbesondere punktweise gegen f konvergieren und dann wäre $f \equiv 0$.

Definition XI.1.4. Es sei $f\colon D \subset \mathbb{R} \to \mathbb{R}$ gegeben. Wir definieren

$$\|f\|_\infty := \sup\bigl\{|f(x)| \bigm| x \in D\bigr\} = \sup_{x\in D} |f(x)| \in \mathbb{R}_{\geqslant 0} \cup \{\infty\}$$

und nennen $\|f\|_\infty$ die *Supremumsnorm von f* (über D).

Bemerkungen XI.1.5.
(i) f ist genau dann beschränkt, wenn $\|f\|_\infty < \infty$.

(ii) f_n konvergiert genau dann gleichmäßig gegen f, wenn $\|f_n - f\|_\infty \to 0$. Beachten Sie, dass letzteres so zu verstehen ist, dass insbesondere ein n_0 existiert, sodass $\|f_n - f\|_\infty < \infty$ für $n \geqslant n_0$ gilt, weil $(\|f_n - f\|_\infty)_{n\in\mathbb{N}}$ a priori eine Folge in $\mathbb{R}_{\geqslant 0} \cup \{\infty\}$ ist.

XI.2. Gleichmäßige Konvergenz und Stetigkeit

Satz XI.2.1. *Es sei $D \subset \mathbb{R}$, $f_n\colon D \to \mathbb{R}$ für $n \in \mathbb{N}$ und $f\colon D \to \mathbb{R}$. Falls $(f_n)_{n\in\mathbb{N}}$ gleichmäßig gegen f konvergiert und alle f_n stetig sind, dann ist f ebenfalls stetig.*

Beweis. Es sei $x_0 \in D$. Wir müssen zeigen:

$$\forall\, \varepsilon > 0\ \exists\, \delta > 0\ \forall\, x \in D\colon |x - x_0| < \delta \ \Rightarrow\ |f(x) - f(x_0)| < \varepsilon.$$

Es sei $\varepsilon > 0$ gegeben. Wir wissen:

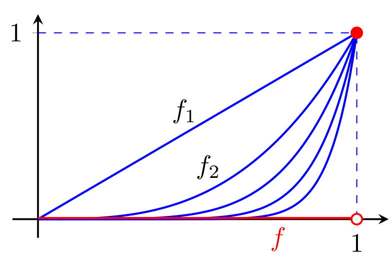
- Wenn f_n gleichmäßig gegen f konvergiert, so gilt: $\exists\, n_0\ \forall\, \tilde{x} \in D\colon |f_{n_0}(\tilde{x}) - f(\tilde{x})| < \frac{\varepsilon}{3}$.
- f_{n_0} stetig $\implies \exists\, \delta > 0\ \forall\, x \in D\colon |x - x_0| < \delta \ \Rightarrow\ |f_{n_0}(x) - f_{n_0}(x_0)| < \frac{\varepsilon}{3}$.

Wir wählen nun $\delta > 0$ wie oben und geben uns $x \in D$ beliebig mit $|x - x_0| < \delta$ vor. Dann folgt mit der Dreiecksungleichung

$$\begin{aligned} |f(x) - f(x_0)| &\leqslant |f(x) - f_{n_0}(x)| + |f_{n_0}(x) - f_{n_0}(x_0)| + |f_{n_0}(x_0) - f(x_0)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$ □

Bemerkung XI.2.2. Hat man nur punktweise Konvergenz, so stimmt es im Allgemeinen *nicht*, dass sich Stetigkeit von den f_n auf die Grenzfunktion f vererbt:

Beispiel XI.2.3. Für $n \geqslant 1$ definieren wir $f_n\colon [0,1] \to \mathbb{R}$, $f_n(x) = x^n$.



Es liegt punktweise Konvergenz vor: $\lim\limits_{n\to\infty} f_n(x) = \lim\limits_{n\to\infty} x^n = \left.\begin{cases} 0, & \text{für } x \in [0,1) \\ 1, & \text{für } x = 1 \end{cases}\right\} =: f(x).$

Man sieht direkt, dass gleichmäßige Konvergenz nicht gelten kann: Wir sehen, dass f nicht stetig ist, so dass mit XI.2.1 folgt, dass f_n nicht gleichmäßig gegen f konvergieren kann. Direkt kann man argumentieren, dass es für $\varepsilon < 1$ und n beliebig ein $x \in [0,1)$ gibt, sodass $|f_n(x) - f(x)| = x^n \geqslant \varepsilon$ gilt.

XI.3. Gleichmäßige Konvergenz und Integration

Satz XI.3.1. *Es seien $f_n\colon [a,b] \to \mathbb{R}$, $n \in \mathbb{N}$, stetig und $(f_n)_{n\in\mathbb{N}}$ konvergiere gleichmäßig gegen eine (dann nach XI.2.1 stetige) Funktion $f\colon [a,b] \to \mathbb{R}$. Dann gilt*

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x.$$

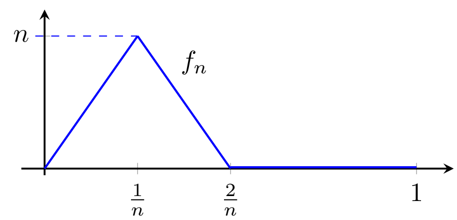
Beweis. Wir bemerken, dass f sowie die f_n als stetige Funktionen auf einem kompakten Intervall R–integrierbar sind. Nun gilt:

$$\begin{aligned}\Big|\int_a^b f(x)\,\mathrm{d}x - \int_a^b f_n(x)\,\mathrm{d}x\Big| &\leqslant \int_a^b |f(x)-f_n(x)|\,\mathrm{d}x\\ &\leqslant (b-a)\sup_{x\in[a,b]}|f(x)-f_n(x)|\\ &= (b-a)\|f-f_n\|_\infty \xrightarrow{\text{XI.1.5(ii)}} 0.\end{aligned}$$

□

Bemerkung XI.3.2. XI.3.1 besagt, dass bei gleichmäßig konvergenten Folgen stetiger Funktion auf kompakten Intervallen Grenzwert und Integral vertauscht werden können $\big(\lim_{n\to\infty}\int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n\big)$. Liegt nur punktweise Konvergenz vor, so ist das im Allgemeinen falsch:

Beispiel XI.3.3. Wir betrachten wieder $f_n\colon [0,1]\to\mathbb{R}$ aus XI.1.3:



Dann ist $\displaystyle\int_0^1 \lim_{n\to\infty} f_n(x)\,\mathrm{d}x = 0$, aber

$$\int_0^1 f_n(x)\,\mathrm{d}x = 2\int_0^{1/n} n^2x\,\mathrm{d}x = 2n^2\frac{x^2}{2}\Big|_0^{1/n} = 1,$$

also $\displaystyle\lim_{n\to\infty}\int_0^1 f_n(x)\,\mathrm{d}x = 1$.

XI.4. Gleichmäßige Konvergenz und Ableitungen

Satz XI.4.1. *Für $n\in\mathbb{N}$ sei $f_n\in \mathrm{C}^1[a,b]$. Es konvergiere f_n punktweise gegen f und f_n' konvergiere gleichmäßig gegen g mit Funktionen $f,g\colon [a,b]\to\mathbb{R}$. Dann ist $f\in\mathrm{C}^1[a,b]$ und für $x\in[a,b]$ gilt $f'(x)=g(x)$.*

Beweis. Wir bemerken zunächst, dass g nach XI.2.1 stetig ist. Da die f_n' auch stetig sind, können wir für jedes n den Hauptsatz anwenden, d.h. wir bekommen für jedes $x\in[a,b]$ die Gleichung

$$f_n(x) = f_n(a) + \int_a^x f_n'(t)\,\mathrm{d}t.$$

Hier nehmen wir nun den Grenzwert $n\to\infty$. Dann gilt $f_n(x)\to f(x)$, $f_n(a)\to f(a)$. Da f_n' gleichmäßig gegen g konvergiert, folgt mit XI.3.1 für jedes $x\in[a,b]$, dass $\int_a^x f_n'(t)\,\mathrm{d}t \to \int_a^b g(t)\,\mathrm{d}t$ gilt. Zusammen erhalten wir also

$$f(x) = f(a) + \int_a^x g(t)\,\mathrm{d}t.$$

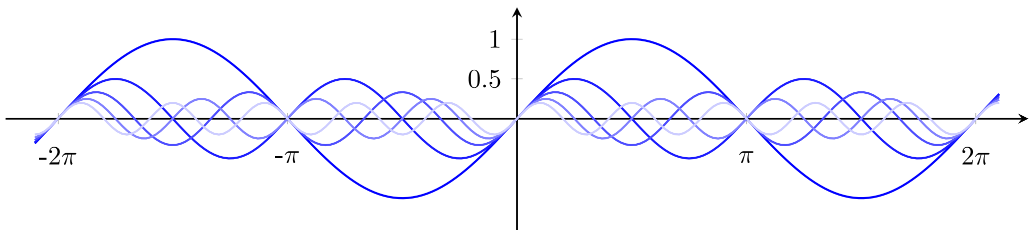
Durch Differenzieren erhalten wir

$$f'(x) = 0 + \frac{\mathrm{d}}{\mathrm{d}x}\Big(\int_a^x g(t)\,\mathrm{d}t\Big) = g(x),$$

weil nach VIII.4.1 $x\mapsto\int_a^x g(t)\,\mathrm{d}t$ Stammfunktion von g ist. □

Bemerkung XI.4.2. XI.4.1 besagt, dass Grenzwert und Ableitung bei einer Folge stetig differenzierbarer Funktionen vertauscht werden können $\big(\frac{\mathrm{d}}{\mathrm{d}x}(\lim_{n\to\infty} f_n(x)) = \lim_{n\to\infty}\frac{\mathrm{d}}{\mathrm{d}x}f_n(x)\big)$, falls die Folge selbst punktweise und die Folge der Ableitungen gleichmäßig konvergiert. Gleichmäßige Konvergenz der $(f_n)_{n\in\mathbb{N}}$ selbst ist hierfür nicht hinreichend:

Beispiel XI.4.3. Es sei $f_n\colon \mathbb{R} \to \mathbb{R}$ definiert durch $f_n(x) = \frac{1}{n}\sin(nx)$ für $n \geqslant 1$. Im Bild sind wachsende Indizes durch heller werdendes Blau angedeutet, d.h. die dunkelblaue Funktion ist $f_1(x) = \sin(x)$, die nächsthellere ist $f_2(x) = \frac{1}{2}\sin(2x)$ usw.:



Da $\|f_n\|_\infty = \sup_{x\in\mathbb{R}} |\frac{1}{n}\sin(nx)| = \frac{1}{n} \to f \equiv 0$, folgt mit XI.2.1(ii), dass f_n gleichmäßig gegen die Nullabbildung konvergiert.

Andererseits gilt $f_n'(0) = 1$ für alle $n \in \mathbb{N}$, d.h. nicht einmal $f_n'(0) \to f'(0)$ gilt.

XI.5. Gleichmäßige Konvergenz und Potenzeihen

Ist $R > 0$ der Konvergenzradius der Potenzreihe

$$\sum_{k=0}^{\infty} c_k(x-x_0)^k$$

mit Entwicklungspunkt $x_0 \in \mathbb{R}$, dann konvergiert die Reihe auf jedem kompakten Intervall $[a,b] \subset (x_0 - R, x_0 + R)$ gleichmäßig gegen die Grenzfunktion $f\colon (x_0 - R, x_0 + R) \to \mathbb{R}$. Da jede Partialsumme

$$f_n := \sum_{k=0}^{n} c_k(x-x_0)^k$$

ein Polynom ist, kann man aus XI.2.1 die Stetigkeit der Grenzfunktion schließen und mit XI.3.1 über $[a,b]$ wie oben gliedweise integrieren. In der Tat kann man auch gliedweise differenzieren: Auf $[a,b] \subset (x_0 - R, x_0 + R)$ ist f_n stetig differenzierbar und

$$f_n'(x) = \sum_{k=0}^{n} kc_k(x-x_0)^{k-1} = \sum_{k=0}^{n-1} (k+1)c_{k+1}(x-x_0)^k =: \sum_{k=0}^{n-1} b_k(x-x_0)^k$$

sind Partialsummen einer neuen Potenzreihe

$$\sum_{k=0}^{\infty} b_k(x-x_0)^k,$$

welche ebenfalls den Konvergenzradius R hat. In der Tat gilt

$$\limsup_{k\to\infty} \sqrt[k]{|b_k|} = \limsup_{k\to\infty} \sqrt[k]{k+1}\Big(\sqrt[k+1]{|c_{k+1}|}\Big)^{\frac{k+1}{k}} = \frac{1}{R},$$

weil $\limsup_{k\to\infty} \sqrt[k+1]{|c_{k+1}|} = \frac{1}{R}$, sowie $\sqrt[k]{k+1} \to 1$ und $\frac{k+1}{k} \to 1$ konvergieren. Damit folgt, dass $g := \lim_{n\to\infty} f_n'$ gleichmäßig auf jedem $[a,b] \subset (x_0 - R, x_0 + R)$ konvergiert, und wir können XI.4.1 anwenden. Induktiv folgt, das jede (reelle) Potenzreihe mit Entwicklungspunkt $x_0 \in \mathbb{R}$ und Konvergenzradius $R > 0$ auf $(x_0 - R, x_0 + R)$ eine C^∞–Funktion definiert.

KAPITEL XII

Taylorentwicklung

Es sei $I \subset \mathbb{R}$ im Folgenden ein offenes, nichtleeres Intervall. Bei der Taylorentwicklung approximiert man geeignete Funktionen durch Polynome.

XII.1. Taylorpolynome und Restglieder

Satz XII.1.1. *Es sei $f\colon I \to \mathbb{R}$ $(n+1)$–mal stetig differenzierbar und $x_0 \in I$. Dann gilt*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{1}{n!}\int_{x_0}^{x} (x-t)^n f^{(n+1)}(t)\,\mathrm{d}t.$$

Definition XII.1.2. Unter obigen Voraussetzungen nennt man

(i) $\mathrm{T}_n[f,x_0](x) := \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$ das *Taylorpolynom n–ter Ordnung von f im Punkt x_0*,

(ii) $R_n[f,x_0](x) := f(x) - \mathrm{T}_n[f,x_0](x)$ das zugehörige *Restglied*, sowie

(iii) $R_n[f,x_0](x) = \frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,\mathrm{d}t$ die *Integraldarstellung des Restglieds* oder kurz das *Integralrestglied*.

Beweis. (von XII.1.1) Wir zeigen die Aussage durch Induktion.

Für den Induktionsanfang bei $n = 0$ haben wir für $f \in \mathrm{C}^1(I)$ mit dem Hauptsatz die gewünschte Gleichung $f(x) = f(x_0) + \int_{x_0}^{x} f'(t)\,\mathrm{d}t$.

Für den Induktionsschritt von n nach $n+1$ geben wir uns $f \in \mathrm{C}^{n+2}(I)$ und $x_0 \in I$ vor und benutzen partielle Integration:

$$\begin{aligned}\frac{1}{n!}\int_{x_0}^{x}(x-t)^n\, f^{(n+1)}(t)\,\mathrm{d}t &= \frac{1}{n!}\Big[\,\frac{-(x-t)^{n+1}}{n+1}\, f^{(n+1)}(t)\,\Big|_{x_0}^{x} - \int_{x_0}^{x}\frac{-(x-t)^{n+1}}{n+1}\, f^{(n+2)}(t)\,\mathrm{d}t\Big]\\ &= \frac{(x-x_0)^{n+1}}{(n+1)!} f^{(n+1)}(x_0) + \frac{1}{(n+1)!}\int_{x_0}^{x}(x-t)^{n+1} f^{(n+2)}(t)\,\mathrm{d}t.\end{aligned}$$

Nun gilt mit der Induktionsannahme und der zweiten Gleichung von oben:

$$\begin{aligned}f(x) &= \mathrm{T}_n[f,x_0](x) + \frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,\mathrm{d}t\\ &= \mathrm{T}_n[f,x_0](x) + \frac{(x-x_0)^{n+1}}{(n+1)!} f^{(n+1)}(x_0) + \frac{1}{(n+1)!}\int_{x_0}^{x}(x-t)^{n+1} f^{(n+2)}(t)\,\mathrm{d}t\\ &= \mathrm{T}_{n+1}[f,x_0](x) + \frac{1}{(n+1)!}\int_{x_0}^{x}(x-t)^{n+1} f^{(n+2)}(t)\,\mathrm{d}t\end{aligned}$$

□

Satz XII.1.3. *(Lagrange–Restglied) Es sei $f\colon I \to \mathbb{R}$ $(n+1)$–mal stetig differenzierbar und $x_0 \in I$. Dann existiert ein ξ zwischen x_0 und x mit*

$$f(x) = \mathrm{T}_n[f,x_0](x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}.$$

Beweis. Wir benutzen den verallgemeinerten Mittelwertsatz der Integralrechnung VIII.3.4 und finden ein ξ zwischen x_0 und x mit

$$\begin{aligned} f(x) - \mathrm{T}_n[f,x_0](x) &= \frac{1}{n!}\int_{x_0}^{x}(x-t)^n f^{(n+1)}(t)\,\mathrm{d}t \\ &= f^{n+1}(\xi)\int_{x_0}^{x}\frac{(x-t)^n}{n!}\,\mathrm{d}t \\ &= f^{(n+1)}(\xi)\Big(-\frac{(x-t)^{n+1}}{(n+1)!}\Big|_{x_0}^{x}\Big) \\ &= \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1} \end{aligned}$$

□

Korollar XII.1.4. *Ist $f \in \mathrm{C}^{n+1}(I)$ und gilt $f^{(n+1)} \equiv 0$, dann ist f ein Polynom vom Grad kleiner gleich n.*

Beweis. Das Lagrange-Restglied ist in diesem Fall konstant Null, d.h. die Funktion $f = \mathrm{T}_n[f,x_0]$ stimmt mit ihrem Taylorpolynom überein. □

Bemerkung XII.1.5. (Fehlerabschätzung/Restgliedabschätzung) Das Lagrange-Restglied kann für gegebene I, f und x_0 abgeschätzt werden, um den Fehler zu kontrollieren, den man macht, wenn man f durch ein Taylorpolynom approximiert. Zum Beispiel ist

$$\mathrm{T}_n[\exp,0](x) = \sum_{k=0}^{n}\frac{x^k}{k!}$$

und

$$\begin{aligned} \big|e^x - \mathrm{T}_n[\exp,0](x)\big| = \big|R_n[f,x_0](x)\big| &\leqslant \sup_{\xi\in[-|x|,|x|]}\left|\frac{\exp^{(n+1)}(\xi)}{(n+1)!}(x-0)^{n+1}\right| \\ &= \sup_{\xi\in[-|x|,|x|]}\frac{|x|^{n+1}}{(n+1)!}e^{\xi} = \frac{|x|^{n+1}}{(n+1)!}e^{|x|}. \end{aligned}$$

Dies liefert genau die Fehlerabschätzung, die wir bereits in der Mathematik 2 gezeigt haben. Ähnliche Abschätzungen werden wir auch für sin und cos noch zeigen (siehe XII.2.6). Interessanter ist hier aber natürlich, dass wir auch Funktionen, für die wir a priori keine Potenzreihendarstellung kennen, unter bestimmten Voraussetzungen trotzdem durch Taylorpolynome approximieren können. Ein Beispiel hierfür sehen wir in XII.2.7 und ein weiteres in einer Aufgabe.

Satz XII.1.6. *Es sei $f\colon I \to \mathbb{R}$ n–mal stetig differenzierbar und $x_0 \in I$. Dann gibt es eine Funktion $\varphi\colon I \to \mathbb{R}$, sodass für alle $x \in I$ gilt*

$$f(x) = \mathrm{T}_n[f,x_0](x) + \varphi(x)(x-x_0)^n \quad \textit{und} \quad \lim_{x\to x_0}\varphi(x) = 0.$$

Beweis. XII.1.3 mit $n-1$ statt n liefert, dass für jedes $x \in I$ ein ξ zwischen x und x_0 existiert mit

$$\begin{aligned} f(x) - \mathrm{T}_{n-1}[f,x_0](x) &= \frac{f^{(n)}(\xi)}{n!}(x-x_0)^n \\ &= \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{f^{(n)}(\xi)-f^{(n)}(x_0)}{n!}(x-x_0)^n \\ &= \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \varphi(x)(x-x_0)^n, \end{aligned}$$

wobei wir im zweiten Schritt eine Null ergänzt haben und für den dritten Schritt beachten, dass ξ von x abhängt, und wir damit $\varphi(x) := \frac{f^{(n)}(\xi)-f^{(n)}(x_0)}{n!}$ als Funktion von x betrachten können. Hierzu wählt man eine Funktion $\xi\colon I \to \mathbb{R}$, indem man für jedes $x \in I$ ein ξ mit XII.1.3 auswählt, denn es könnte ja auch mehrere solcher ξ geben. Es folgt nun

$$f(x) = \mathrm{T}_n[f,x_0](x) + \varphi(x)(x-x_0)^n$$

und da $\xi \to x_0$ wenn $x \to x_0$ haben wir

$$\lim_{x\to x_0} \varphi(x) = \lim_{\xi\to x_0} \frac{f^{(n)}(\xi) - f^{(n)}(x_0)}{n!} = 0,$$

wobei wir im letzten Schritt benutzen, dass $f^{(n)}$ in x_0 stetig ist. $\square$

Definition XII.1.7. (Das Landau–Symbol „klein–o“) Es seien $f,\, g\colon (a,b)\backslash\{x_0\} \to \mathbb{R}$ mit $x_0 \in [a,b]$, wobei $a, b \in \mathbb{R}$ sind. Wir schreiben $f(x) = \mathrm{o}(g(x))$ für $x \to x_0$ (gelesen „f ist klein–o von g“), falls gilt

$$\forall\, \varepsilon > 0\; \exists\, \delta > 0\; \forall\, x \in (a,b)\backslash\{x_0\}\colon |x - x_0| < \delta \implies |f(x)| \leqslant \varepsilon |g(x)|.$$

Bemerkung XII.1.8.

(i) Ist $g(x) \neq 0$, so gilt

$$f(x) = o(g(x)) \iff \lim_{x\to x_0} \frac{f(x)}{g(x)} = 0.$$

Stellt man sich hier vor, dass $g(x) \to 0$ für $x \to x_0$ gilt, dann bedeutet obiges, dass ebenfalls $f(x) \to 0$ für $x \to x_0$ gilt und dass f echt schneller gegen Null konvergiert als g, denn der Quotient $\frac{f}{g}$ geht ja immer noch gegen Null.

(ii) Es gibt weitere Landau–Symbole und jeweils Versionen für $x \to \pm\infty$ statt $x \to x_0 \in \mathbb{R}$. Im Folgenden meinen wir immer $x \to x_0$ und erwähnen dies daher nicht extra.

(iii) Ist g gegeben, so definiert die Bedingung in XII.1.7 eigentlich eine Menge von Funktionen f mit $f = \mathrm{o}(g)$. Insbesondere implizieren $f_1 = \mathrm{o}(g)$ und $f_2 = \mathrm{o}(g)$ nicht $f_1 = f_2$.

(iv) Mit dem Landau–Symbol läßt sich XII.1.6 als

$$f(x) = \mathrm{T}_n[f, x_0](x) + \mathrm{o}(|x - x_0|^n)$$

schreiben: Bis auf einen Fehlerterm, der für $x \to x_0$ echt schneller gegen Null geht als $|x - x_0|^n$, stimmt f mit $\mathrm{T}_n[f, x_0]$ überein.

XII.2. Die Taylorreihe

Definition XII.2.1. Es sei $f\colon I \to \mathbb{R}$ beliebig oft differenzierbar und $x_0 \in I$. Dann heißt die Potenzreihe

$$\mathrm{T}[f, x_0](x) := \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

die *Taylorreihe* von f an der Stelle x_0.

Bemerkung XII.2.2. Die natürliche Frage ist jetzt, ob $f = \mathrm{T}[f, x_0]$ gilt. Klar ist, dass die Taylorreihe $\mathrm{T}[f, x_0]$ genau für diejenigen $x \in I$ konvergiert, für die das Restglied $R_n[f, x_0](x)$ gegen Null geht.

Satz XII.2.3. *Es sei $x_0 \in \mathbb{R}$ und sei $\sum_{k=0}^{\infty} c_k(x - x_0)^k$ eine Potenzreihe mit Konvergenzradius $R \in \mathbb{R}_{>0} \cup \{\infty\}$, d.h. $f\colon (x_0 - R, x_0 + R) \to \mathbb{R}$, $f(x) = \sum_{k=0}^{\infty} c_k(x - x_0)^k$ ist wohldefiniert und nach XI.5 eine C^∞–Funktion. Dann gilt $c_k = \frac{1}{k!} f^{(k)}(x_0)$ für $k \in \mathbb{N}$ und $\mathrm{T}[f, x_0] = f$ auf $(x_0 - R, x_0 + R)$.*

Beweis. Sukzessive Anwendung von XI.5 zeigt mit gliedweisem Ableiten

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)(n-2)\cdots(n-k+1)c_n(x - x_0)^{n-k}.$$

Auswerten in $x = x_0$ impliziert $f^{(k)}(x_0) = k! c_k$, also $c_k = \frac{1}{k!} f^{(k)}(x_0)$. Damit folgt

$$f(x) = \sum_{k=0}^{\infty} c_k(x - x_0)^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k = \mathrm{T}[f, x_0](x)$$

für $x \in (x_0 - R, x_0 + R)$. $\square$

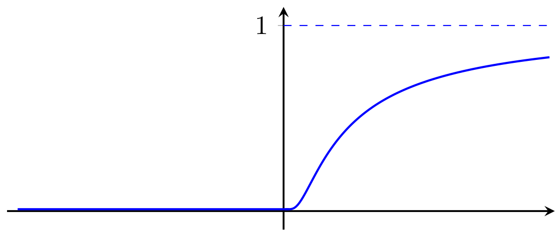
Bemerkung XII.2.4. XII.2.3 besagt, dass bei Funktionen, die als Potenzreihe dargestellt werden können oder sogar über eine solche definiert wurden, die Taylorreihe wieder genau diese Reihe zurückliefert.

Wie bei der Approximation auch ist aber die spannende Frage, ob man Funktionen, bei denen a priori keine Reihendarstellung bekannt ist, durch eine Taylorreihe darstellen kann. Man spricht dann davon, eine Funktion *in eine Taylorreihe zu entwickeln*. Die Approximation durch ein Taylorpolynom nennt man ebenfalls *Taylorentwicklung*.

Bemerkungen XII.2.5.

(1) Das folgende Beispiel ist berühmt und Sie sollten sich dieses merken. Wir setzen

$$f\colon \mathbb{R} \to \mathbb{R}, \quad f(x) := \begin{cases} e^{-1/x}, & \text{für } x > 0, \\ 0, & \text{sonst.} \end{cases}$$



Wir haben in einer Aufgabe gezeigt, dass $f \in \mathrm{C}^\infty(\mathbb{R})$ und dass $f^{(k)}(0) = 0$ für alle $k \in \mathbb{N}_0$ gilt. D.h. $\mathrm{T}[f,0] = \sum_{k=0}^\infty 0 \cdot x^k$ hat Konvergenzradius ∞, aber $\mathrm{T}[f,0] \equiv 0 \neq f$.

(2) Es gibt Beispiele von Funktionen, bei denen der Konvergenzradius der Taylorreihe Null ist, d.h. die Taylorreihe konvergiert nur in x_0.

Beispiele XII.2.6.

(1) Die Taylorreihe der Exponentialfunktion entwickelt in $x_0 = 0$ ist nach XII.2.3

$$\mathrm{T}[\exp, 0](x) = \sum_{k=0}^\infty \frac{x^k}{k!} = \exp(x).$$

(2) Die Taylorreihe der Exponentialfunktion in $x_0 \neq 0$ ist

$$\mathrm{T}[\exp, x_0](x) = \sum_{k=0}^\infty \frac{e^{x_0}}{k!}(x - x_0)^k = e^{x_0} \sum_{k=0}^\infty \frac{(x-x_0)^k}{k!} = e^{x_0} e^{x - x_0} = \exp(x).$$

(3) Die Taylorreihe des Sinus ist (nach XII.2.3 oder 'zu Fuß' durch Berechnung von $\sin^{(k)}(0)$)

$$\mathrm{T}[\sin, 0](x) = \sum_{k=0}^\infty \frac{\sin^{(k)}(0)}{k!}(x-0)^k = \sum_{k=0}^\infty (-1)^k \frac{x^{2k+1}}{(2k+1)!} = \sin x.$$

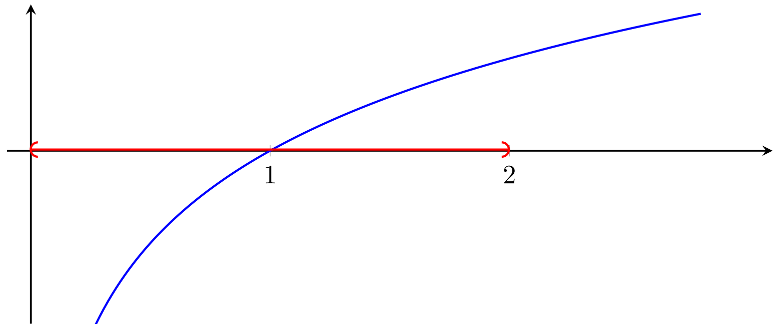
Das Lagrange-Restglied (nach XII.1.2) kann dann (mit ξ zwischen x und Null) wie folgt abgeschätzt werden

$$\begin{aligned} \big|\sin x - \mathrm{T}_{2n+1}[\sin, 0](x)\big| &= \Big|\sin x - \sum_{k=0}^n (-1)^k \frac{x^{(2k+1)}}{(2k+1)!}\Big| \\ &= \big|R_{2n+1}[\sin, 0](x)\big| \\ &= \frac{|\sin^{(2n+2)}(\xi)|}{(2n+2)!}|x|^{2n+2} \\ &\leqslant \frac{|x|^{2n+2}}{(2n+2)!}, \end{aligned}$$

was fast das Gleiche liefert wie in Kapitel V.

(4) Für den Kosinus gilt Ähnliches.

Beispiel XII.2.7. Es sei $\log\colon (0,\infty) \to \mathbb{R}$. Da wir log als Inverse der Exponentialfunktion definiert haben, haben wir a priori keine Potenzreihendarstellung. Aber log ist eine C^∞–Funktion. Wir entwickeln in $x_0 = 1$.



Dann gilt

$$\log^{(0)} x = \log x,\ \ \log^{(1)} x = \frac{1}{x},\ \ \log^{(2)} x = -\frac{1}{x^2},\ \ \log^{(3)} x = \frac{2\cdot 1}{x^3},$$
$$\log^{(4)} x = -\frac{3\cdot 2\cdot 1}{x^4},\ \ \log^{(5)} x = \frac{4\cdot 3\cdot 2\cdot 1}{x^5},\ \ldots$$

und damit

$$\begin{aligned}\mathrm{T}[\log,1](x) &= \sum_{k=0}^{\infty} \frac{\log^{(k)}(1)}{k!}(x-1)^k \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}(k-1)!}{k!}(x-1)^k \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(x-1)^k.\end{aligned}$$

Man sieht nun, dass der Konvergenzradius dieser Reihe 1 ist. Zu klären bleibt, ob für $x \in (0,2)$ auch $\mathrm{T}[\log,1](x) = \log x$ gilt (vgl. XII.2.5). In der Tat finden wir mit XII.1.3 für $x \in (\frac{1}{2}, 2)$ ein ξ zwischen $x_0 = 1$ und x, sodass gilt

$$\begin{aligned}\big|R_n[\log,1](x)\big| &= \Big|\frac{\log^{(n)}(\xi)}{n!}(x-1)^n\Big| \\ &\Big|\frac{(n-1)!\frac{1}{\xi^n}}{n!}(x-1)^n\Big| \\ &= \frac{1}{n}\frac{|x-1|^n}{\xi^n} \\ &\leqslant \frac{1}{n} \to 0,\end{aligned}$$

wobei man mit der Fallunterscheidung ($\frac{1}{2} < x < \xi < 1$ oder $1 < \xi < x < 2$) die letzte Abschätzung sieht. Es folgt

$$\log x = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(x-1)^k$$

für $x \in (\frac{1}{2}, 2)$. Dies gilt aber auch für $x \in (0,2)$, wie man folgendermaßen sieht: Mit $y := x - 1 \in (-1,1)$ gilt

$$\log(x) = \log(1+y) = \log(1+t)\big|_0^y = \int_0^y \frac{1}{1+t}\,\mathrm{d}t = \int_0^y \sum_{k=0}^{\infty}(-1)^k t^k\,\mathrm{d}t,$$

wobei wir im letzten Schritt eine Potenzreihe vorliegen haben, die auf $[-|y|,|y|]$ gleichmäßig konvergiert. Daher folgt mit XI.3.1 nun

$$\log(x) = \sum_{k=0}^{\infty}(-1)^k \int_0^y t^k\,\mathrm{d}t = \sum_{k=0}^{\infty}(-1)^k \frac{y^{k+1}}{k+1} = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}y^k = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}(x-1)^k$$

tatsächlich für $x \in (0,2)$ bzw.

$$\log(1+y) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} y^k$$

für $y \in (-1,1)$.

KAPITEL XIII

Gewöhnliche Differentialgleichungen

In diesem Kapitel behandeln wir Anwendungsprobleme, die wir mit sogenannten Differentialgleichungen untersuchen. Bei Funktionen, die von der Zeit abhängen, werden in Anwendungsfächern (Physik, Biologie, Chemie...) oft ein Punkt über dem Buchstaben, der die Funktion repräsentiert, gemacht statt des Striches, den wir bisher für die Ableitung benutzt haben.

XIII.1. Beispiele

Beispiel XIII.1.1. Wir untersuchen das **Abkühlen von Kaffee**. Wir suchen die Temperatur $T(t)$ des Kaffees zum Zeitpunkt $t \geqslant 0$. Wir nehmen das Folgende an:

- $T(0) = T_0$ bezeichne die Temperatur beim Aufgießen.
- Die Funktion $T = T(t)$ fällt monoton.
- Wartet man lange genug, so nimmt der Kaffee Raumtemperatur R_0 an, also $\lim\limits_{t\to\infty} T(t) = R_0$.
- Die Änderung der Temperatur $\dot{T}(t) = \frac{\mathrm{d}T}{\mathrm{d}t}(t) =$ passiert schnell, wenn Differenz zur Raumtemperatur groß ist. Ist die Differenz klein, ändert sich die Temperatur des Kaffees langsamer. Daher nehmen wir an, dass $\dot{T}(t) = \kappa(T(t) - R_0)$ gilt mit $\kappa < 0$.
- Die Konstante $\kappa =$ kann durch ein Experiment bestimmt werden.

Wir suchen also eine Lösung $T\colon [0,\infty) \to \mathbb{R}$ des *Anfangswertproblems*

$$\text{(AWP)} \quad \begin{cases} \dot{T}(t) = \kappa(T(t) - R_0), \\ T(0) = T_0. \end{cases}$$

Beispiel XIII.1.2. Wir suchen die Wasserhöhe $h(t)$ zum Zeitpunkt $t \geqslant 0$ in einer **Badewanne**. Wir nehmen das Folgende an:

- Die Wasserhöhe zum Zeitpunkt 0 sei bekannt: $h(0) = h_0$.
- Für $t \geqslant 0$ gilt nach dem Energieerhaltungssatz
$$m \cdot \mathrm{g} \cdot h(t) = \frac{1}{2} m \cdot v(t)^2,$$
wobei m die Masse des Wassers in der Badewanne, g die Gravitationskonstante und v die Geschwindigkeit ist, mit der das Wasser ausströmt.
- Für t' nahe bei t mit $t' > t$ gilt $F \cdot (h(t) - h(t')) = f \cdot v(t) \cdot (t' - t)$ nach dem Satz über die Massenerhaltung, wobei F die Grundfläche der Badewanne und f die Fläche des Abflusses bezeichnet. Hieraus folgt durch Division durch $t' - t$ und durch F
$$\frac{h(t') - h(t)}{t' - t} = -\frac{f}{F} v(t) = -\frac{F}{f}\sqrt{2\,\mathrm{g}\,h(t)}.$$
Durch Betrachtung des Grenzübergangs $t' \to t$ folgt also
$$\dot{h}(t) = \lim_{t' \to t} \frac{h(t') - h(t)}{t' - t} = \alpha\sqrt{h(t)}.$$
- Die Konstante $\alpha < 0$ hängt von der Badewanne ab und kann durch ein Experiment bestimmt werden.

Wir suchen also eine Lösung $h\colon [0,\infty) \to \mathbb{R}$ des

$$\text{(AWP)}\quad \begin{cases} \dot{h}(t) = \alpha\sqrt{h(t)}, \\ h(0) = h_0. \end{cases}$$

Beispiel XIII.1.3. Wir suchen die Anzahl der Mitglieder $N(t)$ einer Population zum Zeitpunkt $t \geqslant 0$ und wollen dazu das **Populationswachstum** verstehen. Wir nehmen das Folgende an:

- Der Populationsbestand zum Zeitpunkt 0 sei $N(0) = N_0$.
- Für $t \geqslant 0$ und kleines $\Delta t > 0$ ist $\Delta N = N(t+\Delta t) - N(t)$ proportional zu $N(t)$ und zu Δt[1], also $\Delta N = \omega \cdot N(t) \cdot \Delta t$ mit $\omega \in \mathbb{R}$. Damit gilt

$$\dot{N}(t) = \lim_{\Delta t \to 0} \frac{N(t+\Delta t) - N(t)}{\Delta t} = \omega N(t).$$

Wir suchen also eine Lösung $N\colon [0,\infty) \to \mathbb{R}$ des

$$\text{(AWP)}\quad \begin{cases} \dot{N}(t) = \omega N(t), \\ N(0) = N_0. \end{cases}$$

XIII.2. Gewöhnliche Differentialgleichungen erster Ordnung

Die obigen Beispiele lassen sich allesamt mit einer bestimmten Art von Differentialgleichungen beschreiben.

Definition XIII.2.1. Eine *gewöhnliche Differentialgleichung erster Ordnung* ist eine Gleichungen der Form $\dot{x} = f(t,x)$ [bzw. $y' = f(x,y)$] mit $f\colon D \subset \mathbb{R} \to \mathbb{R}$. Unter einer *Lösung* verstehen wir eine Funktion $\varphi \in \mathrm{C}^1(I,\mathbb{R})$ mit einem Intervall I, und $\dot{\varphi}(t) = f(t,\varphi(t))$ [bzw. $\varphi'(x) = f(x,\varphi(x))$] für alle $x \in I$. Ist eine *Anfangsbedingung* $x(t_0) = x_0$ [bzw. $y(x_0) = y_0$] gegeben, so fordern wir zusätzlich $t_0 \in I$ und $\varphi(t_0) = x_0$ [bzw. $\varphi(x_0) = y_0$].

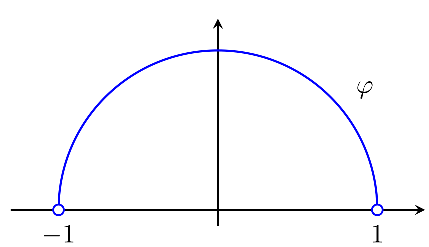
Beispiele XIII.2.2.

(1) Nach VI.1.9 ist $\varphi\colon \mathbb{R} \to \mathbb{R}$, $\varphi(t) = N_0 e^{\omega t}$ die eindeutig bestimmte und auf ganz $\mathbb{R}$ existierende Lösung des Anfangswertproblems in XIII.1.3.

(2) Das Anfangswertproblem $y' = \sqrt{y}$, $y(0) = 0$ hat zwei Lösungen $\varphi(x) \equiv 0$, $\varphi(x) = \frac{x^2}{4}$.

(3) Die Differentialgleichung $(y')^2 + 1 = 0$ hat keine (reelle) Lösung.

(4) Das Anfangswertproblem $y' = -\frac{x}{y}$, $y(0) = 1$ hat die Lösung $\varphi(x) = \sqrt{1-x^2}$, die aber nur auf $(-1,1)$ existiert:



[1]Stellen Sie sich vor, Sie zählen, wieviele Individuen in einem kurzen Zeitraum hinzukommen, z.B. wieviele Babys an einem Tag geboren werden. Wenn Sie dann statt einem Tag zwei oder drei Tage lang zählen, ist es naheliegend anzunehmen, dass sich die Zahl verdoppeln bzw. verdreifachen wird. Zählen Sie über den gleichen Zeitraum, wieviele Babys in Hamburg geboren werden, und dann, wieviele in Bayern, dann werden Sie naheliegenderweise annehmen, dass die Zahl sich um das 7.1–fache erhöhen sollte, wenn Bayern 7.1-mal so viele Einwohner:innen hat wie Hamburg. Dieser Ansatz ignoriert natürlich weitere Einflüsse.

Definition XIII.2.3. Wir betrachten ein Anfangswertproblem der Form

$$\text{(AWP)} \quad \begin{cases} y' = g(x)h(y), \\ y(x_0) = y_0, \end{cases}$$

mit zwei Funktionen $g = g(x)$, $h = h(x)$. Solche Differentialgleichungen nennt man *Differentialgleichungen mit getrennten Variablen* und für geeignete g, h kann man sie mit der folgenden Methode lösen, die wir zunächst anhand eines Beispiels erklären.

Beispiel XIII.2.4. (Methode der Trennung der Variablen (TdV))
Wir betrachten $y' = xy$, $y(0) = -1$.

(1) Wir formulieren die Gleichung um (für $y \neq 0$)

$$\frac{1}{y}\frac{\mathrm{d}y}{\mathrm{d}x} = x.$$

(2) Wir trennen die Variablen x und y und integrieren auf beiden Seiten: $\int \frac{1}{y}\,\mathrm{d}y = \int x\,\mathrm{d}x$.
(3) Wir ermitteln die Stammfunktionen: $\log|y| = \frac{x^2}{2} + c$ für eine Konstante c.
(4) Wir lösen nach y auf: $y = \pm e^{x^2/2+c} = \pm e^c e^{x^2/2} = Ce^{x^2/2}$ mit einer neuen Konstante C.
(5) Wir wählen C so, dass die Anfangsbedingung erfüllt ist: $-1 = y(0) = Ce^0 = C$.
(6) Prüfen Sie, ob $y(x) := -e^{x^2/2}$ auf dem natürlichen Definitionsbereich, hier $x \in \mathbb{R}$, das Anfangswertproblem auch wirklich löst: $y'(x) = -xe^{x^2/2} = x \cdot y(x)$ und $y(0) = -1$.

Das dies funktioniert, zeigt der folgende Satz:

Satz XIII.2.5. *Es seien $g\colon I \subset \mathbb{R} \to \mathbb{R}$, $h\colon J \subset \mathbb{R} \to \mathbb{R}$ stetig auf offenen Intervallen und sei $h(y) \neq 0$ für alle $y \in J$. Es sei $(x_0, y_0) \in I \times J$. Definiere*

$$G\colon I \to \mathbb{R},\ G(x) := \int_{x_0}^{x} g(t)\,\mathrm{d}t \quad \textit{und} \quad H\colon J \to \mathbb{R},\ H(y) := \int_{y_0}^{y} \frac{1}{h(t)}\,\mathrm{d}t.$$

Es sei $I' \subset I$ ein Intervall mit $x_0 \in I'$ und $G(I') \subset H(J)$. Dann existiert genau eine Lösung $\varphi\colon I' \to \mathbb{R}$ des

$$\textit{(AWP)} \quad \begin{cases} y' = g(x)h(y), \\ y(x_0) = y_0. \end{cases}$$

Für diese Lösung gilt $\varphi(I') \subset J$ und

$$\forall\, x \in I'\colon H(\varphi(x)) = G(x). \tag{XIII.2.1}$$

Beweis.

(1) Wir zeigen zuerst: Ist $\varphi\colon I' \to \mathbb{R}$ eine Lösung des AWP, dann gilt (XIII.2.1). In der Tat gilt für $t \in I'$ dann $\varphi'(t) = g(t)h(\varphi(t))$, was durch Teilen durch $h(\varphi(t))$ zu

$$g(t) = \frac{\varphi'(t)}{h(\varphi(t))}$$

führt. Da nach Annahme $h(y) \neq 0$ ist für alle $y \in J$, ist dies wohldefiniert. Integration liefert dann für $x \in I'$:

$$\int_{x_0}^{x} g(t)\,\mathrm{d}t = \int_{x_0}^{x} \frac{\varphi'(t)}{h(\varphi(t))}\,\mathrm{d}t = \int_{\varphi(x_0)}^{\varphi(x)} \frac{1}{h(u)}\,\mathrm{d}u,$$

wobei wir im letzten Schritt die Substitutionsregel verwendet haben. Nach Voraussetzung ist $\varphi \in \mathrm{C}^1(I')$ und $\frac{1}{h}$ ist stetig. Da der linke Ausdruck oben gerade $G(x)$ ist und der rechte $H(\varphi(x))$, haben wir also (XIII.2.1).

(2) Jetzt zeigen wir die Eindeutigkeit: Für $y \in J$ gilt

$$H'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\int_{y_0}^{y} \frac{1}{h(t)}\,\mathrm{d}t = \frac{1}{h(y)} \neq 0,$$

d.h. H ist streng monoton und C^1, besitzt also nach der Umkehrregel eine stetig differenzierbare Umkehrfunktion $\phi\colon H(J) \to J$. Ist $\varphi\colon I' \to \mathbb{R}$ nun eine Lösung, so folgt aus (1), dass (XIII.2.1) gilt, also $H(\varphi(x)) = G(x)$ für $x \in I'$ gilt. Nun wenden wir ϕ auf beiden Seiten an und erhalten

$$\varphi(x) = \phi(H(x)) = \phi(G(x))$$

für alle $x \in I'$ und damit die Eindeutigkeit.

(3) Nun zur Existenz. Da wir nach Voraussetzung $G(I') \subset H(J)$ haben, können wir $\varphi\colon I' \to \mathbb{R}$, $\varphi(x) := \phi(G(x))$ definieren, wobei ϕ die Funktion aus (2) ist. Dann wissen wir, dass $\varphi \in C^1$ ist, und erhalten

$$\varphi(x_0) = \phi(G(x_0)) = \phi(0) = y_0,$$

wobei wir im letzten Schritt $H(y_0) = 0$ benutzt haben, sowie

$$\varphi'(x) = \phi'(G(x))G'(x) = \frac{1}{H'(\phi(G(x)))}G'(x) = \frac{1}{\frac{1}{h(\varphi(x))}}g(x) = g(x)h(\varphi(x)).$$

Dies zeigt, dass $\varphi\colon I' \to \mathbb{R}$ das Anfangswertproblem löst.

□

Definition XIII.2.6. Es sei $I \subset \mathbb{R}$ und seien $a, b\colon I \to \mathbb{R}$ stetige Funktionen. Dann nennt man

$$y' = a(x)y + b(x)$$

lineare Differentialgleichung erster Ordnung und zwar *homogen*, falls $b \equiv 0$ ist und sonst *inhomogen*.

Im Folgenden kürzen wir das Wort Differentialgleichung mit DGL ab.

Die homogene Gleichung ist eine DGL mit getrennten Variablen, könnte also mit XIII.2.4 oder XIII.2.5 behandelt werden. In der Tat kann man hier aber auch eine Lösung explizit direkt angeben:

Satz XIII.2.7. *Es sei $a\colon I \subset \mathbb{R} \to \mathbb{R}$ stetig auf einem Intervall I und $x_0 \in I$, $c \in \mathbb{R}$. Dann gibt es genau eine Lösung $\varphi\colon I \to \mathbb{R}$ der DGL $y' = a(x)y$ mit $\varphi(x_0) = c$, nämlich*

$$\varphi(x) := c\exp\Big(\int_{x_0}^{x} a(t)\,\mathrm{d}t\Big).$$

Beweis. Es gilt

$$\varphi'(x) = c\exp\Big(\int_{x_0}^{x} a(t)\,\mathrm{d}t\Big) \cdot a(x) = a(x)\varphi(x)$$

und

$$\varphi(x_0) = c\exp(0) = c,$$

d.h. es muss nur die Eindeutigkeit gezeigt werden. Wie oben erfüllt

$$\varphi_0(x) := \exp\Big(-\int_{x_0}^{x} a(t)\,\mathrm{d}t\Big)$$

die Gleichung $\varphi_0'(x) = -a(x)\varphi_0(x)$. Es sei nun $\psi\colon I \to \mathbb{R}$ eine beliebige Lösung der DGL $y' = a(x)y(x)$ mit Anfangsbedingung $\psi(x_0) = c$. Wir betrachten $\psi_0(x) := \psi(x)\varphi_0(x)$. Dann gilt

$$\psi_0'(x) = \psi'(x)\varphi_0(x) + \psi(x)\varphi_0'(x) = a(x)\big(\psi(x)\varphi_0(x) - \psi(x)\varphi_0(x)\big) = 0$$

für alle $x \in I$, d.h. ψ_0 ist konstant und daher

$$\forall\, x \in I\colon \psi_0(x) = \psi_0(x_0) = \psi(x_0)\varphi_0(x_0) = c\exp(0) = c.$$

Da $\varphi_0 \neq 0$ gilt, erhalten wir

$$\psi(x) = \frac{\psi_0(x)}{\varphi_0(x)} = \frac{c}{\exp(-\int_{x_0}^{x} a(t)\,\mathrm{d}t)} = c\exp\Big(\int_{x_0}^{x} a(t)\,\mathrm{d}t\Big) = \varphi(x)$$

für alle $x \in I$.

□

Inhomogene Gleichungen bearbeitet man mit einer Methode die als *Variation der Konstanten* (VdK) bekannt ist.

Satz XIII.2.8. *Es seien $a, b\colon I \subset \mathbb{R} \to \mathbb{R}$ stetig, $x_0 \in I$ und $c \in \mathbb{R}$. Dann existiert genau eine Lösung $\varphi\colon I \to \mathbb{R}$ der DGL $y' = a(x)y + b(x)$ mit $\varphi(x_0) = c$, nämlich*

$$\varphi(x) = \varphi_0(x)\Big(c + \int_{x_0}^{x} \varphi_0(t)^{-1} b(t)\,\mathrm{d}t\Big) \quad \textit{mit} \quad \varphi_0(x) = \exp\Big(\int_{x_0}^{x} a(t)\,\mathrm{d}t\Big).$$

Beweis. Die Lösungen der homogenen Differentialgleichung $y' = a(x)y$ sind alle von der Form $\varphi(x) = k\varphi_0(x)$ mit $k \in \mathbb{R}$. Für die inhomogene machen wir den *Ansatz* $\varphi(x) := k(x)\varphi_0(x)$, d.h. wir ersetzen die Konstante k durch eine Funktion $k(x)$ und versuchen nun, $k(x)$ so zu bestimmen, dass φ die inhomogene Gleichung löst:

$$\begin{aligned}\varphi'(x) &= k'(x)\varphi_0(x) + \varphi_0'(x)k(x) = k'(x)\varphi_0(x) + a(x)\varphi_0(x)k(x)\\ &= k'(x)\varphi_0(x) + a(x)\varphi(x) \stackrel{!}{=} a(x)\varphi(x) + b(x).\end{aligned}$$

Damit löst φ genau dann die inhomogene DGL, wenn $k'(x)\varphi_0(x) = b(x)$, oder anders ausgedrückt wenn $k'(x) = \varphi_0(x)^{-1}b(x)$ gilt. Damit die Anfangsbedingung erfüllt ist, muss außerdem gelten $c = \varphi(x_0) = k(x_0)\cdot 1 = k(x_0)$. Der Hauptsatz zeigt, dass dies genau dann der Fall ist, wenn

$$k(x) = \int_{x_0}^{x} \varphi_0(t)^{-1} b(t)\,\mathrm{d}t + c.$$

Die Eindeutigkeit folgt, weil eine beliebige Lösung ψ als $\psi(x) = \tilde{k}(x)\varphi_0(x)$ mit $\tilde{k}(x) = \frac{\psi(x)}{\varphi_0(x)}$ geschrieben werden kann. Der obige Beweis zeigt dann $\tilde{k}(x) = k(x)$ für $x \in I$. □

Bemerkung XIII.2.9. Man muss sich die Formeln aus XIII.2.7 und XIII.2.8 nicht merken, sondern kann $y' = a(x)y$ durch TdV lösen (mit Lösung $y_{\text{homog.}}$) und dann durch den VdK–Ansatz $y(x) = k(x)\cdot y_{\text{homog.}}(x)$ die Lösung von $y' = a(x)y + b(x)$ bestimmen.

Beispiel XIII.2.10. Wir betrachten das Anfangswertproblem

$$\text{(AWP)} \quad \begin{cases} y' = 2xy + x^3,\\ y(0) = c.\end{cases}$$

- Wir lösen die homogene Gleichung durch TdV:
$$\frac{\mathrm{d}y}{\mathrm{d}x} = 2xy \implies \int \frac{\mathrm{d}y}{y} = \int 2x\,\mathrm{d}x \implies \log|y| = x^2 + k \implies y_{\text{homog.}} = ke^{x^2}.$$
- Der Ansatz $y = k(x)e^{x^2}$ liefert
$$y' = k'(x)e^{x^2} + k(x)2xe^{x^2} \stackrel{!}{=} 2xk(x)e^{x^2} + x^3$$
und es folgt $k'(x)e^{x^2} = x^3$ und damit $k'(x) = x^3e^{-x^2}$. Integrieren, erst mit der Substitution $u = x^2$ und dann mit partieller Integration, liefert
$$\begin{aligned}k(x) &= \int x^3 e^{-x^2}\,\mathrm{d}x\\ &= \frac{1}{2}\int ue^{-u}\,\mathrm{d}u\\ &= \frac{1}{2}\Big(u(-e^{-u}) - \int 1(-e^{-u})\,\mathrm{d}u\Big)\\ &= -\frac{1}{2}(u+1)e^{-u} + \tilde{c}\\ &= -\frac{1}{2}(x^2+1)e^{-x^2} + \tilde{c},\end{aligned}$$
wobei wegen $c = y(0) = k(0)e^0 = k(0) = -\frac{1}{2} + \tilde{c}$ folgt, dass $\tilde{c} = c + \frac{1}{2}$ gelten muss.

- Wir erhalten die Lösung
$$y(x) = \Big(-\frac{1}{2}(x^2+1)e^{-x^2} + \frac{1}{2} + c\Big)e^{x^2} = \big(\frac{1}{2} + c\big)e^{x^2} - \frac{1}{2}\big(x^2+1\big)$$
für $x \in \mathbb{R}$. Die Methode aus XIII.2.8 mit $a, b\colon \mathbb{R} \to \mathbb{R}$, $a(x) = 2x$, $b(x) = x^3$ liefert genau das Gleiche. Prüfen Sie das bitte nach.

Bemerkung XIII.2.11. Die Differentialgleichung aus XIII.1.1 zur Kaffeetemperatur kann mit obiger Methode gelöst werden; die aus XIII.1.2 zum Wasserstand in der Badewanne mit Trennung der Variablen.

Literaturverzeichnis

[1] S. Bosch, *Lineare Algebra.* Springer, 2014.
[2] T. Bröcker, *Lineare Algebra und Analytische Geometrie.* Birkhäuser, 2004.
[3] H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Lamotke, K. Mainzer, J. Neukirch, A. Prestel, R. Remmert, *Zahlen.* Springer, 1992.
[4] G. Fischer, *Lernbuch Lineare Algebra und Analytische Geometrie.* Springer, 2019.
[5] G. Fischer, B. Springborn, *Lineare Algebra.* Springer, 2020.
[6] O. Forster, *Analysis 1.* Springer, 2015.
[7] G. Greefrath, R. Oldenburg, H.-S. Siller, V. Ulm, H.-G. Weigand, *Didaktik der Analysis.* Springer, 2016.
[8] H.-W. Henn, A. Filler, *Didaktik der Analytischen Geometrie und Linearen Algebra.* Springer, 2015.
[9] H. Heuser, *Lehrbuch der Analysis. Teil 1.* Vieweg+Teubner, 2009.
[10] S. Hildebrandt, *Analysis 1.* Springer, 2006.
[11] K. Jänich, *Lineare Algebra.* Springer, 2008.
[12] K. Königsberger, *Analysis 1.* Springer, 2004.
[13] W. Walter, *Analysis 1.* Springer, 2009.