M A T H E M A T I K I

2024/2025

Ernst Kuwert

Mathematisches Institut
Universität Freiburg

4. Februar 2025

Inhaltsverzeichnis

1	Zahlen und Vektoren	1
1	Mengen und Abbildungen	1
2	Reelle Zahlen	4
3	Vollständige Induktion	9
4	Geometrie im $\mathbb{R}^n$	14
5	Komplexe Zahlen	21
2	Funktionen, Grenzwerte, Stetigkeit	23
1	Rationale Funktionen	23
2	Grenzwerte von Folgen und Funktionen	26
3	Anwendungen der Stetigkeit	32
3	Differentialrechnung für Funktionen einer Variablen	35
1	Die Ableitung: Definition und Regeln	35
2	Mittelwertsatz und Anwendungen	39
3	Die Exponentialfunktion	41
4	Die trigonometrischen Funktionen	45
4	Integralrechnung	51
1	Das Riemannsches Integral	51
2	Ableitung und Integral	55
5	Approximation von Funktionen	63
1	Taylorapproximation	63
2	Darstellung von Funktionen durch Reihen	67
3	Schwingung mit Dämpfung	73
4	Fourierapproximation	77
6	Lineare Algebra	78
1	Vektorräume	78
2	Lineare Abbildungen und Matrizen	85

Kapitel 1

Zahlen und Vektoren

1 Mengen und Abbildungen

Wenn es um die Formulierung und Kommunikation mathematischer Sachverhalte geht, ist die Umgangssprache nur teilweise geeignet. Die Sprache der Mengenlehre ist oft nützlich. Keine Angst, wir wollen uns nicht inhaltlich mit der Mengenlehre befassen, es geht nur um ein paar Vokabeln: Angabe von Mengen, Teilmenge und Obermenge, Vereinigung, Durchschnitt, Differenz, Komplement, leere Menge. Funktionen lassen sich ebenfalls mit Begriffen der Mengenlehre beschreiben, einschließlich der Konzepte Bild und Urbild, Einschränkung, Verkettung und Graph einer Funktion. Weiter gehören dazu die Begriffe injektiv, surjektiv, bijektiv und Umkehrfunktion.

Der Begriff der Menge wurde 1895 von Georg Cantor eingeführt. Hier eine kurze Fassung:

Eine Menge ist eine Zusammenfassung von bestimmten, wohlunterschiedenen Objekten zu einem Ganzen.

Das ist für unsere Zwecke ausreichend. Die einzelnen Objekte nennen wir *Elemente der Menge*. Wir schreiben $a \in M$ wenn a ein Element der Menge M ist; andernfalls schreiben wir $a \notin M$. Die Entscheidung, ob irgendein Objekt a Element von M ist oder nicht, muss anhand der Beschreibung der Menge M immer möglich sein. Beispiele:

- (a) Das griechische Alphabet ist die Menge

$$M = \{\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta, \theta, \iota, \kappa, \lambda, \mu, \nu, \xi, \omicron, \pi, \rho, \sigma, \tau, \upsilon, \phi, \chi, \psi, \omega\}$$

Die Elemente sind die einzelnen Buchstaben, und wir haben M durch *Aufzählen aller Elemente* angegeben. Die Menge M bleibt gleich, wenn wir die Reihenfolge der Aufzählung ändern oder Elemente doppelt aufführen. Oft werden die Elemente nur unvollständig aufgezählt in der Erwartung, dass man sich den Rest irgendwie denken kann: $M = \{\alpha, \beta, \dots, \omega\}$.

- (b) Betrachte die Menge $N = \{1, 2, \dots, 29, 30\}$ sowie

$$M = \{p \in N \mid p \text{ ist Primzahl}\}.$$

Hier entsteht M durch Auswahl von Elementen der Menge N mittels einer Eigenschaft, nämlich Primzahl zu sein. Es ist $M = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29\}$.

(c) Eine Variante der aufzählenden Beschreibung haben wir bei

$$M = \{n^2 | n = 1, 2, 3, \dots\}.$$

Hier wird M durch die Menge der natürlichen Zahlen “parametrisiert”.

Eine Menge M heißt Teilmenge der Menge N , wenn jedes Element von M auch Element von N ist (Notation: $M \subset N$). Gibt es außerdem mindestens ein Element von N , das nicht zu M gehört, so ist M eine echte Teilmenge von N . Beispiel: unter den Studierenden der Mikrosystemtechnik bilden die männlichen eine echte Teilmenge, denn es gibt ja auch Studentinnen. Statt $M \subset N$ schreiben wir manchmal $N \supset M$, also N ist Obermenge von M .

Zur Veranschaulichung ist es praktisch, Mengen als Gebiete in der Ebene aufzufassen (Venn-Diagramm). Für zwei Mengen A und B können so folgende Mengen darstellen:

$$\begin{aligned} A \cap B &= \{x | x \in A \text{ und } x \in B\} && \text{(Schnittmenge)} \\ A \cup B &= \{x | x \in A \text{ oder } x \in B\} && \text{(Vereinigungsmenge)} \\ A \setminus B &= \{x | x \in A \text{ und } x \notin B\} && \text{(Differenzmenge)}. \end{aligned}$$

Wir führen das Symbol $\emptyset$ ein für die Menge, die kein Element enthält, die *leere Menge*. Das ist praktisch zum Beispiel bei der Bildung von Schnittmengen, es gilt

$$A \cap B = \emptyset \quad \text{wenn } A, B \text{ kein gemeinsames Element haben.}$$

Ist X eine gegebene Grundmenge und $A \subset X$, so nennen wir $A^C = X \setminus A$ das Komplement von A . Das Komplement hängt von der Grundmenge X ab: die Aussage *ich trinke alles außer Ganter* hat verschiedene Bedeutung, je nachdem ob sie sich nur auf die badischen Biere bezieht oder auf alle alkoholischen Getränke.

Oben haben wir die Worte *und* sowie *oder* benutzt, die Aussagen miteinander logisch verknüpfen. Unter einer *Aussage* verstehen wir einen Satz, bei dem wir sinnvoll nach dem Wahrheitswert fragen können, das heißt ist der Satz *wahr* oder *falsch*? Zum Beispiel ist *Der Mars besteht aus grünem Käse* eine Aussage, während *Geh nach Hause!* oder *Was gibts heute in der Mensa?* keine Aussagen sind. Für Aussagen P, Q setze

$\neg P$	<i>nicht P</i>	wahr genau wenn P falsch falsch genau wenn P wahr
$P \wedge Q$	<i>P und Q</i>	wahr genau wenn P, Q beide wahr falsch genau wenn mindestens eine der Aussagen P, Q falsch
$P \vee Q$	<i>P oder Q</i>	wahr genau wenn mindestens eine der Aussagen P, Q wahr falsch genau wenn P, Q beide falsch
$P \Rightarrow Q$	<i>aus P folgt Q</i>	falsch genau wenn P wahr und Q falsch, wahr genau wenn P, Q beide wahr oder wenn P falsch.

Beachten Sie:

- das *oder* ist nicht ausschließend, kein entweder/oder.
- das Äquivalenzzeichen $P \Leftrightarrow Q$ bedeutet $P \Rightarrow Q$ und $Q \Rightarrow P$. Es ist also wahr genau wenn P, Q beide wahr sind oder beide falsch.

- die Implikation $P \Rightarrow Q$ ist gleichbedeutend mit $\neg Q \Rightarrow \neg P$. Dazu folgende Tabelle:

P	Q	$P \Rightarrow Q$	$\neg Q \Rightarrow \neg P$	$Q \Rightarrow P$
w	w	w	w	w
w	f	f	f	w
f	w	w	w	f
f	f	w	w	w

Wir sehen hier dass $Q \Rightarrow P$ nicht gleichbedeutend ist mit $P \Rightarrow Q$, das wird oft nicht beachtet. Wir kommen zum zweiten Begriff aus der Überschrift, den Abbildungen. Seien X, Y Mengen. Eine Abbildung von X nach Y (gleichbedeutend: eine Funktion auf X mit Werten in Y) schreiben wir in der Form

$$f : X \rightarrow Y, x \mapsto f(x).$$

Jedem $x \in X$ wird genau ein Bildpunkt (Funktionswert) $f(x) \in Y$ zugeordnet. Die Menge X heißt Definitionsbereich von f . Mit dem Bild von f meinen wir die Menge der Bildpunkte (Menge der Funktionswerte)

$$f(X) = \{f(x) : x \in X\} \subset Y.$$

Das Urbild einer Menge $B \subset Y$ unter f ist die Menge

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

Verkleinern des Definitionsbereichs ergibt eine Einschränkung von f . Für $A \subset X$ ist

$$f|_A : A \rightarrow Y, (f|_A)(x) = f(x) \quad \text{für alle } x \in A.$$

Beispiel: sei X die Menge aller Waren in einem Supermarkt und Y die Menge aller möglichen Preise in Euro. Ordnen wir jeder Ware $x \in X$ einen Preis $f(x) \in Y$ zu, so haben wir eine Funktion $f : X \rightarrow Y$. Die Menge der Waren, die maximal 1 Euro kosten, ist genau das Urbild des Intervalls $[0, 1]$. Betrachten wir statt aller Waren nur die Menge V der veganen Waren, so ergibt sich die Einschränkung $f|_V$.

Eine naheliegende Abbildung auf X ist die Identität, also $\text{id}_X : X \rightarrow X, \text{id}_X(x) = x$. Die Verkettung (Hintereinanderschaltung) von Abbildungen $f : X \rightarrow Y$ und $g : Y \rightarrow Z$ ist

$$g \circ f : X \xrightarrow{f} Y \xrightarrow{g} Z, (g \circ f)(x) = g(f(x)).$$

Eine aus der Schule bekannte Möglichkeit, sich Funktionen bzw. Abbildungen vorzustellen, bietet der Graph. Dazu bilden wir das kartesische Produkt¹

$$X \times Y = \{(x, y) : x \in X, y \in Y\}.$$

Es gilt $(x, y) = (x', y')$ genau wenn $x = x'$ und $y = y'$. Der Graph von f ist die Teilmenge

$$G_f = \{(x, f(x)) : x \in X\} \subset X \times Y.$$

Als nächstes drei Begriffe zum Abbildungsverhalten: $f : X \rightarrow Y$ heißt

<i>injektiv</i>	aus $f(x) = f(x')$ folgt $x = x'$
<i>surjektiv</i>	zu jedem $y \in Y$ gibt es ein $x \in X$ mit $f(x) = y$
<i>bijektiv</i>	f ist injektiv und surjektiv.

¹René Descartes, 1596–1650

Eine äquivalente Definition von f injektiv ist: aus $x \neq x'$ folgt $f(x) \neq f(x')$, siehe oben. Wir haben die obige Formulierung gewählt weil sie leichter zu checken ist.

Ordnen wir jedem Kind seine Mutter zu, so erhalten wir eine Abbildung von der Menge K aller Kinder in die Menge F aller Frauen. Diese Abbildung ist nicht surjektiv, denn nicht jede Frau ist Mutter. Sie ist auch nicht injektiv, denn es gibt Mütter mit mehr als einem Kind.

Was bedeuten diese drei Begriffe für die Lösbarkeit einer Gleichung $f(x) = y$, wobei $f : X \rightarrow Y$ und $y \in Y$ gegeben?

f injektiv es gibt höchstens eine Lösung
f surjektiv es gibt mindestens eine Lösung
f bijektiv es gibt genau eine Lösung.

Die drei Aussagen gelten für jedes $y \in Y$. Sei nun $f : X \rightarrow Y$ bijektiv. Dann können wir $y \in Y$ die eindeutig bestimmte Lösung von $f(x) = y$ zuordnen. Wir kriegen eine Abbildung $g : Y \rightarrow X$, $y \mapsto g(y)$, wobei $g(y)$ eben diese Lösung ist, das heißt

$$f(g(y)) = y \text{ für alle } y \in Y, \text{ also } f \circ g = \text{id}_Y.$$

Wir nennen $g = f^{-1}$ die zu f inverse Abbildung oder einfach Umkehrfunktion. Wählen wir in der letzten Gleichung $y = f(x)$ so folgt $f(g(f(x))) = f(x)$. Wegen f injektiv ergibt sich

$$g(f(x)) = x, \text{ und damit } g \circ f = \text{id}_X.$$

Umgekehrt: angenommen zu $f : X \rightarrow Y$ gibt es eine Funktion $g : Y \rightarrow X$ mit $f \circ g = \text{id}_Y$ und $g \circ f = \text{id}_X$. Nach der ersten Gleichung ist $g(y)$ eine Lösung der Gleichung $f(x) = y$:

$$f(g(y)) = \text{id}_Y(y) = y.$$

Nach der zweiten Gleichung ist diese Lösung eindeutig, denn aus $f(x) = y = f(x')$ folgt

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

Somit ist f bijektiv, und g ist die Umkehrfunktion von f . Durch Vertauschen der Rollen von f und g folgt: ist g die Umkehrfunktion von f , so ist f auch die Umkehrfunktion von g , insbesondere ist g ebenfalls bijektiv. Wir erhalten für den Graph der Umkehrfunktion, indem wir $y = f(x)$ substituieren,

$$G_g = \{(y, g(y)) : y \in Y\} = \{(f(x), g(f(x))) : x \in X\} = \{(f(x), x) : x \in X\} \subset Y \times X.$$

Der Graph ergibt sich also durch „Spiegelung an der Winkelhalbierenden“.

2 Reelle Zahlen

Wir wiederholen aus der Schule die Gesetze der Addition und Multiplikationen reeller Zahlen, und die Anordnungsgesetze mit Definition des Betrags einer reellen Zahl. Die reellen Zahlen werden durch die Existenz des Supremums charakterisiert. Die rationalen Zahlen bilden eine echte Teilmenge, die aber dicht in den reellen Zahlen liegt.

Aus der Schule kennen wir die Zahlen $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$:

$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, \dots\} && \text{natürliche Zahlen} \\ \mathbb{Z} &= \{\dots, -2, -1, 0, 1, 2, \dots\} && \text{ganze Zahlen} \\ \mathbb{Q} &= \left\{ \frac{p}{q} \mid q \in \mathbb{N}, p \in \mathbb{Z} \right\} && \text{rationale Zahlen} \end{aligned}$$

Manchmal brauchen wir auch $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. Für $x \in \mathbb{Q}$ ist die Darstellung $x = p/q$ nicht eindeutig, eine eindeutige Darstellung ergibt sich durch Kürzen. Die reellen Zahlen $\mathbb{R}$ fassen wir als Punkte auf der Zahlengeraden auf. Anstatt eine abstrakte Konstruktion von $\mathbb{R}$ durchzuführen, stellen wir im Folgenden die Regeln (Axiome) in $\mathbb{R}$ zusammen. Die Gesetze der Addition und Multiplikation lauten wie folgt:

$$\begin{array}{l} + \qquad \qquad \qquad \bullet \\ \text{Assoziativgesetz:} \quad (a + b) + c = a + (b + c) \quad (a \cdot b) \cdot c = a \cdot (b \cdot c) \\ \text{Kommutativgesetz:} \quad a + b = b + a \quad a \cdot b = b \cdot a \\ \text{Neutrales Element:} \quad a + 0 = a \quad a \cdot 1 = a \\ \text{Inverses Element:} \quad a + (-a) = 0 \quad a \cdot \frac{1}{a} = 1 \quad \text{falls } a \neq 0 \\ \text{Distributivgesetz:} \quad a \cdot (b + c) = a \cdot b + a \cdot c. \end{array}$$

Alle bekannten Rechenregeln wie zum Beispiel $(-a)(-b) = ab$ können daraus hergeleitet werden. Wir erwähnen die *Nullteilerfreiheit*:

$$\text{Aus } a \cdot b = 0 \text{ folgt } a = 0 \text{ oder } b = 0.$$

Denn ist $a \neq 0$, so folgt

$$b = \left(\frac{1}{a} \cdot a \right) \cdot b = \frac{1}{a} \cdot (a \cdot b) = \frac{1}{a} \cdot 0 = 0.$$

Den Punkt bei der Multiplikation lassen wir meistens weg. Auch die Regeln der Bruchrechnung lassen sich aus den Körperaxiomen herleiten, wir listen sie hier auf:

Satz 2.1 (Bruchrechnung) Für $a, b, c, d \in \mathbb{R}$ mit $c, d \neq 0$ gilt:

- (1) $\frac{a}{c} + \frac{b}{d} = \frac{ad + bc}{cd}$ (Gleichnamig machen von Brüchen),
- (2) $\frac{a}{c} \cdot \frac{b}{d} = \frac{ab}{cd}$ (Multiplikation von Brüchen),
- (3) $\frac{a/c}{b/d} = \frac{ad}{bc}$, falls zusätzlich $b \neq 0$ (Division von Brüchen).

Als nächstes erklären wir die Anordnung von $\mathbb{R}$. Die Ungleichung $a < b$ bedeutet auf der Zahlengeraden, dass a links von b liegt bzw. $a - b$ links von Null. Offenbar gilt genau eine der Relationen $a < b$, $a > b$ oder $a = b$. Hier einige Regeln:

- Aus $a, b > 0$ folgt $a + b > 0$ und $a \cdot b > 0$.
- Aus $a > b, b > c$ folgt $a > c$ (*Transitivität*).
- Aus $a > b$ und $c > d$ folgt $a + c > b + d$ (*Addition von Ungleichungen*)
- Aus $a > b$ folgt $\begin{cases} ac > bc & \text{falls } c > 0 \\ ac < bc & \text{falls } c < 0 \end{cases}$ (*Multiplikation mit einer Zahl*).

Dass sich in der letzten Regel das Vorzeichen umdreht ist eine häufige Fehlerquelle. Wir wollen diese und weitere Regeln nicht herleiten. Eine wichtige Konsequenz ist jedenfalls: *Quadrate in $\mathbb{R}$ sind nichtnegativ*, denn

$$a^2 = \begin{cases} a \cdot a > 0 & \text{im Fall } a > 0, \\ (-a) \cdot (-a) > 0 & \text{im Fall } -a > 0. \end{cases}$$

Definition 2.1 *Der Betrag einer Zahl $a \in \mathbb{R}$ ist*

$$|a| = \begin{cases} a & \text{falls } a \geq 0 \\ -a & \text{falls } a \leq 0. \end{cases} \quad (2.1)$$

Für $a = 0$ ergibt sich jeweils $|a| = 0$, aus praktischen Gründen haben wir den Overlap der beiden Fälle hier erlaubt. Auf der Zahlengeraden ist $|a|$ der Abstand zum Nullpunkt. Die Definition hat folgende Konsequenz: Wenn Sie ein Betragszeichen beseitigen möchten, müssen Sie eine Fallunterscheidung machen, je nachdem ob der Term in den Betragsstrichen positiv oder negativ ist.

Satz 2.2 (Rechnen mit Beträgen) *Für $a, b \in \mathbb{R}$ gelten folgende Aussagen:*

- (1) $a \leq |a|$.
- (2) $|-a| = |a|$.
- (3) $|ab| = |a| \cdot |b|$.
- (4) $|a + b| \leq |a| + |b|$.
- (5) $|a - b| \geq ||a| - |b||$.

BEWEIS: Im Fall $a \geq 0$ ist $a = |a|$, andernfalls ist $a \leq 0 \leq |a|$, dies zeigt Aussage (1). Auch (2) folgt direkt aus Definition 2.1, denn

$$|-a| = \begin{cases} -a & \text{falls } -a \geq 0 \\ -(-a) & \text{falls } -a \leq 0 \end{cases} = \begin{cases} -a & \text{falls } a \leq 0 \\ a & \text{falls } a \geq 0 \end{cases} = |a|.$$

In (3) bleiben die linke und rechte Seite gleich, wenn wir a durch $-a$ ersetzen, dasselbe gilt bezüglich b . Also können wir $a, b \geq 0$ annehmen, und erhalten $|ab| = ab = |a| \cdot |b|$ wie verlangt. Für (4) schätzen wir mit (1) wie folgt ab:

$$|a + b| = \pm(a + b) = \pm a + (\pm b) \leq |a| + |b|.$$

Schließlich gilt $|a| = |a - b + b| \leq |a - b| + |b|$ nach (4), also $|a - b| \geq |a| - |b|$. Durch Vertauschen von a und b folgt (5). $\square$

Wie liegen die rationalen Zahlen in $\mathbb{R}$ drin? Das regelt das *Archimedische Axiom*:

Zu jedem $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, existiert ein $n \in \mathbb{N}$ mit $1/n < \varepsilon$.

Es gibt dann auch zu jedem $K \in \mathbb{R}$ ein $n \in \mathbb{N}$ mit $n > K$. Für $K \leq 0$ ist das sowieso klar. Im Fall $K > 0$ wähle oben $\varepsilon = 1/K > 0$, und erhalte ein $n \in \mathbb{N}$ mit $1/n < 1/K$ bzw. $n > K$.

Eine wichtige Konsequenz ist, dass die rationalen Zahlen auf der Zahlengerade $\mathbb{R}$ dicht verteilt sind. Um das zu formulieren führen wir folgende Bezeichnungen für Intervalle mit Grenzen $a, b \in \mathbb{R}$ ein:

$(a, b) = \{x \in \mathbb{R} : a < x < b\}$	offenes Intervall
$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$	abgeschlossenes Intervall
$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$	linksseitig abgeschlossen, rechtsseitig offen
$(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$	linksseitig offen, rechtsseitig abgeschlossen
$ I = b - a$ für ein Intervall I	Intervalllänge

Es ist praktisch, $+\infty$ und $-\infty$ als *offene* Intervallgrenzen zuzulassen, zum Beispiel ist $(-\infty, 1] = \{x \in \mathbb{R} : -\infty < x \leq 1\}$.

Satz 2.3 ($\mathbb{Q}$ ist dicht in $\mathbb{R}$) Für alle $a, b \in \mathbb{R}$, $a < b$, gibt es ein $q \in \mathbb{Q}$ mit $q \in (a, b)$.

BEWEIS: Sei $n \in \mathbb{N}$ mit $1/n < b - a$. Die Zahlen k/n , $k \in \mathbb{Z}$, sind rational und bilden ein Gitter, mit Gitterlänge kleiner als $b - a$. Sei $k \in \mathbb{Z}$ minimal mit $k/n > a$. Dann folgt

$$\frac{k}{n} = \underbrace{\frac{k-1}{n}}_{\leq a} + \underbrace{\frac{1}{n}}_{< b-a} < a + (b-a) = b, \quad \text{also } q = \frac{k}{n} \in (a, b).$$

Warum gibt es das minimale k ? Es gibt $k_+ \in \mathbb{N}$ mit $k_+ > na$, also $k_+/n > a$. Es gibt weiter $k_- \in \mathbb{N}$ mit $k_- > -na$, also $-k_-/n < a$. Es kommen nur k mit $-k_- < k \leq k_+$ infrage, das sind endlich viele, unter diesen wählen wir das kleinste k mit $k/n > a$. $\square$

Der zentrale Begriff der Analysis ist der Grenzwert. Wir betrachten hier den Grenzwert einer Zahlenfolge a_n , also $a_1, a_2, a_3, \dots$ mit $a_n \in \mathbb{R}$.

Definition 2.2 (Konvergenz) Die Folge a_n konvergiert gegen $a \in \mathbb{R}$, falls gilt:

Zu jedem $\varepsilon > 0$ gibt es ein $K \geq 0$, so dass für alle $n > K$ gilt: $|a_n - a| < \varepsilon$.

Die Zahl a heißt Grenzwert der Folge, wir schreiben

$$\lim_{n \rightarrow \infty} a_n = a \quad \text{oder} \quad a_n \rightarrow a \text{ für } n \rightarrow \infty.$$

In der Schule wird das eher in Worten ausgedrückt: zum Beispiel a_n *strebt gegen* a , oder etwas genauer *die Abweichung zwischen a_n und a wird beliebig klein*. Wir brauchen aber eine präzise Formulierung, um damit zu argumentieren. Die Zahl $\varepsilon > 0$ gibt die Genauigkeit vor, wie groß der Abstand $|a_n - a|$ höchstens sein soll. Die Zahl K gibt an, ab wann diese Genauigkeit erfüllt ist. In der Regel werden die ersten Folgenglieder das nicht leisten. Betrachten

wir zum Beispiel $a_n = \frac{1}{n}$. Die Folge sollte gegen $a = 0$ konvergieren, der Abstand ist dann $|a_n - a| = 1/n$. Wie muss K gewählt werden, damit $|a_n - a| < \varepsilon$ ist für alle $n > K$?

$$\begin{array}{rcccccc} \varepsilon = & 1 & 0,5 & 0,2 & 0,1 & 0,01 & \dots \\ K = & 1 & 2 & 5 & 10 & 100 & \dots \end{array}$$

Wir haben hier K optimal gewählt, aber das ist im allgemeinen unnötig, Hauptsache die Abschätzung stimmt für alle $n > K$. Wie man am Beispiel sieht, muss für bessere Genauigkeit, also $\varepsilon > 0$ kleiner, typischerweise ein größeres K genommen werden. Der Punkt ist nun, wir müssen für *jede* gegebene Genauigkeit $\varepsilon > 0$ ein passendes K finden (ausrechnen).

Beispiel 2.1 Die Folge $a_n = 1/n$ konvergiert gegen $a = 0$. Sei dazu $\varepsilon > 0$ gegeben, wir wählen $K = 1/\varepsilon$. Es folgt für alle $n > K$

$$|a_n - a| = |1/n - 0| = 1/n < 1/K = \varepsilon.$$

Wir werden den Begriff der Konvergenz noch später studieren, jetzt geht es erstmal um die reellen Zahlen. Der Unterschied zwischen $\mathbb{Q}$ und $\mathbb{R}$ besteht darin, dass wir in $\mathbb{R}$ beliebige Grenzprozesse durchführen können. Wir können damit Lösungen von Gleichungen als Grenzwerte konstruieren.

Als Beispiel betrachten wir die Gleichung $x^2 = 2$. Folgendes Verfahren bestimmt schrittweise Approximationen durch Dezimalzahlen. Im ersten Schritt ist $1^2 < 2$ und $2^2 > 2$, wir setzen $a_0 = 1$ und $b_0 = 2$. Im nächsten Schritt ist $1,4^2 = 1,96 < 2$ und $1,5^2 = 2,25 > 2$, also kommt $a_1 = 1,4$ und $b_1 = 1,5$. Allgemein wählen wir die Ziffer so dass $a_n^2 \leq 2$ und $b_n^2 > 2$, wobei $b_n = a_n + 10^{-n}$. Wir erhalten die wachsende Folge a_n und die fallende Folge b_n :

$$\begin{array}{rcccccc} & a_n & 1 & 1,4 & 1,41 & 1,414 & 1,4142 & \dots \\ \text{Näherungen von } \sqrt{2} : & b_n & 2 & 1,5 & 1,42 & 1,415 & 1,4143 & \dots \\ & b_n - a_n & 1 & 0,1 & 0,01 & 0,001 & 0,0001 & \dots \end{array}$$

Wir wollen $\sqrt{2}$ als Grenzwert der Folgen a_n bzw. b_n definieren. *Wenn* diese Folgen konvergieren so haben sie denselben Grenzwert, denn es gilt $b_n - a_n = 10^{-n} \rightarrow 0$. Und angenommen wir hätten $a_n, b_n \rightarrow x$ mit $n \rightarrow \infty$, dann folgt

$$x^2 = \lim_{n \rightarrow \infty} a_n^2 \leq 2 \quad \text{und} \quad x^2 = \lim_{n \rightarrow \infty} b_n^2 \geq 2.$$

Also ist x tatsächlich Lösung der Gleichung $x^2 = 2$. Bleibt die Frage ob die Folgen a_n, b_n konvergieren. In der Definition der Konvergenz geht es um den Abstand von a_n zum Grenzwert a . Das können wir hier nicht brauchen, denn wir wollen $a = \sqrt{2}$ ja erst definieren. Was wir aber machen können ist, den Abstand von a_n und a_m anzuschauen, ohne Einschränkung für $m > n$. Dieser Abstand wird klein für n, m hinreichend gross. Und zwar gilt $a_n \leq \dots \leq a_m \leq b_m \leq \dots \leq b_n$, daraus folgt

$$0 \leq a_m - a_n \leq b_m - a_n \leq b_n - a_n = 10^{-n} \rightarrow 0 \quad \text{mit } n \rightarrow \infty.$$

Folgen mit dieser Eigenschaft heißen Cauchyfolgen². Formal ist a_n Cauchyfolge wenn gilt:

$$\text{Zu jedem } \varepsilon > 0 \text{ gibt es ein } K \in \mathbb{R}, \text{ so dass } |a_m - a_n| < \varepsilon \text{ für alle } m, n > K.$$

Die Existenz von Grenzwerten wird nun durch das **Vollständigkeitsaxiom** garantiert:

²A.L. Cauchy, 1789–1857

Jede Cauchyfolge in $\mathbb{R}$ konvergiert, hat also einen Grenzwert in $\mathbb{R}$.

Insbesondere haben unsere Folgen a_n und b_n einen Grenzwert x , und wie oben erklärt ist $x^2 = 2$. Die a_n und b_n sind endliche Dezimalzahlen und damit in $\mathbb{Q}$, aber der Grenzwert x liegt nicht in $\mathbb{Q}$. Das sieht man wie folgt.

Angenommen es gibt ein $x \in \mathbb{Q}$ mit $x^2 = 2$. Wir können $x > 0$ annehmen (ersetze x durch $-x$), also $x = p/q$ mit $p, q \in \mathbb{N}$. Wir können weiter annehmen, dass p und q nicht beide gerade sind. Andernfalls kürzen wir den Bruch so oft durch 2 bis das der Fall ist. Bei jedem Kürzen verkleinern sich Zähler und Nenner, also ist nach endlich vielen Schritten der Zähler oder Nenner ungerade. Nun folgt aus $x^2 = 2$:

$$\begin{aligned} \frac{p^2}{q^2} = 2 &\Rightarrow p^2 = 2q^2, \text{ also } p^2 \text{ gerade} \\ &\Rightarrow p \text{ gerade, also } p = 2p' \text{ mit } p' \in \mathbb{N} \\ &\Rightarrow q^2 = \frac{p^2}{2} = 2(p')^2, \text{ also } q^2 \text{ gerade} \\ &\Rightarrow q \text{ gerade.} \end{aligned}$$

Also sind p, q doch beide gerade, ein Widerspruch. Wir haben benutzt, dass für p^2 gerade auch p gerade ist. Denn wäre p ungerade, also Endziffer 1,3,5,7,9, dann hätte p^2 die Endziffer 1,9,5,9,1, wäre also nicht gerade.

Wir haben nun alle Axiome in $\mathbb{R}$ zusammen: die algebraischen Rechenregeln, die Anordnungsaxiome, das Archimedische Axiom und das Vollständigkeitsaxiom. Die Begriffe der Konvergenz und der Vollständigkeit sind überhaupt nicht trivial, sie werden uns weiter beschäftigen.

3 Vollständige Induktion

Wir behandeln das Prinzip der vollständigen Induktion. Erste Beispiele sind die Formeln für die arithmetische und geometrische Summe. Danach beschäftigen wir uns mit dem Zählen, und bestimmen die Anzahl von Permutationen und Kombinationen.

Das Beweisverfahren der vollständigen Induktion beruht auf folgendem Prinzip:

Sei $A(n)$ eine Folge von Aussagen für $n \in \mathbb{N}$. Es gelte:

- (1) $A(1)$ ist wahr.
- (2) $A(n-1)$ ist wahr $\Rightarrow A(n)$ ist wahr.

Dann sind alle Aussagen $A(n)$ wahr.

Ein Induktionsbeweis funktioniert also in zwei Schritten: erst wird die Aussage $A(n)$ für den Fall $n = 1$ verifiziert (Induktionsanfang). Dann wird *vorausgesetzt*, dass $A(n-1)$ richtig ist (Induktionsannahme), und daraus $A(n)$ hergeleitet (Induktionsschluss). Dazu kann man $n \geq 2$ annehmen, denn $A(1)$ ist ja schon bekannt. Das Verfahren spielt eine wesentliche Rolle

bei Rekursionen. Betrachten wir zum Beispiel die Summe der ersten n natürlichen Zahlen $S_n = 1 + 2 + \dots + n$. Diese ergibt sich rekursiv wie folgt:

$$S_1 = 1, \quad S_n = S_{n-1} + n \quad \text{für } n \geq 2.$$

Offenbar liefert die Rekursion

$$S_1 = 1, \quad S_2 = S_1 + 2 = 3, \quad S_3 = S_2 + 3 = 6, \dots$$

Wir behaupten dass folgende allgemeine Formel gilt:

$$S_n = \frac{n(n+1)}{2}.$$

Für $n = 1$ stimmt das. Und wenn die Formel für $n - 1$ gilt so folgt

$$S_n = S_{n-1} + n = \frac{(n-1)((n-1)+1)}{2} + n = \frac{n(n-1+2)}{2} = \frac{n(n+1)}{2}.$$

Damit ist die Formel bewiesen, und sogar ziemlich einfach. Das ist durchaus typisch für die Induktion: das Problem ist eher die richtige Formel zu finden, der Beweis per Induktion geht dann nach Schema F. Übrigens hat Gauß im Alter von 5 Jahren die Formel so gezeigt :

$$\begin{array}{rcl} S_n & = & 1 + 2 + \dots + n \\ S_n & = & n + (n-1) + \dots + 1 \quad \text{also } 2S_n = n(n+1). \\ 2S_n & = & n+1 + n+1 + \dots + n+1 \end{array}$$

Betrachten wir als zweites Beispiel die geometrische Summe, das ist die Summe der ersten $n + 1$ Potenzen einer Zahl x , genauer $S_n = x^0 + x^1 + \dots + x^n$. Es ist praktisch bei $x^0 = 1$ anzufangen. Der Fall $x = 1$ ist klar, also nehmen wir $x \neq 1$ an. Die Rekursion lautet

$$S_0 = x^0 = 1, \quad S_n = S_{n-1} + x^n.$$

Es geht also statt bei $n = 1$ bei $n = 0$ los, die ersten paar Schritte lauten

$$S_1 = S_0 + x^1 = 1 + x, \quad S_2 = S_1 + x^2 = 1 + x + x^2, \quad S_3 = S_2 + x^3 = 1 + x + x^2 + x^3, \dots,$$

Wir zeigen nun die allgemeine Formel

$$S_n = \frac{1 - x^{n+1}}{1 - x} \quad \text{für } x \neq 1.$$

Für $n = 0$ steht rechts Eins, das stimmt. Angenommen die Formel gilt für S_{n-1} , dann folgt

$$S_n = S_{n-1} + x^n = \frac{1 - x^{(n-1)+1}}{1 - x} + x^n = \frac{1 - x^n + (1 - x)x^n}{1 - x} = \frac{1 - x^{n+1}}{1 - x}.$$

Auch hier gibt es einen Trick, wie man die Summe ohne Induktion ausrechnen kann:

$$\begin{aligned} 1 - x^{n+1} &= (1 + x + \dots + x^{n-1} + x^n) - (x + x^2 + \dots + x^n + x^{n+1}) \\ &= (1 - x) + (x - x^2) + \dots + (x^n - x^{n+1}) \\ &= (1 - x)(1 + x + \dots + x^n). \end{aligned}$$

Wir notieren die beiden Formeln nochmal mit dem Summenzeichen:

$$\sum_{k=1}^n k = \frac{n(n+1)}{2},$$

$$\sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x} \quad \text{wobei } x \neq 1.$$

Hinter dem Summenzeichen steht die Formel für den k -ten Summanden. Unter dem Summenzeichen ist angegeben, wo der Laufindex k startet, über dem Summenzeichen steht die obere Grenze für den Laufindex k .

Wir wollen als nächstes die Elemente gewisser Mengen zählen. Als erstes zeigen wir: wenn man vier Paar Socken in eine Kommode mit drei Schubladen packt, müssen zwei Paare in dieselbe Schublade. Das geht natürlich auch mit Briefen und Briefkästen. Dieses sogenannte *Schubfachprinzip* wurde zuerst von G.L. Dirichlet (1805-1859) festgestellt.

Satz 3.1 *Ist $f : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$ injektiv mit $m, n \in \mathbb{N}$, so folgt $m \leq n$.*

BEWEIS: durch Induktion nach $n \in \mathbb{N}$. Für $n = 1$ haben wir $f : \{1, \dots, m\} \rightarrow \{1\}$ injektiv, das geht nur für $m = 1$. Sei nun eine injektive Abbildung $f : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$ gegeben; zu zeigen ist $m \leq n$. Kommt n im Bild von f nicht vor, so bildet f die Menge $\{1, \dots, m\}$ nach $\{1, \dots, n-1\}$ injektiv ab. Nach Induktion folgt dann $m \leq n-1$. Andernfalls gibt es genau ein $k \in \{1, \dots, m\}$ mit $f(k) = n$. Sei $\varphi : \{1, \dots, m-1\} \rightarrow \{1, \dots, m\}$ mit

$$\varphi(i) = \begin{cases} i & \text{für } i = 1, \dots, k-1, \\ i+1 & \text{für } i = k, \dots, m. \end{cases}$$

Da $\varphi(i) \neq k$ für alle i bildet $f \circ \varphi$ von $\{1, \dots, m-1\}$ nach $\{1, \dots, n-1\}$ injektiv ab.

$$\begin{array}{ccccccc} & 1 & \dots & k-1 & - & k & \dots & m-1 \\ \varphi \downarrow & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & 1 & \dots & k-1 & & k+1 & \dots & m \\ f \downarrow & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ & f(1) & \dots & f(k-1) & & f(k) = n & & f(k+1) & \dots & f(m) \end{array}$$

Per Induktion folgt $m-1 \leq n-1$ und damit $m \leq n$. □

Jetzt kommen wir zum Zählen. Man könnte meinen dass wir dazu gar nichts sagen müssen, wir können das ja alle seit wir kleine Kinder sind. Aber wir brauchen ein Konzept dass wir zum Beispiel auch als Programm implementieren können, und in mathematischen Argumenten benutzen können. Darum müssen wir die Sache formalisieren. Letztlich machen wir aber nichts anderes als mein Enkel Theo, der stolz ist dass er schon 10 Dinge abzählen kann.

Definition 3.1 (Zahl der Elemente) *Sei M eine Menge. Gibt es $f : \{1, \dots, n\} \rightarrow M$ bijektiv für ein $n \in \mathbb{N}$, so heißt M endlich und $n = \#M$ ist die Anzahl seiner Elemente. Die leere Menge ist ebenfalls endlich mit $\#\emptyset = 0$.*

Dass die Zahl der Elemente eindeutig definiert ist, folgt aus dem Schubfachprinzip.

$$\begin{array}{ccc}
 \{1, \dots, n\} & & \\
 | & \searrow f & \\
 g^{-1} \circ f & & M \\
 \downarrow & \nearrow g & \\
 \{1, \dots, m\} & &
 \end{array}$$

Sind f und g bijektiv, so ist $g^{-1} \circ f$ injektiv und damit $n \leq m$ nach dem Schubfachprinzip. Analog folgt $m \leq n$.

Jetzt wollen wir das anwenden. Wir wollen als erstes die Anzahl der Anordnungen (oder Umordnungen oder Permutationen) von n Dingen bestimmen. Für $n = 2$ gibt es zwei Möglichkeiten, für $n = 3$ sechs:

$$\begin{array}{cccc}
 12 & 123 & 231 & 312 \\
 21 & 213 & 321 & 132
 \end{array}$$

Für die allgemeine Antwort brauchen wir folgende Notation (wir setzen $0! = 1$):

$$n! = 1 \cdot 2 \cdot \dots \cdot n \quad (n\text{-Fakultät}).$$

Satz 3.2 (Zahl der Permutationen) Für $n \in \mathbb{N}$ sei S_n die Menge der bijektiven Abbildungen $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Dann gilt $\#S_n = n!$.

BEWEIS: Wir zeigen die Behauptung durch Induktion, wobei der Induktionsanfang $n = 1$ offensichtlich ist. Sei $S_{n,k}$ die Menge der Permutationen von $\{1, \dots, n\}$, die die Nummer k auf n abbilden. Da die restlichen $n-1$ Nummern beliebig bijektiv auf $\{1, \dots, n-1\}$ abgebildet werden, hat $S_{n,k}$ genau $(n-1)!$ Elemente nach Induktion. Es folgt durch Summation

$$\#S_n = \#S_{n,1} + \dots + \#S_{n,n} = n \cdot (n-1)! = n!.$$

□

Definition 3.2 (Binomialkoeffizienten) Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}$ setzen wir

$$\binom{\alpha}{k} = \frac{\alpha \cdot (\alpha - 1) \cdot \dots \cdot (\alpha - k + 1)}{1 \cdot 2 \cdot \dots \cdot k}, \quad \text{sowie} \quad \binom{\alpha}{0} = 1.$$

Lemma 3.1 (Additionstheorem für Binomialkoeffizienten) Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}$ erfüllen die Binomialkoeffizienten die Formel

$$\binom{\alpha}{k} = \binom{\alpha - 1}{k} + \binom{\alpha - 1}{k - 1}.$$

BEWEIS: Für $k = 1$ ist die Formel leicht zu sehen. Für $k \geq 2$ berechnen wir

$$\begin{aligned}
 \binom{\alpha - 1}{k} + \binom{\alpha - 1}{k - 1} &= \frac{(\alpha - 1) \cdot \dots \cdot ((\alpha - 1) - k + 1)}{1 \cdot 2 \cdot \dots \cdot k} + \frac{(\alpha - 1) \cdot \dots \cdot ((\alpha - 1) - (k - 1) + 1)}{1 \cdot 2 \cdot \dots \cdot (k - 1)} \\
 &= \frac{(\alpha - 1) \cdot \dots \cdot (\alpha - k + 1) \cdot (\alpha - k + k)}{1 \cdot 2 \cdot \dots \cdot k} \\
 &= \frac{\alpha \cdot (\alpha - 1) \cdot \dots \cdot (\alpha - k + 1)}{1 \cdot 2 \cdot \dots \cdot k} \\
 &= \binom{\alpha}{k}.
 \end{aligned}$$

Satz 3.4 (Binomische Formel) Für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (3.4)$$

BEWEIS: Multiplizieren wir das n -fache Produkt $(a + b)^n = (a + b) \cdot (a + b) \cdot \dots \cdot (a + b)$ aus und ordnen nach den Potenzen von a, b so ergibt sich eine Summe von Termen der Form $a^k b^{n-k}$ mit $0 \leq k \leq n$. Wir erhalten den Term $a^k b^{n-k}$ genau dann, wenn wir in k Klammern a nehmen und in den anderen b . Für die Wahl der a -Klammern gibt es $\binom{n}{k}$ Möglichkeiten nach Satz 3.3, so oft kommt also der Term $a^k b^{n-k}$ Term vor. $\square$

4 Geometrie im $\mathbb{R}^n$

In diesem Abschnitt erklären wir ein paar geometrische Grundbegriffe, die in den Anwendungsfächern vorkommen. Dazu gehören Geraden und Ebenen, sowie allgemeiner Unterräume im $\mathbb{R}^n$, die Euklidische Länge, das Skalarprodukt und der Winkel zwischen zwei Vektoren. Wir zeigen auch die Ungleichung von Cauchy-Schwarz und die Dreiecksungleichung. Auf $\mathbb{R}^3$ definieren wir schließlich das Kreuzprodukt.

Im folgenden sei $n \in \mathbb{N}$, zum Beispiel $n = 2$ oder $n = 3$. Wir bezeichnen mit $\mathbb{R}^n$ das n -fache kartesische Produkt $\mathbb{R} \times \dots \times \mathbb{R}$, also die Menge aller n -Tupel

$$\mathbb{R}^n = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_i \in \mathbb{R} \right\}.$$

Zwei Punkte $x, y \in \mathbb{R}^n$ sind genau dann gleich, wenn alle Komponenten gleich sind, also $x_i = y_i$ für alle $i = 1, \dots, n$. Größen wie zum Beispiel Geschwindigkeit, Impuls, Beschleunigung oder Feldstärke werden auch durch n -Tupel im $\mathbb{R}^n$ beschrieben, jedenfalls im Fall $n = 3$. Man spricht dann von Vektoren statt von Punkten, und oft verwendet man die Notation $\vec{x}$ statt x . Die Unterscheidung zwischen Punkten und Vektoren ist in den Anwendungen unter Umständen wichtig, darum wird die Notation mit dem Pfeil benutzt. Für uns ist das weniger relevant, und die Schreibweise mit den Pfeilen wäre schwierig durchzuhalten, daher schreibe ich Vektoren ohne den Pfeil. Zum Beispiel können wir Vektoren addieren und mit reellen Zahlen multiplizieren (zum Beispiel strecken):

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad \lambda x = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

In diesem Zusammenhang nennt man die Zahl λ auch einen Skalar, und spricht von Skalarmultiplikation. Die Standard-Einheitsvektoren sind

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \text{für } i = 1, \dots, n.$$

Mit ihnen ergibt sich für jeden Vektor $x \in \mathbb{R}^n$ die Darstellung

$$x = \sum_{i=1}^n x_i e_i.$$

Die Vektoraddition und Skalarmultiplikation erfüllen folgende Regeln, die alle direkt aus den Definitionen folgen:

- (1) $(x + y) + z = x + (y + z)$
- (2) $x + 0 = x$ für alle x , wobei $0 = (0, \dots, 0)$ (Nullvektor)
- (3) Es gilt $x + (-x) = 0$
- (4) $x + y = y + x$
- (5) $(\lambda + \mu)x = \lambda x + \mu x$
- (6) $(\lambda\mu)x = \lambda(\mu x)$
- (7) $\lambda(x + y) = \lambda x + \lambda y$
- (8) $1 \cdot x = x$.

Ist $u \in \mathbb{R}^n$ nicht der Nullvektor, so nennen wir die Menge der Punkte $x = \lambda u$, $\lambda \in \mathbb{R}$, die von u aufgespannte Gerade. Hier ist λ eindeutig. Denn sei $\lambda_1 u = \lambda_2 u$ mit $\lambda_1 \neq \lambda_2$, dann folgt

$$0 = \frac{1}{\lambda_1 - \lambda_2}(\lambda_1 u - \lambda_2 u) = \frac{1}{\lambda_1 - \lambda_2}(\lambda_1 - \lambda_2)u = u, \quad \text{Widerspruch.}$$

Wir nennen λ die Koordinate von x bezüglich u . Ähnliches gilt mit zwei Vektoren: sei $u_1 \neq 0$ und u_2 sei nicht auf der Geraden die durch u_1 aufgespannt wird. Dann nennen wir die Menge der Vektoren $\lambda_1 u_1 + \lambda_2 u_2$ mit $\lambda_{1,2} \in \mathbb{R}$ die durch u_1, u_2 aufgespannte Ebene. Die Koordinaten $\lambda_{1,2}$ sind wieder eindeutig: angenommen es ist $\lambda_1 u_1 + \lambda_2 u_2 = \mu_1 u_1 + \mu_2 u_2$, bzw.

$$(\lambda_1 - \mu_1)u_1 + (\lambda_2 - \mu_2)u_2 = 0.$$

Wäre $\lambda_2 \neq \mu_2$ so hätten wir

$$u_2 = -\frac{\lambda_1 - \mu_1}{\lambda_2 - \mu_2}u_1.$$

Das bedeutet, u_2 liegt auf der von u_1 aufgespannten Geraden, Widerspruch. Also ist $\lambda_2 = \mu_2$, und weiter folgt leicht $\lambda_1 = \mu_1$. Das geht auch mit drei Vektoren: angenommen u_1, u_2 spannen eine Ebene auf, und u_3 liegt nicht auf dieser Ebene. Dann nennen wir die Menge der Punkte $x = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3$ mit $\lambda_{1,2,3} \in \mathbb{R}$ den durch u_1, u_2, u_3 aufgespannten Unterraum U . Im Fall $n = 3$ ist das der ganze $\mathbb{R}^3$, im Fall $n \geq 4$ eben ein drei-dimensionaler Unterraum. Die Koordinaten sind wieder eindeutig: angenommen es ist

$$(\lambda_1 - \mu_1)u_1 + (\lambda_2 - \mu_2)u_2 + (\lambda_3 - \mu_3)u_3 = 0.$$

Wäre $\lambda_3 \neq \mu_3$ so hätten wir

$$u_3 = -\frac{\lambda_1 - \mu_1}{\lambda_3 - \mu_3}u_1 - \frac{\lambda_2 - \mu_2}{\lambda_3 - \mu_3}u_2.$$

Das heißt u_3 liegt in der durch u_1, u_2 aufgespannten Ebene, Widerspruch. Also gilt $\lambda_3 = \mu_3$, und $\lambda_{1,2} = \mu_{1,2}$ folgt wie oben. Ist G die von u aufgespannte Gerade, so nennen wir u eine Basis von G . Analog sind u_1, u_2 eine Basis der aufgespannten Ebene E , und u_1, u_2, u_3 eine Basis des aufgespannten Unterraums Raums U . Eine Basis ist nicht eindeutig bestimmt, zum Beispiel wird G auch durch $2u$ oder $-u$ aufgespannt. Die Koordinaten sind nicht die Standardkoordinaten des Punkts x , sondern sie hängen von der Basis ab. Ist x ein Punkt auf G, E oder U , so stellt sich die Frage wie man die Koordinaten von x bezüglich der gegebenen Basis bestimmt. Dazu muss ein lineares Gleichungssystem gelöst werden, das werden wir später machen.

Die Euklidische Länge (oder die Euklidische Norm) eines Vektors ist definiert durch

$$|x| = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \quad \text{für } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Die Definition leitet sich aus dem Satz des Pythagoras ab (Bild), es gilt

$$|x| = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{für } n = 2, \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{für } n = 3. \end{cases}$$

Offenbar gelten für $x \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ die Regeln

$$|x| \geq 0 \quad (\text{Gleichheit nur für } x = 0) \quad \text{sowie} \quad |\lambda x| = |\lambda| |x|.$$

Der Euklidische Abstand von zwei Punkten $x, y \in \mathbb{R}^n$ ist

$$|x - y| = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

Die Euklidische Länge $|x|$ ist also der Abstand zum Nullpunkt. Eine weitere nützliche Größe ist das Skalarprodukt zwischen Vektoren im $\mathbb{R}^n$:

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \quad \text{für } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

Man schreibt das Skalarprodukt auch in der Form $x \cdot y$, also mit einem Punkt. Wir bleiben aber bei den Klammern. Folgende Regeln ergeben sich direkt aus der Definition, wobei $x, y, z \in \mathbb{R}^n$ und $\lambda, \mu \in \mathbb{R}$.

$$\begin{aligned} \langle x, y \rangle &= \langle y, x \rangle && \text{(Symmetrie)} \\ \langle \lambda x + \mu y, z \rangle &= \lambda \langle x, z \rangle + \mu \langle y, z \rangle && \text{(Bilinearität)} \\ \langle x, x \rangle &= |x|^2 \geq 0 && \text{(Positivität)}. \end{aligned}$$

Zwei Vektoren $x, y \in \mathbb{R}^n$ heißen orthogonal, wenn $\langle x, y \rangle = 0$. Im Dreieck mit Eckpunkten $0, x$ und y gilt dann der Satz des Pythagoras:

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

das heißt im Nullpunkt ist ein rechter Winkel. Für eine Gerade, eine Ebene oder einen Unterraum wie oben können wir mithilfe des Skalarprodukts spezielle Basen wählen. Wir erklären das zuerst für eine Ebene E . Eine Basis v_1, v_2 von E heißt Orthonormalbasis, wenn

$$|v_1| = |v_2| = 1 \quad \text{und} \quad \langle v_1, v_2 \rangle = 0.$$

Die Vektoren sind also auf Eins normiert und zueinander senkrecht. Sei u_1, u_2 irgendeine Basis von E , dann können wir daraus eine Orthonormalbasis konstruieren. Und zwar sei

$$v_1 = \frac{u_1}{|u_1|}.$$

Das geht denn u_1 ist nicht der Nullvektor. Dann bestimmen wir $\lambda \in \mathbb{R}$ so dass $u_2 + \lambda v_1$ senkrecht zu v_1 ist:

$$0 \stackrel{!}{=} \langle u_2 + \lambda v_1, v_1 \rangle = \langle u_2, v_1 \rangle + \lambda, \quad \text{also } \lambda = -\langle u_2, v_1 \rangle.$$

Durch Normierung ergibt sich schließlich

$$v_2 = \frac{u_2 - \langle u_2, v_1 \rangle v_1}{|u_2 - \langle u_2, v_1 \rangle v_1|}.$$

Hier brauchen wir dass $u_2 - \langle u_2, v_1 \rangle v_1$ nicht der Nullvektor ist. Aber wäre das der Fall, so liegt u_2 auf der durch v_1 bzw. u_1 aufgespannten Geraden, und u_1, u_2 wäre keine Basis von E . Bezüglich einer Orthonormalbasis v_1, v_2 von E können wir die Koordinaten eines Punkts $x \in E$ direkt angeben. Denn ist $x = \lambda_1 v_1 + \lambda_2 v_2$ so folgt

$$\begin{aligned} \langle x, v_1 \rangle &= \langle \lambda_1 v_1 + \lambda_2 v_2, v_1 \rangle = \lambda_1, \\ \langle x, v_2 \rangle &= \langle \lambda_1 v_1 + \lambda_2 v_2, v_2 \rangle = \lambda_2. \end{aligned}$$

Das ist also einfacher als bei einer beliebigen Basis. Es folgt die Darstellung

$$x = \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 \quad \text{für alle } x \in E.$$

Für eine Gerade G ist eine Orthonormalbasis einfach ein Vektor auf G mit Länge $|v| = 1$, und jedes $x \in G$ hat die Darstellung $x = \langle x, v \rangle v$. Eine Orthonormalbasis eines drei-dimensionalen Unterraums V sind drei Vektoren v_1, v_2, v_3 in V mit durch drei Vektoren v_1, v_2, v_3 in V gegeben mit

$$\langle v_i, v_j \rangle = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Die v_i sind also normiert und stehen paarweise aufeinander senkrecht. Für $x \in V$ ergibt sich wie im Fall der Ebene die Darstellung

$$x = \langle x, v_1 \rangle v_1 + \langle x, v_2 \rangle v_2 + \langle x, v_3 \rangle v_3.$$

Analog zur Ebene kann man aus einer beliebigen Basis eine Orthonormalbasis herstellen.

Wir wollen nun den Winkel $\sphericalangle(v, w) \in [0, \pi]$ zwischen zwei Vektoren $v, w \in \mathbb{R}^n$ definieren. Zur Motivation betrachten wir im $\mathbb{R}^2$ die Kreisbewegung

$$c : \mathbb{R} \rightarrow \mathbb{R}^2, c(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Es gilt $|c(t)| = \sqrt{(\cos t)^2 + (\sin t)^2} = 1$ und $c(0) = (1, 0)$. Der Punkt $c(t)$ bewegt sich auf dem Einheitskreis entgegen dem Uhrzeigersinn mit Absolutgeschwindigkeit Eins:

$$c'(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} \quad \text{also} \quad |c'(t)| = \sqrt{(-\sin t)^2 + (\cos t)^2} = 1.$$

Ein Kreisbogen $c : [\alpha, \beta] \rightarrow \mathbb{R}^2$ hat also genau die Länge $\beta - \alpha$, und das ist der Winkel zwischen $c(\alpha)$ und $c(\beta)$. Mit dem Additionstheorem des Kosinus (Schulwissen) folgt

$$\cos \sphericalangle(c(\alpha), c(\beta)) = \cos(\beta - \alpha) = \cos \alpha \cos \beta + \sin \alpha \sin \beta = \langle c(\alpha), c(\beta) \rangle.$$

Wir nehmen hier $\beta - \alpha \leq \pi$ an, damit c höchstens einen Halbkreis durchläuft.

Definition 4.1 (Winkel zwischen Vektoren) *Seien $x, y \in \mathbb{R}^n$ beide nicht null. Dann definieren wir den Winkel $\phi = \sphericalangle(x, y) \in [0, \pi]$ durch die Gleichung*

$$\cos \phi = \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle.$$

Es ist zu begründen, dass der Winkel ϕ durch die Gleichung eindeutig definiert ist. Die Funktion Kosinus ist auf dem Intervall $[0, \pi]$ streng monoton fallend mit $\cos 0 = 1$ und $\cos \pi = -1$ (Bild). Sie nimmt daher jeden Wert in $[-1, 1]$ genau einmal an. Es bleibt zu zeigen, dass die rechte Seite in der Definition des Winkels im Intervall $[-1, 1]$ liegt. Dies ergibt sich aus folgender Ungleichung.

Satz 4.1 (Ungleichung von Cauchy-Schwarz) *Für $x, y \in \mathbb{R}^n$ gilt*

$$|\langle x, y \rangle| \leq |x| |y|.$$

Gleichheit genau dann, wenn x, y (gleichsinnig oder entgegengesetzt) parallel sind.

BEWEIS: Wir betrachten erst Vektoren a, b mit $|a| = |b| = 1$. Es gilt

$$1 \pm \langle a, b \rangle = \frac{1}{2} (|a|^2 + |b|^2 \pm 2\langle a, b \rangle) = \frac{1}{2} |a \pm b|^2 \geq 0.$$

Also ist $|\langle a, b \rangle| \leq 1$. Für $x, y \neq 0$ beliebig wenden wir das an mit $a = \frac{x}{|x|}$, $b = \frac{y}{|y|}$:

$$|\langle x, y \rangle| = |x| |y| |\langle a, b \rangle| \leq |x| |y|.$$

Bei Gleichheit gilt für die Einheitsvektoren

$$0 = a \pm b = \frac{x}{|x|} \pm \frac{y}{|y|},$$

das heißt x und y sind parallel. □

Definition 4.1 ergibt sofort den Kosinussatz, im Dreieck mit Ecken $0, x, y$ gilt nämlich

$$|x - y|^2 = |x|^2 + |y|^2 - 2\langle x, y \rangle = |x|^2 + |y|^2 - 2|x||y| \cos \sphericalangle(x, y).$$

Wir zeigen nun eine fundamentale Eigenschaft der Euklidischen Länge.

Satz 4.2 (Dreiecksungleichung) Für $x, y \in \mathbb{R}^n$ gilt die Ungleichung

$$|x + y| \leq |x| + |y|,$$

Gleichheit genau wenn x, y gleichsinnig parallel.

BEWEIS: Aus der Cauchy-Schwarz Ungleichung folgt

$$|x + y|^2 = |x|^2 + 2\langle x, y \rangle + |y|^2 \leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2.$$

Durch Wurzelziehen folgt $|x + y| \leq |x| + |y|$. Bei Gleichheit müssen x und y parallel sein nach Satz 4.1. Es muss aber auch $\langle x, y \rangle \geq 0$ sein, also sind x, y gleichsinnig parallel. $\square$

In einem Dreieck ist eine Seite immer kürzer als die Summe der anderen beiden. Hier bilden die Punkte $0, x$ und $x + y$ ein Dreieck mit Seiten $|x|, |y|$ und $|x + y|$, daher die Bezeichnung.

Für eine beliebige Basis v_1, v_2 einer Ebene E haben wir die Abbildung

$$\phi: \mathbb{R}^2 \rightarrow E, \phi(x) = x_1 v_1 + x_2 v_2 \quad \text{wobei } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Diese Abbildung ist linear, das heißt es gilt

$$\phi(x + y) = \phi(x) + \phi(y) \quad \text{und} \quad \phi(\lambda x) = \lambda \phi(x).$$

Da v_1, v_2 eine Basis ist, hat jedes $v \in E$ eine eindeutige Koordinatendarstellung $v = \phi(x)$, also ist ϕ bijektiv. Sei nun v_1, v_2 eine Orthonormalbasis. Dann ist ϕ eine Isometrie, das heißt für alle $x, y \in \mathbb{R}^2$ gilt

$$|\phi(x)| = |x| \quad \text{und} \quad \langle \phi(x), \phi(y) \rangle = \langle x, y \rangle.$$

Das lässt sich leicht nachrechnen, und zwar gilt

$$\begin{aligned} |\phi(x)|^2 &= |x_1 v_1 + x_2 v_2|^2 = x_1^2 + x_2^2 = |x|^2, \\ \langle \phi(x), \phi(y) \rangle &= \langle x_1 v_1 + x_2 v_2, y_1 v_1 + y_2 v_2 \rangle = x_1 y_1 + x_2 y_2 = \langle x, y \rangle. \end{aligned}$$

Da der Winkel durch das Skalarprodukt bestimmt ist, folgt auch $\sphericalangle(\phi(x), \phi(y)) = \sphericalangle(x, y)$.

Wir kommen schließlich zu einer speziellen Struktur im $\mathbb{R}^3$, dem Kreuzprodukt (oder Vektorprodukt). Auf der Standardbasis setzen wir

$$\begin{aligned} e_1 \times e_2 &= -e_2 \times e_1 = e_3 \\ e_2 \times e_3 &= -e_3 \times e_2 = e_1 \\ e_3 \times e_1 &= -e_1 \times e_3 = e_2. \end{aligned}$$

Man nennt die Permutationen 123, 231, 312 zyklisch, und 213, 321, 132 antizyklisch. Das Kreuzprodukt beliebiger Vektoren x, y definieren wir durch Ausmultiplizieren, also

$$\begin{aligned} x \times y &= (x_1 e_1 + x_2 e_2 + x_3 e_3) \times (y_1 e_1 + y_2 e_2 + y_3 e_3) \\ &= (x_2 y_3 - x_3 y_2) e_1 + (x_3 y_1 - x_1 y_3) e_2 + (x_1 y_2 - x_2 y_1) e_3 \\ &= \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}. \end{aligned}$$

Beachten Sie dass das Skalarprodukt von zwei Vektoren eine reelle Zahl ist, das Kreuzprodukt aber wieder ein Vektor. Folgende Regeln sind offensichtlich:

$$\begin{aligned}x \times y &= -y \times x && \text{(Schiefsymmetrie)} \\(\lambda x + \mu y) \times z &= \lambda x \times z + \mu y \times z && \text{(Bilinearität)}.\end{aligned}$$

Insbesondere gilt $x \times x = 0$. Wir wollen den Vektor $x \times y$ jetzt geometrisch charakterisieren. Als erstes berechnen wir dazu seine Länge (nachprüfen)

$$\begin{aligned}|x \times y|^2 &= (x_1y_2 - x_2y_1)^2 + (x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 \\&= (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2 \\&= |x|^2|y|^2 - \langle x, y \rangle^2.\end{aligned}$$

Mit $\langle x, y \rangle = |x||y| \cos \sphericalangle(x, y)$ (der Definition des Winkels) ergibt sich

$$|x \times y| = |x||y|\sqrt{1 - \cos^2 \sphericalangle(x, y)} = |x||y| \sin \sphericalangle(x, y).$$

Insbesondere ist $x \times y$ genau dann null wenn x, y parallel sind (oder x bzw. y ist null). Sind x, y orthogonal so folgt $|x \times y| = |x||y|$. Ab jetzt seien x, y nicht parallel und damit $x \times y \neq 0$. Wir berechnen

$$\langle x \times y, z \rangle = x_1y_2z_3 + x_2y_3z_1 + x_3y_1z_2 - x_2y_1z_3 - x_3y_2z_1 - x_1y_3z_2. \quad (4.1)$$

Ist $z = x$ oder $z = y$, so ist die rechte Seite Null. Also steht $x \times y$ senkrecht auf die von x, y aufgespannte Ebene. Da die Länge schon bekannt ist, bleiben nur zwei Möglichkeiten, die sich um den Faktor -1 unterscheiden.

Die rechte Seite der Gleichung (4.1) ist die Determinante $\det(x, y, z)$. Um sich die Formel zu merken, kann man die Vektoren in ein Schema schreiben und jeweils die Produkte über die Diagonalen bilden. Dabei werden Diagonalen nach Südost positiv, Diagonalen nach Südwest negativ gezählt (Regel von Sarrus).

$$\left(\begin{array}{ccc|cc} x_1 & x_2 & x_3 & x_1 & x_2 \\ y_1 & y_2 & y_3 & y_1 & y_2 \\ z_1 & z_2 & z_3 & z_1 & z_2 \end{array} \right).$$

Die Determinante $\det(x, y, z)$ wird auch als Spatprodukt der drei Vektoren bezeichnet. Sie ist gleich dem Volumen des von den Vektoren gebildeten Spats (Parallelepiped), bis auf Vorzeichen. Wir vereinbaren:

$$x, y, z \text{ sind positiv orientiert, genau wenn } \det(x, y, z) > 0.$$

Hier kommt es auf die Reihenfolge der Vektoren an: werden zwei Vektoren vertauscht, so ändert die Determinante ihr Vorzeichen. Die Richtung von $x \times y \neq 0$ ist nun dadurch bestimmt, dass $x, y, x \times y$ in dieser Reihenfolge positiv orientiert sind. Denn wählen wir in der Formel für die Determinante $z = x \times y$, so ergibt sich

$$\det(x, y, x \times y) = \langle x \times y, x \times y \rangle = |x \times y|^2 > 0.$$

Anschaulich kann die Orientierung mithilfe der Rechte-Hand-Regel geprüft werden, danach bilden die zueinander senkrechten Daumen, Zeigefinger und Mittelfinger der rechten Hand eine positiv orientierte Basis. Die Charakterisierung des Kreuzprodukts hat folgende Konsequenz: ist v_1, v_2, v_3 irgendeine positiv orientierte Orthonormalbasis von $\mathbb{R}^3$ so gilt

$$v_1 \times v_2 = v_3, v_2 \times v_3 = v_1, v_3 \times v_1 = v_2.$$

Das ist aus der Definition mittels Koordinaten nicht direkt ersichtlich.

5 Komplexe Zahlen

Wir beginnen mit einer anderen Schreibweise für Punkte des $\mathbb{R}^2$. Und zwar setzen wir $\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1$ und $\begin{pmatrix} 0 \\ 1 \end{pmatrix} = i$, und haben dann die allgemeine Darstellung

$$z = x + iy \quad \text{für alle } z = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

In diesem Zusammenhang heißen die Punkte des $\mathbb{R}^2$ komplexe Zahlen (Symbol $\mathbb{C}$). Realteil und Imaginärteil einer komplexen Zahl sind gegeben durch

$$\operatorname{Re} z = x \quad \operatorname{Im} z = y \quad \text{für } z = x + iy.$$

Die Addition in $\mathbb{C}$ ist die übliche komponentenweise Addition, also

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2).$$

Für die Multiplikation setzen wir $i^2 = -1$ und multiplizieren aus:

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 - y_1y_2 + i(x_1y_2 + x_2y_1).$$

Um dies zu veranschaulichen, schreiben wir die komplexen Zahlen in Polarkoordinaten: zu jedem $z \in \mathbb{C}$ mit $z \neq 0$ gibt es ein $r > 0$ und ein $\varphi \in \mathbb{R}$ mit

$$z = r(\cos \varphi + i \sin \varphi).$$

Dabei ist $r = |z| = \sqrt{x^2 + y^2}$, und $\varphi \in \mathbb{R}$ ist der Winkel, den z mit der positiven x -Achse einschließt (Bild). Da die Funktionen $\cos, \sin$ die Periode 2π haben, ist der Winkel nur bis auf ganzzahlige Vielfache von 2π eindeutig bestimmt. Er ist eindeutig, wenn wir $\varphi \in [0, 2\pi)$ verlangen.

Für die komplexen Zahlen $\mathbb{C}$ gelten dieselben Rechenregeln wie für die reellen Zahlen. Bezüglich der Addition ist $0 = 0 + i0$ das neutrale Element, und $-z = -x - iy$ das zu $z = x + iy$ inverse Element. Kommutativgesetz und Assoziativgesetz der Addition sind klar, denn diese ist einfach die komponentenweise Addition.

Das neutrale Element der Multiplikation ist $1 = 1 + i0$, denn

$$z \cdot 1 = (x + iy)(1 + i0) = x + iy.$$

Wir nennen $\bar{z} = x - iy$ die zu $z = x + iy$ konjugiert komplexe Zahl. Wegen $i^2 = -1$ gilt

$$z\bar{z} = (x + iy)(x - iy) = x^2 - i^2y^2 + i(yx - xy) = x^2 + y^2 = |z|^2.$$

Damit können wir das zu $z \neq 0$ inverse Element angeben, und zwar ist

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}, \quad \text{denn } z \frac{\bar{z}}{|z|^2} = \frac{z\bar{z}}{|z|^2} = 1.$$

In Polarkoordinaten lautet nun die Multiplikation (Additionstheoreme für $\cos, \sin$)

$$\begin{aligned} z_1 z_2 &= r_1 r_2 (\cos \varphi_1 + i \sin \varphi_1)(\cos \varphi_2 + i \sin \varphi_2) \\ &= r_1 r_2 (\cos \varphi_1 \cos \varphi_2 - \sin \varphi_1 \sin \varphi_2 + i(\cos \varphi_1 \sin \varphi_2 + \sin \varphi_1 \cos \varphi_2)) \\ &= r_1 r_2 (\cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2)). \end{aligned}$$

Das Kommutativgesetz ist damit klar, und Multiplikation mit $z_3 = r_3(\cos \varphi_3 + i \sin \varphi_3)$ liefert

$$(z_1 z_2) z_3 = r_1 r_2 r_3 (\cos(\varphi_1 + \varphi_2 + \varphi_3) + i \sin(\varphi_1 + \varphi_2 + \varphi_3)).$$

Für $z_1(z_2 z_3)$ kommt dasselbe raus, also gilt das Assoziativgesetz. Schließlich berechnen wir

$$\begin{aligned} (a + ib)((x_1 + iy_1) + (x_2 + iy_2)) &= (a + ib)(x_1 + x_2 + i(y_1 + y_2)) \\ &= a(x_1 + x_2) - b(y_1 + y_2) + i(b(x_1 + x_2) + a(y_1 + y_2)) \\ &= ax_1 - by_1 + i(bx_1 + ay_1) + ax_2 - by_2 + i(bx_2 + ay_2) \\ &= (a + ib)(x_1 + iy_1) + (a + ib)(x_2 + iy_2). \end{aligned}$$

Wir notieren im Vorbeigehen folgende nützliche Formeln, die sich leicht verifizieren lassen. Für die zweite Gleichung in (1) und für (3) kann man die Polardarstellung von $z_{1,2}$ verwenden.

$$(1) \overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \quad \overline{z_1 z_2} = \overline{z_1} \overline{z_2},$$

$$(2) \text{ Für } z = x + iy \text{ ist } \operatorname{Re} z = \frac{1}{2}(z + \bar{z}) \text{ und } \operatorname{Im} z = \frac{1}{2i}(z - \bar{z})$$

$$(3) |z_1 z_2| = |z_1| |z_2|.$$

Es gibt Gleichungen die in $\mathbb{R}$ keine Lösung haben, aber in $\mathbb{C}$ lösbar sind. Zum Beispiel hat die Gleichung $x^2 = -q$ für $q > 0$ keine Lösung in $\mathbb{R}$. Aber in $\mathbb{C}$ gibt es die beiden Lösungen

$$z_{\pm} = \pm i\sqrt{q} \quad (\sqrt{q} \text{ bezeichnet die positive Wurzel in } \mathbb{R}).$$

Tatsächlich ist jede Gleichung $x^2 + px + q = 0$ mit $p, q \in \mathbb{R}$ lösbar in $\mathbb{C}$. Substituiere dazu $x = y + a$ und berechne $0 \stackrel{!}{=} (y + a)^2 + p(y + a) + q = y^2 + (2a + p)y + a^2 + pa + q$. Durch Wahl von $a = -p/2$ ergibt sich $y^2 = p^2/4 - q$, das hat die Lösungen

$$y_{\pm} = \begin{cases} \pm\sqrt{p^2/4 - q} & \text{falls } p^2/4 - q > 0 \\ \pm i\sqrt{-(p^2/4 - q)} & \text{falls } p^2/4 - q < 0 \\ 0 & \text{falls } p^2/4 - q = 0. \end{cases}$$

Durch Rückeinsetzen ergeben sich die beiden Lösungen $x_{\pm} = -p/2 + y_{\pm}$. Das ist nicht der einzige Vorteil der komplexen Zahlen, zum Beispiel werden wir sehen dass sie bei Schwingungen eine Rolle spielen.

Kapitel 2

Funktionen, Grenzwerte, Stetigkeit

1 Rationale Funktionen

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt Polynom vom Grad n , wenn es reelle Zahlen $a_0, a_1, \dots, a_n$ gibt mit $a_n \neq 0$, so dass

$$f(x) = a_0 + a_1x + \dots + a_nx^n = \sum_{k=0}^n a_kx^k \quad \text{für alle } x \in \mathbb{R}. \quad (1.1)$$

Die $a_i \in \mathbb{R}$ heißen Koeffizienten des Polynoms, a_n heißt Leitkoeffizient. Als erstes diskutieren wir das Abspalten von Linearfaktoren in einer Nullstelle, genauer zeigen wir: ist $f(\lambda) = 0$ so gibt es ein Polynom $g(x)$ vom Grad $n - 1$ mit Leitkoeffizient a_n so dass

$$f(x) = (x - \lambda)g(x) \quad \text{für alle } x \in \mathbb{R}.$$

Im Fall $\lambda = 0$ ist $a_0 = f(0) = 0$ und es folgt wie behauptet

$$f(x) = x(a_1 + \dots + a_nx^{n-1}) = xg(x) \quad \text{für alle } x \in \mathbb{R}.$$

Für $\lambda \in \mathbb{R}$ beliebig verwenden wir folgende Tatsache: $f(x + \lambda)$ ist auch ein Polynom vom Grad n mit Leitkoeffizient a_n . Und zwar gilt

$$f(x + \lambda) = a_0 + a_1(x + \lambda) + \dots + a_n(x + \lambda)^n.$$

Wir lösen die Klammern auf und ordnen nach Potenzen von x . Die höchste Potenz ist x^n , mit Koeffizient $a_n \neq 0$. An der Stelle $x = 0$ ist $f(x + \lambda) = f(\lambda) = 0$. Wie gezeigt gibt es ein Polynom $g(x)$ vom Grad $n - 1$ mit Leitkoeffizient a_n so dass

$$f(x + \lambda) = xg(x) \quad \text{für alle } x \in \mathbb{R}.$$

Es folgt $f(x) = f(x - \lambda + \lambda) = (x - \lambda)g(x - \lambda)$, die Behauptung ist bewiesen.

Lemma 1.1 (Zahl der Nullstellen) Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ ein Polynom vom Grad $n \in \mathbb{N}_0$. Dann hat f höchstens n verschiedene Nullstellen.

BEWEIS: Per Induktion über $n \in \mathbb{N}_0$. Im Fall $n = 0$ ist $f(x) = a_0$ für alle $x \in \mathbb{R}$, wobei $a_0 \neq 0$, also hat f keine Nullstelle. Ist f Polynom vom Grad $n \in \mathbb{N}$, so hat entweder f keine Nullstelle oder es gilt $f(x) = (x - \lambda)g(x)$ für alle $x \in \mathbb{R}$, mit einem Polynom g vom Grad $n - 1$. Nach Induktion hat g höchstens $n - 1$ Nullstellen, also f höchstens n Nullstellen. $\square$

Lemma 1.2 (Koeffizientenvergleich) Seien $f, g : \mathbb{R} \rightarrow \mathbb{R}$ Polynome vom Grad m bzw. n , das heißt es gilt mit $a_m, b_n \neq 0$

$$f(x) = \sum_{i=0}^m a_i x^i \quad \text{und} \quad g(x) = \sum_{i=0}^n b_i x^i \quad \text{für alle } x \in \mathbb{R}.$$

Ist $f(x) = g(x)$ an mehr als $\max(m, n)$ Stellen, so ist $m = n$ und $a_i = b_i$ für $i = 0, \dots, n$.

BEWEIS: Wäre $m \neq n$ oder $a_i \neq b_i$ für ein i , so ist $f - g$ Polynom vom Grad höchstens $\max(m, n)$, und hat nach Lemma 1.1 höchstens $\max(m, n)$ Nullstellen, Widerspruch. $\square$

All das hilft nicht weiter, wenn ein Polynom einfach keine Nullstellen hat, zum Beispiel $f(x) = x^2 + 1$. Um die Sache wirklich zu verstehen, müssen wir ins Komplexe gehen. Ein komplexes Polynom vom Grad n hat die Form

$$f : \mathbb{C} \rightarrow \mathbb{C}, f(z) = a_0 + a_1 z + \dots + a_n z^n \quad \text{wobei } a_i \in \mathbb{C}, a_n \neq 0.$$

Das Abspalten von Linearfaktoren und der Koeffizientenvergleich gelten in $\mathbb{C}$ ganz analog, weil sie nur die Algebra verwenden. Insbesondere hat auch ein komplexes Polynom höchstens n Nullstellen. Im Unterschied zum Reellen gilt aber der

Satz 1.1 (Fundamentalsatz der Algebra) ¹ Jedes komplexe Polynom vom Grad $n \geq 1$ hat mindestens eine Nullstelle $\lambda \in \mathbb{C}$.

Folgerung 1.1 Ein komplexes Polynom $f(z)$ vom Grad $n \geq 1$ hat eine Zerlegung

$$f(z) = a_n (z - \lambda_1) \cdot \dots \cdot (z - \lambda_n) \quad \text{für alle } z \in \mathbb{C}.$$

Die $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ sind die Nullstellen. Eine Nullstelle λ kann hier mehrfach auftreten, die Anzahl nennen wir die Vielfachheit der Nullstelle. Die Zahl a_n ist der Leitkoeffizient des Polynoms $f(z)$.

BEWEIS: Nach dem Fundamentalsatz der Algebra hat $f(z)$ eine Nullstelle $\lambda_1 \in \mathbb{C}$. Durch Abspalten folgt $f(z) = (z - \lambda_1) f_1(z)$, wobei $f_1(z)$ komplexes Polynom vom Grad $n - 1$ ist. Nun hat $f_1(z)$ wieder eine Nullstelle $\lambda_2 \in \mathbb{C}$, und so weiter. Der Prozess stoppt genau nach n Schritten, denn dann hat das Restpolynom den Grad Null, ist also konstant. $\square$

Wir kommen jetzt auf $\mathbb{R}$ zurück. Jedes reelle Polynom $p(x) = a_0 + a_1 x + \dots + a_n x^n$, also $a_i \in \mathbb{R}$, kann als komplexes Polynom aufgefasst werden, indem wir für x auch komplexe Zahlen z einsetzen. Damit wird die reelle Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ zu einer komplexen Funktion $f : \mathbb{C} \rightarrow \mathbb{C}$ fortgesetzt. Das Endergebnis lautet so.

Folgerung 1.2 Ein reelles Polynom $f(x)$ vom Grad $n \geq 1$ ist ein Produkt von Linearfaktoren $x - \lambda$ und quadratischen Faktoren $x^2 - 2\alpha x + \alpha^2 + \beta^2$. Die Zahlen λ sind die Nullstellen in $\mathbb{R}$, die Zahlen $\alpha \pm i\beta$ sind die Nullstellen in $\mathbb{C} \setminus \mathbb{R}$. Es gilt $n = k + 2m$, wenn k die Zahl der Linearfaktoren und m die Zahl der quadratischen Faktoren ist.

BEWEIS: Wir können annehmen, dass $f(x)$ keine reellen Nullstellen hat, sonst spalten wir die Linearfaktoren ab. Ist $\lambda = \alpha + i\beta$ eine Nullstelle in $\mathbb{C} \setminus \mathbb{R}$, so ist auch $\bar{\lambda} = \alpha - i\beta$ eine Nullstelle. Wegen $a_i \in \mathbb{R}$ gilt nämlich

$$0 = \overline{f(\lambda)} = \overline{a_0 + a_1 \lambda + \dots + a_n \lambda^n} = a_0 + a_1 \bar{\lambda} + \dots + a_n \bar{\lambda}^n = f(\bar{\lambda}).$$

¹C. F. Gauß 1799

Wir können also beide Faktoren $z - \lambda$ und $z - \bar{\lambda}$ abspalten. Es gilt

$$(z - \lambda)(z - \bar{\lambda}) = z^2 - (\lambda + \bar{\lambda})z + |\lambda|^2 = z^2 - 2\alpha z + \alpha^2 + \beta^2.$$

Es gibt also ein Polynom $g(z)$ vom Grad $n - 2$ mit

$$f(z) = (z - \lambda)(z - \bar{\lambda})g(z) = (z^2 - 2\alpha z + \alpha^2 + \beta^2)g(z).$$

Jetzt machen wir mit dem Restpolynom $g(z)$ weiter und erhalten schließlich die gewünschte Zerlegung in quadratische Faktoren. Ein Detail fehlt: es ist zu begründen, dass $g(z)$ wieder reelle Koeffizienten hat! Sei

$$g(z) = b_0 + b_1 z + \dots + b_{n-2} z^{n-2}, \quad \text{mit zunächst } b_i \in \mathbb{C}.$$

Wir setzen in $f(z)$ nun reelle Zahlen x ein und erhalten

$$\begin{aligned} 0 &= \operatorname{Im} f(x) \quad (\text{da } f \text{ reelle Koeffizienten hat}) \\ &= \operatorname{Im} \left((x^2 - 2\alpha x + \alpha^2 + \beta^2) g(x) \right) \\ &= (x^2 - 2\alpha x + \alpha^2 + \beta^2) \operatorname{Im} g(x) \\ &= (x^2 - 2\alpha x + \alpha^2 + \beta^2) \sum_{i=0}^{n-2} (\operatorname{Im} b_i) x^i. \end{aligned}$$

Die linke Klammer ist für $x \in \mathbb{R}$ niemals Null. Also ist die Summe null für alle $x \in \mathbb{R}$, und es folgt $\operatorname{Im} b_i = 0$ für alle $i = 0, \dots, n - 2$ durch Koeffizientenvergleich, siehe Lemma 1.1. $\square$

Die Existenz einer Nullstelle ist in $\mathbb{C}$ durch Satz 1.1 gesichert. Dieser Satz liefert aber kein Verfahren, wie die Nullstellen berechnet werden sollen. Mit Substitutionen und Ziehen von k -ten Wurzeln kommt man ab Grad $n \geq 5$ nicht zum Ziel (Abel 1825). Deshalb spielen numerische Lösungsverfahren eine Rolle, durch die die Nullstelle approximativ bestimmt wird.

Eine rationale Funktion f ist ein Quotient zweier Polynome. Seien genauer $p(x), q(x)$ reelle Polynome vom Grad m, n , also beide nicht null. Dann ist die Menge der Nullstellen N_q endlich, und wir bilden die Funktion

$$f : \mathbb{R} \setminus N_q \rightarrow \mathbb{R}, \quad f(x) = \frac{p(x)}{q(x)}.$$

Im Fall $m \geq n$ gibt es eindeutig bestimmte Polynome $g(x)$ vom Grad $m - n$ und $r(x)$ vom Grad kleiner n , eventuell die Nullfunktion, so dass

$$f(x) = g(x) + \frac{r(x)}{q(x)} \quad \text{für alle } x \in \mathbb{R} \setminus N_q.$$

Dies folgt durch Division mit Rest. Und zwar schreiben wir

$$\frac{p(x)}{q(x)} = \frac{a_m}{b_n} x^{m-n} + \frac{p_1(x)}{q(x)} \quad \text{mit } p_1(x) = p(x) - \frac{a_m}{b_n} x^{m-n} q(x).$$

Das Polynom $p_1(x)$ hat Grad $m_1 < m$. Ist $m_1 < n$ (oder ist $p_1(x)$ null) so gilt die Darstellung mit $g(x) = \frac{a_m}{b_n} x^{m-n}$ und $r(x) = p_1(x)$. Ist $m_1 \geq n$ so haben wir per Induktion über m

$$\frac{p_1(x)}{q(x)} = g_1(x) + \frac{r_1(x)}{q(x)}.$$

Dabei ist $g_1(x)$ ein Polynom vom Grad $m_1 - n$, und $r_1(x)$ ist ein Polynom vom Grad kleiner n oder null. Es folgt

$$\frac{p(x)}{q(x)} = \frac{a_m}{b_n} x^{m-n} + g_1(x) + \frac{r_1(x)}{q(x)}.$$

Das ist dann die gewünschte Darstellung. Zur Eindeutigkeit: angenommen die Zerlegung gilt mit g_1, r_1 und g_2, r_2 . Dann folgt durch Subtraktion und Multiplikation mit $q(x)$

$$(g_1(x) - g_2(x))q(x) + r_1(x) - r_2(x) = 0 \quad \text{für alle } x \in \mathbb{R} \setminus N_q.$$

Wäre $g_1 - g_2$ nicht Null, so ist die linke Seite ein Polynom vom Grad mindestens n , aber mit unendlich vielen Nullstellen. Das kann nicht sein wegen Lemma 1.1. Analog sehen wir, dass die Polynome $r_{1,2}$ gleich sind. $\square$

Was ist mit den Ausnahmestellen $\lambda \in N_q$? Sei λ eine k -fache Nullstelle von $p(x)$ und eine ℓ -fache Nullstelle von $q(x)$. Im Fall $p(\lambda) \neq 0$ setzen wir $k = 0$. Es gibt nun Polynome $p_1(x), q_1(x)$ mit $p_1(\lambda), q_1(\lambda) \neq 0$ und

$$f(x) = \frac{p(x)}{q(x)} = \frac{(x - \lambda)^k p_1(x)}{(x - \lambda)^\ell q_1(x)} = (x - \lambda)^{k-\ell} \frac{p_1(x)}{q_1(x)} \quad \text{für } x \in \mathbb{R} \setminus N_q.$$

Abhängig von $k - \ell$ ergeben sich zwei Fälle:

- (1) Ist $k \geq \ell$, so kann $f(x)$ im Punkt λ sinnvoll definiert werden, und zwar

$$f(x) = \begin{cases} \frac{p_1(x)}{q_1(x)} & \text{falls } k = \ell, \\ 0 & \text{falls } k > \ell. \end{cases}$$

λ ist dann eine *hebbare Singularität* von $f(x)$.

- (2) Ist $k < \ell$, so nennen wir λ eine *Polstelle* von $f(x)$. Das Verhalten von $|f(x)|$ für $x \rightarrow \lambda$ wird durch den Faktor $|x - \lambda|^{k-\ell}$ bestimmt, der gegen Unendlich geht.

2 Grenzwerte von Folgen und Funktionen

In diesem Abschnitt beschäftigen wir uns genauer mit dem Konzept des Grenzwerts. Wir erinnern an die Definition: a_n konvergiert gegen a wenn es zu jedem $\varepsilon > 0$ ein $K \geq 0$ gibt mit

$$|a_n - a| < \varepsilon \quad \text{für alle } n > K.$$

Zum Beispiel hat die konstante Folge $a_n = a$ den Grenzwert a , denn es ist $|a_n - a| = 0 < \varepsilon$ sogar für alle $n \in \mathbb{N}$. Hier ein weiteres instruktives Beispiel.

Beispiel 2.1 (Plusminusfolge) $a_n = (-1)^n$ ist nicht konvergent. Denn für jedes $a \in \mathbb{R}$ gilt

$$2 = |a_n - a_{n+1}| \leq |a_n - a| + |a - a_{n+1}| \quad \text{für alle } n \in \mathbb{N}.$$

Hätte a_n den Grenzwert a , so würde die rechte Seite gegen null gehen, Widerspruch.

Eine Umformulierung der Konvergenz verwendet den Begriff der ε -Umgebung von a ,

$$U_\varepsilon(a) = \{x \in \mathbb{R} : |x - a| < \varepsilon\} = \{x \in \mathbb{R} : a - \varepsilon < x < a + \varepsilon\}.$$

Eine Folge a_n konvergiert genau dann gegen a wenn die Folgenglieder ab einer gewissen Nummer in der ε -Umgebung von a liegen, egal wie klein $\varepsilon > 0$ gewählt ist. Diese äquivalente Formulierung ist manchmal praktisch.

Satz 2.1 (Eindeutigkeit) *Der Grenzwert einer Folge a_n ist eindeutig bestimmt.*

BEWEIS: Angenommen a_n hat zwei Grenzwerte $a \neq a'$. Setze $\varepsilon = |a - a'|/2 > 0$. Es gilt

$$a_n \in U_\varepsilon(a) \cap U_\varepsilon(a') = \emptyset \quad \text{für } n \text{ groß, Widerspruch.}$$

□

Beispiel 2.2 (Geometrische Folge) Für $q \in \mathbb{R}$ mit $|q| < 1$ gilt $\lim_{n \rightarrow \infty} q^n = 0$. Für $q = 0$ ist das klar. Sonst ist $1/|q| > 1$, wir setzen dann $1/|q| = 1 + x$ mit $x > 0$. Es gilt

$$(1 + x)^n = \sum_{k=0}^n \binom{n}{k} x^k \geq \binom{n}{1} x = nx, \quad \text{also } |q|^n = \frac{1}{(1+x)^n} \leq \frac{1}{nx} \rightarrow 0.$$

□

Definition 2.1 *Eine reelle Folge a_n heißt beschränkt, wenn es ein $K \geq 0$ gibt mit*

$$|a_n| \leq K \quad \text{für alle } n.$$

Genauer kann noch wie folgt differenziert werden:

$$\begin{aligned} a_n \text{ nach oben beschränkt} &\Leftrightarrow \exists K \in \mathbb{R} \text{ mit } a_n \leq K \text{ für alle } n \in \mathbb{N} \\ a_n \text{ nach unten beschränkt} &\Leftrightarrow \exists K \in \mathbb{R} \text{ mit } a_n \geq K \text{ für alle } n \in \mathbb{N}. \end{aligned}$$

a_n beschränkt bedeutet a_n ist nach oben *und* unten beschränkt.

Beispiel 2.3 Die Folge $a_n = n$ ist nach unten beschränkt, denn es ist zum Beispiel $a_n \geq 0$ für alle n . Sie ist aber nicht nach oben beschränkt.

Satz 2.2 (konvergent $\Rightarrow$ beschränkt) *Konvergente Folgen sind beschränkt.*

BEWEIS: Sei $a_n \rightarrow a$ mit $n \rightarrow \infty$, also gilt $|a_n - a| < 1$ für $n > N$. Es folgt

$$|a_n| = |a_n - a + a| \leq |a_n - a| + |a| \leq 1 + |a| \quad \text{für } n > N.$$

Also $|a_n| \leq \max(|a_1|, \dots, |a_N|, 1 + |a|)$ für alle $n \in \mathbb{N}$.

□

Satz 2.3 (Rechenregeln für Grenzwerte) *Es gelte $a_n \rightarrow a, b_n \rightarrow b$ mit $n \rightarrow \infty$.*

- $\lim_{n \rightarrow \infty} (\lambda a_n + \mu b_n) = \lambda a + \mu b$ für alle $\lambda, \mu \in \mathbb{R}$
- $\lim_{n \rightarrow \infty} (a_n \cdot b_n) = a \cdot b$.
- $\lim_{n \rightarrow \infty} a_n/b_n = a/b$, falls $b \neq 0$.

BEWEIS: Die Dreiecksungleichung liefert für a)

$$|(\lambda a_n + \mu b_n) - (\lambda a + \mu b)| \leq |\lambda| |a_n - a| + |\mu| |b_n - b| \rightarrow 0.$$

Für b) wählen wir $K \geq 0$ mit $|a_n| \leq K$ für alle n , siehe Satz 2.2, und mit $|b| \leq K$. Dann gilt

$$\begin{aligned} |a_n b_n - ab| &= |a_n b_n - a_n b + a_n b - ab| \\ &\leq |a_n| \cdot |b_n - b| + |a_n - a| \cdot |b| \\ &\leq K(|a_n - a| + |b_n - b|) \rightarrow 0 \quad \text{mit } n \rightarrow \infty. \end{aligned}$$

Für c) sei zuerst $a_n = b = 1$. Für n groß ist $|b_n| = |1 - (1 - b_n)| \geq 1 - |1 - b_n| \geq \frac{1}{2}$. Also folgt

$$\left| \frac{1}{b_n} - 1 \right| = \frac{|1 - b_n|}{|b_n|} \leq 2|1 - b_n| \rightarrow 0 \quad \text{mit } n \rightarrow \infty.$$

Schießlich folgt für a_n, b allgemein aus b) und dem Spezialfall von c)

$$\frac{a_n}{b_n} = \frac{a_n}{b} \frac{1}{b_n/b} \rightarrow \frac{a}{b}.$$

□

Hier zwei Anwendungen der Rechenregeln für Grenzwerte.

Beispiel 2.4 (geometrische Reihe) Für $-1 < q < 1$ betrachten wir die Folge

$$S_n = 1 + q + \dots + q^n = \sum_{k=0}^n q^k.$$

Aus der Formel für die geometrische Summe, Beispiel 2.2 und Satz 2.3 folgt

$$\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} \sum_{k=0}^n q^k = \lim_{n \rightarrow \infty} \frac{1 - q^{n+1}}{1 - q} = \frac{1}{1 - q}.$$

Wir schreiben kurz $\sum_{k=0}^{\infty} q^k = 1/(1 - q)$. Folgen, deren Folgenglieder Summen sind, heißen Reihen. Sie werden noch ausführlicher untersucht.

Satz 2.4 (Grenzwerte und Ungleichungen) Sei $\lim_{n \rightarrow \infty} a_n = a$ und $\lim_{n \rightarrow \infty} b_n = b$.

- a) Ist $a_n \leq b_n$ für alle n , so folgt $a \leq b$.
- b) Gilt $c \leq a_n \leq d$ für alle n mit $c, d \in \mathbb{R}$, so folgt $c \leq a \leq d$.
- c) Ist $a_n \leq c_n \leq b_n$ und $a = b$, so konvergiert auch die Folge c_n gegen $a = b$.

BEWEIS: Angenommen es ist $b > a$. Dann ist $b < (a + b)/2 < a$, also $b_n < (a + b)/2 < a_n$ für n groß, Widerspruch. Aussage b) folgt unmittelbar aus a), wenn wir c, d als konstante Folgen auffassen. Für c) sei $\varepsilon > 0$. Dann gilt für n groß

$$a - \varepsilon < a_n \leq c_n \leq b_n < b + \varepsilon = a + \varepsilon,$$

also $\lim_{n \rightarrow \infty} c_n = a$ nach Definition des Grenzwerts.

□

Achtung: aus $a_n < b_n$ folgt nicht $a < b$, sondern nur $a \leq b$. Die Striktheit von Ungleichungen geht beim Übergang zu Grenzwerten im allgemeinen verloren. Zum Beispiel gilt $1/n > 0$ für alle $n \in \mathbb{N}$, aber $\lim_{n \rightarrow \infty} 1/n = 0$.

Um zu zeigen dass eine Folge konvergiert haben wir bisher nur die Definition oder das Konzept der Cauchyfolge. Ein sehr praktisches Kriterium ist das folgende.

Satz 2.5 Ist a_n nach oben beschränkt und monoton wachsend, so hat a_n einen Grenzwert.

BEWEIS: Wir zeigen dass a_n eine Cauchyfolge ist. Sei b obere Schranke, also $a_1 \leq a_n \leq b$ für alle n . Sei $\varepsilon > 0$ gegeben. Wir setzen $\ell = b - a_1$ und zerlegen $[a_1, b]$ in K Teilintervalle

$$I_k = \left[a_1 + (k-1) \frac{\ell}{K}, a_1 + k \frac{\ell}{K} \right] \quad \text{mit } k = 1, \dots, K.$$

Dabei wählen wir K mit $\ell/K < \varepsilon$. Sei k maximal so dass I_k Folgenglieder enthält, und dann N minimal mit $a_N \in I_k$. Für $n \geq N$ gilt dann

$$a_n < a_1 + k \frac{\ell}{K} \quad \text{und} \quad a_n \geq a_N \geq a_1 + (k-1) \frac{\ell}{K}.$$

Es folgt $|a_m - a_n| \leq \ell/K < \varepsilon$ für $m, n \geq N$. □

Beispiel 2.5 (Dezimalbrüche) Jede Dezimalbruchfolge $a_n = k_0, k_1 k_2 \dots k_n$ mit $k_0 \in \mathbb{Z}$ und $k_j \in \{0, 1, \dots, 9\}$ konvergiert gegen eine gewisse reelle Zahl. Denn die Folge a_n ist monoton wachsend und nach oben beschränkt, es gilt (geometrische Reihe)

$$a_n \leq k_0 + \sum_{j=1}^n 9 \cdot 10^{-j} = k_0 + \frac{9}{10} \frac{1 - 10^{-n}}{1 - 10^{-1}} \leq k_0 + 1.$$

Satz 2.6 (Intervallschachtelung) Sei I_n Intervallschachtelung, das heißt $I_n = [a_n, b_n]$ mit $I_{n+1} \subset I_n$ für alle n und $|I_n| = b_n - a_n \rightarrow 0$. Dann gibt es genau ein $x \in \mathbb{R}$ mit $x \in I_n$ für alle n , und zwar $x = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$.

BEWEIS: Wir haben $a_1 \leq \dots \leq a_n \leq b_n \leq \dots \leq b_1$. Die Folge a_n ist monoton wachsend mit oberer Schranke b_1 , die Folge b_n ist monoton fallend mit unterer Schranke a_1 . Beide Folgen konvergieren, wegen $b_n - a_n \rightarrow 0$ mit gleichem Grenzwert $x \in \mathbb{R}$. Es folgt

$$a_n \leq \lim_{m \rightarrow \infty} a_m = x = \lim_{m \rightarrow \infty} b_m \leq b_n, \quad \text{also } x \in I_n.$$

Sei $x' \in I_n$ für alle n , also $a_n \leq x' \leq b_n$. Mit $n \rightarrow \infty$ folgt $x \leq x' \leq x$, also $x' = x$. □

Wir übertragen das Konzept des Grenzwerts nun auf Funktionen $f : D \rightarrow \mathbb{R}$, wobei $D \subset \mathbb{R}$.

Definition 2.2 (Grenzwert für Funktionen) Die Funktion $f : D \rightarrow \mathbb{R}$ konvergiert für $x \rightarrow x_0$ gegen $a \in \mathbb{R}$ (Notation: $\lim_{x \rightarrow x_0} f(x) = a$), falls gilt:

$$f(x_n) \rightarrow a \quad \text{für jede Folge } x_n \rightarrow x_0 \text{ mit } x_n \neq x_0.$$

Hier sind einige Bemerkungen angesagt:

- (1) Für Existenz und Wert des Grenzwerts ist es egal, ob $f(x)$ im Punkt x_0 definiert ist bzw. welchen Funktionswert die Funktion dort hat.
- (2) Beim rechtsseitigen (linksseitigen) Grenzwert betrachtet man nur Folgen $x_n \rightarrow x_0$ mit $x_n > 0$ (bzw. $x_n < 0$). *Notation:* $\lim_{x \searrow x_0} f(x)$ bzw. $\lim_{x \nearrow x_0} f(x)$.
- (3) Analog für Grenzwerte $\lim_{x \rightarrow \infty} f(x)$ bzw. $\lim_{x \rightarrow -\infty} f(x)$.

Beispiel 2.6 $f(x) = \sin \frac{1}{x}$ hat in $x_0 = 0$ keinen Grenzwert. Denn für $n \in \mathbb{N}$ gilt

$$f\left(\frac{1}{n\pi}\right) = \sin n\pi = 0 \quad \text{aber} \quad f\left(\frac{1}{2n\pi + \pi/2}\right) = \sin(2n\pi + \pi/2) = 1.$$

Für $g(x) = x \sin \frac{1}{x}$ gilt dagegen $\lim_{x \rightarrow 0} g(x) = 0$, denn

$$|g(x_n)| \leq |x_n| \rightarrow 0 \quad \text{für } x_n \rightarrow 0, x_n \neq 0.$$

Beispiel 2.7 Die Signumfunktion

$$\text{sign} : \mathbb{R} \rightarrow \mathbb{R}, \text{sign}(x) = \begin{cases} 1 & \text{für } x > 0 \\ -1 & \text{für } x < 0 \\ 0 & \text{für } x = 0 \end{cases}$$

hat die einseitigen Grenzwerte $\lim_{x \searrow 0} \text{sign}(x) = +1$ und $\lim_{x \nearrow 0} \text{sign}(x) = -1$, während $\lim_{x \rightarrow 0} \text{sign}(x)$ nicht existiert.

Folgende Regeln ergeben sich aus den Aussagen für Folgen, siehe Satz 2.3 und Satz 2.4.

Satz 2.7 (Rechenregeln für Grenzwerte)

- (1) Aus $f(x) \rightarrow a$, $g(x) \rightarrow b$ für $x \rightarrow x_0$, wobei $a, b \in \mathbb{R}$ und $x_0 \in \mathbb{R} \cup \{\pm\infty\}$, folgt

$$\begin{aligned} \lambda f(x) + \mu g(x) &\rightarrow \lambda a + \mu b \quad (\lambda, \mu \in \mathbb{R}), \\ f(x)g(x) &\rightarrow ab, \\ f(x)/g(x) &\rightarrow a/b, \quad \text{falls } b \neq 0. \end{aligned}$$

- (2) Sei $f(x) \leq g(x) \leq h(x)$. Aus $f(x), h(x) \rightarrow a$ mit $x \rightarrow x_0$ folgt $\lim_{x \rightarrow x_0} g(x) = a$.

Definition 2.3 (Stetigkeit) Die Funktion $f : D \rightarrow \mathbb{R}$ heißt stetig in $x_0 \in D$, falls gilt:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

Wir haben die Stetigkeit auf die Konvergenz von Folgen $f(x_n) \rightarrow f(x_0)$ zurückgeführt. Alternativ gibt es die sogenannte ε - δ -Definition der Stetigkeit:

$$\text{für alle } \varepsilon > 0 \text{ gibt es ein } \delta > 0 \text{ so das gilt: } |x - x_0| < \delta \quad \Rightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

Es ist nicht schwer zu zeigen dass das zu Definition 2.3 äquivalent ist, wir hatten die Konvergenz von Folgen ja auch mit einem ε -Kriterium definiert. Die Regeln für Grenzwerte implizieren direkt folgende Regeln zur Stetigkeit.

Satz 2.8 (Stetigkeitsregeln) Seien $f, g : D \rightarrow \mathbb{R}$ stetig in $x_0 \in D$. Dann gilt:

- (1) Für beliebige $\lambda, \mu \in \mathbb{R}$ ist die Funktion $\lambda f + \mu g$ stetig in x_0 .
- (2) Die Funktion fg ist stetig in x_0 .
- (3) Ist $g(x_0) \neq 0$, so ist $g(x) \neq 0$ nahe bei x_0 und f/g ist stetig in x_0 .

Angenommen in (3) wäre nicht $g(x) \neq 0$ nahe bei x_0 , dann gibt es eine Folge $x_n \rightarrow x_0$ mit $g(x_n) = 0$, und damit $g(x_0) = \lim_{n \rightarrow \infty} g(x_n) = 0$, Widerspruch.

Beispiel 2.8 Konstante Funktionen $f(x) = c$ sind stetig auf $\mathbb{R}$, denn für sie ist

$$x_n \rightarrow x_0 \quad \Rightarrow \quad f(x_n) = c = f(x_0).$$

Beispiel 2.9 Die Funktion $f(x) = x$ ist stetig auf $\mathbb{R}$, denn es gilt

$$x_n \rightarrow x_0 \quad \Rightarrow \quad f(x_n) = x_n \rightarrow x_0 = f(x_0).$$

Beispiel 2.10 Betrachte für Polynome $p(x)$ und $q(x)$ die rationale Funktion

$$f : \mathbb{R} \setminus N_q \rightarrow \mathbb{R}, \quad f(x) = \frac{p(x)}{q(x)} \quad \text{wobei } N_q = \{x \in \mathbb{R} \mid q(x) = 0\}.$$

Mit den Beispielen und Satz 2.8 folgt, dass $f(x)$ stetig ist auf $\mathbb{R} \setminus N_q$. Sei nun $\lambda \in N_q$ eine k -fache Nullstelle von $p(x)$ und eine ℓ -fache Nullstelle von $q(x)$ (mit $k = 0$ im Fall $p(\lambda) \neq 0$). Dann gibt es Polynome $p_1(x)$ und $q_1(x)$ mit $p_1(\lambda), q_1(\lambda) \neq 0$, so dass gilt:

$$f(x) = (x - \lambda)^{k-\ell} \frac{p_1(x)}{q_1(x)} \quad \text{für alle } x \in \mathbb{R} \setminus N_q.$$

Die Funktion $f_1(x) = p_1(x)/q_1(x)$ ist im Punkt λ definiert und stetig. Im Fall $k \geq \ell$ folgt

$$\lim_{x \rightarrow \lambda} f(x) = \lim_{x \rightarrow \lambda} (x - \lambda)^{k-\ell} f_1(x) = \begin{cases} f_1(\lambda) & \text{falls } k = \ell, \\ 0 & \text{falls } k > \ell. \end{cases}$$

Damit ist $f(x)$ in $x = \lambda$ stetig fortsetzbar. der Punkt ist eine *hebbare Singularität*, siehe Kapitel 2.2. Im Fall $k < \ell$ gilt $|f(x)| \rightarrow \infty$ für $x \rightarrow \lambda$, und $f(x)$ ist nicht stetig fortsetzbar.

Satz 2.9 (Stetigkeit der Verkettung) Seien $f : D \rightarrow \mathbb{R}$, $g : E \rightarrow \mathbb{R}$ mit $f(D) \subset E$. Ist f stetig in x_0 und g stetig in $y_0 = f(x_0)$, so ist $g \circ f : D \rightarrow \mathbb{R}$ stetig in x_0 .

BEWEIS: Sei $x_n \in D$ eine Folge mit $x_n \rightarrow x_0$. Dann folgt $f(x_n) \rightarrow f(x_0)$ wegen f stetig in x_0 , und weiter $g(f(x_n)) \rightarrow g(f(x_0))$ wegen g stetig in $y_0 = f(x_0)$. $\square$

Beispiel 2.11 Die Betragsfunktion $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(x) = |x|$, ist stetig auf $\mathbb{R}$, denn es gilt

$$x_n \rightarrow x_0 \quad \Rightarrow \quad |g(x_n) - g(x_0)| = ||x_n| - |x_0|| \leq |x_n - x_0| \rightarrow 0.$$

Ist weiter $f : D \rightarrow \mathbb{R}$ stetig, so auch $g \circ f = |f| : D \rightarrow \mathbb{R}$.

3 Anwendungen der Stetigkeit

Satz 3.1 (Zwischenwertsatz) Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig. Dann gibt es zu jedem y zwischen $f(a)$ und $f(b)$ ein $x \in [a, b]$ mit $f(x) = y$.

BEWEIS: Sei $f(a) \leq y \leq f(b)$. Wir konstruieren eine Intervallschachtelung $[a_n, b_n]$ so dass $f(a_n) \leq y \leq f(b_n)$. Für $n = 1$ setzen wir $[a_1, b_1] = [a, b]$. Ist I_n schon gefunden, so betrachten wir $m_n = (a_n + b_n)/2$ und definieren

$$I_{n+1} = \begin{cases} [a_n, m_n] & \text{falls } f(m_n) \geq y, \\ [m_n, b_n] & \text{falls } f(m_n) < y. \end{cases}$$

Es folgt $f(a_{n+1}) \leq y \leq f(b_{n+1})$ und $|I_n| = 2^{1-n}|I_1| \rightarrow 0$. Sei $x \in [a, b]$ der durch die Schachtelung definierte Punkt, dann gilt

$$y \leq \lim_{n \rightarrow \infty} f(b_n) = f(x) = \lim_{n \rightarrow \infty} f(a_n) \leq y.$$

□

Satz 3.2 (Monotonie und Umkehrfunktion) Sei $f : I = [a, b] \rightarrow \mathbb{R}$ streng monoton wachsend und stetig. Dann ist $f : [a, b] \rightarrow [f(a), f(b)]$ bijektiv, und die Umkehrfunktion $g : [f(a), f(b)] \rightarrow [a, b]$ ist auch streng monoton wachsend und stetig.

BEWEIS: Da f streng monoton, ist f injektiv und bildet in das Intervall $[f(a), f(b)]$ ab. f ist surjektiv nach dem Zwischenwertsatz. Sei $g : [f(a), f(b)] \rightarrow [a, b]$ die Umkehrfunktion. Wäre g nicht streng monoton wachsend, so gibt es y_1, y_2 mit $y_1 < y_2$ aber $g(y_2) \leq g(y_1)$. Es folgt

$$y_2 = f(g(y_2)) \leq f(g(y_1)) = y_1, \quad \text{Widerspruch.}$$

Wir zeigen jetzt dass g in einem Punkt y_0 linksseitig stetig ist. Sei $y_n \nearrow y_0$ eine gegebene Folge und $\varepsilon > 0$. Da $f(x)$ wachsend, ist $f(g(y_0) - \varepsilon) < f(g(y_0)) = y_0$. Es folgt $y_n > f(g(y_0) - \varepsilon)$ für n groß, und dann wegen $g(y)$ wachsend

$$g(y_0) \geq g(y_n) > g(f(g(y_0) - \varepsilon)) = g(y_0) - \varepsilon.$$

Die rechtsseitige Stetigkeit folgt analog. □

Wir kommen nun zu der wichtigsten Anwendung, nämlich der Existenz von Extremwerten. Im folgenden betrachten wir Minima, für Maxima gilt alles entsprechend. Eine nach unten beschränkte Menge $M \subset \mathbb{R}$ hat im allgemeinen kein kleinstes Element. Zum Beispiel ist das Intervall $(0, 1]$ nach unten beschränkt, aber für jedes $x \in (0, 1]$ ist $x/2$ ein Element von $(0, 1]$ das kleiner ist. Aber wenn eine Menge M eine untere Schranke hat, so hat M auch eine größte untere Schranke, das sogenannte Infimum von M (Notation: $\inf M$).² Dieses können wir durch eine Intervallschachtelung finden. Wähle $[a_1, b_1]$ so dass a_1 untere Schranke ist und $b_1 \in M$. Ist $[a_n, b_n]$ schon bestimmt, so betrachte $m_n = (a_n + b_n)/2$ und setze

$$I_{n+1} = [a_{n+1}, b_{n+1}] = \begin{cases} [m_n, b_n] & \text{falls } m_n \text{ untere Schranke,} \\ [a_n, m_n] & \text{falls } m_n \text{ keine untere Schranke.} \end{cases}$$

Per Induktion sehen wir dass für alle $n \in \mathbb{N}$ gilt:

²Analog existiert die kleinste obere Schranke, das Supremum von M , Notation: $\sup M$

- a_n ist untere Schranke
- Es gibt ein $x \in M$ mit $x \leq b_n$

Sei nun x_0 der durch die Intervallschachtelung bestimmte Punkt, also $a_n \nearrow x_0$ und $b_n \searrow x_0$. Dann ist x_0 eine untere Schranke, denn für alle $x \in M$ gilt $x \geq \lim_{n \rightarrow \infty} a_n = x_0$. Sei andererseits $b > x_0$. Wähle n groß so dass $b_n < b$, und wähle dann $x \in M$ mit $x \leq b_n$. Es folgt $x < b$, das heißt b ist keine untere Schranke. Damit ist x_0 die größte untere Schranke.

Sei nun $f : D \rightarrow \mathbb{R}$. Ein Punkt $x_0 \in D$ heißt Minimalstelle von $f(x)$ wenn

$$f(x_0) = \inf\{f(x) : x \in D\} \quad \Leftrightarrow \quad f(x_0) \leq f(x) \text{ für alle } x \in D.$$

Das Infimum wird nicht immer angenommen, das heißt es gibt nicht immer eine Minimalstelle. Auch könnte es sein, dass die Funktion gar nicht nach unten beschränkt ist.

$$f : (0, 1] \rightarrow \mathbb{R}, f(x) = x \quad \text{oder} \quad f(x) = -\frac{1}{x},$$

$$f : [0, 1] \rightarrow \mathbb{R}, f(x) = \begin{cases} -x & \text{für } 0 \leq x < \frac{1}{2} \\ 0 & \text{für } \frac{1}{2} \leq x \leq 1. \end{cases}$$

Wenn es aber eine Minimalstelle gibt, so sprechen wir auch vom Minimum statt Infimum. Der folgende Satz garantiert die Existenz einer Minimalstelle unter gewissen Voraussetzungen.

Satz 3.3 (Existenz von Extremstellen) *Sei $f : I = [a, b] \rightarrow \mathbb{R}$ stetig. Dann gibt es Punkte $x_0, x_1 \in [a, b]$ mit $f(x_0) \leq f(x) \leq f(x_1)$ für alle $x \in [a, b]$. Insbesondere ist f beschränkt.*

BEWEIS: (Skizze) Wir zeigen das für die Minimalstelle. Wähle $x_k \in I$ mit

$$\lim_{k \rightarrow \infty} f(x_k) = \inf\{f(x) : x \in I\}.$$

Eine solche Folge nennt man eine Minimalfolge. Angenommen die x_k konvergieren gegen ein $x_0 \in I$. Dann folgt wegen der Stetigkeit von $f(x)$

$$f(x_0) = \lim_{k \rightarrow \infty} f(x_k) = \inf\{f(x) : x \in I\}.$$

Somit ist die gesuchte Minimalstelle gefunden. Im allgemeinen muss die Folge x_k nicht konvergieren, zum Beispiel könnte es mehrere Minimalstellen geben. Aber es reicht, wenn wir eine konvergente Teilfolge $x_{k_1}, x_{k_2}, \dots$ mit $k_1 < k_2 < \dots$ finden, diese ist ja wieder eine Minimalfolge, das Argument ist dann anwendbar.

Eine Folge x_k in $[a, b]$ hat nun immer eine konvergente Teilfolge. Wähle als erstes eine Teilfolge so dass die erste Dezimalziffer konstant ist. Dann wähle daraus wieder eine Teilfolge so dass auch die zweite Ziffer konstant ist usw. Wir erzeugen so eine Dezimalzahl in $[a, b]$, die Grenzwert einer Teilfolge der x_k ist. $\square$

Kapitel 3

Differentialrechnung für Funktionen einer Variablen

1 Die Ableitung: Definition und Regeln

Im diesem Abschnitt betrachten wir reellwertige Funktionen einer Variablen, die auf einem offenen Intervall $I \subset \mathbb{R}$ definiert sind.

Definition 1.1 (Ableitung) Die Funktion $f : I \rightarrow \mathbb{R}$ hat im Punkt $x_0 \in I$ die Ableitung $a \in \mathbb{R}$ (Notation: $f'(x_0) = a$ oder $\frac{df}{dx}(x_0) = a$), falls gilt:

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = a. \quad (1.1)$$

Wir nennen f differenzierbar in x_0 , falls es ein $a \in \mathbb{R}^n$ mit (1.1) gibt, falls also der in (1.1) betrachtete Grenzwert existiert.

Eine alternative Formulierung ergibt sich durch die Substitution $x = x_0 + h$:

$$f'(x_0) = a \quad \Leftrightarrow \quad \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = a.$$

Leibniz betrachtete das Problem, die Tangente einer Kurve in einem Punkt zu definieren. Betrachte eine Kurve die als Graph einer Funktion $f : I \rightarrow \mathbb{R}$ gegeben ist, gesucht ist die Tangente im Punkt $(x_0, f(x_0))$. Der Differenzenquotient

$$\frac{f(x) - f(x_0)}{x - x_0} \quad (x_0, x \in I, x \neq x_0)$$

ist geometrisch die Steigung der Sekante durch die Punkte $(x_0, f(x_0))$ und $(x, f(x))$. Die Existenz der Ableitung bedeutet, dass die Sekantensteigungen für $x \rightarrow x_0$ gegen den Wert $f'(x_0)$ konvergieren. Die Tangente wird *definiert* als die Gerade, die durch den Punkt $(x_0, f(x_0))$ geht und dort die Steigung $f'(x_0)$ hat. Daraus ergibt sich ihre Gleichung

$$y = f(x_0) + f'(x_0)(x - x_0) \quad \text{für alle } x \in \mathbb{R}.$$

Newton entwickelte den Differentialkalkül (Englisch: Calculus) unter anderem um die Gesetze für die Planetenbewegung zu begründen, genauer konnte er diese Gesetze alle aus dem

Gravitationsgesetz ableiten. Dazu wird die Bewegung eines Planeten durch seinen Ortsvektor

$$f : I \rightarrow \mathbb{R}^3, \quad f(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$

beschrieben, also durch dessen Koordinaten zur Zeit $t \in I$. Ziel ist zum Beispiel die Definition der Geschwindigkeit als Vektor in $\mathbb{R}^3$. Die vektorielle Durchschnittsgeschwindigkeit auf dem Zeitintervall $[t_0, t]$ ist der Quotient von Weg und Zeit, also gleich

$$\frac{f(t) - f(t_0)}{t - t_0} \in \mathbb{R}^3.$$

Die Momentangeschwindigkeit $v(t_0)$ zum Zeitpunkt $t = t_0$ ist deshalb als vektorielle Ableitung zu definieren, wobei Newton einen Punkt statt eines Strichs benutzt hat:

$$v(t_0) = f'(t_0) = \begin{pmatrix} x'(t_0) \\ y'(t_0) \\ z'(t_0) \end{pmatrix} \in \mathbb{R}^3.$$

Definition 1.2 (Ableitungsfunktion) Die Funktion $f : I \rightarrow \mathbb{R}$ heißt differenzierbar, falls f in jedem $x_0 \in I$ differenzierbar ist. Die hierdurch gegebene Funktion

$$f' : I \rightarrow \mathbb{R}, \quad x_0 \mapsto f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0},$$

heißt Ableitungsfunktion oder schlicht Ableitung von f .

Beispiel 1.1 Für eine konstante Funktion $f(x) = c$ gilt

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{c - c}{x - x_0} = 0 \quad \Rightarrow \quad f'(x_0) = 0 \text{ bzw. } f' = 0.$$

Beispiel 1.2 Für $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x$, gilt

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{x - x_0}{x - x_0} = 1 \quad \text{für alle } x \neq x_0,$$

also folgt $f'(x_0) = 1$ bzw. $f' = 1$.

Beispiel 1.3 Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x|$, ist nicht differenzierbar in $x_0 = 0$, denn

$$\lim_{x \searrow 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \searrow 0} \frac{x}{x} = 1 \quad \text{und} \quad \lim_{x \nearrow 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \nearrow 0} \frac{-x}{x} = -1.$$

Die rechts- und linksseitige Ableitung existieren, sie sind aber nicht gleich.

Satz 1.1 Ist $f : I \rightarrow \mathbb{R}$ differenzierbar in $x_0 \in I$, so ist f auch stetig in x_0 .

BEWEIS: Es gilt mit $x \rightarrow x_0$

$$f(x) = f(x_0) + \frac{f(x) - f(x_0)}{x - x_0}(x - x_0) \rightarrow f(x_0) + f'(x_0) \cdot 0 = f(x_0).$$

□

Es gilt aber nicht die umgekehrte Implikation: die Funktion $f(x) = |x|$ ist stetig in $x_0 = 0$, aber nicht differenzierbar, siehe oben.

Satz 1.2 (Differentiationsregeln) Seien $f, g : I \rightarrow \mathbb{R}$ differenzierbar in $x_0 \in I$. Dann sind auch die Funktionen $\alpha f + \beta g$ ($\alpha, \beta \in \mathbb{R}$), fg und f/g (im Fall $g(x_0) \neq 0$) in x_0 differenzierbar mit folgenden Ableitungen:

(1) Linearität:

$$(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0)$$

(2) Produktregel:

$$(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$$

(3) Quotientenregel:

$$\left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{g(x_0)^2}$$

BEWEIS: Wir müssen jeweils für $x \neq x_0$ die Differenzenquotienten bilden und zeigen, dass diese mit $x \rightarrow x_0$ gegen das gewünschte konvergieren. Für (1) haben wir

$$\begin{aligned} \frac{(\alpha f(x) + \beta g(x)) - (\alpha f(x_0) + \beta g(x_0))}{x - x_0} &= \alpha \frac{f(x) - f(x_0)}{x - x_0} + \beta \frac{g(x) - g(x_0)}{x - x_0} \\ &\rightarrow \alpha f'(x_0) + \beta g'(x_0). \end{aligned}$$

Die Produktregel folgt durch „Mischen der Terme“:

$$\begin{aligned} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} &= \frac{f(x) - f(x_0)}{x - x_0} g(x) + f(x_0) \frac{g(x) - g(x_0)}{x - x_0} \\ &\rightarrow f'(x_0)g(x_0) + f(x_0)g'(x_0), \end{aligned}$$

wobei die Stetigkeit von g in x_0 benutzt wurde (Satz 1.1). Für die Quotientenregel reicht es, die Funktion $1/g$ zu betrachten, also $f \equiv 1$.

$$\frac{1}{x - x_0} \left(\frac{1}{g(x)} - \frac{1}{g(x_0)} \right) = -\frac{1}{g(x)g(x_0)} \frac{g(x) - g(x_0)}{x - x_0} \rightarrow -\frac{g'(x_0)}{g(x_0)^2}.$$

□

Beispiel 1.4 Sei $f_n(x) = x^n$. Dann ist $f'_1 = 1$ nach Beispiel 1.2, und die Produktregel ergibt

$$f'_n(x) = (f_1 f_{n-1})'(x) = f'_1(x) f_{n-1}(x) + f_1(x) f'_{n-1}(x) = x^{n-1} + x f'_{n-1}(x).$$

Per Induktion folgt $f'_n(x) = nx^{n-1}$. Das stimmt für $n = 1$, und für $n \geq 2$ ergibt sich

$$f'_n(x) = x^{n-1} + x(n-1)x^{(n-1)-1} = nx^{n-1}.$$

Weiter folgt mit Satz 1.2(1) für Polynome $p(x) = \sum_{k=0}^n a_k x^k$

$$p'(x) = \sum_{k=1}^n k a_k x^{k-1} = \sum_{j=0}^{n-1} (j+1) a_{j+1} x^j.$$

Beispiel 1.5 Für $f(x) = x^{-n}$, $n \in \mathbb{N}$, gilt nach Quotientenregel

$$f'(x) = \frac{d}{dx} \frac{1}{x^n} = -\frac{nx^{n-1}}{(x^n)^2} = -nx^{-n-1} \quad \text{für } x \neq 0.$$

Satz 1.3 (Kettenregel) Seien $f : I \rightarrow \mathbb{R}$, $g : J \rightarrow \mathbb{R}$ mit $f(I) \subset J$. Ist f in x_0 sowie g in $y_0 = f(x_0)$ differenzierbar, so ist auch $g \circ f : I \rightarrow \mathbb{R}$ in x_0 differenzierbar und

$$(g \circ f)'(x_0) = g'(f(x_0)) f'(x_0).$$

BEWEIS: Wir betrachten für $x \neq x_0$ den Differenzenquotienten:

$$\frac{g(f(x)) - g(f(x_0))}{x - x_0} = \frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \cdot \frac{f(x) - f(x_0)}{x - x_0} \rightarrow g'(f(x_0)) \cdot f'(x_0).$$

Hier haben wir benutzt, dass $f(x) \rightarrow f(x_0)$ mit $x \rightarrow x_0$ nach Satz 1.1. Ein technisches Problem gibt es, wenn $f(x) = f(x_0)$ für x nahe x_0 , aber das wollen wir hier nicht behandeln. $\square$

Satz 1.4 (Ableitung der Umkehrfunktion) Die Funktion $f : (a, b) \rightarrow \mathbb{R}$ sei streng monoton und stetig. Ist $f'(x_0) \neq 0$, so ist die Umkehrfunktion $g = f^{-1}$ differenzierbar in $y_0 = f(x_0)$ mit Ableitung

$$g'(y_0) = \frac{1}{f'(g(y_0))}.$$

BEWEIS: g existiert und ist stetig nach Satz 3.2. Wir berechnen für $y \rightarrow y_0$, also $g(y) \rightarrow g(y_0)$,

$$\frac{g(y) - g(y_0)}{y - y_0} = \frac{g(y) - g(y_0)}{f(g(y)) - f(g(y_0))} = \frac{1}{\frac{f(g(y)) - f(g(y_0))}{g(y) - g(y_0)}} \rightarrow \frac{1}{f'(x_0)}.$$

$\square$

Die Formel für die Ableitung folgt auch aus der Kettenregel:

$$f(g(y)) = y \quad \Rightarrow \quad f'(g(y_0))g'(y_0) = 1.$$

Wir sehen: damit die Umkehrfunktion im Punkt $y_0 = f(x_0)$ differenzierbar ist, muss $f'(x_0) \neq 0$ gelten. Zum Beispiel kann die Umkehrfunktion von $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^3$, im Nullpunkt nicht differenzierbar sein.

Die beiden vorangegangenen Regeln sind in der von Leibniz eingeführten Notation besonders suggestiv. Leibniz schreibt Funktionen in der Form $y = y(x)$ und bezeichnet die Ableitung mit dem Symbol $\frac{dy}{dx}$, das auch als *Differentialquotient* bezeichnet wird. Formal ergeben sich die Kettenregel und die Regel für die Ableitung der Umkehrfunktion dann aus der Bruchrechnung:

$$\begin{aligned} y = y(x), z = z(y) &\Rightarrow \frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}, \\ y = y(x), x = x(y) &\Rightarrow \frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}. \end{aligned}$$

Bei der Anwendung dieser saloppen Notation ist jedoch darauf zu achten, an welchen Stellen die jeweiligen Funktionen auszuwerten sind.

2 Mittelwertsatz und Anwendungen

Zur Erinnerung: die Funktion $f : I \rightarrow \mathbb{R}$ hat in x_0 ein Minimum falls $f(x_0) \leq f(x)$ für alle $x \in I$. Entsprechend für ein Maximum.

Satz 2.1 (notwendige Bedingung für Extrema I) Die Funktion $f : (a, b) \rightarrow \mathbb{R}$ habe in $x_0 \in (a, b)$ ein Extremum. Ist f in x_0 differenzierbar, so gilt $f'(x_0) = 0$.

BEWEIS: Im Fall eines Minimums in x_0 haben wir

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \geq 0 & \text{für } x > x_0, \\ \leq 0 & \text{für } x < x_0. \end{cases}$$

Mit $x \searrow x_0$ folgt $f'(x_0) \geq 0$, mit $x \nearrow x_0$ folgt $f'(x_0) \leq 0$. □

Die Funktion $f(x) = x^3$ erfüllt $f'(0) = 0$, aber in $x = 0$ liegt kein Extremum vor. Die Bedingung $f'(x_0) = 0$ ist notwendig für eine Extremalstelle einer differenzierbaren Funktion, aber sie ist nicht hinreichend.

Satz 2.2 (Mittelwertsatz der Differentialrechnung) Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig und differenzierbar auf (a, b) . Dann gibt es ein $\xi \in (a, b)$ mit

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

BEWEIS: Wir zeigen die Behauptung zuerst im Fall $f(a) = f(b) = 0$ (Satz von Rolle). Wir brauchen dann ein $\xi \in (a, b)$ mit $f'(\xi) = 0$. Nach Kapitel 2, Satz 3.3 gibt es $\xi_1, \xi_2 \in [a, b]$ mit

$$f(\xi_1) = \inf_{x \in [a, b]} f(x) \quad \text{und} \quad f(\xi_2) = \sup_{x \in [a, b]} f(x).$$

Ist $\xi_1 \in (a, b)$, so folgt $f'(\xi_1) = 0$ nach Satz 2.1, also können wir $\xi = \xi_1$ wählen. Analog, wenn $\xi_2 \in (a, b)$. Es bleibt der Fall dass ξ_1, ξ_2 beides Randpunkte sind. Dann gilt $f(\xi_{1,2}) = 0$, also sind Infimum und Supremum gleich Null und f ist die Nullfunktion. Aber dann ist $f'(\xi) = 0$ für jedes $\xi \in (a, b)$. Seien nun $f(a), f(b)$ beliebig. Definiere $h : [a, b] \rightarrow \mathbb{R}$ durch

$$h(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right).$$

Es gilt $h(a) = h(b) = 0$. Also existiert ein $\xi \in (a, b)$ mit

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

□

Satz 2.3 (Monotoniekriterium) Sei f differenzierbar auf (a, b) , stetig auf $[a, b]$. Dann gelten folgende Aussagen:

$$\begin{aligned} f'(x) \geq 0 \text{ für alle } x \in (a, b) &\Rightarrow f \text{ ist wachsend auf } [a, b] \\ f'(x) \leq 0 \text{ für alle } x \in (a, b) &\Rightarrow f \text{ ist fallend auf } [a, b] \\ f'(x) = 0 \text{ für alle } x \in (a, b) &\Rightarrow f \text{ ist konstant.} \end{aligned}$$

Bei strikter Ungleichung folgt strenge Monotonie auf $[a, b]$.

BEWEIS: Sei $a \leq x_1 < x_2 \leq b$. Nach dem Mittelwertsatz gibt es ein $\xi \in (x_1, x_2)$, so dass gilt:

$$f(x_2) - f(x_1) = f'(\xi) \underbrace{(x_2 - x_1)}_{>0} \begin{cases} \geq 0 & \text{wenn } f'(\xi) \geq 0 \\ > 0 & \text{wenn } f'(\xi) > 0 \\ \leq 0 & \text{wenn } f'(\xi) \leq 0 \\ < 0 & \text{wenn } f'(\xi) < 0 \\ = 0 & \text{wenn } f'(\xi) = 0. \end{cases}$$

□

Satz 2.4 (Schrankensatz) Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig und $|f'(x)| \leq L$ für alle $x \in (a, b)$. Dann folgt die Abschätzung

$$|f(b) - f(a)| \leq L |b - a|.$$

BEWEIS: Nach Mittelwertsatz gibt es ein $\xi \in [a, b]$ mit

$$|f(b) - f(a)| = |f'(\xi)(b - a)| \leq L |b - a|.$$

□

Wir kommen jetzt zu höheren Ableitungen.

Definition 2.1 Die zweite Ableitung von $f : (a, b) \rightarrow \mathbb{R}$ in x_0 ist

$$f''(x_0) = (f')'(x_0).$$

Dazu muss f' nahe bei x_0 definiert sein, und f' muss in x_0 differenzierbar sein.

Mit der zweiten Ableitung gewinnen wir genauere Informationen über lokale Extrema.

Satz 2.5 (notwendige Bedingung für Extrema II) Sei $f : (a, b) \rightarrow \mathbb{R}$ differenzierbar, mit einem Minimum in x_0 . Falls $f''(x_0)$ existiert, so gilt

$$f'(x_0) = 0, \quad f''(x_0) \geq 0.$$

Für Maxima gilt analog $f'(x_0) = 0, f''(x_0) \leq 0$.

BEWEIS: $f'(x_0) = 0$ wurde in Satz 2.1 bewiesen. Es folgt

$$f''(x_0) = \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0)}{x - x_0} = \lim_{x \rightarrow x_0} \frac{f'(x)}{x - x_0}.$$

Wäre $f''(x_0) < 0$, so wäre $f'(x) > 0$ links von x_0 und $f'(x) < 0$ rechts von x_0 , jeweils nahe bei x_0 . Nach Satz 2.3 ist f dann streng monoton wachsend links von x_0 und streng monoton fallend rechts von x_0 , hat also in x_0 ein striktes, lokales Maximum im Widerspruch zur Annahme. □

Die Funktion $f(x) = x^4$ zeigt, dass in einem Minimum $f''(x_0) = 0$ gelten kann.

3 Die Exponentialfunktion

Wachstums- oder Zerfallsprozesse spielen in Natur und Anwendungen eine große Rolle. Dabei ist die zeitliche Entwicklung einer Größe $K(t)$, $t \in I$, durch ein Wachstums- oder Zerfallsgesetz und einen Startwert $K(t_0) = K_0$ gegeben. Ein diskretes Modell von diesem Typ ist der Zinseszins. Der Zinssatz $\lambda \in \mathbb{R}$ gibt das Plus an, dass man bei linearer Verzinsung nach einem Jahr pro Euro bekommt. Beim Zinseszins wird das Jahr in Zeitschritte der Länge Δt unterteilt, und in jedem Zeitintervall $[t, t + \Delta t]$ wird anteilig mit der Rate λ verzinst. Das bedeutet

$$K(t + \Delta t) = K(t)(1 + \lambda\Delta t) \quad \text{bzw.} \quad \frac{K(t + \Delta t) - K(t)}{\Delta t} = \lambda K(t).$$

Der Effekt ist, dass die bereits erworbenen Zinsen ihrerseits Zinsen liefern, je kleiner die Zeitschritte gewählt werden desto besser. Um ein kontinuierliches Modell herzuleiten, lassen wir Δt gegen Null gehen. Formal ergibt sich das Wachstumsgesetz, und die Anfangsbedingung,

$$K'(t) = \lambda K(t) \quad \text{für } t \in I, \quad K(t_0) = K_0. \quad (3.1)$$

Ob dieses Modell sinnvoll ist, muss mathematisch untersucht werden. Zum einen ist die Existenz einer Lösungsfunktion $K(t)$ zu klären, sonst wäre alles Quatsch. Zweitens sollte die Lösung eindeutig sein, damit Aussagen über zukünftige oder vergangene Entwicklung möglich sind. Monotonie oder Verhalten für große Zeiten sind natürlich auch von Interesse. Die Aufgabe, zu einem Wachstumsgesetz und Startwert die zeitliche Entwicklung zu bestimmen, bezeichnet man allgemein als Anfangswertproblem. Wir kennen schon ein triviales aber fundamentales Beispiel:

$$K'(t) = 0 \quad \text{für } t \in I, \quad K(t_0) = K_0.$$

Die eindeutige Lösung ist die konstante Funktion $K(t) = K_0$, siehe Satz 2.3. Wenn in einem Wachstumsgesetz die Ableitung der gesuchten Funktion vorkommt, spricht man von einer Differentialgleichung. Speziell ist $K' = \lambda K$ die Differentialgleichung des natürlichen Wachstums mit Rate $\lambda \in \mathbb{R}$. Im folgenden betrachten wir den Spezialfall $\lambda = 1$ und $t_0 = 0$, der allgemeine Fall wird sich später leicht durch Skalierung ergeben.

Satz 3.1 *Es gibt genau eine Lösung $f : \mathbb{R} \rightarrow \mathbb{R}$ des Anfangswertproblems*

$$f' = f \quad \text{auf } \mathbb{R}, \quad f(0) = 1. \quad (3.2)$$

Diese Funktion bezeichnen wir mit $\exp : \mathbb{R} \rightarrow \mathbb{R}$. Es gilt die Reihendarstellung

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad \text{für alle } x \in \mathbb{R}. \quad (3.3)$$

BEWEIS: Unser erstes Ziel ist die Eindeutigkeit. Sei f eine Lösung von (3.2). Dann ist natürlich f differenzierbar, und damit auch stetig. Weiter folgt

$$f(x)f(-x) = 1 \quad \text{für alle } x \in \mathbb{R}. \quad (3.4)$$

Und zwar berechnen wir

$$\frac{d}{dx} f(x)f(-x) = f'(x)f(-x) - f(x)f'(-x) = f(x)f(-x) - f(x)f(-x) = 0.$$

Also gilt $f(x)f(-x) = f(0)^2 = 1$. Es folgt $f(x) \neq 0$ für alle $x \in \mathbb{R}$. Nun ist $f(0) = 1$, wäre $f(x) < 0$ für ein $x \in \mathbb{R}$ so hat f eine Nullstelle nach dem Zwischenwertsatz, Widerspruch. Somit ist f strikt positiv. Sei nun g eine weitere Lösung von (3.2). Dann berechnen wir

$$\left(\frac{g}{f}\right)' = \frac{g'f - gf'}{f^2} = 0.$$

Also ist g/f konstant mit $(g/f)(0) = 1$, das heißt $g = f$.

Zur Existenz machen wir mit Cauchy den Ansatz $f(x) = \sum_{k=0}^{\infty} a_k x^k$. Wir berechnen

$$f'(x) = \frac{d}{dx} \sum_{k=0}^{\infty} a_k x^k = \sum_{k=0}^{\infty} \frac{d}{dx} a_k x^k = \sum_{k=1}^{\infty} k a_k x^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} x^k.$$

Es ist unklar ob (3.2) eine Lösung hat die so als Reihe geschrieben werden kann, auch die Vertauschbarkeit der Ableitung mit der Reihe müsste begründet werden. Aber wir wollen hier gar nichts beweisen, sondern nur auf eine potentielle Formel für die Lösung kommen. Koeffizientenvergleich in der Gleichung $f' = f$ liefert $a_{k+1} = a_k/(k+1)$. Mit $a_0 = f(0) = 1$ folgt per Induktion $a_k = 1/k!$. Wir kommen also zum *educated guess*

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \quad (3.5)$$

Wir behaupten dass diese Formel eine Lösung des Anfangswertproblems (3.2) liefert. Rechts steht eine unendliche Summe, diese ist zu verstehen als Grenzwert der endlichen Summen

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) \quad \text{wobei } f_n(x) = \sum_{k=0}^n \frac{x^k}{k!}.$$

Als erstes müssen wir die Konvergenz der $f_n(x)$ zeigen. Für $K \in \mathbb{N}$ beliebig betrachte die $x \in \mathbb{R}$ mit $|x| \leq K/2$. Für $n \geq K$ gilt die Abschätzung

$$\frac{|x|^n}{n!} \leq \frac{K^n}{n!} 2^{-n} \leq \frac{K^K}{K!} \underbrace{\frac{K}{K+1}}_{\leq 1} \cdots \underbrace{\frac{K}{n}}_{\leq 1} 2^{-n} \leq C_K 2^{-n} \quad \text{mit } C_K = K^K/K!.$$

Also bilden die $f_n(x)$ eine Cauchyfolge, und zwar gilt für $n+1 \geq K$ und $m > n$

$$|f_m(x) - f_n(x)| \leq \sum_{k=n+1}^m \frac{|x|^k}{k!} \leq C_K \sum_{k=n+1}^m 2^{-k} \leq C_K 2^{-n} \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Damit ist $f(x)$ definiert. Offenbar ist $f(0) = 1$, zu zeigen bleibt die Differentialgleichung $f' = f$. Wir verwenden die folgende Abschätzung, die für jedes $n \in \mathbb{N}$ und $|x| \leq R$ gilt:

$$|f_n(x)| \leq \sum_{k=0}^n \frac{|x|^k}{k!} \leq C \quad \text{mit } C = \sum_{k=0}^{\infty} \frac{R^k}{k!} < \infty.$$

Die $f_n(x)$ selbst lösen nicht die Differentialgleichung. Aber doch fast, es gilt

$$f'_{n+1}(x) = \frac{d}{dx} \sum_{k=0}^{n+1} \frac{x^k}{k!} = \sum_{k=1}^{n+1} \frac{x^{k-1}}{(k-1)!} = f_n(x).$$

Seien nun $|x|, |x_0| \leq R$. Dann folgt $|f'_n(x)| = |f'_{n-1}(x)| \leq C$, und weiter aus dem Mittelwertsatz für ein ξ zwischen x, x_0

$$\left| \frac{f_{n+1}(x) - f_{n+1}(x_0)}{x - x_0} - f_n(x_0) \right| = |f'_{n+1}(\xi) - f_n(x_0)| = |f_n(\xi) - f_n(x_0)| \leq C|x - x_0|.$$

Dabei haben wir zuletzt den Schrankensatz benutzt. Mit $n \rightarrow \infty$ folgt

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - f(x_0) \right| \leq C|x - x_0|.$$

Für $x \rightarrow x_0$ folgt $f'(x_0) = f(x_0)$. Da $R > 0$ beliebig gilt $f' = f$. $\square$

Folgerung 3.1 Die eindeutige Lösung von (3.1) ist $K(t) = K_0 \exp(\lambda(t - t_0))$ für $t \in I$.

BEWEIS: Dass die Formel eine Lösung liefert ist klar. Für eine Lösung $K(t)$ berechnen wir

$$\frac{d}{dt} \exp(-\lambda(t - t_0))K(t) = (-\lambda) \exp(-\lambda(t - t_0))K(t) + \exp(-\lambda(t - t_0))\lambda K(t) = 0.$$

Einsetzen von t_0 ergibt $\exp(-\lambda(t - t_0))K(t) = \exp(0)K(t_0) = K_0$. Multiplikation mit $\exp(\lambda(t - t_0))$ liefert nun $K(t) = K_0 \exp(\lambda(t - t_0))$, wobei (3.4) benutzt wurde. $\square$

Jetzt interessieren wir uns für weitere Eigenschaften der Exponentialfunktion. Wir hatten schon gesehen dass sie strikt positiv ist, also $\exp : \mathbb{R} \rightarrow (0, \infty)$ und $\exp(0) = 1$. Wegen $\exp' = \exp > 0$ ist die Funktion streng monoton wachsend. Die Reihendarstellung zeigt $\exp(x) \geq 1 + x$ für $x \geq 0$, insbesondere $\exp(x) \rightarrow \infty$ für $x \nearrow \infty$. Aus (3.4) folgt dann $f(x) \rightarrow 0$ für $x \rightarrow -\infty$.

Satz 3.2 (Funktionalgleichung) Für die Exponentialfunktion gilt

$$\exp(x + y) = \exp(x) \exp(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

BEWEIS: Für $g(y) = \exp(x + y)$, mit $x \in \mathbb{R}$ fest, gilt $g' = g$. Aus Folgerung 3.1 erhalten wir $g(y) = g(0) \exp(y) = \exp(x) \exp(y)$. $\square$

Satz 3.3 (Definition des Logarithmus) $\exp : (-\infty, \infty) \rightarrow (0, \infty)$ ist streng monoton wachsend und bijektiv. Die Umkehrfunktion $\log : (0, \infty) \rightarrow (-\infty, \infty)$ heißt (natürlicher) Logarithmus. Sie ist streng monoton wachsend und bijektiv, und hat die Grenzwerte

$$\lim_{r \searrow 0} \log(r) = -\infty \quad \text{und} \quad \lim_{r \nearrow \infty} \log(r) = \infty. \quad (3.6)$$

Es gilt $\log(1) = 0$, und die Ableitung des Logarithmus ist

$$\log'(r) = \frac{1}{r} \quad \text{für alle } r \in (0, \infty). \quad (3.7)$$

Schließlich gilt die Funktionalgleichung

$$\log(rs) = \log(r) + \log(s) \quad \text{für alle } r, s > 0. \quad (3.8)$$

BEWEIS: Nach Satz 3.2 zur Umkehrfunktion ist $\exp$ bijektiv, und die Umkehrfunktion $\log$ ist streng monoton wachsend. Da $\exp(0) = 1$ ist $\log(1) = 0$. Die Grenzwerte (3.6) folgen wegen

$$\lim_{x \rightarrow \pm\infty} \log(\exp(x)) = \lim_{x \rightarrow \pm\infty} x = \pm\infty.$$

Da $\exp'(x) = \exp(x) > 0$, ist $\log$ differenzierbar mit Ableitung, siehe Satz 1.4,

$$\log'(r) = \frac{1}{\exp'(\log(r))} = \frac{1}{\exp(\log(r))} = \frac{1}{r}.$$

Die Funktionalgleichung (3.8) ergibt sich schließlich aus $\exp(x) \exp(y) = \exp(x + y)$, indem wir $x = \log r$, $y = \log(s)$ einsetzen und dann den Logarithmus nehmen. $\square$

Definition 3.1 (Potenz mit reellen Exponenten) Für $a > 0$, $x \in \mathbb{R}$ setzen wir

$$a^x = \exp(x \log(a)).$$

Die Eulersche Zahl e ist, siehe Satz 3.1 für die Reihendarstellung,

$$e = \exp(1) = \sum_{k=0}^{\infty} \frac{1}{k!} \approx 2,71828\dots \quad (3.9)$$

Es folgt $\log(e) = 1$ und $e^x = \exp(x)$. Die Funktionalgleichungen von $\log$ und $\exp$ liefern

$$\begin{aligned} a^x a^y &= a^{x+y} & (a^x)^y &= a^{xy} \\ \left(\frac{1}{a}\right)^x &= a^{-x} & a^x b^x &= (ab)^x. \end{aligned}$$

Durch Induktion sieht man $a^{nx} = (a^x)^n$. Mit $x = \frac{1}{n}$ folgt $(a^{\frac{1}{n}})^n = a^{n \cdot \frac{1}{n}} = a$, das heißt $a^{\frac{1}{n}} = \sqrt[n]{a}$. Also stimmt für $x \in \mathbb{Q}$ unsere Definition der Potenz mit der Definition aus der Schule überein. Weiter gelten für $x \in \mathbb{R}$ die

$$\frac{d}{dx} a^x = \log(a) a^x \quad \text{und} \quad \frac{d}{da} a^x = x a^{x-1}.$$

Wir haben die Differentialgleichung $f' = f$ als Grenzwert des diskreten Zinsmodells motiviert, wenn der Zeitschritt gegen Null geht. Die diskreten Lösungen sollten dann die kontinuierliche Lösung approximieren. Dies sollte zu einer alternativen Formel für die Exponentialfunktion führen. Unterteilen wir das Zeitintervall $[0, 1]$ in n Abschnitte der Länge $1/n$, so ergibt sich bei Zinsrate x und Anfangswert $f_n(0) = 1$ mittels Induktion

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n.$$

Satz 3.4 (Eulerapproximation) Für alle $x \in \mathbb{R}$ gilt

$$\exp(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

BEWEIS: Wir zeigen $n \log\left(1 + \frac{x}{n}\right) \rightarrow x$, mit $\exp$ folgt daraus die Behauptung. Es gilt

$$\left| \frac{d}{dx} \left(x - n \log\left(1 + \frac{x}{n}\right)\right) \right| = \left| 1 - \frac{1}{1 + \frac{x}{n}} \right| = \frac{1}{\left|1 + \frac{x}{n}\right|} \frac{|x|}{n} \leq \frac{2|x|}{n} \quad \text{für } n \geq 2|x|.$$

Diese Abschätzung gilt für alle x' mit $|x'| \leq |x|$, also liefert der Schrankensatz

$$\left| x - n \log\left(1 + \frac{x}{n}\right) \right| \leq \frac{2|x|^2}{n} \rightarrow 0 \quad \text{mit } n \rightarrow \infty.$$

$\square$

4 Die trigonometrischen Funktionen

Bei der harmonischen Schwingung wird die Auslenkung eines schwingungsfähigen Systems aus seiner Ruhelage durch eine zeitabhängige Funktion $u(t)$, $t \in \mathbb{R}$, beschrieben. Ein klassisches Beispiel ist die Schwingung einer Masse an einer Feder. Nach dem Hookeschen Gesetz (1676) ist die elastische Kraft proportional zur Auslenkung, also $F = -ku$. Die Kraft bewirkt eine Beschleunigung der Masse, nach dem Newtonschen Grundgesetz (1687) ist $F = mu''$. Wir wählen die Konstanten so dass $k/m = 1$, und erhalten die Bewegungsgleichung

$$u'' + u = 0 \quad \text{für } u : \mathbb{R} \rightarrow \mathbb{R}. \quad (4.10)$$

Setzen wir $v = -u'$ so folgt aus (4.10) das System

$$(u, v)' = (-v, u) \quad \text{für } (u, v) : \mathbb{R} \rightarrow \mathbb{R}^2. \quad (4.11)$$

Für jede Lösung $c(t) = (u(t), v(t))$ von (4.11) folgt

$$\frac{d}{dt}|c|^2 = \frac{d}{dt}(u^2 + v^2) = 2(uu' + vv') = 2(u(-v) + vu) = 0. \quad (4.12)$$

Der Punkt $c(t)$ bewegt sich also auf einem Kreis um den Nullpunkt. Das hat physikalisch folgende Interpretation: längs der Bewegung findet ein Austausch zwischen potentieller und kinetischer Energie statt. Die potentielle Energie ist $ku^2/2$, das folgt aus dem Hookeschen Gesetz, die kinetische Energie ist $mv^2/2$. Bei unserer Normierung $k/m = 1$ ist die Gesamtenergie proportional zu $u^2 + v^2 = |c|^2$, also ist Gleichung (4.12) der Energieerhaltungssatz.

Wir gehen jetzt zu einer Formulierung mit komplexen Zahlen über, indem wir $(x, y) \in \mathbb{R}^2$ mit $x + iy \in \mathbb{C}$ identifizieren. Der Vektor $(-y, x)$ entspricht dann $-y + ix = i(x + iy)$. Setzen wir $c(t) = u(t) + iv(t) \in \mathbb{C}$, so erhalten wir folgende äquivalente Fassung von (4.11):

$$c' = ic \quad \text{für } c = u + iv : \mathbb{R} \rightarrow \mathbb{C}. \quad (4.13)$$

Formal ist dies die Gleichung des natürlichen Wachstums mit Rate $\lambda = i$, siehe Gleichung (3.1), nur sind $c(t)$ und i hier komplexe Zahlen. Wir konstruieren die Lösung wieder mit der Exponentialfunktion, aber jetzt in $\mathbb{C}$.

Definition 4.1 (Exponentialfunktion) Die komplexe Exponentialfunktion ist

$$\exp : \mathbb{C} \rightarrow \mathbb{C}, \quad \exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}. \quad (4.14)$$

Wir schreiben auch e^z statt $\exp(z)$.

Die Reihe konvergiert für jedes $z \in \mathbb{C}$, dazu zeigt man wie in Satz 3.1 dass die endlichen Summen eine Cauchyfolge bilden. Also ist die Funktion wohldefiniert.

Satz 4.1 Es gibt genau eine Lösung $c : \mathbb{R} \rightarrow \mathbb{C}$ des Anfangswertproblems

$$c' = ic \quad \text{auf } \mathbb{R}, \quad c(0) = 1. \quad (4.15)$$

Die Lösung ist gegeben durch $c(t) = e^{it}$ für alle $t \in \mathbb{R}$.

BEWEIS: Wir zeigen als erstes die Eindeutigkeit. Seien $c_{1,2} : \mathbb{R} \rightarrow \mathbb{C}$ Lösungen der Gleichung $c'_{1,2} = ic_{1,2}$, dann folgt für $c = c_1 - c_2$

$$c' = (c_1)' - (c_2)' = i(c_1 - c_2) = ic.$$

Der Erhaltungssatz (4.12) liefert $|c(t)|^2 = |c(0)|^2$ für alle $t \in \mathbb{R}$. Haben $c_{1,2}$ denselben Anfangswert, so ist $c(0) = 0$ und damit $c(t) = 0$ für alle t , also $c_1 = c_2$. Wir zeigen nun dass $c(t) = e^{it}$ eine Lösung von (4.15) ist. Klar ist $c(0) = \exp(0) = 1$. Weiter berechnen wir

$$\frac{d}{dt} e^{it} = \frac{d}{dt} \sum_{k=0}^{\infty} \frac{(it)^k}{k!} = \sum_{k=0}^{\infty} \frac{d}{dt} \frac{i^k t^k}{k!} = \sum_{k=1}^{\infty} \frac{i^k k t^{k-1}}{k!} = i \sum_{k=1}^{\infty} \frac{(it)^{k-1}}{(k-1)!} = i e^{it}.$$

Also gilt $c'(t) = ic(t)$ wie behauptet. Allerdings ist da noch eine Frage: war die Vertauschung der Ableitung $\frac{d}{dt}$ mit der unendlichen Summe korrekt? Diese Frage trat auch bei der reellen Exponentialfunktion auf, siehe Satz 3.1. Für die endlichen Summen gilt hier $c'_n(t) = ic_{n-1}(t)$, wir können das dortige Argument leicht anpassen. Wir verzichten auf die Details. $\square$

Satz 4.2 Die Kreisbewegung $c(t) = e^{it}$ hat folgende Eigenschaften:

- (a) $|e^{it}| = 1$ für alle $t \in \mathbb{R}$.
- (b) $e^{i(s+t)} = e^{is}e^{it}$ für alle $s, t \in \mathbb{R}$ (Funktionalgleichung).
- (c) $\overline{e^{it}} = e^{-it}$.

BEWEIS: Nach (4.12) ist $|c(t)|$ konstant, aus $c(0) = 1$ folgt $|c(t)| = 1$ für alle t . Für die Funktionalgleichung (b) berechne

$$\begin{aligned} \frac{d}{ds} c(s+t) &= c'(s+t) = ic(s+t), & c(s+t)|_{s=0} &= c(t), \\ \frac{d}{ds} c(s)c(t) &= c'(s)c(t) = ic(s)c(t), & c(s)c(t)|_{s=0} &= c(t). \end{aligned}$$

Aus der Eindeutigkeit in Satz 4.1 folgt $c(s+t) = c(s)c(t)$ wie behauptet. Es folgt weiter $c(t)c(-t) = c(t+(-t)) = 1$. Andererseits ist auch $c(t)\overline{c(t)} = |c(t)|^2 = 1$, also ist $\overline{c(t)} = c(-t)$. $\square$

Für eine allgemeine Kurve $c : [a, b] \rightarrow \mathbb{R}^n$ definieren wir die Länge wie folgt: wir wählen die Unterteilungspunkte $t_k = a + k(b-a)/n$ für $k = 0, \dots, n$, und setzen

$$L_n(c) = \sum_{k=1}^n |c(t_k) - c(t_{k-1})|.$$

Das ist die Länge des Streckenzugs mit den Punkten $c(t_k)$. Die Länge von c ist der Grenzwert

$$L(c) = \lim_{n \rightarrow \infty} L_n(c).$$

Das funktioniert für jede hinreichend reguläre Kurve.

Lemma 4.1 Die Kurve $c(t) = e^{it}$ bildet längentreu ab: für jedes Intervall $I = [a, b]$ gilt

$$L(c|_I) = |I|.$$

BEWEIS: Sei $t_k = a + k\tau_n$ mit $\tau_n = \frac{b-a}{n}$ die äquidistante Unterteilung. Wir berechnen

$$\begin{aligned} L_n(c|_{[a,b]}) &= \sum_{k=1}^n |e^{i(a+k\tau_n)} - e^{i(a+(k-1)\tau_n)}| \\ &= \sum_{k=1}^n |e^{i(a+(k-1)\tau_n)}(e^{i\tau_n} - 1)| \\ &= n |e^{i\tau_n} - 1| \\ &= \left| \frac{e^{i\tau_n} - 1}{\tau_n} \right| (b - a). \end{aligned}$$

Nun gilt $(e^{it} - 1)/t = (c(t) - c(0))/t \rightarrow c'(0) = i$ mit $t \rightarrow 0$, also folgt

$$L(c|_I) = \lim_{n \rightarrow \infty} L_n(c|_I) = b - a = |I|. \quad (4.16)$$

□

Eine wichtige Eigenschaft der Kreisbewegung ist die Periodizität. Die Kurve $c(t) = e^{it}$ durchläuft den Einheitskreis in einer Richtung, nämlich gegen den Uhrzeigersinn, mit konstanter Geschwindigkeit

$$|c'(t)| = |ic(t)| = |c(t)| = 1.$$

Es gibt deshalb ein erstes $p > 0$ mit $c(p) = c(0)$, und $c(t + p)$ löst das Anfangswertproblem

$$\frac{d}{dt}c(t + p) = c'(t + p) = ic(t + p), \quad c(t + p)|_{t=0} = c(p) = c(0).$$

Aus der Eindeutigkeit folgt $c(t + p) = c(t)$ für alle $t \in \mathbb{R}$, das heißt die Kurve $c(t)$ hat die Periode p . Da die Kurve auf $[0, p]$ den Kreis genau einmal durchläuft folgt aus Lemma 4.1

$$p = L(c|_{[0,p]}) = 2\pi \quad \text{mit } \pi = 3,1415\dots$$

Definition 4.2 (Cosinus und Sinus) Die Funktionen $\cos, \sin : \mathbb{R} \rightarrow \mathbb{R}$ sind definiert als

$$\cos t = \operatorname{Re}(e^{it}) = \frac{e^{it} + e^{-it}}{2}, \quad \sin t = \operatorname{Im}(e^{it}) = \frac{e^{it} - e^{-it}}{2i}. \quad (4.17)$$

Aus der Definition folgt direkt die *Eulersche Formel*

$$e^{it} = \cos t + i \sin t \quad \text{für alle } t \in \mathbb{R}. \quad (4.18)$$

Satz 4.3 (Eigenschaften von cos und sin) Es gelten folgende Eigenschaften:

(1) *cos und sin haben die Reihendarstellungen*

$$\cos t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \quad \text{und} \quad \sin t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}. \quad (4.19)$$

(2) *Es gilt $\cos' = -\sin$ und $\sin' = \cos$.*

(3) $u(t) = \cos t$ und $v(t) = \sin t$ sind die eindeutigen Lösungen der Anfangswertprobleme

$$\begin{aligned} u'' + u &= 0, & u(0) &= 1, & u'(0) &= 0, \\ v'' + v &= 0, & v(0) &= 0, & v'(0) &= 1. \end{aligned}$$

(4) $\cos^2 t + \sin^2 t = 1$ für alle $t \in \mathbb{R}$.

(5) $\cos(-t) = \cos t$ und $\sin(-t) = -\sin t$ für alle $t \in \mathbb{R}$.

(6) Für alle $s, t \in \mathbb{R}$ gelten die Additionstheoreme

$$\begin{aligned} \cos(s+t) &= \cos s \cos t - \sin s \sin t, \\ \sin(s+t) &= \sin s \cos t + \cos s \sin t. \end{aligned}$$

(7) $\cos$ und $\sin$ sind periodisch, jeweils mit kleinster Periode 2π .

BEWEIS: Die Reihendarstellungen folgen aus der Definition der Exponentialfunktion:

$$\cos t + i \sin t = e^{it} = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}.$$

Die Differentialgleichungen (2) folgen mit $c(t) = e^{it}$ aus

$$\cos'(t) + i \sin'(t) = c'(t) = ic(t) = -\sin t + i \cos t.$$

In (3) ergeben sich die Anfangswerte aus $c(0) = 1$ und $c'(0) = ic(0) = i$. Die Differentialgleichungen folgen aus (2) durch Differentiation. Nach Satz 4.2 ist $|e^{it}| = 1$, also gilt (4). Weiter folgt aus Aussage (c) in Satz 4.2

$$\cos(-t) + i \sin(-t) = e^{-it} = \overline{e^{it}} = \cos t - i \sin t.$$

Also gilt Behauptung (5). Berechne nun mit der Funktionalgleichung, siehe Satz 4.2(b),

$$\cos(s+t) + i \sin(s+t) = e^{i(s+t)} = e^{is} e^{it} = (\cos s + i \sin s)(\cos t + i \sin t).$$

Ausmultiplizieren und Vergleich liefert (6). Nach Definition ist jede Periode von e^{it} auch Periode von $\cos$ und $\sin$. Es gilt aber auch die Umkehrung: ist zum Beispiel $\sin(t+p) = \sin t$ für alle $t \in \mathbb{R}$, so folgt durch Ableiten $\cos(t+p) = \cos t$ für alle $t \in \mathbb{R}$, und damit $\exp i(t+p) = \exp it$. Aber 2π ist die kleinste Periode von $c(t) = e^{it}$, also gilt das auch für $\cos t$ und $\sin t$. $\square$

Hier eine (nicht sehr detaillierte) Wertetabelle der Funktionen Cosinus und Sinus. Die vier Quadranten sind gegen den Uhrzeigersinn mit *I* bis *IV* nummeriert.

t	0	$\pi/2$	π	$3\pi/2$	2π				
e^{it}	1	<i>I</i>	<i>i</i>	<i>II</i>	-1	<i>III</i>	- <i>i</i>	<i>IV</i>	1
$\cos$	1	$\searrow$	0	$\searrow$	-1	$\nearrow$	0	$\nearrow$	1
$\sin$	0	$\nearrow$	1	$\searrow$	0	$\searrow$	-1	$\nearrow$	0

Die Nullstellen von Cosinus und Sinus sind wie folgt.

$$\begin{aligned} e^{it} = 1 &\Leftrightarrow t = 2k\pi \quad \text{mit } k \in \mathbb{Z}, \\ \cos t = 0 &\Leftrightarrow t = \pi/2 + k\pi \quad \text{mit } k \in \mathbb{Z}, \\ \sin t = 0 &\Leftrightarrow t = k\pi \quad \text{mit } k \in \mathbb{Z}. \end{aligned} \tag{4.20}$$

Die Funktionen $\cos$ und $\sin$ sind periodisch und damit nicht injektiv. Wir können sie aber auf ein Intervall einschränken, auf dem sie streng monoton sind und damit eine Umkehrfunktion *arcus cosinus* bzw. *arcus sinus*¹ haben. Man spricht genauer von einem Zweig dieser Umkehrfunktionen, denn Einschränkung auf ein anderes Monotonieintervall führt zu einer anderen Umkehrfunktion. Unsere Wahl hier wird oft als Hauptzweig bezeichnet.

Satz 4.4 (Arcusfunktionen) $\cos : [0, \pi] \rightarrow [-1, 1]$ und $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-1, 1]$ haben stetige Umkehrfunktionen, die wie folgt bezeichnet werden:

(a) $\arccos : [-1, 1] \rightarrow [0, \pi]$ (*arcus cosinus*), mit Ableitung

$$\arccos'(x) = -\frac{1}{\sqrt{1-x^2}} \quad \text{für } x \in (-1, 1). \quad (4.21)$$

(b) $\arcsin : [-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$, (*arcus sinus*), mit Ableitung

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{für } x \in (-1, 1). \quad (4.22)$$

BEWEIS: Es gilt $\cos' = -\sin < 0$ auf $(0, \pi)$, und $\sin' = \cos > 0$ auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Also ist $\cos$ auf $[0, \pi]$ streng monoton fallend, und $\sin$ auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$ streng monoton wachsend. Die Existenz und Stetigkeit der Umkehrfunktionen gilt nach Satz 3.2. Die Arcusfunktionen sind differenzierbar auf $(-1, 1)$ nach Satz 1.4. Wir berechnen, wobei wir $\sin(\arccos x) > 0$ haben wegen $\arccos x \in (0, \pi)$:

$$\begin{aligned} 1 &= \frac{d}{dx} \cos(\arccos x) \\ &= -\sin(\arccos x) \arccos'(x) \\ &= -\sqrt{1-\cos^2(\arccos x)} \arccos'(x) \\ &= -\sqrt{1-x^2} \arccos'(x). \end{aligned}$$

Die Formel für $\arcsin'(x)$ folgt analog. □

Folgerung 4.1 (Polarkoordinaten) Zu jedem $z \in \mathbb{C} \setminus \{0\}$ gibt es $r > 0$, $\vartheta \in [0, 2\pi)$ mit $z = re^{i\vartheta}$, und diese sind eindeutig bestimmt.

BEWEIS: Sei $z = x + iy$, und zunächst $|z| = 1$. Wir definieren

$$\vartheta = \begin{cases} \arccos x \in [0, \pi] & \text{für } y \geq 0, \\ 2\pi - \arccos x \in (\pi, 2\pi) & \text{für } y < 0. \end{cases}$$

Verwende $\sin \vartheta = \sqrt{1-\cos^2 \vartheta}$ für $\vartheta \in [0, \pi]$, sowie $\sin(2\pi - \vartheta) = -\sin \vartheta$. Damit folgt

$$e^{i\vartheta} = \begin{cases} \cos(\arccos x) + i \sin(\arccos x) = x + iy & \text{falls } y \geq 0, \\ \cos(\arccos x) - i \sin(\arccos x) = x + iy & \text{falls } y < 0. \end{cases}$$

Für $z \in \mathbb{C} \setminus \{0\}$ beliebig setzen wir $r = |z| > 0$ und definieren $\vartheta \in [0, 2\pi]$ mit $z/r = e^{i\vartheta}$. Die Eindeutigkeit von r ist klar, denn $r = |z|$, für den Polarwinkel ϑ folgt sie aus der Tatsache dass e^{it} die kleinste Periode 2π hat. □

¹Arcus = Bogen (Latein)

Kapitel 4

Integralrechnung

1 Das Riemannsches Integral

Sei $f : I = [a, b] \rightarrow \mathbb{R}$. Für $f \geq 0$ ist das Integral von f der Flächeninhalt des Gebiets zwischen dem Graphen $G_f = \{(x, f(x)) : x \in I\}$ im $\mathbb{R}^2$ und der x -Achse. Das ist auch richtig wenn f Werte in $\mathbb{R}$ hat, nur ist dann der Flächeninhalt unter der x -Achse negativ zu zählen. Als Definition können wir diese Beschreibung allerdings nicht nutzen, denn wir haben den Flächeninhalt von Mengen im $\mathbb{R}^2$ noch gar nicht definiert. Stattdessen approximieren wir das Integral durch Rechtecksummen. Dazu sei Z eine Zerlegung von $[a, b]$, gegeben durch Unterteilungspunkte $a = x_0 \leq \dots \leq x_n = b$. In jedem Teilintervall $I_k = [x_{k-1}, x_k]$ wählen wir eine Stützstelle ξ_k und bilden dann die Riemannsche Summe

$$S(f) = \sum_{k=1}^n f(\xi_k) \Delta x_k \in \mathbb{R} \quad \text{wobei } \Delta x_k = x_k - x_{k-1}. \quad (1.1)$$

Die Zahl $S(f)$ ist ein Näherungswert für das Integral. Eine konkrete Wahl der x_k und ξ_k liefert eine numerische Approximation des Integrals, zum Beispiel könnten wir die äquidistante Zerlegung nehmen, und für ξ_k die Mittelpunkte der Teilintervalle:

$$S(f) = \sum_{k=1}^n f\left(\frac{x_k + x_{k-1}}{2}\right) \frac{b-a}{n} \quad \text{mit } x_k = a + k \frac{b-a}{n}.$$

Aber für die Analysis wäre das zu unflexibel, wir wollen uns nicht so festlegen. Klar ist, dass die Approximation um so besser ist, je kleiner die Teilintervalle sind. Um das zu messen definieren wir allgemein die Feinheit der Zerlegung

$$\Delta(Z) = \max_{1 \leq k \leq n} \Delta x_k. \quad (1.2)$$

Definition 1.1 (Riemannsches Integral) Sei $f : I = [a, b] \rightarrow \mathbb{R}$ stetig. Wir setzen

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} S_n(f), \quad (1.3)$$

wobei $S_n(f)$ Riemannsche Summen sind zu Zerlegungen Z_n mit $\Delta(Z_n) \rightarrow 0$.

Die Notation geht auf Leibniz zurück: das Integralzeichen ist ein stilisiertes S für Summe, der Integrand $f(x) dx$ erinnert an die Rechteckfläche $f(\xi_k) \Delta x_k$.

Ehrlich gesagt, damit die Definition Sinn macht müssen wir zeigen dass der Limes der $S_n(f)$ existiert. Der Beweis davon ist aber etwas technisch, deshalb stellen wir ihn zurück und schauen erst, welche Eigenschaften sich aus der Definition ergeben.

Beispiel 1.1 Für die konstante Funktion $f : I = [a, b] \rightarrow \mathbb{R}$, $f(x) = c$, gilt

$$\int_a^b f(x) dx = c(b - a).$$

Denn für jede Zerlegung Z und jede Wahl der $\xi_k \in I_k$ gilt

$$S(f) = \sum_{k=1}^n f(\xi_k) \Delta x_k = c \sum_{k=1}^n (x_k - x_{k-1}) = c(b - a).$$

Beispiel 1.2 Es gilt $\int_a^b x dx = \frac{1}{2}(b^2 - a^2)$. Wähle als Zerlegung $x_k = a + k(b - a)/n$ und als Stützstellen $\xi_k = x_k$ für $k \geq 1$. Es folgt mit $n \rightarrow \infty$

$$\begin{aligned} S(f) &= \sum_{k=1}^n \left(a + k \frac{b-a}{n} \right) \frac{b-a}{n} \\ &= a(b-a) + \frac{(b-a)^2}{n^2} \frac{n(n+1)}{2} \\ &\rightarrow a(b-a) + \frac{(b-a)^2}{2} = \frac{b^2 - a^2}{2}. \end{aligned}$$

Wir zeigen nun ein paar Eigenschaften des Integrals.

Satz 1.1 (Linearität des Integrals) Für $f, g : I = [a, b] \rightarrow \mathbb{R}$ stetig und $\alpha, \beta \in \mathbb{R}$ gilt

$$\int_a^b (\alpha f + \beta g)(x) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx.$$

BEWEIS: Für jede Zerlegung Z mit Stützstellen ξ_k gilt

$$\begin{aligned} S(\alpha f + \beta g) &= \sum_{k=1}^n (\alpha f(\xi_k) + \beta g(\xi_k)) \Delta x_k \\ &= \alpha \sum_{k=1}^n f(\xi_k) \Delta x_k + \beta \sum_{k=1}^n g(\xi_k) \Delta x_k \\ &= \alpha S(f) + \beta S(g). \end{aligned}$$

Wähle eine Folge Z_n mit $\Delta(Z_n) \rightarrow 0$, dann folgt

$$\begin{aligned} \int_a^b (\alpha f + \beta g)(x) dx &= \lim_{n \rightarrow \infty} S_n(\alpha f + \beta g) \\ &= \alpha \lim_{n \rightarrow \infty} S_n(f) + \beta \lim_{n \rightarrow \infty} S_n(g) \\ &= \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx. \end{aligned}$$

□

Satz 1.2 (Monotonie des Integrals) Seien $f, g : I = [a, b] \rightarrow \mathbb{R}$ stetig. Dann folgt:

$$f(x) \leq g(x) \text{ für alle } x \in I \quad \Rightarrow \quad \int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

BEWEIS: Für jede Zerlegung Z mit Stützstellen ξ_k gilt

$$S(f) = \sum_{k=1}^n f(\xi_k) \Delta x_k \leq \sum_{k=1}^n g(\xi_k) \Delta x_k = S(g).$$

Wähle wieder Z_n mit $\Delta(Z_n) \rightarrow 0$. Es folgt

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} S_n(f) \leq \lim_{n \rightarrow \infty} S_n(g) = \int_a^b g(x) dx.$$

□

Satz 1.3 (Standard-Abschätzung) Für $f : I = [a, b] \rightarrow \mathbb{R}$ stetig gilt

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx \leq |b - a| \max_I |f|.$$

BEWEIS: Für jede Zerlegung Z mit Stützstellen ξ_k gilt nach Dreiecksungleichung

$$\left| \sum_{k=1}^n f(\xi_k) \Delta x_k \right| \leq \sum_{k=1}^n |f(\xi_k)| \Delta x_k \leq \max_I |f| \sum_{k=1}^n \Delta x_k = (b - a) \max_I |f|.$$

Also gilt $|S(f)| \leq S(|f|) \leq (b - a) \max_I |f|$. Jetzt wähle Z_n mit $\Delta(Z_n) \rightarrow 0$, es folgt

$$\left| \int_a^b f(x) dx \right| = \lim_{n \rightarrow \infty} |S_n(f)| \quad \text{und} \quad \int_a^b |f(x)| dx = \lim_{n \rightarrow \infty} S_n(|f|).$$

□

Satz 1.4 (Additivität) Sei $f : [a, c] \rightarrow \mathbb{R}$ stetig. Dann gilt für $b \in [a, c]$

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx.$$

BEWEIS: Seien $S'_n(f)$ und $S''_n(f)$ Riemannsche Summen bezüglich Zerlegungen Z'_n von $[a, b]$ und Z''_n von $[b, c]$ mit $\Delta(Z'_n), \Delta(Z''_n) \rightarrow 0$. Dann ist $S_n(f) = S'_n(f) + S''_n(f)$ Riemannsche Summe bezüglich der Zerlegung $Z_n = Z'_n \cup Z''_n$ von $[a, c]$. Es folgt $\Delta(Z_n) \rightarrow 0$ und weiter

$$\int_a^c f(x) dx = \lim_{n \rightarrow \infty} S_n(f) = \lim_{n \rightarrow \infty} (S'_n(f) + S''_n(f)) = \int_a^b f(x) dx + \int_b^c f(x) dx.$$

□

Bis hier konnten wir die Theorie des Integrals für stetige Funktionen ziemlich mühelos entwickeln, allerdings beruhen alle Aussagen auf Definition 1.1. Diese Definition des Integrals wollen wir nun rechtfertigen. Zu zeigen ist die Konvergenz der Riemannschen Summen $S_n(f)$ für Zerlegungen Z_n mit $\Delta(Z_n) \rightarrow 0$. Ich gebe zu dass diese Aussage in Ihrem

späteren Leben keine explizite Rolle spielen dürfte, aber unser Begriff des Integrals wäre auf Sand gebaut wenn wir das nicht verifizieren würden. Und so schwierig ist es auch nicht.

Sei $f : [a, b] \rightarrow \mathbb{R}$. Für $\delta > 0$ definieren wir die Oszillation von f durch

$$\text{osz}_f(\delta) = \sup_{|x-x'|\leq\delta} |f(x) - f(x')|. \quad (1.4)$$

Es gilt $\text{osz}_f(\delta_1) \leq \text{osz}_f(\delta_2)$ für $\delta_1 \leq \delta_2$, die Funktion $\text{osz}_f(\delta)$ ist monoton wachsend.

Satz 1.5 Sei $f : [a, b] \rightarrow \mathbb{R}$ stetig. Dann gilt $\text{osz}_f(\delta) \rightarrow 0$ mit $\delta \rightarrow 0$.

BEWEIS: Angenommen für eine Folge $\delta_k \rightarrow 0$ gilt $\text{osz}_f(\delta_k) > \varepsilon > 0$. Dann gibt es x_k, x'_k mit $|x_k - x'_k| \leq \delta_k$ aber $|f(x_k) - f(x'_k)| \geq \varepsilon$. Angenommen es gilt $x_k \rightarrow x \in [a, b]$. Dann gilt auch $x'_k \rightarrow x$ und es folgt wegen Stetigkeit

$$0 = |f(x) - f(x)| = \lim_{k \rightarrow \infty} |f(x_k) - f(x'_k)| \geq \varepsilon, \quad \text{Widerspruch.}$$

Nun muss die Folge x_k nicht konvergieren. Es gibt aber eine Teilfolge die konvergiert, siehe Kapitel 3, Satz 3.3. Auf diese Teilfolge können wir das Argument anwenden. $\square$

Wir wollen analysieren, wie die Riemannsummen von der Zerlegung abhängen. Um die Stützstellen loszuwerden, gehen wir zu Ober- und Untersummen über: für eine Zerlegung Z mit Teilintervallen I_k für $k = 1, \dots, N$ betrachten wir

$$\bar{S}_Z(f) = \sum_{k=1}^N (\max_{I_k} f) |I_k| \quad \text{und} \quad \underline{S}_Z(f) = \sum_{k=1}^N (\min_{I_k} f) |I_k|.$$

Offenbar gilt $\underline{S}_Z(f) \leq \bar{S}_Z(f)$, andererseits haben wir mit $|I_k| \leq \Delta(Z)$ die Abschätzung

$$\bar{S}_Z(f) - \underline{S}_Z(f) = \sum_{k=1}^N (\max_{I_k} f - \min_{I_k} f) |I_k| \leq \text{osz}_f(\Delta(Z)) (b - a). \quad (1.5)$$

Für Zerlegungen Z, Z' von $[a, b]$ bezeichnen wir mit $Z \cup Z'$ die Zerlegung, die aus allen Unterteilungspunkten von Z, Z' besteht, wobei die Punkte der Größe nach nummeriert sind.

Fakt. Hat Z Feinheit $\Delta(Z) \leq \delta$ und besteht Z' aus N Unterteilungspunkten, so gilt

$$\bar{S}_{Z \cup Z'}(f) \leq \bar{S}_Z(f) \leq \bar{S}_{Z \cup Z'}(f) + 2N\delta \max_I |f|, \quad (1.6)$$

$$\underline{S}_{Z \cup Z'}(f) \geq \underline{S}_Z(f) \geq \underline{S}_{Z \cup Z'}(f) - 2N\delta \max_I |f|. \quad (1.7)$$

Wir zeigen (1.6), Aussage (1.7) geht analog. Sei Z gegeben durch $a = x_0 \leq \dots \leq x_n = b$. Es reicht einen Punkt $\xi \in [x_{k-1}, x_k]$ dazu zu nehmen, die allgemeine Aussage folgt dann mit Induktion. Durch den Punkt ξ wird I_k zerlegt in $I'_k \cup I''_k$, es gilt

$$\bar{S}_{Z \cup \{\xi\}}(f) - \bar{S}_Z(f) = (\max_{I'_k} f) |I'_k| + (\max_{I''_k} f) |I''_k| - (\max_{I_k} f) |I_k|.$$

Für die linke Ungleichung in (1.6) verwenden wir

$$(\max_{I'_k} f) |I'_k| + (\max_{I''_k} f) |I''_k| - (\max_{I_k} f) |I_k| \leq (\max_{I_k} f) (|I'_k| + |I''_k| - |I_k|) = 0.$$

Für die rechte Ungleichung in (1.6) schätzen wir ab

$$\begin{aligned} (\max_{I'_k} f) |I'_k| + (\max_{I''_k} f) |I''_k| - (\max_{I_k} f) |I_k| &\geq -(\max_I |f|)(|I'_k| + |I''_k|) - (\max_I |f|) |I_k| \\ &= -2(\max_I |f|) |I_k| \\ &\geq -2(\max_I |f|) \delta. \end{aligned}$$

Wähle nun Zerlegungen $Z'_1 \subset Z'_2 \subset \dots$ so dass $\Delta(Z'_k) \rightarrow 0$, zum Beispiel durch fortgesetzte Halbierung der Intervalle. Wegen der linken Ungleichung in (1.6) bzw. (1.7) gilt

$$\bar{S}_{Z'_1}(f) \geq \dots \geq \bar{S}_{Z'_k}(f) \geq \underline{S}_{Z'_k}(f) \geq \dots \geq \underline{S}_{Z'_1}(f).$$

Außerdem liefert (1.5) und Satz 1.5

$$\bar{S}_{Z'_k}(f) - \underline{S}_{Z'_k}(f) \leq \text{osz}_f(\Delta(Z'_k))(b-a) \rightarrow 0.$$

Also konvergieren die Folgen $\bar{S}_{Z'_k}(f)$ und $\underline{S}_{Z'_k}(f)$ gegen dieselbe Zahl $I \in \mathbb{R}$. Wir zeigen jetzt dass I der gesuchte Grenzwert in Definition 1.1 ist. Sei $S_n(f)$ irgendeine Folge von Riemannsummen für Zerlegungen Z_n mit $\delta_n = \Delta(Z_n) \rightarrow 0$. Mit (1.6) haben wir

$$S_n(f) \leq \bar{S}_{Z_n}(f) \leq \bar{S}_{Z_n \cup Z'_k}(f) + 2N_k \delta_n \max_I f \leq \bar{S}_{Z'_k}(f) + 2N_k \delta_n \max_I f.$$

Dabei ist N_k die Anzahl der Unterteilungspunkte von Z'_k . Wähle Z'_k so dass

$$\bar{S}_{Z'_k}(f) < I + \frac{\varepsilon}{2}.$$

Da $\delta_n \rightarrow 0$ folgt $S_n(f) < I + \varepsilon$ und analog $S_n(f) > I - \varepsilon$ für n hinreichend groß. Also existiert der Grenzwert in Definition 1.1. Er hängt auch nicht von der Approximation ab, denn er ist gleich I für jede Wahl von Zerlegungen und Stützstellen.

2 Ableitung und Integral

Wir kommen nun zu dem zentralen, von Newton und Leibniz studierten Zusammenhang zwischen Differentiation und Integration. Im folgenden bezeichnet I ein Intervall.

Definition 2.1 Sei $f : I \rightarrow \mathbb{R}$. Eine Funktion $F : I \rightarrow \mathbb{R}$ heißt *Stammfunktion* von f wenn

$$F' = f \quad \Leftrightarrow \quad F'(x) = f(x) \quad \text{für alle } x \in I.$$

Satz 2.1 (Eindeutigkeit der Stammfunktion) Sind F, G Stammfunktionen von f auf I , so gilt $G = F + c$ mit einer Konstanten $c \in \mathbb{R}$.

BEWEIS: Auf I gilt $(G - F)' = G' - F' = f - f = 0$. Nach Satz 2.3 in Kapitel 3 gibt es eine Konstante $c \in \mathbb{R}$ mit $G - F = c$, also $G = F + c$. $\square$

Mit anderen Worten, eine Lösung F der Gleichung $F' = f$ auf I ist bis auf eine additive Konstante $c \in \mathbb{R}$ eindeutig bestimmt. Es schließt sich die Frage nach der Existenz an:

Für welche f ist die Gleichung $F' = f$ auf dem Intervall I lösbar?

Wir wollen zeigen dass für stetige Funktionen f die Gleichung lösbar ist, und zwar wollen wir dazu das Integral verwenden. Wir verallgemeinern noch die Definition auf den Fall von Integrationsgrenzen $b \leq a$, indem wir setzen

$$\int_a^b f(x) dx := - \int_b^a f(x) dx \quad \text{und} \quad \int_a^a f(x) dx = 0. \quad (2.1)$$

Es folgt dann für beliebige $a, b, c \in \mathbb{R}$

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx. \quad (2.2)$$

Das gilt nach Satz 1.4 wenn $a \leq b \leq c$. Mit (2.1) folgt es dann auch im Fall $b \leq a \leq c$:

$$\int_b^c f(x) dx = \int_b^a f(x) dx + \int_a^c f(x) dx = - \int_a^b f(x) dx + \int_a^c f(x) dx.$$

Analog ergibt sich die Gleichung für alle Anordnungen von a, b, c .

Satz 2.2 (Hauptsatz der Differential- und Integralrechnung) Sei $f : I \rightarrow \mathbb{R}$ stetig. Dann ist für jedes $x_0 \in I$ die Funktion

$$F : I \rightarrow \mathbb{R}, \quad F(x) = \int_{x_0}^x f(\xi) d\xi$$

eine Stammfunktion von f , das heißt es gilt $F'(x) = f(x)$ für alle $x \in I$.

Bemerkung. In einem Randpunkt von I gilt das im Sinne der einseitigen Ableitung.

BEWEIS: Da $f(x)$ stetig ist, ist das Integral und damit $F(x)$ definiert. Wir berechnen

$$\begin{aligned} \left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \frac{1}{|h|} \left| F(x+h) - F(x) - hf(x) \right| \\ &= \frac{1}{|h|} \left| \int_{x_0}^{x+h} f(\xi) d\xi - \int_{x_0}^x f(\xi) d\xi - hf(x) \right| \\ &= \frac{1}{|h|} \left| \int_x^{x+h} (f(\xi) - f(x)) d\xi \right| \\ &\leq \frac{1}{|h|} |h| \max_{|\xi-x| \leq |h|} |f(\xi) - f(x)|. \end{aligned}$$

Für die Abschätzung siehe Satz 1.3. Da f stetig geht die rechte Seite mit $h \rightarrow 0$ gegen Null. $\square$

Folgerung 2.1 Sei $f : I \rightarrow \mathbb{R}$ stetig und F Stammfunktion von f . Dann gilt für $x_0 \in I$

$$F(x) = F(x_0) + \int_{x_0}^x f(\xi) d\xi \quad \text{für alle } x \in I. \quad (2.3)$$

BEWEIS: Nach Satz 2.1 und Satz 2.2 gibt es ein $c \in \mathbb{R}$ mit

$$F(x) = c + \int_{x_0}^x f(\xi) d\xi \quad \text{für alle } x \in I.$$

Setze $x = x_0$ und erhalte $c = F(x_0)$. $\square$

Folgerung 2.2 (Berechnung von bestimmten Integralen mittels Stammfunktion)

Sei $f : I = [a, b] \rightarrow \mathbb{R}$ stetig, und $F : I \rightarrow \mathbb{R}$ Stammfunktion von f . Dann gilt

$$\int_a^b f(x) dx = F(b) - F(a) =: [F(x)]_{x=a}^{x=b}$$

BEWEIS: Folgt aus (2.3) mit $x_0 = a$, $x = b$. □

Die Folgerung eröffnet die Möglichkeit, viele Integrale durch Kenntnis von Stammfunktionen (bzw. Ableitungsregeln) auszurechnen. Hier eine kleine *Aufleit-Tabelle*:

$f(x)$	$F(x)$	Bemerkungen
x^k	$\frac{1}{k+1} x^{k+1}$	$k \in \mathbb{Z} \setminus \{-1\}$, $x \neq 0$ für $k < 0$
$\frac{1}{x}$	$\log x$	$x \in (0, \infty)$
x^α	$\frac{1}{\alpha+1} x^{\alpha+1}$	$\alpha \in \mathbb{R} \setminus \{-1\}$, $x \in (0, \infty)$
$e^{\lambda x}$	$\frac{1}{\lambda} e^{\lambda x}$	$\lambda \neq 0$
$\cos x$	$\sin x$	
$\sin x$	$-\cos x$	
$\frac{1}{\cos^2 x}$	$\tan x$	$x \notin (2\mathbb{Z} + 1)\frac{\pi}{2}$
$\frac{1}{\sin^2 x}$	$-\cot x$	$x \notin \mathbb{Z}\pi$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$	$x \in (-1, 1)$
$\frac{1}{\sqrt{1-x^2}}$	$-\arccos x$	$x \in (-1, 1)$
$\frac{1}{x^2+1}$	$\arctan x$	
$\cosh x$	$\sinh x$	
$\sinh x$	$\cosh x$	
$\frac{1}{\sqrt{x^2+1}}$	$\operatorname{Arsinh} x$	
$\frac{1}{\sqrt{x^2-1}}$	$\operatorname{Arcosh} x$	$x \in (1, \infty)$
$\frac{1}{1-x^2}$	$\operatorname{Artanh} x$	$x \in (-1, 1)$

Weiter können wir mit dem Hauptsatz bzw. Folgerung 2.2 aus den Differentiationsregeln in Kapitel 3 Abschnitt 1 Integrationsregeln herleiten. Dies wird im Folgenden durchgeführt.

Satz 2.3 (Partielle Integration) Seien f, g auf $I = [a, b]$ stetig differenzierbar. Dann gilt

$$\int_a^b f(x)g'(x) dx = [f(x)g(x)]_{x=a}^{x=b} - \int_a^b f'(x)g(x) dx.$$

BEWEIS: Nach Voraussetzung ist $(fg)' = f'g + fg'$ stetig auf I . Folgerung 2.2 liefert

$$[f(x)g(x)]_{x=a}^{x=b} = \int_a^b (fg)'(x) dx = \int_a^b (f'(x)g(x) + f(x)g'(x)) dx.$$

□

Hier ein einfaches Beispiel für die partielle Integration.

Beispiel 2.1

$$\int_1^x \log u du = \int_1^x 1 \cdot \log u du = [u \log u]_{u=1}^{u=x} - \int_1^x u \frac{1}{u} du = x \log x - (x - 1).$$

Eine schöne Anwendung der partiellen Integration ist die Produktformel von Wallis.

Beispiel 2.2 Wir berechnen $A_n = \int_0^{\pi/2} \sin^n x \, dx$. Zunächst gilt

$$A_0 = \pi/2 \quad \text{und} \quad A_1 = [-\cos x]_{x=0}^{x=\pi/2} = 1.$$

Für $n \geq 1$ erhalten wir durch partielle Integration

$$\begin{aligned} A_{n+1} &= \int_0^{\pi/2} \sin x \sin^n x \, dx \\ &= \underbrace{[-\cos x \sin^n x]_{x=0}^{x=\pi/2}}_{=0} + n \int_0^{\pi/2} \cos^2 x \sin^{n-1} x \, dx \\ &= n \int_0^{\pi/2} \sin^{n-1} x \, dx - n \int_0^{\pi/2} \sin^{n+1} x \, dx, \end{aligned}$$

wobei wir $\cos^2 x = 1 - \sin^2 x$ benutzt haben. Es folgt

$$A_{n+1} = \frac{n}{n+1} A_{n-1} \quad (n \geq 1).$$

Durch Induktion erhalten wir

$$\begin{aligned} A_{2n} &= \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdot \dots \cdot \frac{1}{2} \cdot A_0 = \left(\prod_{k=1}^n \frac{2k-1}{2k} \right) \frac{\pi}{2}, \\ A_{2n+1} &= \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdot \dots \cdot \frac{2}{3} \cdot A_1 = \left(\prod_{k=1}^n \frac{2k}{2k+1} \right) 1. \end{aligned}$$

Satz 2.4 (Substitutionsregel) Sei $f : [\alpha, \beta] \rightarrow \mathbb{R}$ stetig, und $\varphi : [a, b] \rightarrow [\alpha, \beta]$ sei stetig differenzierbar. Dann gilt die Formel

$$\int_{\varphi(a)}^{\varphi(b)} f(y) \, dy = \int_a^b f(\varphi(x)) \varphi'(x) \, dx.$$

BEWEIS: Wähle mit Satz 2.2 eine Stammfunktion $F : [\alpha, \beta] \rightarrow \mathbb{R}$ von f . Dann gilt

$$\begin{aligned} \int_{\varphi(a)}^{\varphi(b)} f(y) \, dy &= [F(y)]_{x=\varphi(a)}^{x=\varphi(b)} \quad (\text{Folgerung 2.2}) \\ &= [F(\varphi(x))]_{x=a}^{x=b} \\ &= \int_a^b (F \circ \varphi)'(x) \, dx \quad (\text{Folgerung 2.2}) \\ &= \int_a^b F'(\varphi(x)) \varphi'(x) \, dx \quad (\text{Kettenregel}) \\ &= \int_a^b f(\varphi(x)) \varphi'(x) \, dx \quad (F' = f). \end{aligned}$$

Wir haben benutzt dass $\varphi'(x)$ auf dem Intervall $[a, b]$ existiert und dort stetig ist. In den Randpunkten sind die einseitigen Ableitungen zu nehmen. $\square$

Für die Anwendung der Substitutionsregel ist folgendes *Kochrezept* nützlich: bei einem gegebenen Integral $\int_{\alpha}^{\beta} f(y) dy$ möchten wir $y = y(x)$ substituieren. Dazu berechnen wir

$$y = y(x) \quad \Rightarrow \quad dy = y'(x) dx.$$

Für die Intervallgrenzen bestimmen wir durch Auflösen nach x die Umkehrfunktion

$$x = x(y) \quad \Rightarrow \quad a = x(\alpha), b = x(\beta).$$

Damit gilt

$$\int_{\alpha}^{\beta} f(y) dy = \int_a^b f(y(x)) y'(x) dx.$$

Hier wieder ein paar Beispiele.

Beispiel 2.3 (Lineare Parameterwechsel) Im Integral $\int_{\alpha}^{\beta} f(y) dy$ substituieren wir

$$y = \frac{x - x_0}{\lambda} \quad \text{für } x_0 \in \mathbb{R}, \lambda > 0.$$

Es ist $dy = 1/\lambda dx$ und $x = x_0 + \lambda y$. Also folgt

$$\int_{\alpha}^{\beta} f(y) dy = \frac{1}{\lambda} \int_{x_0 + \lambda\alpha}^{x_0 + \lambda\beta} f\left(\frac{x - x_0}{\lambda}\right) dx.$$

Beispiel 2.4 (Integration von Ableitungen)

$$\begin{aligned} \int_a^b \frac{f'(x)}{f(x)} dx &= \int_a^b (\log f)'(x) dx = [\log f(x)]_{x=a}^{x=b} \quad (f > 0), \\ \int_a^b u \sqrt{1 + u^2} du &= \frac{1}{3} \int_a^b ((1 + u^2)^{\frac{3}{2}})' du = \left[\frac{1}{3} (1 + u^2)^{\frac{3}{2}} \right]_{u=a}^{u=b}, \\ \int_a^b F'(f(x)) f'(x) dx &= [F(f(x))]_{x=a}^{x=b}. \end{aligned}$$

Beispiel 2.5 (Flächeninhalt unter Hyperbel) Zu berechnen ist das Integral

$$A(x) = \int_1^x \sqrt{u^2 - 1} du \quad \text{für } x \geq 1.$$

Wir substituieren $u = \cosh t$ und erhalten $du = \sinh t dt$, $t = \operatorname{Arcosh} u$, also mit $\operatorname{Arcosh} 1 = 0$

$$\begin{aligned} A(x) &= \int_0^{\operatorname{Arcosh} x} \sinh^2 t dt \\ &= \frac{1}{4} \int_0^{\operatorname{Arcosh} x} (e^{2t} + e^{-2t} - 2) dt \\ &= \left[\frac{1}{8} (e^{2t} - e^{-2t}) \right]_{t=0}^{t = \operatorname{Arcosh} x} - \frac{1}{2} \operatorname{Arcosh} x \\ &= \frac{1}{2} \left(\left[\frac{e^t + e^{-t}}{2} \frac{e^t - e^{-t}}{2} \right]_{t=0}^{t = \operatorname{Arcosh} x} - \operatorname{Arcosh} x \right) \\ &= \frac{1}{2} (x \sqrt{x^2 - 1} - \operatorname{Arcosh} x). \end{aligned}$$

Für rationale Funktionen, also Quotienten von Polynomen, hat man ein spezielles Integrationsverfahren, die Partialbruchzerlegung, die wir hier nur an einem Beispiel vorführen:

Beispiel 2.6 (Partialbruchzerlegung) Um das Integral $\int_{-1/2}^{1/2} \frac{dx}{1-x^2}$ zu berechnen, machen wir den Ansatz

$$\frac{1}{1-x^2} = \frac{1}{(1+x)(1-x)} \stackrel{!}{=} \frac{A}{1-x} + \frac{B}{1+x} = \frac{(A-B)x + (A+B)}{1-x^2}.$$

Der Koeffizientenvergleich ergibt $A = B = 1/2$, also folgt

$$\int_{-1/2}^{1/2} \frac{dx}{1-x^2} = \int_{-1/2}^{1/2} \left(\frac{1/2}{1-x} + \frac{1/2}{1+x} \right) dx = \frac{1}{2} \left[\log \frac{1+x}{1-x} \right]_{x=-1/2}^{x=1/2} = \frac{1}{2} (\log 3 - \log 1/3) = \log 3.$$

Wir wollen auch Integrale über unendliche Intervalle berechnen, sowie Integrale mit Ausnahmestellen, wo die Funktion $f(x)$ rechts- oder linksseitig nicht stetig ist. Wir fangen mit der Situation an, wo die Problemstelle die rechte (oder linke, analog) Intervallgrenze ist.

Definition 2.2 (uneigentliches Integral) Sei $-\infty < a < b \leq \infty$. Für eine stetige Funktion $f : I = [a, b) \rightarrow \mathbb{R}$ setzen wir

$$\int_a^b f(x) dx = \lim_{\beta \nearrow b} \int_a^\beta f(x) dx,$$

falls der rechte Grenzwert existiert. Andernfalls heißt das Integral divergent.

Hier einige Beispiele von uneigentlichen Integralen.

Beispiel 2.7

$$\begin{aligned} \int_1^\infty x^\alpha dx &= \begin{cases} \lim_{R \nearrow \infty} \left[\frac{x^{\alpha+1}}{\alpha+1} \right]_{x=1}^{x=R} = -\frac{1}{\alpha+1} & \text{für } \alpha < -1, \\ \text{divergent} & \text{für } \alpha \geq -1. \end{cases} \\ \int_0^1 x^\alpha dx &= \begin{cases} \lim_{\varepsilon \searrow 0} \left[\frac{x^{\alpha+1}}{\alpha+1} \right]_{x=\varepsilon}^{x=1} = \frac{1}{\alpha+1} & \text{für } \alpha > -1, \\ \text{divergent} & \text{für } \alpha \leq -1. \end{cases} \\ \int_0^1 \frac{dx}{\sqrt{1-x^2}} &= \lim_{b \nearrow 1} \int_0^b \frac{dx}{\sqrt{1-x^2}} = \lim_{b \nearrow 1} \arcsin b = \pi/2. \end{aligned}$$

Als nächstes betrachten wir eine Funktion $f : (a, b) \rightarrow \mathbb{R}$, für die sowohl die linke als auch die rechte Intervallgrenze problematisch ist, zum Beispiel $(a, b) = (-\infty, \infty)$. Wir wählen dann einen Zwischenpunkt $c \in (a, b)$ und setzen

$$\int_a^b f(x) dx := \int_a^c f(x) dx + \int_c^b f(x) dx = \lim_{\alpha \searrow a} \int_\alpha^c f(x) dx + \lim_{\beta \nearrow b} \int_c^\beta f(x) dx,$$

sofern die beiden Grenzwerte bzw. uneigentliche Integrale rechts existieren. Diese Definition hängt nicht von der Wahl von c ab, denn es folgt zum Beispiel für $c' > c$

$$\int_\alpha^{c'} f(x) dx = \int_\alpha^c f(x) dx + \int_c^{c'} f(x) dx \xrightarrow{\alpha \searrow a} \int_a^c f(x) dx + \int_c^{c'} f(x) dx,$$

$$\int_{c'}^{\beta} f(x) dx = \int_c^{\beta} f(x) dx - \int_c^{c'} f(x) dx \xrightarrow{\beta \nearrow b} \int_c^b f(x) dx - \int_c^{c'} f(x) dx.$$

Also bleibt die Summe der Integrale gleich.

Beispiel 2.8

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{dx}{1+x^2} &= \lim_{R \rightarrow \infty} \int_{-R}^0 \frac{dx}{1+x^2} + \lim_{R \rightarrow \infty} \int_0^R \frac{dx}{1+x^2} \\ &= \lim_{R \rightarrow \infty} [\arctan x]_{x=-R}^{x=0} + \lim_{R \rightarrow \infty} [\arctan x]_{x=0}^{x=R} \\ &= \frac{\pi}{2} + \frac{\pi}{2} = \pi. \end{aligned}$$

Ganz analog ist die Situation im Fall einer stetigen Funktion $f : [a, b] \setminus \{c\} \rightarrow \mathbb{R}$ mit $a < c < b$, also mit einer Singularität im Innern des Intervalls. Wir definieren

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx,$$

wobei wieder beide uneigentlichen Integrale rechts existieren müssen.

Beispiel 2.9 Für $\alpha > -1$ berechnen wir

$$\int_{-1}^1 |x|^\alpha dx = \int_{-1}^0 (-x)^\alpha dx + \int_0^1 x^\alpha dx = 2 \int_0^1 x^\alpha dx = \frac{2}{\alpha+1}.$$

Beispiel 2.10 Für die Funktion $f(x) = \frac{1}{x}$ ist das uneigentliche Integral auf dem Intervall $[-1, 1]$ nicht definiert, denn die einzelnen Integrale gehen gegen $\pm\infty$:

$$\begin{aligned} \int_{\varepsilon}^1 \frac{dx}{x} &= [\log x]_{x=\varepsilon}^1 = -\log \varepsilon \rightarrow \infty, \\ \int_{-1}^{-\varepsilon} \frac{dx}{x} &= [\log(-x)]_{x=-1}^{x=-\varepsilon} = \log \varepsilon \rightarrow -\infty. \end{aligned}$$

Allerdings existiert immer noch der Grenzwert

$$\lim_{\varepsilon \searrow 0} \left(\int_{-1}^{-\varepsilon} \frac{dx}{x} + \int_{\varepsilon}^1 \frac{dx}{x} \right) = \lim_{\varepsilon \searrow 0} (\log \varepsilon - \log \varepsilon) = 0.$$

Diesen Wert nennt man dann den Cauchy-Hauptwert des Integrals (Englisch: principal value). Solche Hauptwerte sind mit Vorsicht zu genießen, wir wollen sie nicht weiter betrachten.

Es stellt sich die Frage nach Kriterien zur Existenz uneigentlicher Integrale.

Satz 2.5 (Vergleichstest) Sei $f : [0, \infty) \rightarrow \mathbb{R}$ stetig. Es gebe $\alpha > 1$ und $R > 0$ mit

$$|f(x)| \leq \frac{C}{x^\alpha} \quad \text{für alle } x \geq R,$$

wobei $C < \infty$ Konstante. Dann konvergiert das unbestimmte Integral $\int_0^\infty f(x) dx$.

BEWEIS: Sei zunächst $f \geq 0$. Dann gilt für alle $b \in [R, \infty)$

$$\begin{aligned} 0 &\leq \int_0^b f(x) dx \\ &\leq \int_0^R f(x) dx + \int_R^b \frac{C}{x^\alpha} dx \\ &\leq \int_0^R f(x) dx + C \int_R^\infty \frac{dx}{x^\alpha} < \infty. \end{aligned}$$

Die Funktion $F(b) = \int_0^b f(x) dx$ ist monoton wachsend und nach oben beschränkt, also konvergiert das Integral mit $b \rightarrow \infty$. Hat $f : [0, \infty) \rightarrow \mathbb{R}$ kein Vorzeichen, so betrachten wir

$$f^+(x) = \max(f(x), 0) \quad \text{und} \quad f^-(x) = -\min(f(x), 0).$$

Die Funktionen $f^\pm$ sind stetig und erfüllen die Abschätzung

$$0 \leq f^\pm(x) \leq |f(x)| \leq \frac{C}{x^\alpha} \quad \text{für alle } x \geq R.$$

Also existieren die uneigentlichen Integrale von $f^\pm$. Aber $f(x) = f^+(x) - f^-(x)$, somit

$$\int_0^b f(x) dx = \int_0^b f^+(x) dx - \int_0^b f^-(x) dx,$$

und die gewünschte Konvergenz für $b \rightarrow \infty$ folgt. □

Beispiel 2.11 Als letztes Beispiel betrachten wir die Funktion

$$f : \left[\frac{\pi}{2}, \infty\right) \rightarrow \mathbb{R}, \quad f(x) = \frac{\sin x}{x}.$$

Wir können Satz 2.5 nicht direkt anwenden, denn es gilt nur $|f(x)| \leq 1/x$, wir brauchen aber eine Potenz größer als Eins. Der Trick ist, die Konvergenz durch partielle Integration zu verbessern: es gilt

$$\int_{\frac{\pi}{2}}^b \sin x \frac{1}{x} dx = \left[(-\cos x) \frac{1}{x}\right]_{x=\frac{\pi}{2}}^{x=b} - \int_{\frac{\pi}{2}}^b (-\cos x) \left(-\frac{1}{x^2}\right) dx.$$

Das rechte Integral erfüllt die Abschätzung aus Satz 2.5 mit $\alpha = 2 > 1$, also konvergiert es für $b \rightarrow \infty$. Wir erhalten

$$\lim_{b \rightarrow \infty} \int_{\frac{\pi}{2}}^b \frac{\sin x}{x} dx = - \int_{\frac{\pi}{2}}^\infty \frac{\cos x}{x^2} dx \in \mathbb{R}.$$

Meistens wird das Integral von $\frac{\sin x}{x}$ auf $[0, \infty)$ genommen. Im Nullpunkt gibt es aber kein Problem, denn dort konvergiert die Funktion gegen Eins. Für die Frage der Konvergenz des Integrals spielt der Anfangspunkt also keine Rolle.

Kapitel 5

Approximation von Funktionen

Sei $f : I \rightarrow \mathbb{R}$ eine gegebene Funktion. In diesem Kapitel geht es um die Frage, ob oder wie die Funktion f durch einfache Funktionen approximiert werden kann. Dabei besteht die Klasse der einfachen Funktionen entweder aus Polynomen (Brook Taylor 1712) oder aus trigonometrischen Polynomen (Baptiste Fourier 1822).

1 Taylorapproximation

Bevor wir in die Materie einsteigen zeigen wir folgendes Hilfsmittel.

Satz 1.1 (Mittelwertsatz der Integralrechnung) Sei $\varphi : [a, b] \rightarrow [0, \infty)$ stetig und nicht die Nullfunktion. Zu jeder stetigen Funktion $f : [a, b] \rightarrow \mathbb{R}$ gibt es dann ein $\xi \in [a, b]$ mit

$$\int_a^b f(x)\varphi(x) dx = f(\xi) \int_a^b \varphi(x) dx.$$

BEWEIS: Nach Voraussetzung ist $\int_a^b \varphi(x) dx > 0$. Wir betrachten die normierte Funktion

$$\hat{\varphi} = \frac{\varphi}{\int_a^b \varphi(x) dx}, \quad \text{also} \quad \int_a^b \hat{\varphi}(x) dx = 1.$$

Sei nun $m = \min_{x \in I} f(x)$ und $M = \max_{x \in I} f(x)$. Wegen $\hat{\varphi} \geq 0$ gilt $m\hat{\varphi} \leq f\hat{\varphi} \leq M\hat{\varphi}$, und durch Integration folgt

$$m = \int_a^b m\hat{\varphi}(x) dx \leq \int_a^b f(x)\hat{\varphi}(x) dx \leq \int_a^b M\hat{\varphi}(x) dx = M.$$

Nach dem Zwischenwertsatz gibt es ein $\xi \in [a, b]$ mit

$$f(\xi) = \int_a^b f(x)\hat{\varphi}(x) dx = \frac{1}{\int_a^b \varphi(x) dx} \int_a^b f(x)\varphi(x) dx.$$

□

Jetzt zur Taylorentwicklung. Auf einem offenen Intervall I betrachten wir eine Funktion $f \in C^k(I)$, das heißt f hat stetige Ableitungsfunktionen $f^{(j)} : I \rightarrow \mathbb{R}$ der Ordnung $j = 0, 1, \dots, k$. Die Idee der Taylorentwicklung ist es, die Funktion $f(x)$ mit einem Polynom zu

vergleichen, das mit $f(x)$ an einer festen Stelle x_0 „von höherer Ordnung“ übereinstimmt. Genauer betrachten wir

$$P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j. \quad (1.1)$$

Das ist das k -te Taylorpolynom von f im Punkt x_0 , es hat die Eigenschaft

$$P_k^{(i)}(x_0) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} \underbrace{\left(\frac{d}{dx}\right)^i (x - x_0)^j}_{=j! \delta_{ij}} \Big|_{x=x_0} = f^{(i)}(x_0) \quad \text{für } i = 0, \dots, k.$$

Die Differenz zwischen der Funktion $f(x)$ und ihrem Taylorpolynom $P_k(x)$ bezeichnet man als das Restglied k -ter Ordnung, also

$$R_k : I \rightarrow \mathbb{R}, \quad R_k(x) = f(x) - P_k(x). \quad (1.2)$$

Der Knackpunkt ist nun die Abschätzung dieses Restglieds und damit eine Aussage darüber, wie gut das Polynom P_k die Funktion asymptotisch für $x \rightarrow x_0$ approximiert. Für die Abschätzung leiten wir eine Darstellung des Restglieds nach Lagrange¹ her.

Satz 1.2 (Taylorentwicklung) Sei $f \in C^{k+1}(I)$ für ein $k \in \mathbb{N}_0$ und $x_0 \in I$. Dann gilt die Lagrange-Darstellung des Restglieds: zu jedem $x \in I$ gibt es ein ξ zwischen x_0 und x mit

$$R_k(x) = f(x) - P_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1}. \quad (1.3)$$

BEWEIS: Wir zeigen erst die folgende Darstellung von Cauchy.

$$R_k(x) = \int_{x_0}^x \frac{f^{(k+1)}(y)}{k!} (x - y)^k dy. \quad (1.4)$$

Für $k = 0$ gilt das nach dem Hauptsatz der Differential- und Integralrechnung:

$$R_0(x) = f(x) - f(x_0) = \int_{x_0}^x f'(y) dy.$$

Für $k \geq 1$ folgt per Induktion, mit Formel (1.4) für $k - 1$ statt k ,

$$R_k(x) = R_{k-1}(x) - \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k = \int_{x_0}^x \frac{f^{(k)}(y)}{(k-1)!} (x - y)^{k-1} dy - \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k.$$

Durch partielle Integration erhalten wir nun

$$\begin{aligned} \int_{x_0}^x f^{(k)}(y) \frac{(x - y)^{k-1}}{(k-1)!} dy &= \left[f^{(k)}(y) \left(-\frac{(x - y)^k}{k!} \right) \right]_{y=x_0}^{y=x} + \int_{x_0}^x f^{(k+1)}(y) \frac{(x - y)^k}{k!} dy \\ &= \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k + \int_{x_0}^x \frac{f^{(k+1)}(y)}{k!} (x - y)^k dy. \end{aligned}$$

¹Joseph-Louis Lagrange, 1736-1813

Einsetzen liefert (1.4). Die Darstellung (1.3) von Lagrange folgt mit dem Mittelwertsatz der Integralrechnung. Im Fall $x \geq x_0$ nehmen wir $\varphi(y) = (x - y)^k \geq 0$, es folgt

$$R_k(x) = \frac{f^{(k+1)}(\xi)}{k!} \int_{x_0}^x (x - y)^k dy = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1}.$$

Der Fall $x \leq x_0$ folgt analog. □

Mit Satz 1.2 kann man das Restglied abschätzen, wir zeigen das an folgendem

Beispiel 1.1 Betrachte die Funktion $f(x) = (1 - x)^{-1/2}$, mit Ableitungen

$$f'(x) = \frac{1}{2}(1 - x)^{-3/2} \quad \text{und} \quad f''(x) = \frac{3}{4}(1 - x)^{-5/2}.$$

Es gilt $f(0) = 1$ und $f'(0) = 1/2$, also lautet das Taylorpolynom der Ordnung Eins in $x_0 = 0$

$$P_1(x) = f(0) + f'(0)x = 1 + \frac{1}{2}x,$$

mit der Lagrange-Restglieddarstellung im Fall $x \geq 0$

$$R_1(x) = \frac{f''(\xi)}{2}x^2 = \frac{3}{8}(1 - \xi)^{-5/2}x^2 \quad \text{für ein } \xi \in [0, x].$$

Daraus ergibt sich eine Aussage zur Relativitätstheorie. Und zwar folgt für die relativistische Energie eines Teilchens mit Ruhemasse m_0 und Geschwindigkeit v , wenn wir $\beta = v/c$ setzen,

$$E = \frac{m_0 c^2}{\sqrt{1 - \beta^2}} = m_0 c^2 \left(1 + \frac{1}{2}\beta^2 + \frac{f''(\xi)}{2}\beta^4 \right) = m_0 c^2 + \frac{1}{2}m_0 v^2 + \Delta E.$$

Der erste Term ist die Ruheenergie, der zweite ist die klassische kinetische Energie E_{kin} . Für den relativistischen Korrekturterm ΔE ergibt sich die Abschätzung

$$\frac{\Delta E}{E_{kin}} = \frac{m_0 c^2 f''(\xi)\beta^4/2}{m_0 v^2/2} = f''(\xi)\beta^2 \leq f''(\beta^2)\beta^2 < 0,008 \quad \text{für } \beta \leq 0,1.$$

Bei Geschwindigkeiten $v \leq c/10$ beträgt die relativistische Korrektur weniger als ein Prozent der klassischen kinetischen Energie.

Jetzt ändern wir die Zielsetzung: Es geht uns nicht um das asymptotische Verhalten für $x \rightarrow x_0$, sondern um die Approximation einer Funktion $f : I \rightarrow \mathbb{R}$ in allen Punkten $x \in I$. Wir nehmen dazu an dass f unendlich oft differenzierbar ist, dann können wir für festes $x_0 \in I$ die Folge der Taylorpolynome betrachten:

$$P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j \quad \text{für } k = 0, 1, 2, \dots$$

Wir interessieren uns dafür ob die $P_k(x)$ mit $k \rightarrow \infty$ gegen die Funktion $f(x)$ konvergieren. Wenn ja so haben wir die Darstellung als Reihe

$$f(x) = \lim_{k \rightarrow \infty} P_k(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j.$$

Dies ist die Taylorreihe von f mit Zentrum x_0 . Nach Definition des Restglieds $R_k(x) = f(x) - P_k(x)$ gilt offenbar folgende Aussage:

$$f(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j \Leftrightarrow \lim_{k \rightarrow \infty} R_k(x) = 0. \quad (1.5)$$

Es gibt Funktionen die unendlich oft differenzierbar sind, aber nicht durch die Taylorreihe dargestellt werden, das heißt das Restglied $R_k(x)$ geht nicht gegen Null. Es gibt sogar viele solche Funktionen. Hier aber ein paar positive Beispiele.

Beispiel 1.2 Für die Exponentialfunktion gilt $\exp^{(j)}(0) = 1$ für alle $j \in \mathbb{N}_0$, also ist die Taylorreihe mit Zentrum $x_0 = 0$ gleich

$$P(x) = \sum_{j=0}^{\infty} \frac{1}{j!} x^j = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots$$

Wir haben die Exponentialfunktion genau durch diese Reihe definiert, siehe Satz 3.1. Also wird $\exp$ für alle $x \in \mathbb{R}$ durch diese Taylorreihe dargestellt.

Beispiel 1.3 Für die Kosinusfunktion haben wir

$$\cos^{(j)}(0) = \begin{cases} (-1)^i & \text{für } j = 2i, \\ 0 & \text{für } j = 2i + 1. \end{cases}$$

Somit ist die Taylorreihe des Kosinus

$$P(x) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(2i)!} x^{2i} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \mp \dots$$

Auch der Kosinus wird für alle $x \in \mathbb{R}$ durch diese Taylorreihe dargestellt, dies wurde schon in Satz 4.3 gezeigt.

Beispiel 1.4 Für $f(x) = \log(1+x)$ ist $f(0) = 0$, und wir berechnen wir per Induktion

$$f^{(j)}(x) = (-1)^{j-1} (j-1)! (1+x)^{-j} \quad \text{für } j \geq 1.$$

Also gilt $f^{(j)}(0) = (-1)^{j-1} (j-1)!$ für $j \geq 1$, und die Taylorreihe im Punkt $x_0 = 0$ lautet

$$P(x) = \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} x^j = \frac{x}{1} - \frac{x^2}{2} + \frac{x^3}{3} \mp \dots$$

Wir zeigen: für $x \in [0, 1]$ wird $f(x)$ durch die Taylorreihe dargestellt. Und zwar gibt es nach Satz 1.2 ein $\xi \in [0, x]$ mit

$$f(x) - P_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} x^{k+1}.$$

Es gilt $|f^{(k+1)}(\xi)| = k!(1+\xi)^{-(k+1)} \leq k!$, also folgt

$$|f(x) - P_k(x)| \leq \frac{x^{k+1}}{k+1} \leq \frac{1}{k+1} \rightarrow 0 \quad \text{für } k \rightarrow \infty.$$

Tatsächlich wird $f(x) = \log(1+x)$ auch für $x \in (-1, 0)$ durch die Taylorreihe dargestellt. Aber unser Argument liefert das nicht.

2 Darstellung von Funktionen durch Reihen

Viele Funktionen in der Analysis können als unendliche Reihen definiert beziehungsweise dargestellt werden, zum Beispiel kennen wir schon die Taylorreihe. Hier geht es darum, wie man mit solchen unendlichen Reihen umgehen kann.

Definition 2.1 Seien $a_n \in \mathbb{R}$ für $n \in \mathbb{N}_0$. Die Reihe mit den Gliedern a_n ist die Folge $S_0 = a_0$, $S_1 = a_0 + a_1$, $S_2 = a_0 + a_1 + a_2$, $\dots$, bzw. allgemein

$$S_n = \sum_{k=0}^n a_k \quad \text{für alle } n \in \mathbb{N}_0.$$

Die Reihe heißt konvergent mit Wert $S \in \mathbb{R}$, wenn S_n gegen S konvergiert:

$$\sum_{k=0}^{\infty} a_k := \lim_{n \rightarrow \infty} \sum_{k=0}^n a_n = \lim_{n \rightarrow \infty} S_n = S.$$

Die Zahl S_n heißt n -te Partialsumme der Reihe. Oft schreibt man die Reihe in der Form $a_0 + a_1 + a_2 + \dots$, oder einfach $\sum_{k=0}^{\infty} a_k$, egal ob die Reihe konvergiert oder nicht.

Beispiel 2.1 Die Reihe $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots$ bzw. $\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$. Es gilt

$$S_1 = \frac{1}{1 \cdot 2} = \frac{1}{2}, \quad S_2 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} = \frac{2}{3}, \quad S_3 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} = \frac{3}{4}, \dots$$

Durch Induktion folgt $S_n = \frac{n}{n+1}$. Also ist die Reihe konvergent mit Wert $\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$.

Für konvergente Reihen $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ und $\lambda, \mu \in \mathbb{R}$ ist auch $\sum_{k=0}^{\infty} (\lambda a_k + \mu b_k)$ konvergent, und es gilt

$$\sum_{k=0}^{\infty} (\lambda a_k + \mu b_k) = \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k.$$

Dies ergibt sich aus dem entsprechenden Satz für Folgen. Durch Abändern, Hinzufügen oder Weglassen von endlich vielen Summanden ändert sich nichts an der Konvergenz/Divergenz, nur der Wert der Reihe ändert sich. Zum Beispiel haben wir

$$\sum_{k=n}^m a_k = \sum_{k=0}^m a_k - \sum_{k=0}^{n-1} a_k \quad \xrightarrow{m \rightarrow \infty} \quad \sum_{k=n}^{\infty} a_k = \sum_{k=0}^{\infty} a_k - \sum_{k=0}^{n-1} a_k. \quad (2.6)$$

Wie bei Folgen kann der Anfangsindex einer Reihe statt $k = 0$ auch eine andere Zahl sein.

Beispiel 2.2 (Geometrische Reihe) Für die geometrische Reihe gilt bekanntlich

$$\sum_{k=0}^{\infty} x^k = \lim_{n \rightarrow \infty} \sum_{k=0}^n x^k = \lim_{n \rightarrow \infty} \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x} \quad \text{für } |x| < 1.$$

Für $|x| \geq 1$ konvergiert die Reihe nicht, das ergibt sich aus dem Nullfolgentest (siehe unten).

Im allgemeinen kann eine Reihe nicht einfach ausgerechnet werden. Stattdessen gibt es Tests, mit denen die Konvergenz gezeigt oder widerlegt werden kann, abhängig von den a_n . Die folgende Aussage ist von diesem Typ.

Satz 2.1 (Nullfolgentest) Ist $\sum_{k=0}^{\infty} a_k$ konvergent, so folgt $\lim_{k \rightarrow \infty} a_k = 0$.

BEWEIS: Nach Annahme ist die Folge der Summen $S_n = \sum_{k=0}^n a_k$ konvergent mit einem Grenzwert S , also folgt $a_n = S_n - S_{n-1} \rightarrow S - S = 0$. $\square$

Beispiel 2.3 Für die geometrische Reihe (siehe oben) gilt $a_k = x^k$. Für $|x| \geq 1$ ist also $|a_k| = |x|^k \geq 1$, die Reihe kann nicht konvergieren.

Beispiel 2.4 (Harmonische Reihe) Die Bedingung $a_k \rightarrow 0$ ist *nicht hinreichend* für die Konvergenz, zum Beispiel ist $\sum_{k=1}^{\infty} 1/k = \infty$. Betrachte dazu für $m \in \mathbb{N}_0$ die Teilsumme über die k mit $2^m \leq k < 2^{m+1}$. Sie enthält $2^{m+1} - 2^m = 2^m$ Summanden, jeder von diesen ist mindestens $1/2^{m+1}$. Damit ist die Teilsumme mindestens

$$2^m \cdot \frac{1}{2^{m+1}} = \frac{1}{2}.$$

Für jedes $m \in \mathbb{N}_0$ haben wir so eine Teilsumme, die Reihe geht also gegen Unendlich.

Wir wollen nun hinreichende Kriterien angeben, und beginnen mit einem einfachen Fall.

Satz 2.2 (Reihen mit nichtnegativen Gliedern) Eine Reihe $\sum_{k=0}^{\infty} a_k$ mit $a_k \geq 0$ konvergiert genau dann, wenn die Partialsummen $S_n = \sum_{k=0}^n a_k$ nach oben beschränkt sind.

BEWEIS: S_n ist monoton wachsend, die Behauptung folgt aus Satz 2.5 (Kapitel 2). $\square$

Ein ziemlich effektives Kriterium ist der

Satz 2.3 (Integralvergleichstest) Sei $f : [1, \infty) \rightarrow [0, \infty)$ monoton fallend. Dann konvergiert die Reihe $\sum_{k=1}^{\infty} f(k)$ genau dann, wenn $\int_1^{\infty} f(x) dx$ konvergiert.

BEWEIS: Sei $k \geq 2$. Für $k-1 \leq x \leq k$ gilt $f(k) \leq f(x) \leq f(k-1)$. Durch Integration folgt

$$f(k) \leq \int_{k-1}^k f(x) dx \leq f(k-1).$$

Jetzt summieren wir über $k = 2, \dots, n$. Es folgt

$$\sum_{k=2}^n f(k) \leq \int_1^n f(x) dx \leq \sum_{k=1}^{n-1} f(k).$$

Konvergiert das Integral, so ist die linke Summe nach oben beschränkt und die Reihe konvergiert nach Satz 2.2. Ist andererseits die rechte Summe nach oben beschränkt, so ist das Integral als Funktion der oberen Grenze monoton wachsend und nach oben beschränkt, also konvergent. $\square$

Beispiel 2.5 Die Funktion $f(x) = x^\alpha$ ist nichtnegativ, und für $\alpha < 0$ monoton fallend auf $[1, \infty)$. Das Integral $\int_1^{\infty} x^\alpha dx$ konvergiert genau für $\alpha < -1$, siehe Beispiel 2.7. Somit konvergieren auch die Reihen $\sum_{k=1}^{\infty} k^\alpha$ genau für $\alpha < -1$.

Satz 2.4 Falls $\sum_{k=0}^{\infty} |a_k| < \infty$, so konvergiert auch die Reihe $\sum_{k=0}^{\infty} a_k$ und es gilt

$$\left| \sum_{k=0}^{\infty} a_k \right| \leq \sum_{k=0}^{\infty} |a_k|. \quad (2.7)$$

Bezeichnung. Man nennt eine Reihe $\sum_{k=0}^{\infty} a_k$ absolut konvergent, wenn $\sum_{k=0}^{\infty} |a_k| < \infty$.

BEWEIS: Sei $a_k^+ = \max(a_k, 0)$, $a_k^- = -\min(a_k, 0)$. Dann ist $a_k^{\pm} \geq 0$ und $a_k^+ + a_k^- = |a_k|$, also

$$\sum_{k=0}^n a_k^+ + \sum_{k=0}^n a_k^- = \sum_{k=0}^n |a_k| \leq \sum_{k=0}^{\infty} |a_k| < \infty.$$

Nach Satz 2.2 sind $\sum_{k=0}^{\infty} a_k^+$ und $\sum_{k=0}^{\infty} a_k^-$ konvergent. Aber $a_k = a_k^+ - a_k^-$, also ist $\sum_{k=0}^{\infty} a_k$ auch konvergent als Differenz von zwei konvergenten Reihen. Weiter folgt

$$\left| \sum_{k=0}^{\infty} a_k \right| = \lim_{n \rightarrow \infty} \left| \sum_{k=0}^n a_k \right| \leq \lim_{n \rightarrow \infty} \sum_{k=0}^n |a_k| = \sum_{k=0}^{\infty} |a_k|.$$

□

Satz 2.5 (Majorantentest) Es gelte $|a_k| \leq c_k$ wobei $\sum_{k=0}^{\infty} c_k < \infty$. Dann ist die Reihe $\sum_{k=0}^{\infty} a_k$ absolut konvergent.

BEWEIS: Die Summen $\sum_{k=0}^n |a_k|$ sind durch $\sum_{k=0}^{\infty} c_k < \infty$ beschränkt, also ist $\sum_{k=0}^{\infty} |a_k|$ konvergent nach Satz 2.2. □

Wie festgestellt ist die Reihe $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$ nicht konvergent, man kann aber zeigen dass die alternierende Reihe $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$ konvergiert. Aus absoluter Konvergenz folgt Konvergenz, aber nicht umgekehrt. Für unsere Zwecke ist absolute Konvergenz aber meistens passend. Es gibt eine Reihe weiterer Tests für absolute Konvergenz, zum Beispiel das Quotientenkriterium oder das Wurzelkriterium. Aber das ist eher ein Job für MathematikerInnen.

Wir haben bei der Taylorentwicklung bereits Reihen kennengelernt, die Grenzwerte von Polynomen sind. Wir wollen nun die Konvergenz solcher Reihen untersuchen. Der Einfachheit halber betrachten wir nur Reihen im Zentrum $x_0 = 0$.

Definition 2.2 (Potenzreihen) Eine reelle Potenzreihe ist eine Reihe der Form

$$P(x) = \sum_{k=0}^{\infty} a_k x^k \quad \text{mit } a_k \in \mathbb{R}.$$

Unser Ziel ist, mit der Reihe eine Funktion $P(x)$ zu definieren. Die Frage ist, für welche x die Reihe konvergiert. Für $x = 0$ sind alle Summanden mit $k \geq 1$ gleich Null, damit gilt $P(0) = a_0$. Es kann passieren dass die Reihe für kein $x \neq 0$ konvergiert. Ein Beispiel ist $\sum_{k=0}^{\infty} k^k x^k$ (verwende den Nullfolgentest), so eine Reihe ist nicht von Interesse. Sehr viel schöner ist es bei der Exponentialreihe $\sum_{k=0}^{\infty} x^k/k!$, die konvergiert sogar für jedes $x \in \mathbb{R}$. Die geometrische Reihe $\sum_{k=0}^{\infty} x^k$ konvergiert genau für $x \in (-1, 1)$. Allgemein kann die Menge der x , für die eine Potenzreihe konvergiert und damit eine Funktion definiert, wie folgt charakterisiert werden.

Satz 2.6 (Konvergenzradius) Zu jeder Potenzreihe $P(x) = \sum_{k=0}^{\infty} a_k x^k$ gibt es genau ein $R \in [0, \infty]$, den Konvergenzradius, mit folgender Eigenschaft:

$$P(x) \text{ ist } \begin{cases} \text{absolut konvergent} & \text{für } |x| < R, \\ \text{nicht konvergent} & \text{für } |x| > R. \end{cases}$$

Bemerkung. Für $|x| = R$ ist keine allgemeine Aussage möglich.

BEWEIS: Aus absoluter Konvergenz folgt Konvergenz nach Satz 2.4, es kann also nur eine der Alternativen eintreten. Daraus folgt die Eindeutigkeit von R . Zur Existenz definieren wir

$$R = \sup\{|x| : P(x) \text{ konvergiert}\} \in [0, \infty].$$

Für $|x| > R$ ist offensichtlich $P(x)$ nicht konvergent. Sei nun $|x| < R$. Dann gibt es ein x_0 mit $|x_0| > |x|$, so dass $P(x_0)$ konvergiert. Nach Satz 2.1 geht $|a_k||x_0|^k$ gegen Null und ist damit beschränkt, also $|a_k||x_0|^k \leq M$ für alle k . Es folgt

$$|a_k||x|^k = |a_k||x_0|^k \left(\frac{|x|}{|x_0|}\right)^k \leq M \left(\frac{|x|}{|x_0|}\right)^k. \quad (2.8)$$

Wir summieren über k . Rechts steht die geometrische Reihe mit Faktor $|x|/|x_0| < 1$, diese ist konvergent. Nach dem Majorantenkriterium ist die gegebene Reihe absolut konvergent. $\square$

Auf dem Intervall $(-R, R)$ ist die Funktion $P(x) = \sum_{k=0}^{\infty} a_k x^k$ nun definiert. Es stellt sich die Frage, ob $P(x)$ differenzierbar ist und wenn ja, ob wir $P'(x)$ durch gliedweise Differentiation berechnen können: gilt

$$\frac{d}{dx} \sum_{k=0}^{\infty} a_k x^k \stackrel{?}{=} \sum_{k=0}^{\infty} \frac{d}{dx} a_k x^k = \sum_{k=1}^{\infty} k a_k x^{k-1}.$$

Dieser Frage sind wir schon bei der Exponentialfunktion begegnet. Allgemeiner haben wir es mit dem Problem der Vertauschung von Grenzprozessen zu tun, und das ist nicht so einfach.

Seien $f_n : I \rightarrow \mathbb{R}$ stetige Funktionen, mit $f_n(x) \rightarrow f(x)$ für alle $x \in I$.

- (a) Ist dann die Funktion $f : I \rightarrow \mathbb{R}$ stetig?
- (b) Wenn ja, gilt dann $\int_I f(x) dx = \lim_{n \rightarrow \infty} \int_I f_n(x) dx$?
- (c) Falls $f'_n(x) \rightarrow g(x)$ für alle $x \in I$, ist dann $f(x)$ differenzierbar mit $f'(x) = g(x)$?

Leider heißt die Antwort erstmal *nein*, auf alle drei Fragen. Wir schauen mal Beispiele dazu an, nur um zu sehen was so schiefgehen kann.

Beispiel 2.6 Die Funktionen $f_n(x) = \arctan nx$ konvergieren für $n \rightarrow \infty$ gegen

$$f(x) = \begin{cases} \frac{\pi}{2} & \text{für } x > 0, \\ -\frac{\pi}{2} & \text{für } x < 0, \\ 0 & \text{für } x = 0. \end{cases}$$

Die Grenzfunktion ist nicht stetig im Punkt $x = 0$.

Beispiel 2.7 Betrachte die Funktionen

$$f_n(x) = \begin{cases} n & \text{für } 0 < x < \frac{1}{n}, \\ 0 & \text{sonst.} \end{cases}$$

Es gilt $f_n(x) \rightarrow 0$ für alle $x \in \mathbb{R}$, die Grenzfunktion ist die Nullfunktion. Aber

$$\lim_{n \rightarrow \infty} \int_0^1 f_n(x) dx = \lim_{n \rightarrow \infty} \frac{1}{n} \cdot n = 1 \neq 0.$$

Beispiel 2.8 Betrachte $f_n(x) = \sqrt{\frac{1}{n} + x^2}$, also $f_n(x) \rightarrow f(x) = |x|$ für $n \rightarrow \infty$. Es gilt

$$f'_n(x) = \frac{x}{\sqrt{\frac{1}{n} + x^2}} \quad \text{also } f'_n(x) \rightarrow \begin{cases} 1 & \text{für } x > 0, \\ -1 & \text{für } x < 0, \\ 0 & \text{für } x = 0, \end{cases}$$

insbesondere $\lim_{n \rightarrow \infty} f'_n(0) = 0$. Aber es gilt nicht $f'(0) = 0$, tatsächlich ist $f(x)$ gar nicht differenzierbar im Punkt $x = 0$.

Wir haben bisher punktweise Konvergenz betrachtet, also $f_n(x) \rightarrow f(x)$ bzw. $f'_n(x) \rightarrow g(x)$ für jedes einzelne $x \in I$. Es stellt sich heraus dass sich die Probleme beheben lassen, wenn die Funktionen besser konvergieren. Die Funktionen $f_n : I \rightarrow \mathbb{R}$ konvergieren gleichmäßig (oder bezüglich der Supremumsnorm) gegen $f : I \rightarrow \mathbb{R}$ wenn gilt

$$\|f_n - f\|_I = \sup_{x \in I} |f_n(x) - f(x)| \rightarrow 0 \quad \text{mit } n \rightarrow \infty. \quad (2.9)$$

Für jedes $x \in I$ folgt $f_n(x) \rightarrow f(x)$, das heißt aus gleichmässiger Konvergenz folgt punktweise Konvergenz. Aber nicht umgekehrt: bei gleichmäßiger Konvergenz muss die *maximale* Abweichung von $f_n(x)$ zu $f(x)$ gegen Null gehen. In Beispiel 2.7 gilt $f_n(x) \rightarrow 0$, aber die maximale Abweichung ist $\sup |f_n(x)| = n$.

Satz 2.7 (Gleichmäßige Konvergenz und Stetigkeit) Seien $f_n : I \rightarrow \mathbb{R}$ stetig. Falls die f_n gleichmäßig gegen $f : I \rightarrow \mathbb{R}$ konvergieren, so ist f ebenfalls stetig auf I .

BEWEIS: Sei $x_k \rightarrow x_0 \in I$ gegeben, zu zeigen ist $f(x_k) \rightarrow f(x_0)$. Wir schätzen ab:

$$\begin{aligned} |f(x_k) - f(x_0)| &\leq |f(x_k) - f_n(x_k)| + |f_n(x_k) - f_n(x_0)| + |f_n(x_0) - f(x_0)| \\ &\leq \|f_n - f\|_I + 2\|f_n - f\|_I. \end{aligned}$$

Zu $\varepsilon > 0$ wähle n mit $\|f_n - f\|_I < \frac{\varepsilon}{3}$. Da f_n stetig gilt $|f_n(x_k) - f_n(x_0)| < \varepsilon/3$ für k groß. $\square$

Satz 2.8 (Vertauschung von Konvergenz und Integral) Seien $f_n : [a, b] \rightarrow \mathbb{R}$ stetig. Falls die f_n gleichmäßig gegen $f : [a, b] \rightarrow \mathbb{R}$ konvergieren, so ist f stetig und

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x) dx.$$

BEWEIS: Die Stetigkeit von f gilt nach dem Satz davor. Die Standardabschätzung des Integrals, siehe Satz 1.3, liefert

$$\left| \int_a^b f_n(x) dx - \int_a^b f(x) dx \right| = \left| \int_a^b (f_n(x) - f(x)) dx \right| \leq \|f_n - f\|_I (b - a) \rightarrow 0.$$

□

Satz 2.9 (Vertauschung von Konvergenz und Ableitung) *Seien $f_n : I \rightarrow \mathbb{R}$ stetig differenzierbar, und $f_n \rightarrow f$ punktweise auf I . Falls die Folge f'_n gleichmäßig gegen eine Funktion g konvergiert, so ist f stetig differenzierbar mit $f' = g$.*

BEWEIS: Für $x_0 \in I$ gilt nach Satz 2.2, dem Hauptsatz,

$$f_n(x) = f_n(x_0) + \int_{x_0}^x f'_n(\xi) d\xi \quad \text{für alle } x \in I.$$

Da f'_n gleichmäßig gegen g konvergiert, folgt nach Satz 2.8 mit $n \rightarrow \infty$

$$f(x) = f(x_0) + \int_{x_0}^x g(\xi) d\xi \quad \text{für alle } x \in I.$$

Die Funktion g ist stetig nach Satz 2.7, es folgt wieder mit dem Hauptsatz $f' = g$. □

Wir kommen jetzt auf die Potenzreihen zurück. Die finale Aussage ist wie folgt.

Satz 2.10 (Differenzierbarkeit von Potenzreihen) *Sei $P(x) = \sum_{k=0}^{\infty} a_k x^k$ eine Potenzreihe mit Konvergenzradius $R > 0$. Dann ist $P(x)$ auf dem Intervall $(-R, R)$ differenzierbar, und die Ableitung kann gliedweise berechnet werden, das heißt*

$$P'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1} \quad \text{für alle } x \in (-R, R). \quad (2.10)$$

Dies folgt letztlich aus der Vertauschbarkeit von Ableitung und Grenzwert der Reihe, siehe Satz 2.9. Allerdings sind dazu ein paar Schritte notwendig:

- zuerst wird gezeigt, dass eine Potenzreihe mit Konvergenzradius $R > 0$ lokal (also auf Teilintervallen $[-r, r] \subset (-R, R)$) gleichmäßig konvergiert. Dazu verwendet man den Vergleich mit der geometrischen Reihe wie in Satz 2.6.
- Dann zeigt man, dass die gliedweise differenzierte Reihe $Q(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$ denselben Konvergenzradius R hat. Das geht wieder mit einem Vergleich wie in Satz 2.6.
- Aus den ersten beiden Schritten folgt, dass die Reihe $Q(x)$ lokal gleichmäßig konvergiert. Damit lässt sich Satz 2.9 anwenden.

Wir verzichten auf die Details, und machen stattdessen eine Anwendung.

Beispiel 2.9 Für $x \in (-1, 1) \subset \mathbb{R}$ gilt die Reihendarstellung

$$\log(1+x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1} = x - \frac{x^2}{2} + \frac{x^3}{3} - + \dots \quad (2.11)$$

Und zwar folgt mit der Formel für die geometrische Reihe

$$\log(1+x)' = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k \quad \text{für } x \in (-1, 1).$$

Die geometrische Reihe konvergiert gleichmäßig auf dem Intervall $[0, x]$, also können wir Satz 2.8 anwenden und das Integral mit der Reihe vertauschen:

$$\log(1+x) = \int_0^x \frac{dy}{1+y} = \int_0^x \sum_{k=0}^{\infty} (-1)^k y^k dy = \sum_{k=0}^{\infty} (-1)^k \int_0^x y^k dy = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1}.$$

3 Schwingung mit Dämpfung

Bei der Herleitung der harmonischen Schwingung waren wir von der Gleichung $mu'' + ku = 0$ ausgegangen. Der rechte Term stellt die elastische Kraft dar, der linke ergibt sich aus dem dynamischen Grundgesetz. Hier betrachten wir die Verallgemeinerung

$$mu'' + ru' + ku = 0. \quad (3.12)$$

Der mittlere Term stellt eine Reibungskraft dar, wir nehmen hier an dass $m, r, k > 0$. Man kann sich vorstellen, dass zum Beispiel eine elastische Feder durch Lagerung in Öl gedämpft wird oder so. Wir betrachten zunächst den Fall dass keine äußere Kraft wirkt, die rechte Seite der Gleichung ist also Null. Die Gleichung (3.12) ist linear, das heißt für zwei Lösungen $u(t)$ und $v(t)$ sind alle Funktionen $\lambda u(t) + \mu v(t)$ für $\lambda, \mu \in \mathbb{R}$ wieder Lösungen.

Eine zentrale Tatsache bei der harmonischen Schwingung war die Eulersche Formel, wonach die reellen Lösungen $\cos t$ und $\sin t$ die komplexe Lösung e^{it} liefern und umgekehrt. Diese Idee spielt auch hier eine Rolle: eine komplexe Lösung der Gleichung ist eine Funktion $w(t) = u(t) + iv(t)$ mit

$$mw'' + rw' + kw = 0.$$

Dabei werden die Ableitungen komponentenweise gebildet, also $w' = u' + iv'$ und $w'' = u'' + iv''$. Die Gleichung ist somit gleichbedeutend mit dem System

$$mu'' + ru' + ku = 0, \quad (3.13)$$

$$mv'' + rv' + kv = 0. \quad (3.14)$$

Aus der komplexen Lösung $w(t) = u(t) + iv(t)$ erhalten wir die beiden reellen Lösungen

$$u(t) = \operatorname{Re} w(t) \quad \text{und} \quad v(t) = \operatorname{Im} w(t).$$

Wir suchen nun komplexe Lösungen von (3.12) mit dem Ansatz

$$w(t) = e^{\lambda t} \quad \text{mit } \lambda = \alpha + i\beta \in \mathbb{C}.$$

Um die Ableitung zu berechnen, schreibe $w(t) = e^{\alpha t} e^{i\beta t}$ und verwende die Produktregel. Damit folgt $w'(t) = \lambda e^{\lambda t}$ und $w''(t) = \lambda^2 e^{\lambda t}$, es ergibt sich

$$0 = e^{\lambda t} (m\lambda^2 + r\lambda + k).$$

Wegen $|e^{\lambda t}| = e^{\alpha t} > 0$ muss λ Lösung der folgenden quadratischen Gleichung sein:

$$\lambda^2 + \frac{r}{m}\lambda + \frac{k}{m} = 0. \quad (3.15)$$

Durch Rückverfolgen der Rechnung sieht man: $e^{\lambda t}$ ist eine komplexe Lösung wenn λ die Gleichung (3.15) erfüllt. Die Diskriminante der Gleichung ist

$$\left(\frac{r}{2m}\right)^2 - \frac{k}{m} = \frac{1}{4m^2}(r^2 - 4mk).$$

Je nach Vorzeichen ergeben sich drei Fälle:

Kriechfall $r^2 - 4mk > 0$ Dann hat (3.15) die beiden reellen Lösungen

$$\lambda_{1,2} = \frac{1}{2m}(-r \pm \sqrt{r^2 - 4km}) < 0.$$

Beide Funktionen $u_{1,2}(t) = e^{\lambda_{1,2}t}$ sind fallend und konvergieren für $t \searrow \infty$ gegen Null.

aperiodischer Grenzfall $r^2 - 4mk = 0$.

Dann hat (3.15) die eine reelle Lösung $\lambda = -\frac{r}{2m}$, die Funktion $u_1(t) = e^{\lambda t}$ ist wieder fallend und konvergiert gegen Null. Es gibt aber die weitere Lösung $u_2(t) = tu_1(t)$, denn mit $u_2' = tu_1' + u_1$ und $u_2'' = tu_1'' + 2u_1'$ folgt

$$u_2'' + \frac{r}{m}u_2' + \frac{k}{m}u_2 = t\left(u_1'' + \frac{r}{m}u_1' + \frac{k}{m}u_1\right) + 2u_1' + \frac{r}{m}u_1.$$

Die Klammer ist bekanntlich Null, und weiter gilt

$$2u_1' + \frac{r}{m}u_1 = e^{\lambda t}\left(2\lambda + \frac{r}{m}\right) = 0.$$

Man kann $u_2(t)$ auch herleiten durch einen Grenzübergang von Fall 1, wo die beiden Eigenwerte zusammen kommen. Es gilt

$$u_2'(t) = e^{-\frac{r}{2m}t}\left(-\frac{r}{2m}t + 1\right),$$

also ist diese Lösung wachsend für $t \leq 2m/r$ und fallend für $t \geq 2m/r$ mit Grenzwert Null.

Schwingfall $r^2 - 4mk < 0$.

Wir setzen $r^2 - 4mk = -4m^2\nu^2$ mit $\nu > 0$, dann hat (3.15) die komplexen Lösungen

$$\lambda_{1,2} = -\frac{r}{2m} \pm i\nu.$$

Die zugehörigen komplexen Lösungen sind

$$w_{1,2}(t) = e^{-\frac{r}{2m}t} e^{\pm i\nu t}.$$

Durch Übergang zu Real- und Imaginärteil ergeben sich die reellen Lösungen

$$u(t) = e^{-\frac{r}{2m}t} \cos \nu t \quad \text{und} \quad v(t) = e^{-\frac{r}{2m}t} \sin \nu t.$$

Die Lösungen beschreiben eine Schwingung mit Frequenz

$$\nu = \sqrt{\frac{k}{m} - \left(\frac{r}{2m}\right)^2}.$$

Die Amplitude ist $e^{-\frac{r}{2m}t}$, sie ist monoton fallend und geht gegen Null mit $t \nearrow \infty$.

Wir wollen jetzt zeigen, dass die Gleichung (3.12) zu gegebenen Anfangsdaten eindeutig lösbar ist. Dass eine Lösung zu gegebenen Anfangsdaten existiert folgt aus den Rechnungen oben, denn die jeweiligen Anfangswerte der Lösungen bilden eine Basis des $\mathbb{R}^2$:

$$\begin{aligned} \text{Kriechfall : } & \begin{pmatrix} u_1(0) & u_2(0) \\ u_1'(0) & u_2'(0) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \\ \text{Grenzfall : } & \begin{pmatrix} u_1(0) & u_2(0) \\ u_1'(0) & u_2'(0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \\ \text{Schwingfall : } & \begin{pmatrix} u(0) & v(0) \\ u'(0) & v'(0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{r}{2m} & \nu \end{pmatrix} \end{aligned}$$

Alle Anfangsbedingungen lassen sich durch Linearkombination realisieren.

Angenommen wir haben auf einem Intervall I mit $0 \in I$ zwei Lösungen $u_{1,2}(t)$ der Gleichung (3.12) mit denselben Anfangsdaten. Durch Subtraktion erhalten wir eine Lösung $u(t)$ mit Anfangsdaten Null, wir müssen zeigen dass $u(t)$ die Nullfunktion ist. Dazu multiplizieren wir die Gleichung mit $u'(t)$ und integrieren auf $[0, T]$. Es gilt

$$\begin{aligned} \int_0^T m u''(t) u'(t) dt &= \frac{m}{2} \int_0^T \frac{d}{dt} u'(t)^2 dt = \frac{m}{2} u'(T)^2, \\ \int_0^T k u(t) u'(t) dt &= \frac{k}{2} \int_0^T \frac{d}{dt} u(t)^2 dt = \frac{k}{2} u(T)^2. \end{aligned}$$

Hier haben wir $u(0) = 0$ und $u'(0) = 0$ benutzt. Es ergibt sich

$$\frac{m}{2} u'(T)^2 + \frac{k}{2} u(T)^2 + r \int_0^T u'(t)^2 dt = 0.$$

Unter den Annahmen $m, k, r > 0$ folgt $u(t) = 0$ auf $[0, T]$. Die Eindeutigkeit kann auch ohne diese Annahmen gezeigt werden, aber darauf verzichten wir hier.

Wir kommen nun zur inhomogenen Gleichung, das heißt wir betrachten sogenannte erzwungene Schwingungen. Uns interessiert erst der Fall einer periodischen rechten Seite $f(t) = ce^{i\omega t}$. Die komplexe Gleichung ist gleichbedeutend mit den beiden reellen Gleichungen zu $f_1(t) = c \cos \omega t$ und $f_2(t) = c \sin \omega t$. Die Vermutung ist nun dass eine Lösung existiert mit der gegebenen Periode ω , man stellt sich vor dass sich der Oszillator nach einem Einschwingvorgang auf die gegebene Frequenz einstellt. Wir machen also den Ansatz

$$w(t) = \sigma e^{i\omega t} \quad \text{mit } \sigma = \sigma(m, r, k, \omega).$$

Die Amplitude σ hängt also nur von den Daten ab, nicht von der Zeit. Mit dem Ansatz gilt

$$w'(t) = i\omega \sigma e^{i\omega t} \quad \text{und} \quad w''(t) = -\omega^2 \sigma e^{i\omega t}.$$

Die inhomogene Gleichung lautet also

$$-m\omega^2\sigma e^{i\omega t} + i r\omega\sigma e^{i\omega t} + k\sigma e^{i\omega t} = c e^{i\omega t}.$$

Kürzen von $e^{i\omega t}$ und Auflösen nach σ liefert

$$\sigma = \frac{c}{k - m\omega^2 + i r\omega} =: \frac{c}{z}.$$

Durch Zurückverfolgen sieht man, dass diese Formel eine Lösung der inhomogenen Gleichung liefert. Um das Ergebnis zu interpretieren erweitern wir mit $\bar{z}$ und schreiben

$$\sigma = c \frac{\bar{z}}{|z|^2} = c\alpha e^{-i\omega\delta}.$$

Dabei bezeichnet man $\alpha > 0$ als den Verzerrungsfaktor, es ergibt sich

$$\alpha = \frac{1}{|z|} = \frac{1}{\sqrt{(k - m\omega^2)^2 + r^2\omega^2}}.$$

Die Zahl $\omega\delta \in \mathbb{R}$ heißt Phasenverschiebung, sie ist gegeben durch

$$e^{i\omega\delta} = \frac{z}{|z|} = \alpha(k - m\omega^2 + i r\omega).$$

Lösen $w_1(t)$ und $w_2(t)$ die inhomogene Gleichung, so ist $w(t) = w_1(t) - w_2(t)$ Lösung der homogenen Gleichung. Somit lautet die allgemeine Lösung der inhomogenen Gleichung

$$w(t) = c\alpha e^{i\omega(t-\delta)} + w_0(t),$$

wobei $w_0(t)$ Lösung der homogenen Gleichung. Diese Lösungen wurden oben bestimmt.

Wir wollen jetzt die Funktion $\alpha(\omega)$ studieren, also die Abhängigkeit des Verzerrungsfaktors von der Frequenz. Es ist praktisch folgende charakteristischen Größen einzuführen:

$$\begin{aligned} \omega_0 &= \sqrt{\frac{k}{m}} && \text{Eigenfrequenz des ungedämpften Oszillators} \\ \nu &= \sqrt{\frac{k}{m} - \frac{r^2}{4m^2}} && \text{Eigenfrequenz des gedämpften Oszillators} \end{aligned}$$

Wir machen jetzt eine kurze Kurvendiskussion der Funktion $\alpha(\omega)$. Es gilt $\alpha(0) = \frac{1}{k}$, und für $\omega \nearrow \infty$

$$\alpha(\omega) = \frac{1}{m\omega^2} \frac{1}{\sqrt{1 + r^2/m^2\omega^2 - 2k/m\omega^2}} \searrow 0.$$

Weiter berechnen wir die Ableitung

$$\begin{aligned} 0 &= \frac{d}{d\omega} \alpha(\omega)^2 (m^2\omega^4 + (r^2 - 2mk)\omega^2 + k^2) \\ &= \frac{2\alpha'(\omega)}{\alpha(\omega)} + \alpha(\omega)^2 (4m^2\omega^3 + 2(r^2 - 2mk)\omega). \end{aligned}$$

Nach Umstellen haben wir

$$\alpha'(\omega) = \frac{1}{2}\alpha(\omega)^3 (4m^2\omega^3 + 2(r^2 - 2mk)\omega).$$

Somit ist $\alpha'(\omega) = 0$ genau für $\omega = 0$ und für

$$\omega = \sqrt{\frac{k}{m} - \frac{r^2}{2m^2}} = \sqrt{\omega_0^2 - \frac{r^2}{2m^2}} = \omega_1.$$

Genauer ist $\alpha'(\omega) < 0$ für $\omega > \omega_1$ und $\alpha'(\omega) > 0$ für $0 < \omega < \omega_1$, das heißt bei $\omega = \omega_1$ ist ein globales Maximum. Der Maximalwert ist

$$\alpha(\omega_1) = \frac{1}{\sqrt{\frac{k}{m} - \frac{r^2}{4m^2}}} \frac{1}{r}.$$

Für $r > 0$, also wenig Dämpfung, ist dieser Wert asymptotisch zu ω_0/r , insbesondere geht er für $r \searrow 0$ gegen Unendlich. Das beschreibt die sogenannte Resonanzkatastrophe.

4 Fourierapproximation

Bevor wir dieses Thema behandeln brauchen wir einige Dinge zur Linearen Algebra.

Kapitel 6

Lineare Algebra

1 Vektorräume

Ein Vektorraum (genauer $\mathbb{R}$ -Vektorraum) ist eine Menge V , auf der eine Addition und eine Skalarmultiplikation gegeben sind:

$$\begin{aligned} V \times V &\rightarrow V & (v, w) &\mapsto v + w && \text{(Vektoraddition)} \\ \mathbb{R} \times V &\rightarrow V & (\lambda, v) &\mapsto \lambda v && \text{(Skalarmultiplikation)} \end{aligned}$$

Dabei werden folgende Rechenregeln verlangt:

$(V, +)$ ist eine additive Gruppe:

$$\begin{aligned} (u + v) + w &= u + (v + w) && \text{(Assoziativgesetz)} \\ v + w &= w + v && \text{(Kommutativgesetz)} \\ v + 0 &= v \text{ für alle } v && \text{(Nullvektor)} \\ v + (-v) &= 0 && \text{(inverses Element)}. \end{aligned}$$

Weitere Rechenregeln für alle $v, w \in V$ und $\lambda, \mu \in \mathbb{R}$:

$$\begin{aligned} \lambda(\mu v) &= (\lambda\mu)v \\ 1 \cdot v &= v \\ \lambda(v + w) &= \lambda v + \lambda w \\ (\lambda + \mu)v &= \lambda v + \mu v. \end{aligned}$$

Aus dieser Liste lassen sich alle weiteren Rechenregeln herleiten. Zum Beispiel folgt, dass der Nullvektor $0 \in V$ eindeutig bestimmt ist. Das zentrale Beispiel für uns ist natürlich der $\mathbb{R}^n$ mit der komponentenweise Addition und Skalarmultiplikation:

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad \lambda x = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}, \quad \text{wobei } x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}.$$

Im $\mathbb{R}^n$ hatten wir die Regeln bereits in Kapitel 1, Abschnitt 4, formuliert.

Ein zweites wichtiges Beispiel ist der Vektorraum $\mathcal{F}(M)$ aller Funktionen $f : M \rightarrow \mathbb{R}$, dabei ist M irgendeine Menge. Die Addition und Skalarmultiplikation sind wie folgt:

$$\begin{aligned} f + g \in \mathcal{F}(M) & \quad (f + g)(x) = f(x) + g(x) \\ \lambda f \in \mathcal{F}(M) & \quad (\lambda f)(x) = \lambda f(x). \end{aligned}$$

Der Nullvektor ist die Nullfunktion, also $f(x) = 0$ für alle $x \in M$. Das zu f inverse Element ist die Funktion $-f$, also $(-f)(x) = -f(x)$. Es ist leicht die obigen Regeln zu verifizieren.

In Kapitel 1.4 hatten wir bereits das Konzept des Unterraums im $\mathbb{R}^n$ ein wenig diskutiert, speziell für Geraden oder Ebenen. Hier die allgemeine Definition.

Definition 1.1 (Untervektorraum) Sei V ein $\mathbb{R}$ -Vektorraum. Eine Menge $X \subset V$ heißt Untervektorraum, wenn $0 \in X$ und wenn folgende Regeln gelten:

$$\begin{aligned} x, y \in X & \Rightarrow x + y \in X \\ x \in X, \lambda \in \mathbb{R} & \Rightarrow \lambda x \in X. \end{aligned}$$

Bemerkung. X ist dann selbst ein $\mathbb{R}$ -Vektorraum, mit den Operationen $+$ und $\cdot$ von V .

Beispiel 1.1 Sei X die Menge aller Lösungen $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ der Gleichung¹

$$x_1 + 2x_2 - x_3 = 0.$$

Der Nullvektor gehört dazu, und mit x, y sind auch $x + y$ und λx Lösungen:

$$\begin{aligned} x_1 + y_1 + 2(x_2 + y_2) - (x_3 + y_3) &= x_1 + 2x_2 - x_3 + y_1 + 2y_2 - y_3 = 0, \\ \lambda x_1 + 2\lambda x_2 - \lambda x_3 &= \lambda(x_1 + 2x_2 - x_3) = 0. \end{aligned}$$

Also ist X ein Untervektorraum des $\mathbb{R}^3$. Dagegen ist die Lösungsmenge der Gleichung

$$x_1 + 2x_2 - x_3 = 3$$

kein Untervektorraum. Das ergibt sich schon daraus, dass der Nullvektor nicht dazugehört. Auch lösen $x + y$ sowie λx (außer für $\lambda = 1$) nicht die Gleichung.

Beispiel 1.2 Für jedes $v \in V$ ist die Menge

$$\mathbb{R} \cdot v = \{\lambda v : \lambda \in \mathbb{R}\}$$

ein Untervektorraum.

Beispiel 1.3 (Die Räume $C^k(I)$) Sei $I \subset \mathbb{R}$ ein Intervall. Dann ist der Raum der stetigen Funktionen $C^0(I)$ ein Untervektorraum des Raums $\mathcal{F}(I)$ aller Funktionen auf I . Denn die Nullfunktion ist stetig, und nach Satz 2.8 in Kapitel 2 gilt

$$\begin{aligned} f, g \text{ stetig auf } I & \Rightarrow f + g \text{ stetig auf } I \\ \lambda \in \mathbb{R}, f \text{ stetig auf } I & \Rightarrow \lambda f \text{ stetig auf } I. \end{aligned}$$

¹Zur Notation: aus Platzgründen schreiben wir Vektoren $x \in \mathbb{R}^n$ im Text als Zeilen, in eingerückten Formeln dagegen als Spalten. Es sollte keine Missverständnisse geben.

Nach Satz 1.2 in Kapitel 3 ist die Summe von differenzierbaren Funktionen wieder differenzierbar mit $(f + g)'(x) = f'(x) + g'(x)$, ebenso $(\lambda f)'(x) = \lambda f'(x)$. Es folgt, dass auch der Raum $C^1(I)$ der stetig differenzierbaren Funktionen ein Unterraum ist, sowie allgemeiner $C^k(I)$ für $k \in \mathbb{N}$ und $C^\infty(I)$. Wir haben also eine Sequenz von Unterräumen

$$C^\infty(I) \subset \dots \subset C^k(I) \subset \dots \subset C^1(I) \subset C^0(I) \subset \mathcal{F}(I).$$

Beispiel 1.4 (Lösungsraum einer linearen Differentialgleichung) Sei $I \subset \mathbb{R}$ ein Intervall, und es seien $a_0, a_1 \in C^0(I)$ gegeben. Betrachte

$$X = \{f \in C^2(I) : f''(t) + a_1(t)f'(t) + a_0(t)f(t) = 0 \text{ für alle } t \in I\}.$$

Lösen f und g die Differentialgleichung, so auch $f + g$ und λf . Daher ist X ein Untervektorraum von $C^2(I)$.

Beispiel 1.5 (Durchschnitt von Unterräumen) Seien $X_i, i \in I$, Unterräume des Vektorraums V . Dann ist der Durchschnitt der X_i , also

$$X = \bigcap_{i \in I} X_i \subset V$$

wieder ein Unterraum von V . Denn es ist $0 \in X_i$ für alle $i \in I$, und mit $x, y \in X_i$ für alle $i \in I$ folgt auch $x + y \in X_i$ sowie $\lambda x \in X_i$ für alle $i \in I$.

Im Gegensatz zum Durchschnitt $X \cap Y$ ist die Vereinigung $X \cup Y$ von zwei Unterräumen im allgemeinen kein Unterraum. Daher folgender Begriff.

Satz 1.1 (Summe von Unterräumen) Sei V ein $\mathbb{R}$ -Vektorraum, und $X, Y \subset V$ seien Unterräume. Dann ist

$$X + Y = \{x + y : x \in X, y \in Y\}$$

wieder ein Unterraum von V .

BEWEIS: X und Y enthalten beide den Nullvektor, also auch $X + Y$. Sind $z_1 = x_1 + y_1, z_2 = x_2 + y_2 \in X + Y$, so folgt

$$\begin{aligned} z_1 + z_2 &= (x_1 + x_2) + (y_1 + y_2) \in X + Y, \\ \lambda z_1 &= \lambda x_1 + \lambda y_1 \in X + Y. \end{aligned}$$

□

Definition 1.2 (direkte Summe) Zwei Unterräume $X, Y \subset V$ heißen komplementär, bzw. V heißt direkte Summe von X und Y (Notation: $V = X \oplus Y$), wenn gilt:

$$X + Y = V \quad \text{und} \quad X \cap Y = \{0\}.$$

Beispiel 1.6 Betrachte im $\mathbb{R}^2$ die Geraden

$$X = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{und} \quad Y = \mathbb{R} \cdot \begin{pmatrix} a \\ b \end{pmatrix}.$$

Wir behaupten $\mathbb{R}^2 = X \oplus Y$ wenn $b \neq 0$. Zuerst zeigen wir $X \cap Y = \{0\}$. Ist $v \in X \cap Y$ so gibt es $\lambda, \mu \in \mathbb{R}$ mit

$$v = \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mu \begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow \mu = 0 \Rightarrow v = 0.$$

Sei nun $v \in \mathbb{R}^2$ gegeben. Wir suchen λ, μ mit

$$v = \lambda \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} a \\ b \end{pmatrix} \Leftrightarrow \begin{array}{l} v_1 = \lambda + \mu a \\ v_2 = \mu b. \end{array}$$

Das gilt mit $\lambda = v_1 - \frac{a}{b}v_2$ und $\mu = \frac{1}{b}v_2$, damit gilt $\mathbb{R}^2 = X \oplus Y$.

Definition 1.3 (erzeugter Unterraum) Sei $\emptyset \neq M \subset V$. Der von M erzeugte Unterraum oder Span von M (Notation: $\text{Span } M$) ist die Menge aller $x \in V$ der Form

$$x = \lambda_1 x_1 + \dots + \lambda_k x_k \quad \text{wobei } x_i \in M, \lambda_i \in \mathbb{R}, k \in \mathbb{N}. \quad (1.1)$$

Eine (endliche) Summe wie in (1.1) heißt Linearkombination von $x_1, \dots, x_k$.

Wir wollen verifizieren dass $\text{Span } M$ ein Unterraum ist. Den Nullvektor kriegen wir mit $0 = 0 \cdot x \in \text{Span } M$ für irgendein $x \in M$. Seien $x, y \in \text{Span } M$. Wir können annehmen, dass x, y Linearkombinationen derselben Menge $x_1, \dots, x_k$ in M sind, sonst vereinigen wir die beiden Mengen und setzen die zusätzlichen Koeffizienten gleich Null. Sei also $x = \lambda_1 x_1 + \dots + \lambda_k x_k$ und $y = \mu_1 x_1 + \dots + \mu_k x_k$. Dann folgt

$$\begin{aligned} x + y &= (\lambda_1 + \mu_1)x_1 + \dots + (\lambda_k + \mu_k)x_k \in \text{Span } M, \\ \alpha x &= (\alpha\lambda_1)x_1 + \dots + (\alpha\lambda_k)x_k \in \text{Span } M. \end{aligned}$$

Damit sind die Unterraumkriterien gezeigt. Enthält ein Unterraum $X \subset V$ die Menge M , so enthält er schon ganz $\text{Span } M \subset X$. Somit ist $\text{Span } M$ der Durchschnitt aller Unterräume, die M enthalten. Verwenden wir diese Charakterisierung im Fall $M = \emptyset$, so folgt $\text{Span } \emptyset = \{0\}$ als sinnvolle Festsetzung.

Definition 1.4 (endlichdimensional) Eine Vektorraum V heißt endlichdimensional, wenn es eine endliche Menge $M \subset V$ gibt mit $V = \text{Span } M$. Andernfalls heißt V unendlichdimensional.

Beispiel 1.7 (Polynomraum) Eine Funktion $P : \mathbb{R} \rightarrow \mathbb{R}$ heißt Polynom vom Grad n , wenn es Koeffizienten $a_0, a_1, \dots, a_n \in \mathbb{R}$, $a_n \neq 0$, gibt mit

$$P(x) = a_0 + a_1 x + \dots + a_n x^n \quad \text{für alle } x \in \mathbb{R}.$$

Der Grad n und die $a_0, \dots, a_n$ sind eindeutig bestimmt. Denn gäbe es eine andere Darstellung mit ggf. anderem Grad m , so ergibt sich durch Abziehen eine Darstellung der Nullfunktion als Polynom vom Grad höchstens $\max(m, n)$. Nach Lemma 1.2 hat diese höchstens $\max(m, n)$ Nullstellen, Widerspruch.

Sind P, Q Polynome vom Grad höchstens n mit Koeffizienten a_i, b_i , so sind $P + Q$ und λP ebenfalls Polynome vom Grad höchstens n , mit Koeffizienten $a_i + b_i$ bzw. λa_i . Somit

ist die Menge $\mathcal{P}_n(\mathbb{R})$ aller Polynome vom Grad höchstens n ein Untervektorraum von $\mathcal{F}(\mathbb{R})$. Jedes Element von $\mathcal{P}_n(\mathbb{R})$ ist Linearkombination der Monome x^i mit $i = 0, \dots, n$, somit ist der Raum endlichdimensional. Der Raum aller reellen Polynome

$$\mathcal{P}(\mathbb{R}) = \bigcup_{n=0}^{\infty} \mathcal{P}_n(\mathbb{R})$$

ist ebenfalls ein Unterraum von $\mathcal{F}(\mathbb{R})$. Denn die Summe von zwei Polynomen vom Grad höchstens m bzw. n ist Polynom vom Grad höchstens $\max(m, n)$. Der Raum $\mathcal{P}(\mathbb{R})$ ist jedoch nicht endlichdimensional: angenommen er ist Span der Polynome $P_i(x) \in \mathcal{P}_{n_i}(\mathbb{R})$ mit $i = 1, \dots, r$. Ist $n > \max_{i=1, \dots, r} n_i$, so ist $P(x) = x^n$ nicht als Linearkombination der P_i darstellbar. Denn sonst gibt es $\lambda_i \in \mathbb{R}$ mit

$$x^n - \sum_{i=1}^r \lambda_i P_i(x) = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Das Polynom hat den Grad n , kann also nur n Nullstellen haben, Widerspruch.

Definition 1.5 (lineare Abhängigkeit) Die Vektoren $v_1, \dots, v_k \in V$ heißen linear abhängig, wenn es $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ gibt, die nicht alle Null sind und so dass

$$\lambda_1 v_1 + \dots + \lambda_k v_k = 0.$$

Andernfalls heißen $v_1, \dots, v_k$ linear unabhängig.

Bemerkung. Oft geht es darum nachzuweisen dass gegebene Vektoren $v_1, \dots, v_k$ linear unabhängig sind. Dazu muss folgende Implikation gezeigt werden:

$$\lambda_1 v_1 + \dots + \lambda_k v_k = 0 \quad \Rightarrow \quad \lambda_1 = \dots = \lambda_k = 0.$$

Beispiel 1.8 Sind die Vektoren $v_1, v_2, v_3 \in \mathbb{R}^3$ linear unabhängig?

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix}.$$

Dazu ist zu klären, ob die Gleichung $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0$ von Null verschiedene Lösungen $\lambda_1, \lambda_2, \lambda_3$ hat. Ausgeschrieben lautet die Gleichung

$$\begin{array}{rcl} \lambda_1 & + \lambda_3 & = 0 \\ 2\lambda_1 & + 2\lambda_2 & + 4\lambda_3 = 0 \\ 3\lambda_1 & + \lambda_2 & + 4\lambda_3 = 0 \end{array}$$

Das System hat die Lösung $(\lambda_1, \lambda_2, \lambda_3) = (1, 1, -1)$, also gilt $v_1 + v_2 - v_3 = 0$. Die Vektoren sind somit linear abhängig.

Unser nächstes Ziel ist die Definition der Dimension eines Vektorraums. Anschaulich ist das klar, eine Gerade hat Dimension 1, eine Ebene Dimension 2 etc. Eine allgemeine Definition ist aber gar nicht so einfach, wir brauchen den folgenden Austauschsatz von Steinitz ².

²Ernst Steinitz, 1871-1928

Satz 1.2 (Austauschsatz) V sei durch $v_1, \dots, v_m$ erzeugt, und $x_1, \dots, x_k \in V$ seien linear unabhängig. Dann gibt es $v_{j_1}, \dots, v_{j_k}$ so dass V auch erzeugt wird durch $x_1, \dots, x_k$ und die v_j mit $j \neq j_1, \dots, j_k$. Insbesondere gilt $k \leq m$.

BEWEIS: Induktion über k , wobei im Fall $k = 0$ nichts zu tun ist. Sei nun V erzeugt durch $x_1, \dots, x_{k-1}$ und die v_j mit $j \neq j_1, \dots, j_{k-1}$. Der Vektor x_k hat dann eine Darstellung

$$x_k = \sum_{i=1}^{k-1} \alpha_i x_i + \sum_{j \neq j_1, \dots, j_{k-1}} \lambda_j v_j.$$

Die λ_j sind nicht alle Null, sonst wären $x_1, \dots, x_k$ linear abhängig. Sei $\lambda_{j_k} \neq 0$, es folgt

$$v_{j_k} = \frac{1}{\lambda_{j_k}} \left(x_k - \sum_{i=1}^{k-1} \alpha_i x_i - \sum_{j \neq j_1, \dots, j_k} \lambda_j v_j \right). \quad (1.2)$$

Wir behaupten dass V durch $x_1, \dots, x_k$ und die v_j mit $j \neq j_1, \dots, j_k$ aufgespannt wird. Schreibe dazu $v \in V$ als Linearkombination von $x_1, \dots, x_{k-1}$ und v_j mit $j \neq j_1, \dots, j_{k-1}$. Falls v_{j_k} in der Darstellung vorkommt, ersetze es durch die rechte Seite von (1.2). Damit ist der Satz gezeigt. $\square$

Definition 1.6 (Basis) Sei V ein endlichdimensionaler Vektorraum. Eine nummerierte Menge $v_1, \dots, v_n \in V$ heißt Basis von V , wenn gilt

- (1) $\text{Span} \{v_1, \dots, v_n\} = V$
- (2) $v_1, \dots, v_n$ sind linear unabhängig.

Bemerkung. Änderung der Reihenfolge ergibt eine andere Basis.

Satz 1.3 (Existenz einer Basis und Dimension) Jeder endlichdimensionale Vektorraum V hat eine Basis, und je zwei Basen von V haben gleichviele Elemente. Diese Anzahl nennen wir die Dimension $\dim V$ von V .

BEWEIS: Da V endlichdimensional ist, gilt $V = \text{Span} \{v_1, \dots, v_m\}$ für ein $m \in \mathbb{N}$. Nach dem Austauschsatz hat eine linear unabhängige Menge höchstens m Elemente. Seien $x_1, \dots, x_k$ maximal viele linear unabhängige Vektoren, also $k \leq m$. Wir zeigen dass dies eine Basis ist. Für jedes $v \in V$ ist $v, x_1, \dots, x_k$ linear abhängig, das heißt es gibt $\lambda, \lambda_1, \dots, \lambda_k$ nicht alle Null mit

$$\lambda v + \lambda_1 x_1 + \dots + \lambda_k x_k = 0.$$

Es muss $\lambda \neq 0$ sein, sonst wären $x_1, \dots, x_k$ linear abhängig. Es folgt

$$v = -\frac{1}{\lambda} (\lambda_1 x_1 + \dots + \lambda_k x_k) \in \text{Span} \{x_1, \dots, x_k\}.$$

Also gilt $V = \text{Span} \{x_1, \dots, x_k\}$, und $x_1, \dots, x_k$ ist eine Basis. Sei nun $v_1, \dots, v_m$ auch eine Basis. Wegen Maximalität folgt $k \geq m$, und damit $k = m$. $\square$

Beispiel 1.9 Im $\mathbb{R}^n$ haben wir die Standardbasis

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \text{für } i = 1, \dots, n.$$

Dabei steht die 1 an der i -ten Stelle. Es ist klar dass $\mathbb{R}^n = \text{Span}\{e_1, \dots, e_n\}$, denn

$$x = \sum_{i=1}^n x_i e_i \quad \text{für jedes } x \in \mathbb{R}^n.$$

Andererseits sind die $e_1, \dots, e_n$ linear unabhängig, denn

$$0 = \sum_{i=1}^n \lambda_i e_i = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \Rightarrow \lambda_i = 0 \text{ für } i = 1, \dots, n.$$

Die Dimension von $\mathbb{R}^n$ ist somit gleich n , wie es sein sollte.

Beispiel 1.10 Sei $X = \{x = (x_1, x_2, x_3) \in \mathbb{R}^3 : 3x_1 + x_2 + x_3 = 0\}$. Wir behaupten, dass folgende Vektoren eine Basis von X bilden:

$$v = \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} \quad w = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}.$$

Zunächst gilt $v, w \in X$, und die Vektoren sind linear unabhängig:

$$0 = \lambda v + \mu w = \begin{pmatrix} -\lambda - \mu \\ 3\lambda \\ 3\mu \end{pmatrix} \Rightarrow \lambda = \mu = 0.$$

Weiter gilt für $x \in X$ die Darstellung

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda \begin{pmatrix} -1 \\ 3 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} \quad \text{mit } \lambda = \frac{x_2}{3} \text{ und } \mu = \frac{x_3}{3}.$$

Beachte dabei $x_1 = -(x_2 + x_3)/3$ wegen $x \in X$. Also ist v, w eine Basis.

Beispiel 1.11 Die Monome $P_i(x) = x^i$, $i = 0, \dots, n$, sind Basis des Raums $\mathcal{P}_n(\mathbb{R})$ der Polynome auf $\mathbb{R}$ vom Grad höchstens n . Denn zu $P \in \mathcal{P}_n(\mathbb{R})$ gibt es $a_0, \dots, a_n \in \mathbb{R}$ mit

$$P = \sum_{i=0}^n a_i P_i \Leftrightarrow P(x) = a_0 + a_1 x + \dots + a_n x^n \quad \text{für alle } x \in \mathbb{R}.$$

Außerdem sind die Monome linear unabhängig, denn sei

$$a_0 + a_1 x + \dots + a_n x^n = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Dann verschwinden alle Koeffizienten, sonst hätte das Polynom höchstens n Nullstellen.

Satz 1.4 (Dimension einer Summe) Für endlichdimensionale Unterräume X, Y eines Vektorraums V gilt die Formel

$$\dim(X + Y) = \dim X + \dim Y - \dim(X \cap Y).$$

BEWEIS: Seien $x_1, \dots, x_m$ und $y_1, \dots, y_n$ Basen von X bzw. Y . Wähle weiter eine Basis $v_1, \dots, v_k$ von $X \cap Y$. Nach dem Austauschsatz, Satz 1.2, können wir annehmen dass

$$x_i = y_i = v_i \quad \text{für } i = 1, \dots, k.$$

Wir zeigen, dass $v_1, \dots, v_k, x_{k+1}, \dots, x_m, y_{k+1}, \dots, y_n$ eine Basis von $X + Y$ ist. Daraus folgt

$$\dim(X + Y) = k + (m - k) + (n - k) = m + n - k = \dim X + \dim Y - \dim(X \cap Y).$$

Es ist klar, dass die Vektoren $X + Y$ aufspannen, zu zeigen ist die Unabhängigkeit. Sei

$$v + x + y = 0 \quad \text{wobei } v = \sum_{i=1}^k \alpha_i v_i, \quad x = \sum_{j=k+1}^m \lambda_j x_j, \quad y = \sum_{\ell=k+1}^n \mu_\ell y_\ell.$$

Es folgt $x = -v - y \in Y$, also ist $x \in X \cap Y$. Es gibt dann $\beta_1, \dots, \beta_k$ mit

$$\sum_{i=1}^k \beta_i v_i = x = \sum_{j=k+1}^m \lambda_j x_j.$$

Da $v_1, \dots, v_k, x_{k+1}, \dots, x_m$ Basis von X ist, folgt $\lambda_j = 0$ (und $\beta_i = 0$). Analog sehen wir $\mu_\ell = 0$ und schließlich $\alpha_i = 0$. Damit ist die lineare Unabhängigkeit bewiesen. $\square$

2 Lineare Abbildungen und Matrizen

Definition 2.1 (lineare Abbildung) Seien V, W $\mathbb{R}$ -Vektorräume. $A : V \rightarrow W$ heißt lineare Abbildung (Notation: $A \in L(V, W)$), wenn gilt:

$$\begin{aligned} A(x + y) &= A(x) + A(y) && \text{für alle } x, y \in V, \\ A(\lambda x) &= \lambda A(x) && \text{für alle } x \in V, \lambda \in \mathbb{R}. \end{aligned}$$

Bei linearen Abbildungen schreibt man oft Ax statt $A(x)$. Die Linearität bedeutet, dass A mit beliebigen Linearkombinationen vertauscht:

$$A(\lambda_1 x_1 + \dots + \lambda_k x_k) = \lambda_1 A x_1 + \dots + \lambda_k A x_k. \quad (2.3)$$

Beispiel 2.1 Die Abbildung $A : V \rightarrow W$, $Av = 0$ für alle $v \in V$, ist linear.

Beispiel 2.2 Die Abbildung $\text{Id}_V : V \rightarrow V$, $\text{Id}_V v = v$, ist linear.

Definition 2.2 Für eine lineare Abbildung $A : V \rightarrow W$ definieren wir:

- Der Kern von A ist $\ker A = \{v \in V : Av = 0\}$,
- Das Bild von A ist $\text{Bild } A = \{Av : v \in V\}$.

$\ker A$ ist ein Unterraum von V : für $x_1, x_2 \in \ker A$ und $\lambda_1, \lambda_2 \in \mathbb{R}$ gilt

$$A(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 A x_1 + \lambda_2 A x_2 = 0, \quad \text{also } \lambda_1 x_1 + \lambda_2 x_2 \in \ker A.$$

Bild A ist ein Unterraum von W : für $w_1, w_2 \in \text{Bild } A$ wähle $v_1, v_2 \in V$ mit $Av_1 = w_1$, $Av_2 = w_2$. Es folgt für $\lambda_1, \lambda_2 \in \mathbb{R}$

$$\lambda_1 w_1 + \lambda_2 w_2 = \lambda_1 Av_1 + \lambda_2 Av_2 = A(\lambda_1 v_1 + \lambda_2 v_2) \in \text{Bild } A.$$

Definition 2.3 (Projektion) $P : V \rightarrow V$ linear heißt *Projektion*, wenn $P \circ P = P$.

Satz 2.1 Ist $P : V \rightarrow V$ Projektion, so gilt $V = \text{Bild } P \oplus \ker P$ und $P|_{\text{Bild } P} = \text{Id}_{\text{Bild } P}$.

BEWEIS: Sei $v \in \text{Bild } P \cap \ker P$. Wähle $x \in V$ mit $Px = v$, dann folgt

$$0 = Pv = P(Px) = Px = v, \quad \text{also } \text{Bild } P \cap \ker P = \{0\}.$$

Sei nun $v \in V$ gegeben. Dann gilt $v = v - Pv + Pv \in \ker P + \text{Bild } P$, denn $P(v - Pv) = Pv - P(Pv) = Pv - Pv = 0$. Ist $w = Pv$ so folgt $Pw = P(Pv) = Pv = w$. $\square$

Beispiel 2.3 (Projektion) Sei $V = X \oplus Y$. Jedes $v \in V$ hat also eine Zerlegung $v = x + y$ mit $x \in X$ und $y \in Y$. Diese Zerlegung ist eindeutig, denn ist $x_1 + y_1 = x_2 + y_2$ mit $x_{1,2} \in X$ und $y_{1,2} \in Y$, so folgt

$$x_1 - x_2 = y_2 - y_1 \in X \cap Y = \{0\},$$

also $x_1 = x_2$ und $y_1 = y_2$. Wir erhalten damit eine Abbildung

$$P : V \rightarrow X, \quad Pv = x \quad \text{für } v = x + y.$$

Die Abbildung P ist linear, denn für $v_{1,2} = x_{1,2} + y_{1,2}$ ergibt sich

$$P(v_1 + v_2) = x_1 + x_2 = Pv_1 + Pv_2 \quad \text{und} \quad P(\lambda v_1) = \lambda x_1 = \lambda Pv_1.$$

Nach Definition ist $Px = x$ für $x \in X$ und $Pv = 0$ für $v \in Y$. Für $v = x + y$ folgt $Pv = x$ und $P(Pv) = Px = x = Pv$. Also ist P eine Projektion. Es gilt $Pv = 0$ genau wenn $x = 0$, also ist $\ker P = Y$. Weiter ist $\text{Bild } P = X$, wie man leicht sieht.

Lemma 2.1 Für eine lineare Abbildung $A : V \rightarrow W$ gilt:

$$\begin{aligned} A \text{ injektiv} &\Leftrightarrow \ker A = \{0\} \\ A \text{ surjektiv} &\Leftrightarrow \text{Bild } A = W. \end{aligned}$$

BEWEIS: Die zweite Aussage ist trivial. Für die erste Aussage beachte

$$Av_1 = Av_2 \Leftrightarrow A(v_1 - v_2) = 0 \Leftrightarrow v_1 - v_2 \in \ker A.$$

Ist $\ker A = \{0\}$, so folgt aus $Av_1 = Av_2$ auch $v_1 = v_2$, das heißt A ist injektiv. Ist umgekehrt A injektiv, so folgt $Av \neq A \cdot 0 = 0$ für $v \neq 0$, also $\ker A = \{0\}$. $\square$

Satz 2.2 Sei $A : X \rightarrow Y$ eine lineare Abbildung.

- (a) Ist A injektiv, also $\ker A = \{0\}$, und sind $x_1, \dots, x_k \in X$ linear unabhängig, so sind $Ax_1, \dots, Ax_k$ auch linear unabhängig.
- (b) Ist A surjektiv, also $\text{Bild } A = Y$, und wird X durch $x_1, \dots, x_k$ erzeugt, so wird Y durch $Ax_1, \dots, Ax_k$ erzeugt.
- (c) Ist A ein Isomorphismus, also bijektiv, und ist $x_1, \dots, x_k$ Basis von X , so ist $Ax_1, \dots, Ax_k$ Basis von Y . Insbesondere gilt $\dim X = \dim Y$.

BEWEIS: (a): Für $\lambda_i \in \mathbb{R}$ gilt

$$\sum_{i=1}^k \lambda_i Ax_i = A\left(\sum_{i=1}^k \lambda_i x_i\right).$$

Ist die linke Seite Null, so folgt $\sum_{i=1}^k \lambda_i x_i = 0$ wegen $\ker A = \{0\}$, und dann $\lambda_1, \dots, \lambda_k = 0$.

(b): Zu $y \in Y$ wähle $x \in X$ mit $Ax = y$. Es gibt $\lambda_i \in \mathbb{R}$ mit $x = \sum_{i=1}^k \lambda_i x_i$. Also

$$y = Ax = A\left(\sum_{i=1}^k \lambda_i x_i\right) = \sum_{i=1}^k \lambda_i Ax_i.$$

(c): Aus (a) und (b) folgt, dass $Ax_1, \dots, Ax_k$ wieder eine Basis ist, und weiter $\dim X = \dim Y$. $\square$

Satz 2.3 (Dimensionsformel) Sei $A \in L(V, W)$ mit $\dim V < \infty$. Dann gilt

$$\dim V = \dim \text{Bild } A + \dim \ker A.$$

BEWEIS: Die Abbildung $A : V \rightarrow \text{Bild } A$ ist surjektiv. Wird V durch $v_1, \dots, v_n$ erzeugt, so wird $\text{Bild } A$ nach Satz 2.2(b) durch $Av_1, \dots, Av_n$ erzeugt, insbesondere ist $\dim \text{Bild } A < \infty$. Sei nun $w_1, \dots, w_r$ eine Basis von $\text{Bild } A$. Wähle $x_1, \dots, x_r \in V$ mit $Ax_j = w_j$ für $j = 1, \dots, r$, und setze $X = \text{Span}\{x_1, \dots, x_r\}$. Die Abbildung $A|_X : X \rightarrow \text{Bild } A$ ist injektiv, denn aus $Ax = 0$ folgt mit $x = \sum_{j=1}^r \lambda_j x_j$

$$0 = Ax = \sum_{j=1}^r \lambda_j Ax_j = \sum_{j=1}^r \lambda_j w_j \quad \Rightarrow \quad \lambda_1, \dots, \lambda_r = 0.$$

Die Abbildung ist auch surjektiv: sei $w = \sum_{j=1}^r \lambda_j w_j \in \text{Bild } A$, dann ist

$$w = \sum_{j=1}^r \lambda_j Ax_j = Ax \quad \text{mit } x = \sum_{j=1}^r \lambda_j x_j \in X.$$

Zu $v \in V$ wähle nun $x \in X$ mit $Ax = Av$, es folgt $v = x + (v - x) \in X + \ker A$, wobei $X \cap \ker A = \{0\}$. Mit Satz 1.4 ergibt sich

$$\dim V = \dim X + \dim \ker A = \dim \text{Bild } A + \dim \ker A.$$

$\square$

Die Zahl $r = \dim \text{Bild } A$ wird als Rang von A bezeichnet.

Definition 2.4 Jedem $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ ordnen wir eine Matrix $(a_{ij}) \in \mathbb{R}^{m \times n}$ zu durch

$$Ae_j = \sum_{i=1}^m a_{ij}e_i \quad \text{für } j = 1, \dots, n. \quad (2.4)$$

Die Definition verwendet die Standardbasen des $\mathbb{R}^n$ und $\mathbb{R}^m$, bei Wahl anderer Basen $v_1, \dots, v_n$ und $w_1, \dots, w_m$ würde nicht dieselbe Matrix rauskommen.

Man nennt (a_{ij}) die Darstellungsmatrix von A . Jeder Bildvektor Ax kann mit (a_{ij}) ausgedrückt werden. Denn mit $x = \sum_{j=1}^n x_j e_j$ folgt $Ax = \sum_{j=1}^n x_j Ae_j$, und mit (2.4) folgt

$$Ax = \sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} e_i. \quad (2.5)$$

Die rechte Seite ist die Linearkombination der n Spalten von (a_{ij}) mit Faktoren x_j .

$$Ax = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Das lässt sich auch folgendermaßen schreiben:

$$Ax = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 a_{11} + \dots + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + \dots + x_n a_{mn} \end{pmatrix}. \quad (2.6)$$

Umgekehrt kann man jeder Matrix $a \in \mathbb{R}^{m \times n}$ eine lineare Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ zuordnen, und zwar setzt man wie in (2.5):

$$Ax = \sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} e_i.$$

Es ist leicht zu sehen dass damit A linear ist:

$$\begin{aligned} A(x+y) &= \sum_{j=1}^n \sum_{i=1}^m (x_j + y_j) a_{ij} e_i = \sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} e_i + \sum_{j=1}^n \sum_{i=1}^m y_j a_{ij} e_i = Ax + Ay, \\ A(\lambda x) &= \sum_{j=1}^n \sum_{i=1}^m \lambda x_j a_{ij} e_i = \lambda \sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} e_i = \lambda Ax. \end{aligned}$$

Hat $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ die Darstellungsmatrix (a_{ij}) , so gehört zu (a_{ij}) wieder die Abbildung A , denn

$$\sum_{j=1}^n \sum_{i=1}^m x_j a_{ij} e_i = \sum_{j=1}^n x_j Ae_j = Ax.$$

Umgekehrt: gehört zu $(a_{ij}) \in \mathbb{R}^{m \times n}$ die lineare Abbildung $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, so hat A die Matrix $(a_{ij}) \in \mathbb{R}^{m \times n}$. Denn es gilt

$$Ae_j = \sum_{i=1}^m a_{ij} e_i.$$

Die Zuordnungen zwischen $L(\mathbb{R}^n, \mathbb{R}^m)$ und $\mathbb{R}^{m \times n}$ sind also zueinander invers, insbesondere bijektiv. Wir werden im folgenden oft nicht unterscheiden zwischen der linearen Abbildung A und der Matrix $A = (a_{ij})$. Für die folgende Aussage erinnern wir an die Definition des Standardskalarprodukts im $\mathbb{R}^n$ (siehe Abschnitt 1.4):

$$\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n.$$

Satz 2.4 (Orthogonales Komplement) Für einen Unterraum $X \subset \mathbb{R}^n$ definiere

$$X^\perp = \{y \in \mathbb{R}^n : \langle x, y \rangle = 0 \text{ für alle } x \in X\}.$$

Dann ist $X^\perp$ auch ein Unterraum von $\mathbb{R}^n$ und es gilt $\mathbb{R}^n = X \oplus X^\perp$.

BEWEIS: $X^\perp$ ist ein Untervektorraum, denn sind $y_1, y_2 \in X^\perp$ und $\lambda_1, \lambda_2 \in \mathbb{R}$ so folgt

$$\langle x, \lambda_1 y_1 + \lambda_2 y_2 \rangle = \lambda_1 \langle x, y_1 \rangle + \lambda_2 \langle x, y_2 \rangle = 0 \quad \text{für alle } x \in X.$$

Ist $x \in X \cap X^\perp$ so folgt $0 = \langle x, x \rangle = |x|^2$, also $x = 0$. Sei nun $x_1, \dots, x_k$ eine Basis von X . Betrachte die lineare Abbildung

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^k, Fv = \sum_{i=1}^k \langle x_i, v \rangle e_i.$$

Es gilt $\ker F = X^\perp$, denn $\langle x_i, v \rangle = 0$ für $i = 1, \dots, k$ ist äquivalent zu $\langle x, v \rangle = 0$ für alle $x \in X$. Weiter ist $\dim \text{Bild } F \leq k = \dim X$, es folgt mit der Dimensionsformel

$$\dim X + \dim X^\perp \geq \dim \text{Bild } F + \dim \text{Ker } F = \dim \mathbb{R}^n \quad \text{bzw. } X \oplus X^\perp = \mathbb{R}^n.$$

□

Definition 2.5 (transponierte Matrix) Für $A \in \mathbb{R}^{m \times n}$ bezeichnet $A^T \in \mathbb{R}^{n \times m}$ die Matrix mit den Einträgen $a_{ji}^T = a_{ij}$ für $1 \leq j \leq n, 1 \leq i \leq m$.

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad A^T = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{pmatrix}$$

Lemma 2.2 Für $A \in \mathbb{R}^{m \times n}$ gilt, bezüglich der Standardskalarprodukte auf $\mathbb{R}^m$ bzw. $\mathbb{R}^n$,

$$\langle Av, w \rangle = \langle v, A^T w \rangle \quad \text{für alle } v \in \mathbb{R}^n, w \in \mathbb{R}^m.$$

BEWEIS: Es gilt $Av = \sum_{j=1}^n \sum_{i=1}^m v_j a_{ij} e_i$ und $A^T w = \sum_{i=1}^m \sum_{j=1}^n w_i a_{ij} e_j$. Somit gilt

$$\langle Av, w \rangle = \sum_{i=1}^m \sum_{j=1}^n v_j a_{ij} w_i = \sum_{j=1}^n \sum_{i=1}^m w_i a_{ij} v_j = \langle v, A^T w \rangle.$$

□

Das Bild von A ist der Span der Spaltenvektoren von A , darum nennt man $\dim \text{Bild } A$ den Spaltenrang von A . Das Bild von A^T ist der Span der Zeilenvektoren von A . Die Dimension von $\text{Bild } A^T$ heißt daher Zeilenrang von A . Es ist nicht direkt offensichtlich dass die beiden Zahlen gleich sind.

Satz 2.5 (Zeilenrang gleich Spaltenrang) Für $A \in \mathbb{R}^{m \times n}$ gilt

$$\dim \text{Bild } A^T = \dim \text{Bild } A.$$

BEWEIS: Wir berechnen für $w \in \mathbb{R}^m$

$$\begin{aligned} w \in \ker A^T &\Leftrightarrow \langle v, A^T w \rangle = 0 \quad \text{für alle } v \in \mathbb{R}^n \\ &\Leftrightarrow \langle Av, w \rangle = 0 \quad \text{für alle } v \in \mathbb{R}^n \\ &\Leftrightarrow w \in (\text{Bild } A)^\perp. \end{aligned}$$

Also ergibt sich mit der Dimensionsformel für A^T

$$\dim \text{Bild } A^T = m - \dim \ker A^T = m - \dim(\text{Bild } A)^\perp = \dim \text{Bild } A.$$

□