Grundlagen der Mathematik

Inhalt

Literatur

[1] O. Forster, "Analysis 1," Vieweg (2006).

[2] K. Königsberger, "Analysis 1," Springer (2004).

[3] G. Fischer, "Lineare Algebra," Vieweg (2005).

Teil I
Grundlagen

1 Mengen und Abbildungen

Für die Formulierung der Mathematik haben sich *Mengen* als zweckmäßig erwiesen. Endliche Mengen kann man durch eine vollständige Liste ihrer Elemente angeben, Schreibweise $X = \{x_1, x_2, \dots, x_n\}$. Die x_i, $i = 1, \dots, n$, heißen *Elemente* der Menge X, Schreibweise $x_i \in X$. Dabei ist die Reihenfolge beliebig, und das gleiche Element darf nicht mehrmals auftreten. Wenn x kein Element von X ist, schreiben wir $x \notin X$. Die *leere Menge* $\varnothing$ enthält kein Element. Eine Menge Y heißt *Teilmenge* einer Menge X, geschrieben $Y \subseteq X$, wenn für jedes Element $y \in Y$ gilt $y \in X$. Ist $Y \subseteq X$, dann ist das *Komplement* $X \setminus Y$ definiert als die Menge der $x \in X$ mit $x \notin Y$. Die *Vereinigung* $X_1 \cup X_2$ zweier Mengen X_1, X_2 ist die Menge jener Elemente, die in X_1 oder X_2 enthalten sind (nicht "entweder oder"). Wir schreiben dafür

$$X_1 \cup X_2 := \{x \ : \ x \in X_1 \text{ oder } x \in X_2\} \,.$$

Durch ":=" bezeichnen wir, daß der links stehende Ausdruck über den rechts stehenden definiert wird; es gibt auch "=:" im getauschten Sinn. Die Konstruktion auf der rechten Seite wird häufig benutzt: Links des ":" stehen allgemeine Elemente, die einer Einschränkung unterworfen werden, welche rechts von ":" steht. Viele Autoren bevorzugen "|" statt ":". Entsprechend wird der *Durchschnitt* zweier Mengen X_1, X_2 erklärt als

$$X_1 \cap X_2 := \{x \ : \ x \in X_1 \text{ und } x \in X_2\} \,.$$

Zwei Mengen mit $X_1 \cap X_2 = \varnothing$ heißen *disjunkt.* Das *direkte Produkt* von Mengen $X_1, \dots, X_n$ ist erklärt als die Menge aller geordneten n-Tupel:

$$X_1 \times X_2 \times \cdots \times X_n := \{(x_1, x_2, \dots, x_n) \ : \ x_i \in X_i\} \,.$$

Speziell schreiben wir $X^n := \underbrace{X \times \cdots \times X}_{n}$.

Viele der Konstruktionen aus endlichen Mengen lassen sich auch auf unendliche Mengen übertragen. Wir benötigen vor allem folgende Zahlbereiche: die Menge $\mathbb{N} = \{0, 1, 2 \dots, \}$ der *natürlichen Zahlen*, die Menge $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ der *ganzen Zahlen*, die Menge $\mathbb{Q}$ der *rationalen Zahlen* sowie die Menge $\mathbb{R}$ der reellen Zahlen. Wir werden später im Detail darauf eingehen und insbesondere auch die Menge $\mathbb{C}$ der komplexen Zahlen definieren.

Eine rationale Zahl denken wir uns als Bruch $r = \frac{p}{q}$ mit $p \in \mathbb{Z}$ und $q \in \mathbb{Z}^\times := \mathbb{Z} \setminus \{0\}$. Wir können die rationalen Zahlen jedoch *nicht* als Menge dieser Brüche auffassen, da wir z.B. $\frac{2}{3}$ und $\frac{4}{6}$ als dieselbe rationale Zahl ansehen. Die Identifikation von $\frac{2}{3}$ und $\frac{4}{6}$ ist ein häufig benutztes Verfahren:

Definition 1.1 Eine *Relation* auf einer Menge X ist eine Teilmenge $R \subseteq X \times X$. Für $(x, y) \in R \subseteq X \times X$ schreiben wir $x \sim y$.

Eine Relation $\sim$ auf X heißt *Äquivalenzrelation*, falls (mit $x, y, z \in X$)

(R1) (reflexiv): $x \sim x$ für jedes $x \in X$.

(R2) (symmetrisch): Für jedes Paar $x \sim y$ ist auch $y \sim x$.

(R3) (transitiv): Ist $x \sim y$ und $y \sim z$, so ist auch $x \sim z$.

Ist $\sim$ eine Äquivalenzrelation auf X, so heißt eine Teilmenge $A \subset X$ *Äquivalenzklasse*, falls gilt:

(Ä1) $A \neq \varnothing$.

(Ä2) Falls $x, y \in A$, so ist $x \sim y$.

(Ä3) Falls $x \in A$ und $y \in X$ mit $x \sim y$, so ist auch $y \in A$.

Wir werden uns überlegen:

Satz 1.2 *Ist $\sim$ eine Äquivalenzrelation auf X, so gilt für zwei beliebige Äquivalenzklassen $A_1, A_2 \subseteq X$ entweder $A_1 = A_2$ oder $A_1 \cap A_2 = \varnothing$.*

Insbesondere liegt jedes $x \in X$ in genau einer Äquivalenzklasse, die oft mit $[x]$ bezeichnet wird. Eine Äquivalenzrelation auf X liefert damit eine Zerlegung von X in disjunkte Teilmengen. Die Menge dieser Äquivalenzklassen heißt *Quotientenmenge* und wird mit $X/\sim$ bezeichnet. Ein konkretes Element einer Äquivalenzklasse heißt *Repräsentant*. Insbesondere ist $x \in [x]$.

Beispiel 1.3 Es sei $X = \mathbb{Z} \times \mathbb{Z}^\times$. Wir vereinbaren auf X die Relation $(p, q) \sim (p', q')$, falls $pq' = p'q$. Man überprüft, daß $\sim$ eine Äquivalenzrelation ist. Die zugehörige Quotientenmenge $\mathbb{Q} := (\mathbb{Z} \times \mathbb{Z}^\times)/\sim$ heißt *Menge der rationalen Zahlen.* ◁

Mathematische Sätze wie Satz 1.2 müssen *bewiesen* werden, d.h. es ist zu zeigen, daß die darin enthaltene Aussage *wahr* ist. Eine mathematische Aussage ist entweder wahr oder falsch, niemals beides. Ein solcher Beweis besteht im Zurückführen der zu zeigenden Aussage auf als wahr vorausgesetzte *Axiome* (hier Definition 1.1) oder auf bereits bewiesene (also als wahr erkannte) Aussagen. Diese Argumentationskette wird mit $\Rightarrow$ symbolisiert.

Beweis von Satz 1.2. Gegeben seien zwei Äquivalenzklassen $A_1, A_2 \subseteq X$ bezüglich einer Äquivalenzrelation $\sim$. Im Fall $A_1 \cap A_2 = \varnothing$ ist nichts weiter zu zeigen. Sei also $A_1 \cap A_2 \neq \varnothing$. Damit gibt es ein $a \in X$ mit $a \in A_1$ und $a \in A_2$. Wir zeigen $A_2 \subseteq A_1$. Sei dazu $b \in A_2$ beliebig. Wegen (Ä2) in der Äquivalenzklasse A_2 gilt $b \sim a$ als Relation in X. Dann sagt aber (Ä3) bezüglich A_1, daß $b \in A_1$ liegt. Somit ist $A_2 \subseteq A_1$ gezeigt. Durch Vertauschen der Rollen von A_1 und A_2 ergibt sich $A_1 \subseteq A_2$ und deshalb $A_1 = A_2$. □

Meist sind in Beweisen mehrere Aussagen zu verknüpfen. Solche Verknüpfungen lassen sich formalisieren. Der Vollständigkeit führen wir ein[1]:

Definition 1.4 Es seien A, B Aussagen. Dann werden Verknüpfungen $A \wedge B$ (*und*), $A \vee B$ (*oder*), $\neg A$ (*nicht*), $A \Rightarrow B$ (*folgt*) und $A \Leftrightarrow B$ (*äquivalent*) definiert durch die *Wahrheitstafel*

A	B	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$	$\neg A$
w	w	w	w	w	w	f
w	f	f	w	f	f	f
f	w	f	w	w	f	w
f	f	f	f	w	w	w

Wir werden $\wedge$ und $\vee$ entweder ausformulieren oder durch die folgenden *Quantoren* ersetzen, die sich bei der Arbeit mit unendlichen Mengen als unverzichtbar erweisen:

Definition 1.5 (für alle; es gibt) Es sei I eine beliebige Indexmenge und $\{A_i \ : \ i \in I\}$ ein System von Aussagen.

i) Die Aussage ($\forall i \in I$ gilt A_i) ist wahr genau dann, wenn alle Aussagen A_i wahr sind.

ii) Die Aussage ($\exists i \in I$ mit A_i) ist wahr genau dann, wenn mindestens eine der Aussagen A_i wahr ist.

Damit ist $\forall$ die Verallgemeinerung von $\wedge$ und $\exists$ die Verallgemeinerung von $\vee$. Neben dieser logischen Bedeutung werden wir $\exists$ und $\forall$ im wörtlichen Sinn nutzen.

Einige Folgerungen aus der Definition sind:

- $(A \wedge B) \Rightarrow A \vee B$
- $(A \Leftrightarrow B) \Rightarrow (A \Rightarrow B)$
- $(A \Leftrightarrow B) \Leftrightarrow ((A \wedge B) \vee (\neg A \wedge \neg B))$
- $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$
- $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$
- $(\neg A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow A)$

Die letzte Eigenschaft ist die Grundlage des *indirekten Beweises*: Um B zu beweisen, suchen wir eine falsche Aussage A und zeigen, daß A aus der Verneinung von B folgt, also $\neg B \Rightarrow A$.

[1] Wir werden diese Sprache aber vermeiden und uns eher per Durchführung genügend vieler konkreter Beweise erarbeiten, was ein korrekter Beweis ist.

Beispiel 1.6 (Irrationalität von $\sqrt{2}$) Es sei B die Aussage: $\sqrt{2}$ ist eine irrationale Zahl. Wir gehen von der Negation $\neg B$ aus: Angenommen, $\sqrt{2}$ wäre eine rationale Zahl, also $\sqrt{2} = \frac{p}{q}$ mit $p, q \in \mathbb{Z}$ und $q \neq 0$. Durch Kürzen kann erreicht werden, daß p oder q ungerade ist. Es gilt $2 = (\sqrt{2})^2 = \frac{p^2}{q^2}$, also $p^2 = 2q^2$. Damit ist p^2 und dann auch p gerade. Somit ist nach Voraussetzung q ungerade. Da p gerade ist, gibt es ein $r \in \mathbb{Z}$ mit $p = 2r$. Nun ergibt sich $2 = \frac{4r^2}{q^2}$, also $q^2 = 2r^2$, d.h. q^2 und damit auch q ist auch gerade! Wir haben also gezeigt:

$$\underbrace{(\sqrt{2} \in \mathbb{Q})}_{\neg B} \Rightarrow \underbrace{(\exists q \in \mathbb{Z}, q \neq 0, \text{ mit } q \text{ ist gerade und ungerade })}_{A} \quad \text{ist wahr.}$$

Da A falsch ist, kann $\neg B \Rightarrow A$ nur dann wahr sein, wenn $\neg B$ falsch ist, also B wahr, d.h. $\sqrt{2}$ ist irrational. □

Wichtig in indirekten Beweisen ist die korrekte Verneinung der Aussage. Dabei gelten:

- $\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$
- $\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$
- $\neg(\forall i \in I \text{ gilt } A_i) \Leftrightarrow \exists i \in I \text{ mit } \neg A_i$
- $\neg(\exists i \in I \text{ mit } A_i) \Leftrightarrow \forall i \in I \text{ gilt } \neg A_i$

Beispiel 1.7 Es sei $(a_n)_{n\in\mathbb{N}}$ eine Folge reeller Zahlen. Die Folge $(a_n)_{n\in\mathbb{N}}$ heißt *konvergent*, wenn gilt:

Es gibt ein $a \in \mathbb{R}$, so daß für alle $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert mit $|a_n - a| < \epsilon$ für alle $n \geq N$.

Die Verneinung ist nicht: “Es gibt kein $a \in \mathbb{R} \ldots$”, sondern:

Die Folge $(a_n)_{n\in\mathbb{N}}$ ist nicht konvergent, wenn für alle $a \in \mathbb{R}$ gilt: Es gibt ein $\epsilon > 0$, so daß für alle $N \in \mathbb{N}$ ein $n \geq N$ existiert mit $|a_n - a| \geq \epsilon$. ◁

Definition 1.8 Seien X, Y Mengen, dann versteht man unter einer *Abbildung von X nach Y* eine Vorschrift f, die jedem $x \in X$ eindeutig ein $f(x) \in Y$ zuordnet. Wir schreiben $f : X \to Y$ für die Abbildung zwischen Mengen und $f : x \mapsto f(x)$ für die Zuordnung der Elemente. Die Menge

$$G(f) := \{(x, f(x)) \ : \ x \in X\} \subseteq X \times Y$$

heißt der *Graph* von f. Mit $\mathrm{Abb}(X, Y)$ werde die Menge der Abbildungen von X nach Y bezeichnet.

Beispiel 1.9 Für $X = Y = \mathbb{R}$ betrachten wir die Abbildung $f : x \mapsto x^2$. Der Graph $\{(x, x^2) \ : \ x \in \mathbb{R}\}$ beschreibt im kartesischen Koordinatensystem eine Parabel. ◁

Definition 1.10 Sei $f : X \to Y$ eine Abbildung zwischen Mengen und seien $M \subseteq X$ und $N \subseteq Y$ Teilmengen. Dann heißt

$$f(M) := \{y \in Y \ : \ \exists x \in M \text{ mit } y = f(x)\} \subseteq Y$$

das *Bild* von M in Y unter f und

$$f^{-1}(N) := \{x \in X \ : \ f(x) \in N\} \subseteq X$$

das *Urbild* von N in X.

Zu beachten ist, daß für eine einelementige Teilmenge $N = \{y\}$ das Urbild $f^{-1}(y) := f^{-1}(\{y\}) \subseteq X$ aus mehreren Elementen bestehen kann oder auch leer sein kann. Deshalb ist f^{-1} im allgemeinen *keine Abbildung* von Y nach X.

Definition 1.11 Eine Abbildung $f : X \to Y$ heißt

- *injektiv*, falls für $x, x' \in X$ aus $f(x) = f(x')$ stets $x = x'$ folgt
- *surjektiv*, falls es zu jedem $y \in Y$ ein $x \in X$ gibt mit $y = f(x)$
- *bijektiv*, falls f injektiv und surjektiv ist

Falls f bijektiv ist, dann ist die *Umkehrabbildung* $f^{-1} : Y \to X$ gegeben durch $f^{-1} : y \mapsto x = f^{-1}(y)$ mit $y = f(x)$.

Seien X, Y, Z Mengen und $f : X \to Y$ sowie $g : Y \to Z$ Abbildungen, so ist die *Komposition* dieser Abbildungen gegeben durch

$$g \circ f : X \to Z \ , \quad x \mapsto (g \circ f)(x) := g(f(x)) \ .$$

Satz 1.12 *Die Komposition von Abbildungen ist assoziativ, d.h. für Abbildungen $f : X \to Y$, $g : Y \to Z$ und $h : Z \to W$ gilt*

$$h \circ (g \circ f) = (h \circ g) \circ f : X \to W \ .$$

Beweis. Für $x \in X$ gilt

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

□

2 Natürliche Zahlen und vollständige Induktion

Die Struktur der Menge $\mathbb{N}$ der natürlichen Zahlen wird als bekannt vorausgesetzt. In dieser Vorlesung ist $0 \in \mathbb{N}$, für die natürlichen Zahlen ohne 0 schreiben wir $\mathbb{N}^\times := \mathbb{N} \backslash \{0\}$. Die wichtigste Eigenschaft von $\mathbb{N}$ ist, daß für jede Zahl $n \in \mathbb{N}$ genau ein Nachfolger $n+1$ existiert, und startend mit 0 wird mit $0, 1, 2, \ldots, n, n+1, \ldots$ jede natürliche Zahl genau einmal durchlaufen. Daraus ergibt sich das

2.1 Beweisprinzip der vollständigen Induktion. *Es sei $N \in \mathbb{N}$, und zu jeder Zahl $n \in \mathbb{N}$ mit $n \geq N$ sei eine Aussage A_n gegeben. Alle Aussagen A_n sind richtig, wenn man* (IA) *und* (IS) *beweisen kann:*

(IA) *A_N ist wahr (Induktionsanfang).*

(IS) *Für beliebiges $n \geq N$ gilt: Falls A_n wahr ist, so ist auch A_{n+1} wahr (Induktionsschritt).*

Satz 2.2 Für jedes $n \in \mathbb{N}$ gilt die Aussage A_n: $0 + 1 + \cdots + n = \frac{1}{2}n(n+1)$.

Beweis. *i) Wähle $N = 0$. Offenbar ist A_0 wahr.*
ii) Unter der Annahme, daß A_n gilt, folgt

$$\underbrace{0 + 1 + \cdots + n}_{A_n} + (n+1) = \frac{1}{2}n(n+1) + (n+1) = \frac{1}{2}(n+1)(n+2) ,$$

also die Aussage A_{n+1}. □

Da Summen wie in Satz 2.2 häufig vorkommen, vereinbart man das *Summenzeichen* Σ: Ist für jede ganze Zahl $k \in \mathbb{Z}$ mit $m \leq k \leq n$ eine reelle Zahl a_k gegeben, dann setzt man

$$\sum_{k=m}^{n} a_k := \begin{cases} a_m + a_{m+1} + \cdots + a_n & \text{falls } n \geq m \\ 0 & \text{falls } n < m \end{cases}$$

Insbesondere ist $\sum\limits_{k=m}^{m} a_k = a_m$ und

$$\sum_{k=m}^{n+1} a_k = \sum_{k=m}^{n} a_k + a_{n+1} . \tag{*}$$

Durch letzte Eigenschaft kann man das Summenzeichen *rekursiv* ähnlich zum Prinzip der vollständigen Induktion definieren: Man wählt als Induktionsanfang $\sum\limits_{k=m}^{m-1} a_k := 0$ und (*) als Induktionsschritt. Weitere Beispiele rekursiver Definitionen sind die Potenzen einer reellen Zahl als $x^1 := x$ (Induktionsanfang) und $x^{n+1} := x \cdot x^n$ (Induktionsschritt). Für $x \neq 0$ definiert man[2] außerdem $x^0 := 1$. Mit diesen Vorbereitungen beweisen wir

Satz 2.3 (geometrische Summenformel) *Es sei x eine von* 0 *und* 1 *verschiedene reelle Zahl und $n \in \mathbb{N}$. Dann gilt* $\sum\limits_{k=0}^{n} x^k = \dfrac{1 - x^{n+1}}{1 - x}$.

[2] Der Ausdruck "0^0" ist nicht definiert. Jedoch werden wir im weiteren Verlauf des Semesters x^y definieren und x^y für $x, y \to 0$ untersuchen.

Beweis. i) Induktionsanfang: Für $n = 0$ steht auf der linken Seite $x^0 = 1$ und auf der rechten Seite $\frac{1-x}{1-x} = 1$, die Aussage ist also wahr *unter der Voraussetzung* $x \neq 0$ *und* $x \neq 1$.
ii) Induktionsschritt:

$$\sum_{k=0}^{n+1} x^k = \sum_{k=0}^{n} x^k + x^{n+1} = \frac{1-x^{n+1}}{1-x} + x^{n+1}\frac{1-x}{1-x} = \frac{1-x^{n+1}}{1-x} + \frac{x^{n+1}-x^{n+2}}{1-x}$$
$$= \frac{1-x^{n+2}}{1-x} \,. \qquad \square$$

(Die linke Seite ist auch für $x = 1$ sinnvoll und ergibt $\sum_{k=0}^{n} 1 = n + 1$. Fragen wie nach dem Wert der rechten Seite für $x \to 1$ sind typisch für die Analysis.)

Eine Auswahl von Formeln, die durch vollständige Induktion bewiesen werden, ist:

Übung 2.4

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2} \qquad (\textit{Satz } 2.2), \qquad \sum_{k=0}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \,,$$
$$\sum_{k=0}^{n} k^3 = \frac{n^2(n+1)^2}{4} \,, \qquad \sum_{k=0}^{n} (2k+1) = (n+1)^2 \,. \qquad \diamond$$

Ähnlich zum Summenzeichen wird auch das Produktzeichen (mit sonst gleichen Bezeichnungen wie zuvor) eingeführt als

$$\prod_{k=m}^{n} a_k := \begin{cases} a_m \cdot a_{m+1} \cdots a_n & \text{falls } n \geq m \\ 1 & \text{falls } n < m \end{cases}$$

Die rekursive Definition ist $\displaystyle\prod_{k=m}^{m-1} a_k := 1$ und $\displaystyle\prod_{k=m}^{n+1} a_k := a_{n+1} \cdot \prod_{k=m}^{n} a_k$. Besonders wichtig ist für $n \in \mathbb{N}$ der Ausdruck $n!$ (gesprochen "n-Fakultät") definiert durch

$$n! := \begin{cases} \displaystyle\prod_{k=1}^{n} k & \text{falls } n \geq 1 \,, \\ 1 & \text{falls } n = 0 \,, \end{cases}$$

also $n! = 1 \cdot 2 \cdots n$ für $n \geq 1$. Außerdem definieren wir für $n \in \mathbb{Z}$ und $\alpha \in \mathbb{R}$ die *Binomialkoeffizienten*

$$\binom{\alpha}{n} := \begin{cases} \displaystyle\frac{1}{n!}\prod_{k=0}^{n-1} (\alpha - k) \,, & \text{falls } n \geq 0 \,, \\ 0 & \text{falls } n < 0 \,. \end{cases}$$

Insbesondere gilt $\binom{n}{0} = \binom{n}{n} = 1$ für alle $n \in \mathbb{N}$ und $\binom{m}{n} = 0$ für $m \in \mathbb{N}$ mit $m < n$. Ist $m \in \mathbb{N}$ mit $m \geq n$, so ergibt sich alternativ die Darstellung $\binom{m}{n} = \frac{m!}{n!(m-n)!}$.

Lemma 2.5 *Für $n, k \in \mathbb{N}$ gilt* $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$.

Beweis. Für $k = 0$ ist auf der linken Seite $\binom{n+1}{1} = n + 1$ und auf der rechten Seite $\binom{n}{0} + \binom{n}{1} = 1 + n$, d.h. die Formel gilt für $k = 0$. Für $n = k$ sind beide Seiten der Gleichung gleich 1 und für $k > n$ verschwinden beide Seiten der Gleichung. Damit verbleibt der Fall $n > k \geq 1$:

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-(k+1))!} \\
&= \frac{n!(k+1)}{k!(k+1)(n-k)!} + \frac{n!(n-k)}{(k+1)!(n-k)(n-k-1)!} \\
&= \frac{n!(k+1) + n!(n-k)}{(k+1)!(n-k)!} = \frac{(n+1)!}{(k+1)!(n+1-(k+1))!} \\
&= \binom{n+1}{k+1}.
\end{aligned}$$

□

Das Lemma ist der Hintergrund für das Pascalsche Dreieck zur Veranschaulichung der Binomialkoeffizienten. Insbesondere ist $\binom{n}{k} \in \mathbb{N}$ für $n, k \in \mathbb{N}$.

Satz 2.6 (Binomialentwicklung) *Für reelle Zahlen $x, y \neq 0$ und $n \in \mathbb{N}$ gilt* $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$.

Beweis. Durch Induktion nach n. i) Induktionsanfang: Für $n = 0$ ist $(x+y)^0 = 1$ und $\sum_{k=0}^{0} \binom{0}{k} x^k y^{-k} = 1$.

ii) Induktionsschritt:

$$\begin{aligned}
(x+y)^{n+1} = (x+y)(x+y)^n &= \underbrace{\sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k}}_{k+1 \to l} + \underbrace{\sum_{k=0}^{n} \binom{n}{k} x^k y^{n+1-k}}_{k \to l} \\
&= \sum_{l=1}^{n+1} \binom{n}{l-1} x^l y^{n+1-l} + \sum_{l=0}^{n} \binom{n}{l} x^l y^{n+1-l}
\end{aligned}$$

$$= \sum_{l=1}^{n} \underbrace{\left\{ \binom{n}{l-1} + \binom{n}{l} \right\}}_{=\binom{n+1}{l}} x^l y^{n+1-l} + \underbrace{\binom{n}{n}}_{=\binom{n+1}{n+1}} x^{n+1} y^0 + \underbrace{\binom{n}{0}}_{=\binom{n+1}{0}} x^0 y^{n+1}$$

$$= \sum_{l=0}^{n+1} \binom{n+1}{l} x^l y^{n+1-l} \ . \qquad \square$$

3 Gruppen und Körper

3.1 Gruppen

Die Erweiterung von $\mathbb{N}$ zu den ganzen Zahlen $\mathbb{Z}$ bewirkt, daß die Subtraktion stets ausführbar ist. Dadurch bildet $\mathbb{Z}$ eine sogenannte *kommutative Gruppe*.

Unter einer *Verknüpfung* oder *Komposition* auf einer Menge G versteht man eine Abbildung

$$* : G \times G \to G \ , \qquad * : (a, b) \mapsto a * b \ .$$

Beispiele sind die Addition $* = +$ und Multiplikation $* = \cdot$ in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$. Ein weiteres wichtiges Beispiel ist die Menge $G = \mathrm{Abb}(X, X) = \{f : X \to X\}$ der Selbstabbildungen von X mit der Komposition $* = \circ$ als Verknüpfung, $f, g \in \mathrm{Abb}(X, X) \ \Rightarrow \ f \circ g \in \mathrm{Abb}(X, X)$.

Definition 3.1 Eine Menge G zusammen mit einer Verknüpfung $*$ heißt *Gruppe*, falls die folgenden Axiome erfüllt sind:

(G1) $(a * b) * c = a * (b * c)$ für alle $a, b, \in G$ (Assoziativgesetz)

(G2) Es gibt ein neutrales Element $e \in G$ mit $e * a = a$ für alle $a \in G$

(G3) Zu jedem $a \in G$ gibt es ein $a^{-1} \in G$ (das zu a inverse Element) mit $a^{-1} * a = e$

Die Gruppe heißt *kommutativ* (oder *abelsch*), falls außerdem $a * b = b * a$ für alle $a, b \in G$. In diesem Fall schreibt man meist $+$ für die Verknüpfung, 0 für das neutrale Element und $-a$ für das zu a inverse Element.

Gruppen sind *sehr wichtig* in Mathematik und Naturwissenschaften. Beispiele für Gruppen sind, neben den Zahlbereichen, z.B. Drehungen, Verschiebungen, Spiegelungen; allgemein Bewegungen und Zeitentwicklung, Umordnungen (Permutationen), Symmetrieoperationen, etc.

Wir stellen wichtige Eigenschaften von Gruppen zusammen, die direkt aus den Definitionen folgen.

Satz 3.2 *Ist G eine Gruppe, so gilt:*

i) *Das neutrale Element ist eindeutig bestimmt, und außerdem gilt $a * e = a$ für alle $a \in G$.*

ii) *Das inverse Element* a^{-1} *ist für jedes* $a \in G$ *eindeutig bestimmt, und außerdem gilt* $a * a^{-1} = e$ *sowie* $(a^{-1})^{-1} = a$ *und* $(a * b)^{-1} = b^{-1} * a^{-1}$.

iii) *Es gelten die Kürzungsregeln* $a * b' = a * b \;\Rightarrow\; b = b'$ *und* $b' * a = b * a \;\Rightarrow\; b = b'$.

Beweis. Sei $e \in G$ eines der neutralen Elemente und $a \in G$. Zum Inversen $a^{-1} \in G$ gibt es wieder ein Inverses $(a^{-1})^{-1} \in G$, und es gilt

$$\begin{aligned} a * a^{-1} &= e * (a * a^{-1}) = ((a^{-1})^{-1} * a^{-1}) * (a * a^{-1}) = (((a^{-1})^{-1} * a^{-1}) * a) * a^{-1} \\ &= ((a^{-1})^{-1} * (a^{-1} * a)) * a^{-1} = ((a^{-1})^{-1} * e) * a^{-1} = (a^{-1})^{-1} * (e * a^{-1}) \\ &= (a^{-1})^{-1} * a^{-1} = e \end{aligned}$$

und daraus

$$a * e = a * (a^{-1} * a) = (a * a^{-1}) * a = e * a = a\ .$$

Sei $\tilde{e} \in G$ ein weiteres neutrales Element, so gilt aus Sicht von e die Gleichung $e * \tilde{e} = \tilde{e}$ und aus Sicht von $\tilde{e}$ die Gleichung $e * \tilde{e} = e$, also $e = \tilde{e}$. Damit ist i) bewiesen und $a * a^{-1} = e$ aus ii).

Sei $\widetilde{a^{-1}}$ ein weiteres Inverses zu a, so gilt

$$\widetilde{a^{-1}} = \widetilde{a^{-1}} * e = \widetilde{a^{-1}} * (a * a^{-1}) = (\widetilde{a^{-1}} * a) * a^{-1} = e * a^{-1} = a^{-1}\ ,$$

also ist das Inverse eindeutig. Damit ist wegen $a * a^{-1} = e$ das Inverse von a^{-1} durch a selbst gegeben. Schließlich gilt

$$(b^{-1}*a^{-1})*(a*b) = ((b^{-1}*a^{-1})*a)*b = (b^{-1}*(a^{-1}*a))*b = (b^{-1}*e)*b = b^{-1}*b = e$$

so daß $b^{-1} * a^{-1}$ das Inverse zu $a * b$ ist.

Die Kürzungsregeln folgen nach Verknüpfung von links/rechts mit a^{-1} unter Verwendung der Assoziativität. □

Definition 3.3 Sei G eine Gruppe mit Verknüpfung $*$. Eine nichtleere Teilmenge $H \subseteq G$ heißt *Untergruppe*, wenn für alle $a, b \in H$ auch $a * b \in H$ und $a^{-1} \in H$ gilt.

Es folgt automatisch $e \in H$ (durch Wahl von $b = a^{-1}$). Ist G eine Gruppe, so sind die Teilmengen $H = G$ und $H = \{e\}$ Untergruppen. Sie heißen *triviale Untergruppen.*

Beispiele sind $\mathbb{Z} \subseteq \mathbb{Q}$ bezüglich der Addition oder G als Gruppe der Drehungen eines Körpers und H als Untergruppe der Drehungen um eine feste Achse.

3.2 Körper

Die Erweiterung von $\mathbb{Z}$ zu $\mathbb{Q}$ bewirkt, daß die Division durch von Null verschiedene Elemente ausführbar wird. Außerdem sind die beiden Verknüpfungen $+$ und $\cdot$ kompatibel. Man sagt, $\mathbb{Q}$ ist ein Körper:

Definition 3.4 Eine Menge K zusammen mit zwei Verknüpfungen

$$\begin{aligned} + : &K \times K \to K\ , \qquad + : (a,b) \mapsto a+b \qquad &\text{(Addition)} \\ \cdot : &K \times K \to K\ , \qquad \cdot : (a,b) \mapsto a \cdot b \qquad &\text{(Multiplikation)} \end{aligned}$$

heißt *Körper*, wenn folgendes gilt:

(K1) K zusammen mit der Addition ist eine kommutative Gruppe. Ihr neutrales Element wird mit 0 bezeichnet und das zu $a \in K$ inverse Element mit $-a$.

(K2) Sei $K^\times := K \setminus \{0\}$, dann ist für $a, b \in K^\times$ auch $a \cdot b \in K^\times$, und $K^\times$ mit dieser Multiplikation ist eine kommutative Gruppe. Ihr neutrales Element wird mit 1 bezeichnet und das zu $a \in K^\times$ inverse Element mit $a^{-1} = 1/a$. Man schreibt auch $a/b = a^{-1} \cdot b = b \cdot a^{-1}$.

(K3) Es gelten die Distributivgesetze

$$a \cdot (b + c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a + b) \cdot c = a \cdot c + b \cdot c$$

für alle $a, b, c \in K$.

Offenbar sind $\mathbb{Q}$ und $\mathbb{R}$ Körper, $\mathbb{Z}$ ist es nicht. Später werden wir den Körper $\mathbb{C}$ der komplexen Zahlen einführen. In der Algebra werden noch viele weitere Körper untersucht, z.B. kann man die zweielementige Menge $\mathbb{Z}_2 := \{0, 1\}$ zu einem Körper machen. In $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ schreiben wir die Multiplikation meist als ab statt $a \cdot b$.

In einem Körper gilt

i) $1 \neq 0$ (da $1 \in K^\times$ und $0 \notin K^\times$)

ii) $a \cdot 0 = 0 \cdot a = 0$ (verwende $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$)

iii) $a \cdot b = 0 \quad \Rightarrow \quad a = 0$ oder $b = 0$ (aus (K2))

iv) $a \cdot (-b) = -(a \cdot b)$ und $(-a) \cdot (-b) = a \cdot b$

v) $a \cdot b = a \cdot c$ und $a \neq 0 \quad \Rightarrow \quad b = c$

3.3 Angeordnete Körper

Die rationalen und reellen Zahlen sind dadurch gekennzeichnet, daß gewisse Elemente als *positiv* ausgezeichnet sind:

Definition 3.5 (Anordnungs-Axiome) Ein Körper K heißt *angeordnet*, falls eine Teilmenge $K_+^\times \subset K$ existiert mit folgenden Eigenschaften:

(A1) Für $a \in K^\times$ ist entweder $a \in K_+^\times$ oder $-a \in K_+^\times$.

(A2) Aus $a, b \in K_+^\times$ folgen $a + b \in K_+^\times$ und $ab \in K_+^\times$.

Die Elemente $a \in K_+^\times$ heißen *positiv*, Schreibweise $a > 0$.

Offenbar sind $\mathbb{Q}$ und $\mathbb{R}$ angeordnete Körper. Man vereinbart

$$\begin{array}{ll} a > b, & \text{falls } a - b > 0 \\ b < a, & \text{falls } a > b \\ a \leq b, & \text{falls } a < b \text{ oder } a = b \\ a \geq b, & \text{falls } a > b \text{ oder } a = b \end{array}$$

Ist $-a$ positiv, also $a < 0$, so heißt a negativ. Ist $a \geq 0$, so heißt a nichtnegativ. Die Menge der nichtnegativen rellen Zahlen wird mit $\mathbb{R}_+$ bezeichnet, die der positiven rellen Zahlen mit $\mathbb{R}_+^\times$. Außerdem ist $\mathbb{R}^\times := \mathbb{R} \setminus \{0\}$. Die Anordnung ermöglicht die Darstellung der rellen Zahlen auf der Zahlengeraden.

Alle Regeln für das Rechnen mit Ungleichungen folgen aus diesen Axiomen. Hier eine Auswahl ohne Beweis.

Lemma 3.6 *Es sei K ein angeordneter Körper (z.B. $\mathbb{R}$ oder $\mathbb{Q}$). Dann gelten:*

i) *Für beliebige $a, b \in K$ gilt genau eine der Relationen $a > b$, $a = b$, $a < b$.*

ii) *Aus $a > b$ und $b > c$ folgt $a > c$ (Transitivität).*

iii) *Aus $a > b$ folgen $\frac{1}{a} < \frac{1}{b}$, $a + c > b + c$ für alle $c \in K$ sowie $ac > bc$ für $c \in K_+^\times$ und $ac < bc$ für $-c \in K_+^\times$.*

iv) *Aus $a > b$ und $c > d$ folgt $a+c > b+d$ und im Fall $b, d > 0$ auch $ac > bd$.*

v) *Für $a \neq 0$ gilt $a^2 > 0$.*

Analoge Regeln gelten für "$\geq$". Indem man $n \in \mathbb{N}$ mit $n1 := \underbrace{1 + \cdots + 1}_{n \text{ mal}} \in K$ identifiziert, kann man $\mathbb{N}$ als Teilmenge jedes angeordneten Körpers K auffassen, analog auch $\mathbb{Z}$. Indem man $\frac{p}{q} \in \mathbb{Q}_+^\times$ für $p \in \mathbb{N}$, $q \in \mathbb{N} \setminus \{0\}$ mit $(p1)(q1)^{-1} \in K$ identifiziert, kann man $\mathbb{Q}_+^\times$ und analog auch $\mathbb{Q}$ als Teilmenge jedes angeordneten Körpers K auffassen.

Definition 3.7 (Absolutbetrag) Es sei K ein angeordneter Körper und $a \in K$. Dann heißt

$$|a| \in K_+ , \qquad |a| := \begin{cases} a & \text{falls } a \geq 0 \\ -a & \text{falls } a < 0 \end{cases}$$

der *Absolutbetrag* von a.

Satz 3.8 *Für den Absolutbetrag in einem angeordneten Körper K gilt (mit $a, b \in K$)*

i) $|a| = |-a|$

ii) $a \leq |a|, \qquad -a \leq |a|$

iii) $|a + b| \leq |a| + |b|$ *(Dreiecksungleichung)*

iv) $|ab| = |a||b|$

v) $||a| - |b|| \leq |a - b|$

Beweis. i) und ii) folgen aus der Definition.
iii) Aus ii) und Lemma 3.6.iv folgen $a + b \leq |a| + |b|$ und $-(a + b) \leq |a| + |b|$, aus der Definition des Betrages dann iii).
iv) Durch Fallunterscheidung nach $a \gtreqless 0$ und $b \gtreqless 0$.
v) Aus der Dreiecksungleichung folgt $|a| = |a - b + b| \leq |a - b| + |b|$ bzw. $|a| - |b| \leq |a - b|$. Vertauschen von a, b liefert $|b| - |a| \leq |b - a| = |a - b|$, zusammengefaßt damit v). □

Die Dreiecksungleichung (und ihre Verallgemeinerungen) werden wir oft benötigen.

Lemma 3.9 (Bernoullische Ungleichung) *Es sei K ein angeordneter Körper (z.B. $\mathbb{R}$ oder $\mathbb{Q}$).*

i) *Für alle $x \in K$ mit $x > -1$ und $x \neq 0$ und $n \in \mathbb{N}$ mit $n \geq 2$ gilt $(1 + x)^n > 1 + nx$.*

ii) *Für alle $x \in K$ mit $x > -1$ und $n \in \mathbb{N}$ gilt $(1 + x)^n \geq 1 + nx$.*

Beweis. Wir beweisen i) durch Induktion nach n. (1) Induktionsanfang: Für $n = 2$ ergibt sich $(1 + x)^2 = 1 + 2x + x^2 > 1 + 2x$ wegen $x^2 > 0$.
(2) Induktionsschritt. Die Bernoullische Ungleichung gelte für n. Dann folgt unter Verwendung von $1 + x > 0$ und $x^2 > 0$

$$(1+x)^{n+1} = \underbrace{(1 + x)^n}_{>1+nx}(1+x) > (1+nx)(1+x) = 1+(n+1)x+nx^2 > 1+(n+1)x\ .$$

Der Beweis von ii) ist analog. □

Definition 3.10 Ein angeordneter Körper K heißt *archimedisch*, wenn zusätzlich gilt:

(A3) Zu jedem Element $a \in K$ gibt es eine natürliche Zahl n, so daß $n - a > 0$.

Offenbar sind $\mathbb{Q}$ und $\mathbb{R}$ archimedisch angeordnet.

Satz 3.11 *Es sei K ein archimedisch angeordneter Körper und $q, R, \epsilon \in K$ und $n \in \mathbb{N}$. Dann gilt:*

i) *Ist $q > 1$, so gibt es zu jedem $R \in K$ ein $N \in \mathbb{N}$ mit $q^n > R \quad \forall n \geq N$.*

ii) *Ist $0 < q < 1$, so gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $q^n < \epsilon \quad \forall n \geq N$.*

iii) *Zu jedem $\epsilon > 0$ gibt es ein $n \in \mathbb{N}^\times$ mit $0 < \frac{1}{n} < \epsilon$.*

Beweis. i) Setze $q = 1 + x$ mit $x > 0$. Nach (A3) gibt es ein $N \in \mathbb{N}$ mit $\frac{R}{x} < N$, also $Nx > R$. Nach Bernoulli gilt $q^N > 1 + Nx$ für $N \geq 2$, $N \in \mathbb{N}$. Für $Nx > R$ ergibt sich $q^N > 1 + R > R$ und damit $q^n > R$ für alle $n \geq N \geq 2$.

ii) Für $0 < q < 1$ ist $\frac{1}{q} > 1$. Nach i) gibt es ein $N \in \mathbb{N}$ mit $(\frac{1}{q})^n > R := \frac{1}{\epsilon}$ für alle $n \geq N$. Damit ergibt sich $0 < q^n < \frac{1}{R} = \epsilon$ für alle $n \geq N$.

iii) Nach (A3) gibt es ein $N \in \mathbb{N}$ mit $0 < \frac{1}{\epsilon} =: R < N$. Dann ist $N \geq 1$, und es folgt $0 < \frac{1}{n} \leq \frac{1}{N} < \frac{1}{R} = \epsilon$ für alle $n \geq N$. □

4 Reelle Zahlen

Alle bisher eingeführten algebraischen Strukturen gelten gleichermaßen für beide archimedisch angeordneten Körper $\mathbb{Q}$ und $\mathbb{R}$. Der wesentliche analytische Unterschied ist die *Vollständigkeit* von $\mathbb{R}$. Wir definieren die Vollständigkeit zunächst geometrisch über Intervallschachtelungen, später über sogenannte Fundamentalfolgen (was sich auf andere Beispiele verallgemeinern läßt).

Definition 4.1 i) Beschränkte Intervalle. Seien $a, b \in \mathbb{R}$ mit $a < b$. Wir setzen

$$\begin{array}{rll} [a,b] := \{x \in \mathbb{R} \ : \ a \leq x \leq b\} & & \text{(abgeschlossenes Intervall)} \\ ]a,b[:= \{x \in \mathbb{R} \ : \ a < x < b\} & & \text{(offenes Intervall)} \\ [a,b[:= \{x \in \mathbb{R} \ : \ a \leq x < b\} & & \text{(rechts halboffenes Intervall)} \\ ]a,b] := \{x \in \mathbb{R} \ : \ a < x \leq b\} & & \text{(links halboffenes Intervall)} \end{array}$$

Die Intervalle $[a, b]$ nennt man auch *kompakt*. Für alle diese Intervalle $I = [a, b],]a, b[, \dots$ heißen die Punkte a, b die *Randpunkte* und die $x \in \mathbb{R}$ mit $a < x < b$ die *inneren Punkte*. Für jedes dieser Intervalle I ist $\bar{I} := [a, b]$ der Abschluß. Die relle Zahl $|I| := b - a$ heißt Länge des Intervalls I.

ii) Unbeschränkte Intervalle. Sei $a \in \mathbb{R}$. Wir setzen

$$\begin{array}{rr} [a,\infty[:= \{x \in \mathbb{R} \ : \ a \leq x\} &]a,\infty[:= \{x \in \mathbb{R} \ : \ a < x\} \\ ]-\infty,a] := \{x \in \mathbb{R} \ : \ x \leq a\} &]-\infty,a[:= \{x \in \mathbb{R} \ : \ x < a\} \end{array}$$

und $]-\infty, \infty[= \mathbb{R}$.

Die in i) und ii) angegebenen Intervalle heißen auch *echte Intervalle* im Unterschied zum unechten Intervall $[a, a] = \{x \in \mathbb{R} \ : \ a \leq x \leq a\} = \{a\}$.

Manchmal werden offene Intervalle auch als (a, b) geschrieben.

Definition 4.2 Eine *Intervallschachtelung* ist eine Folge $I_0, I_1, I_2, \dots,$ kompakter Intervalle, kurz $(I_n)_{n\in\mathbb{N}}$, mit folgenden Eigenschaften:

(I1) $I_{n+1} \subseteq I_n$ für alle $n \in \mathbb{N}$,

(I2) zu jedem $\epsilon > 0$ gibt es ein Intervall I_n mit der Länge $|I_n| < \epsilon$.

Eine Formulierung der Vollständigkeit der reellen Zahlen besteht in der Gültigkeit des

4.3 Intervallschachtelungsprinzip. *Zu jeder Intervallschachtelung* $(I_n)_{n\in\mathbb{N}}$ *in* $\mathbb{R}$ *existiert ein* $x \in \mathbb{R}$ *mit* $x \in I_n$ *für alle* $n \in \mathbb{N}$.

Die so erhaltene reelle Zahl x ist eindeutig bestimmt: Gäbe es zwei solche Zahlen x_1, x_2 mit $x_1 < x_2$, so müßte das Intervall $[x_1, x_2]$ in allen I_n enthalten sein, und jedes dieser I_n hätte eine Länge $\geq x_2 - x_1$ im Widerspruch zu (I2). Wir betrachten zwei Intervallschachtelungen $(I_n)_{n\in\mathbb{N}}$ und $(J_n)_{n\in\mathbb{N}}$ als äquivalent, wenn es ein $N \in \mathbb{N}$ gibt, so daß für alle $n \geq N$ natürliche Zahlen k, l existieren mit $J_k \subseteq I_n \subseteq J_0$ und $I_l \subseteq J_n \subseteq I_0$. Die zugehörige Quotientenmenge, d.h. die Menge der Äquivalenzklassen von Intervallschachtelungen, heißt dann Menge der reellen Zahlen und wird mit $\mathbb{R}$ bezeichnet.

Eine erste Anwendung der Vollständigkeit ist die Existenz der k-ten Wurzeln positiver reeller Zahlen.

Satz 4.4 *Zu jeder reellen Zahl* $x > 0$ *und jeder natürlichen Zahl* $k \geq 1$ *gibt es genau eine reelle Zahl* $y > 0$ *mit* $y^k = x$. *(Diese wird als* $y = x^{\frac{1}{k}}$ *oder* $y = \sqrt[k]{x}$ *geschrieben.)*

Beweis. Da aus $y_1 > y_2 > 0$ die Relation $y_1^k > y_2^k > 0$ folgt, kann es höchstens eine positive Lösung geben. Für $x = 1$ ist $y = 1$ die einzige Lösung. Wir betrachten zunächst den Fall $x > 1$ und konstruieren eine Intervallschachtelung $(I_n)_{n\in\mathbb{N}}$ mit folgenden Eigenschaften:

(1_n) $\quad a_n^k \leq x \leq b_n^k$

(2_n) $\quad |I_n| = \frac{1}{2^n}|I_0|$

Wir beginnen mit $I_0 := [1, x]$ als Induktionsanfang. Dann ist (2_0) klar, und aus $1 < x$ folgt $1^k \leq x \leq x^k$. Induktionsschritt von n auf $n+1$: Gegeben ist $I_n = [a_n, b_n]$ mit (1_n) und (2_n). Setze $m_n := \frac{1}{2}(a_n + b_n) = a_n + \frac{1}{2}(b_n - a_n)$ (Mittelpunkt) und

$$\begin{aligned} a_{n+1} &:= m_n \ , \ b_{n+1} := b_n \quad \text{falls } m_n^k \leq x \ , \\ a_{n+1} &:= a_n \ , \ b_{n+1} := m_n \quad \text{falls } m_n^k > x \ . \end{aligned}$$

Diese Konstruktion erfüllt (1_{n+1}) und wegen $|I_{n+1}| = \frac{1}{2}|I_n|$ auch (2_{n+1}). Nach Satz 3.11 erfüllt $(I_n)_{n\in\mathbb{N}}$ die Eigenschaften einer Intervallschachtelung (Definition 4.2), und nach dem Intervallschachtelungsprinzip gibt es genau ein $y \in \mathbb{R}$ mit $y \in I_n$ für alle $n \in \mathbb{N}$.

Bleibt zu beweisen, daß $y^k = x$. Dazu zeigt man, daß auch $(J_n)_{n\in\mathbb{N}}$ mit $J_n := [a_n^k, b_n^k]$ eine Intervallschachtelung ist. Der Einschluß $J_{n+1} \subseteq J_n$ ist klar, und für die Längen gilt für $k > 1$

$$|J_n| = b_n^k - a_n^k = \underbrace{(b_n - a_n)}_{=|I_n|}\underbrace{(b_n^{k-1}a_n^0 + b_n^{k-2}a_n^1 + \cdots + b_n^0 a_n^{k-1})}_{<kx^{k-1}} < \frac{kx^{k-1}}{2^n}|I_0| \ .$$

Nun liegen nach Konstruktion sowohl x als auch y^k in jedem Intervall J_n. Da diese in allen Intervallen der Schachtelung liegende reelle Zahl eindeutig ist, gilt $y^k = x$.

Ist $0 < x < 1$, so konstruiert man zu $\frac{1}{x} > 1$ die k-te Wurzel $y > 1$ mit $y^k = \frac{1}{x}$. Dann ist $\frac{1}{y}$ die k-te Wurzel aus x, denn $(\frac{1}{y})^k = \frac{1}{y^k} = \frac{1}{1/x} = x$. □

Übung 4.5 In den Übungen werden das arithmetische, geometrische und harmonische Mittel definiert und untersucht. Mit diesen lassen sich sehr zweckmäßige Intervallschachtelungen konstruieren. ◊

Definition 4.6 (obere und untere Schranken) Eine Teilmenge $M \subseteq \mathbb{R}$ heißt *nach oben beschränkt* bzw. *nach unten beschränkt*, wenn es ein $s \in \mathbb{R}$ gibt, so daß für alle $x \in M$ gilt

$$x \leq s \qquad \text{bzw.} \qquad x \geq s \; .$$

In diesem Fall heißt s eine *obere Schranke* bzw. eine *untere Schranke*. Die Menge M heißt *beschränkt*, wenn M nach oben und unten beschränkt ist.

Beispiel 4.7 Für $a, b \in \mathbb{R}$ sind die offenen, halboffenen und abgeschlossenen Intervalle $]a, b[$, $[a, b[$, $]a, b]$ und $[a, b]$ beschränkt, damit auch nach oben oder unten beschränkt. Aus dem archimedischen Axiom folgt: Die Intervalle $[a, \infty]$ und $]a, \infty]$ sind nach unten beschränkt, aber nicht beschränkt. Die Intervalle $] - \infty, b[$ und $] - \infty, b]$ sind nach oben beschränkt, aber nicht beschränkt. Das Intervall $] - \infty, \infty[$ ist weder nach oben noch nach unten beschränkt. ◁

In einer beschränkten Menge braucht es eine größte oder kleinste Zahl nicht zu geben. Ist $I =]0, 1[$, dann gibt es zu jeder Zahl $x \in I$ eine größere, z.B. $\frac{1+x}{2} > x$. Damit ist $s = 1$ eine obere Schranke von I (und jede reelle Zahl $s > 1$ ist ebenfalls eine obere Schranke), aber $1 \notin I$. Aber 1 ist die kleinste obere Schranke von I:

Definition 4.8 (Supremum und Infimum) Eine Zahl $s \in \mathbb{R}$ heißt *Supremum* der Menge $M \subseteq \mathbb{R}$, geschrieben $s = \sup M$, wenn s die kleinste obere Schranke für M ist, d.h.

i) M ist beschränkt und s ist eine obere Schranke für M.

ii) Jede Zahl $s' < s$ ist keine obere Schranke für M.

Entschrechend ist das *Infimum* von $M \subseteq \mathbb{R}$ die größte untere Schranke, geschrieben $\inf M$.

Beispiel 4.9 Für die Intervalle I gegeben durch $]a, b[$, $[a, b[$, $]a, b]$ und $[a, b]$ mit $a < b$ gilt jeweils $\sup I = b$ und $\inf I = a$. Auf der nach oben abgeschlossenen Seite von $[a, b]$ und $]a, b]$ wird das Supremum b in I angenommen, es liegt dann ein Maximum vor, wobei $\max M := \{x \in M \;:\; y \leq x \quad \forall y \in M\}$. Es gilt dann $\sup(]a, b]) = \max(]a, b]) = b$. Dagegen gilt $\sup([a, b[) = b$, aber $[a, b[$ besitzt kein Maximum. ◁

Satz 4.10 *Jede nach oben (unten) beschränkte nichtleere Teilmenge $M \subseteq \mathbb{R}$ besitzt ein Supremum (Infimum).*

Bemerkung: Das ist eine Formulierung der Vollständigkeit und gilt somit nicht für $M \subseteq \mathbb{Q}$!

Beweis (für das Supremum). Analog zum Existenzbeweis der k-ten Wurzel (Satz 4.4) konstruiert man eine Intervallschachtelung $(I_n)_{n\in\mathbb{N}}$ mit $I_n = [a_n, b_n]$ derart, daß $b_n \in \mathbb{R}$ eine obere Schranke ist und $a_n \in M$ keine obere Schranke ist. Man startet mit einem beliebigen $a_0 \in M$ und einer beliebigen oberen Schranke b_0 und konstruiert per Induktion das Intervall I_{n+1} durch Halbierung von I_n. Eine der Hälften hat die geforderten Eigenschaften. □

Umgekehrt folgt das Intervallschachtelungsprinzip aus der Existenz eines Supremums und Infimums für beschränkte Mengen: Es sei $(I_n)_{n\in\mathbb{N}}$ mit $I_n = [a_n, b_n]$ eine Intervallschachtelung. Die Menge $A := \{a_0, a_1, \dots\}$ ist nach oben beschränkt durch jede der oberen Schranken b_i. Dann gibt es das Supremum $s = \sup A \in \mathbb{R}$ (kleinste obere Schranke) mit $a_n \leq s \leq b_n$ für alle $n \in \mathbb{N}$.

5 Komplexe Zahlen

Die komplexen Zahlen stellen eine Erweiterung der reellen Zahlen dar, in der die Gleichung $z^2 + 1 = 0$ lösbar ist. Viele Bereiche der Analysis lassen sich ohne Modifikation von den reellen auf die komplexen Zahlen erweitern, und manche Eigenschaften im Reellen lassen sich über den "Umweg" der komplexen Zahlen besser verstehen.

Für den zu bestimmenden Erweiterungskörper $\mathbb{C}$ der reellen Zahlen fordern wir:

i) $z^2 + 1 = 0$ ist lösbar, und i bezeichne eine Lösung dieser Gleichung.

ii) Für jede reelle Zahl $x \in \mathbb{R}$ gilt auch $x \in \mathbb{C}$.

Dann ist für $x, y \in \mathbb{R}$ auch $z := x + \mathrm{i}y \in \mathbb{C}$. Die Körperaxiome (insbesondere Kommutativität und Distributivität) sowie $\mathrm{i}^2 = -1$ liefern für $x, y, u, v \in \mathbb{R}$:

$$\begin{aligned}(x + \mathrm{i}y) + (u + \mathrm{i}v) &= (x + u) + \mathrm{i}(y + v)\ ,\\ (x + \mathrm{i}y) \cdot (u + \mathrm{i}v) &= (xu - yv) + \mathrm{i}(xv + yu)\ . \end{aligned}\tag{*}$$

Damit ist die Menge der Elemente der Form $x + \mathrm{i}y$ abgeschlossen unter Addition und Multiplikation. Aus $(x+\mathrm{i}y) = (u+\mathrm{i}v)$ folgt $(x-u)^2 = -(y-v)^2$, also $x = u$ und $y = v$. Die beiden reellen Zahlen x, y in $z = x + \mathrm{i}y$ sind durch z eindeutig definiert.

Satz 5.1 *Die mit der Addition und Multiplikation aus (*) versehene Menge* $\mathbb{C} = \{z = x + \mathrm{i}y \ :\ x, y \in \mathbb{R}\}$ *bildet einen Körper, den Körper der komplexen Zahlen.*

Beweis. Es sind die Körperaxiome nach Definition 3.4 zu überprüfen. Man findet:

- $0 = 0 + \mathrm{i}0$ ist das neutrale Element der Addition.
- $1 = 1 + \mathrm{i}0$ ist das neutrale Element der Multiplikation.
- $-z = -x - \mathrm{i}y$ ist bezüglich der Addition das zu $z = x + \mathrm{i}y$ inverse Element.
- Nicht so offensichtlich ist das zu $z = x + \mathrm{i}y \neq 0$ bezüglich der Multiplikation inverse Element: Dieses ist gegeben durch $z^{-1} = \frac{x}{x^2+y^2} - \mathrm{i}\frac{y}{x^2+y^2}$. Entscheidend dabei ist $x^2 + y^2 \neq 0$ für alle $z = x + \mathrm{i}y \neq 0$.

Durch Nachrechnen zeigt man die Distributivgesetze. □

Allerdings läßt sich in $\mathbb{C}$ keine Anordung $\mathbb{C}_+^\times$ definieren, so daß je zwei komplexe Zahlen nicht vergleichbar sind. Gäbe es in $\mathbb{C}$ eine Relation $>$, so wäre nach Lemma 3.6.v) sowohl $1 = 1^2 > 0$ als auch $-1 = \mathrm{i}^2 > 0$, und dann $0 = 1+(-1) > 0$, Widerspruch. Es kann jedoch der Abstand komplexer Zahlen über den Betrag verglichen werden.

Definition 5.2 Für $z = x + \mathrm{i}y \in \mathbb{C}$ heißt

- $\mathsf{Re}(z) := x \in \mathbb{R}$ der *Realteil* von z,
- $\mathsf{Im}(z) := y \in \mathbb{R}$ der *Imaginärteil* von z
- $\bar{z} := x - \mathrm{i}y$ die zu z *konjugiert komplexe Zahl*.
- $|z| := \sqrt{z \cdot \bar{z}} = \sqrt{x^2 + y^2}$ der *Betrag* von z.

Die Abbildung $z \mapsto \bar{z}$ heißt *komplexe Konjugation*.

Dabei ist mit $\sqrt{x^2 + y^2} \in \mathbb{R}_+$ die eindeutig bestimmte nichtnegative Wurzel gemeint. Damit folgt für $z \neq 0$ die Identität $\frac{1}{z} = \frac{\bar{z}}{z\cdot\bar{z}} = \frac{\bar{z}}{|z|^2}$. Weiter gilt

$$\begin{aligned} \overline{z + w} &= \bar{z} + \bar{w} \;, & \overline{z \cdot w} &= \bar{z} \cdot \bar{w} \;, \\ z + \bar{z} &= 2\mathrm{Re}(z) \;, & z - \bar{z} &= 2\mathrm{i}\mathrm{Im}(z) \;, \\ \bar{\bar{z}} &= z \;, & z \in \mathbb{R} \quad &\Leftrightarrow \quad z = \bar{z} \;. \end{aligned}$$

Beispiel 5.3 Es sei $z = \frac{3+4\mathrm{i}}{1-2\mathrm{i}}$, gesucht sind $\mathrm{Re}(z)$, $\mathrm{Im}(z)$ und $|z|$. Dazu schreiben wir:

$$z = \frac{3 + 4\mathrm{i}}{1 - 2\mathrm{i}} = \frac{(3 + 4\mathrm{i})(1 + 2\mathrm{i})}{(1 - 2\mathrm{i})(1 + 2\mathrm{i})} = \frac{1}{5}\big((3 \cdot 1 - 4 \cdot 2) + \mathrm{i}(3 \cdot 2 + 4 \cdot 1)\big) = -1 + 2\mathrm{i}\;.$$

Somit $\mathrm{Re}(z) = -1$, $\mathrm{Im}(z) = 2$ und $|z| = \sqrt{5}$. ◁

Satz 5.4 *Für beliebige $z, w \in \mathbb{C}$ gilt*

i) $|z| > 0 \quad \Leftrightarrow \quad z \neq 0$,

ii) $|\bar{z}| = |z|$,

iii) $|\mathrm{Re}(z)| \leq |z|$, $\qquad |\mathrm{Im}(z)| \leq |z|$,

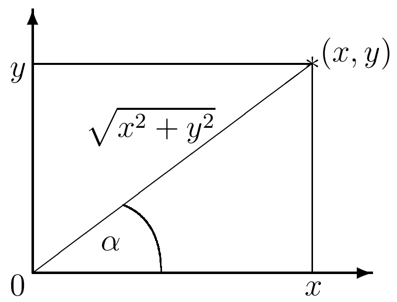
iv) $|z \cdot w| = |z| \cdot |w|$,

v) $|z + w| \leq |z| + |w| \qquad$ *(Dreiecksungleichung).*

Beweis. i),ii),iii) sind klar. Zu (iv) betrachte $|zw|^2 = zw\overline{zw} = z\bar{z}w\bar{w} = |z|^2|w|^2$. Bleibt die Dreiecksungleichung:

$$|z+w|^2 = (z+w)(\bar{z}+\bar{w}) = |z|^2 + z\bar{w} + \bar{z}w + |w|^2 = |z|^2 + 2\mathrm{Re}(z\bar{w}) + |w|^2$$
$$\overset{iii)}{\leq} |z|^2 + 2|z\bar{w}| + |w|^2 = |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2 \ . \qquad \square$$

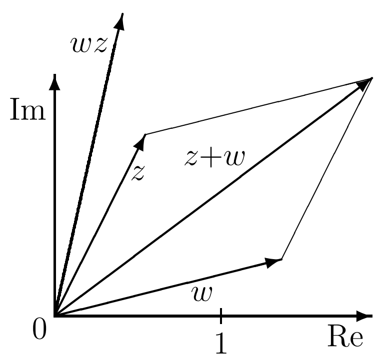
Es ist nun sehr intuitiv, eine komplexe Zahl $z = x + \mathrm{i}y$ als Punkt (x, y) in einem kartesischen Koordinatensystem in der Ebene (Gaußsche Zahlenebene) darzustellen. Nach Pythagoras ist dann $|z|$ gerade der Abstand dieses Punktes vom Ursprung $0 = (0, 0)$.



Ist $z \neq 0$, dann ergeben sich Real- und Imaginärteil in Polarkoordinaten zu $x = |z| \cos\alpha$ und $y = |z| \sin\alpha$, so daß wir folgende Darstellung einer komplexen Zahl erhalten:

$$z = |z|(\cos\alpha + \mathrm{i}\sin\alpha) \ , \qquad 0 \leq \alpha < 2\pi \ .$$

Die Addition $(x, y) + (u, v) := (x + u, y + v)$ entspricht geometrisch der Addition von Vektoren:



Die Dreiecksungleichung $|z+w| \leq |z|+|w|$ entspricht der Tatsache, daß in einem Dreieck die Summe zweier Seitenlängen größer ist als die Länge der verbleibenden Seite.

Die Multiplikation schreibt man am besten in Polarkoordinaten: Ist $z = |z|(\cos\alpha + \mathrm{i}\sin\alpha)$ und $w = |w|(\cos\beta + \mathrm{i}\sin\beta)$, dann ergibt sich

$$
\begin{aligned}
zw &= |z||w|\big((\cos\alpha\cos\beta - \sin\alpha\sin\beta) + \mathrm{i}(\cos\alpha\sin\beta + \sin\alpha\cos\beta)\big) \\
&= |z||w|\big(\cos(\alpha+\beta) + \mathrm{i}\sin(\alpha+\beta)\big) \;,
\end{aligned}
$$

wobei im letzten Schritt die Additionstheoreme verwendet wurden. Folglich ergibt sich in Polarkoordinaten das Produkt zw durch Multiplikation der Radien und Addition der Winkel von z und w.

Über diese geometrische Deutung der Multiplikation findet man die *n-ten Einheitswurzeln*, also komplexe Zahlen mit $z^n = 1$. Nach der Eindeutigkeit der reellen Wurzeln gilt $|z_k| = 1$ für jede Lösung z_k: die Einheitswurzeln liegen auf dem (Einheits-) Kreis mit Radius 1 um den Ursprung. Damit gilt $z_k = \cos\alpha_k + \mathrm{i}\sin\alpha_k$ und nach der Multiplikationsregel $z_k^n = \cos(n\alpha_k) + \mathrm{i}\sin(n\alpha_k)$. Aus $z_k^n = 1$ folgt nun $n\alpha_k = 2\pi k$ mit $k \in \mathbb{Z}$, und die Menge der verschiedenen Lösungswinkel ist somit $\alpha = \{\frac{2\pi k}{n} \;:\; k = 0, 1, 2, \dots, n-1\}$. Folglich hat $z^n = 1$ die n Lösungen $z_k = \cos\frac{2\pi k}{n} + \mathrm{i}\sin\frac{2\pi k}{n}$ mit $k = 0, 1, \dots, n-1$. Insbesondere hat $z^2 = 1$ die Lösungen $z = \{1, -1\}$, und $z^4 = 1$ hat die Lösungen $z = \{1, \mathrm{i}, -1, -\mathrm{i}\}$. Die 5. Einheitswurzeln lassen sich durch den Goldenen Schnitt $g > 1$ ausdrücken, $g - 1 = \frac{1}{g} = h$.

Die Tatsache, daß $z^n = 1$ genau n Lösungen hat, steht in Verbindung zum

5.5 Fundamentalsatz der Algebra. *Es sei $n > 0$. Jede Gleichung*

$$z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0$$

mit komplexen Koeffizienten besitzt in $\mathbb{C}$ mindestens eine Lösung.

Nach Polynomdivision hat die Gleichung dann genau n mit Vielfachheit gezählte Lösungen. Der Beweis wird später gegeben, wenn geeignete Methoden erarbeitet sind.

Wir betrachten den Fall $n = 2$ und

$$0 = z^2 + 2az + b = (z+a)^2 - (a^2 - b) \;, \qquad a, b \in \mathbb{C} \;.$$

Die formale Lösung $z = -a \pm \sqrt{a^2 - b}$ bleibt natürlich richtig, die konkrete Berechnung erfordert jedoch die Umwandlung von $a^2 - b$ in Polarkordinaten und Winkelhalbierung. Eine andere Strategie ist die folgende: Wir suchen Lösungen $w = x + \mathrm{i}y \in \mathbb{C}$ von $w^2 = c$ mit $w = z + a$ und $c = a^2 - b = \alpha + \mathrm{i}\beta \in \mathbb{C}$. Dann ist

$$x^2 - y^2 = \alpha \;, \qquad x^2 + y^2 = |c| = \sqrt{\alpha^2 + \beta^2} \;, \qquad 2xy = \beta \;.$$

Ist $\beta \neq 0$, dann folgt unter Beachtung von $2xy = \beta$

$$x = \pm\sqrt{\frac{1}{2}(\sqrt{\alpha^2+\beta^2} + \alpha)} \;, \qquad y = \pm\frac{\beta}{|\beta|}\sqrt{\frac{1}{2}(\sqrt{\alpha^2+\beta^2} - \alpha)} \;,$$

Ist $\beta = 0$, dann ist $w = \pm\sqrt{\alpha}$ für $\alpha > 0$ und $w = \pm\mathrm{i}\sqrt{-\alpha}$ für $\alpha < 0$.

Zusammenfassung Teil I

- Grundlegende Begriffe zu Mengen und Abbildungen
- $\forall$ (für alle) und $\exists$ (es gibt ein) sowie ihre Negationen
- Beweismethoden: indirekter Beweis und vollständige Induktion
- Geometrische Summenformel, Binomialentwicklung
- Rechnen mit Ungleichungen, Dreiecksungleichung, Bernoullische Ungleichung
- Reelle Zahlen: Intervallschachtelungsprinzip; Supremum und Infimum, Maximum und Minumum
- Komplexe Zahlen: $\mathrm{i}^2 = -1$, Bildung des Inversen, komplexe Konjugation, Betrag, Dreiecksungleichung

Teil II
Vektorräume

6 Lineare Gleichungssysteme und Matrizen

Wir geben hier zunächst eine elementare Einführung in lineare Gleichungssysteme (LGS) und das Standardverfahren zur Bestimmung der Lösung. Eine systematische Untersuchung der Menge der Lösungen erfolgt später, wenn Vektorräume und linearen Abbildungen zwischen ihnen bereitgestellt sind.

Definition 6.1 Es sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Ein *lineares Gleichungssystem* über $\mathbb{K}$ aus m Gleichungen mit n Unbekannten ist ein System der Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{*}$$

wobei die Koeffizienten $(a_{ij}) = \{a_{11}, \dots, a_{mn}\} \in \mathbb{K}$ und die $b_1, \dots, b_m \in \mathbb{K}$ fest vorgegebene Zahlen sind und das n-Tupel $(x_1, \dots, x_n) \in \mathbb{K}^n$ eine zu bestimmende Lösung des Systems ist.

Wichtig ist, daß keine Exponenten wie x_1^2, x_2^{-1}, x_n^3 usw. der x_i auftreten. Offenbar gilt:

Satz 6.2 (elementare Zeilenumformungen) *Gegeben sei ein LGS (*). Die Menge der Lösungen $(x_1, \dots, x_n) \in \mathbb{K}^n$ des Systems bleibt unverändert bei folgenden Typen von Änderungen des Systems:*

Typ I *Vertauschen zweier Gleichungen.*

Typ II *Ersetzen der j-ten Gleichung durch die Summe aus der i-ten und j-ten Gleichung und Beibehalten der i-ten Gleichung.*

Typ III *Multiplikation der i-ten Gleichung mit $\lambda \in \mathbb{K}^\times$ (d.h. $\lambda \neq 0$).*

Die Änderungen vom Typ II und III werden oft zusammen ausgeführt:

Typ IV *Ersetzen der j-ten Gleichung durch die Summe aus dem λ-fachen der i-ten Gleichung und der j-ten Gleichung und Beibehalten der i-ten Gleichung.*

Alle diese elemenaren Zeilenumformungen lassen sich wieder rückgängig machen. Z.B. Typ II durch

i) Multiplikation der i-ten Gleichung mit (-1)

ii) Addition der neuen i-ten Gleichung zur j-ten

iii) Multiplikation der i-ten Gleichung mit (-1)

und Typ III durch

i) Multiplikation der i-ten Gleichung mit $\frac{1}{\lambda}$.

Die Lösungsstrategie besteht darin, durch elementaren Zeilenumformungen das System so umzuformen, daß zumindest eine der Variablen allein steht und abgelesen werden kann. Das Verfahren wird dann für die verbliebenen $n-1$ Variablen wiederholt, usw.

Beispiel 6.3 I_{ij} bedeutet Vertauschen der Gleichungen i und j und $\mathrm{IV}_{ij}(\lambda)$ bedeutet Addition der λ-fachen i-ten Gleichung zur j-ten:

$$\begin{aligned} 1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + \quad\;\; 0\cdot x_4 &= 2 \\ 0\cdot x_1 + 3\cdot x_2 + 0\cdot x_3 + (-1)\cdot x_4 &= 6 \\ 0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + \quad\;\; 0\cdot x_4 &= 0 \end{aligned}$$

$$\mathrm{I}_{23},\ \mathrm{IV}_{23}(-3): \qquad \begin{aligned} 1\cdot x_1 + 0\cdot x_2 + 1\cdot x_3 + \quad\;\; 0\cdot x_4 &= 2 \\ 0\cdot x_1 + 1\cdot x_2 + \tfrac{1}{4}\cdot x_3 + \quad\;\; 0\cdot x_4 &= 0 \\ 0\cdot x_1 + 0\cdot x_2 - \tfrac{3}{4}\cdot x_3 + (-1)\cdot x_4 &= 6 \end{aligned}$$

Geben wir z.B. x_4 beliebig vor, dann lesen wir ab: $x_3 = -8 - \frac{4}{3}x_4$, $x_2 = 2 + \frac{1}{3}x_4$ und $x_1 = 10 + \frac{4}{3}x_4$. ◁

Das ist eine typische Situation: in Lösungen linearer Gleichungssysteme können manche Variablen beliebig gewählt werden, andere sind eindeutig bestimmt, oder es kann auch gar keine Lösung geben. Wir werden im 2. Semster Methoden erarbeiten, mit denen die Menge der Lösungen eines LGS charakterisiert werden kann. Zunächst führen wir eine Matrixschreibweise ein, mit der lineare Gleichungssysteme und die elementaren Zeilenumformungen sich übersichtlicher schreiben lassen.

Definition 6.4 Ein Schema der Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

mit $a_{ij} \in \mathbb{K}$ für alle $1 \leq i \leq m$ und $1 \leq j \leq n$ heißt $m \times n$-*Matrix (über* $\mathbb{K}$). Die Menge aller $m \times n$-Matrizen über $\mathbb{K}$ wird mit $M(m \times n, \mathbb{K})$ bezeichnet.

Man schreibt auch $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ oder, wenn die Größe $m \times n$ klar ist, auch $A = (a_{ij})$, mit dem Eintrag a_{ij} auf der Kreuzung der i-ten Zeile mit der j-ten Spalte. Der erste Index ist also der Zeilenindex, der zweite Index der Spaltenindex. Wir identifizieren $M(n \times 1, \mathbb{K})$ mit $\mathbb{K}^n$, d.h. wir schreiben Elemente aus $\mathbb{K}^n$ in der

Regel als Spalten $x = (x_i)_{i=1,\dots,n} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Sind $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{K}^n$ und ist $\lambda \in \mathbb{K}$, so erklärt man die Addition und die Multiplikation mit Skalaren durch

$$x + y := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \qquad \text{und} \qquad \lambda x := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Schließlich erklären wir eine Matrixmultiplikation $\cdot : M(m \times n, \mathbb{K}) \times \mathbb{K}^n \to \mathbb{K}^m$ wie folgt:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix}.$$

Die Größe der Matrizen muß dabei passen, d.h. ein Produkt $M(m \times n, \mathbb{K}) \times \mathbb{K}^l$ ist für $l \neq n$ nicht erklärt! Ist also $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ und $x = (x_j)_{j=1,\dots,n}$, so ist $A \cdot x = b = (b_i)_{i=1,\dots,m}$ mit $b_i = \sum_{j=1}^{n} a_{ij}x_j$.

Beispiel 6.5 Sei $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in M(2 \times 3, \mathbb{R})$ und $x = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \in \mathbb{R}^3$, so ist $A \cdot x = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot (-1) \\ 4 \cdot 1 + 5 \cdot 2 + 6 \cdot (-1) \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix} \in \mathbb{R}^2$. $\triangleleft$

Damit läßt sich ein lineares Gleichungssystem aus m Gleichungen mit n Unbekannten schreiben als $A \cdot x = b$, wobei $A \in M(m \times n, \mathbb{K})$ und $b \in \mathbb{K}^m$ gegeben sind und die Lösung $x \in \mathbb{K}^n$ gesucht ist. Wir können nun die elementaren Zeilenumformungen des LGS $A \cdot x = b$ in Matrixschreibweise formulieren:

- Typ I Vertauschen zweier Zeilen von A und der entsprechenden Einträge von b.
- Typ II Ersetzen der j-ten Zeile von A durch die Summe aus der i-ten und j-ten Zeile und Beibehalten der i-ten Zeile, sowie Ersetzen des j-ten Eintrags von b durch die Summe aus i-tem und j-tem Eintrag von b und Beibehalten des i-ten Eintrags von b.
- Typ III Multiplikation der i-ten Zeile von A und des i-ten Eintrags von b mit $\lambda \in \mathbb{K}^\times$.

Typ IV Ersetzen der j-ten Zeile von A durch die Summe aus dem λ-fachen der i-ten Zeile und der j-ten Zeile und Beibehalten der i-ten Zeile, und entsprechend für die Einträge von b.

Man sieht dabei, daß die elementaren Zeilenumformungen an der Lösung $x = (x_j)$ nichts ändern. Deshalb kann x gefahrlos weggelassen werden und die Zeilenumformungen einheitlich für die *erweiterte Koeffizientenmatrix*

$$(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

durchgeführt werden.

Beispiel 6.6 Wir wiederholen die Schritte aus Beispiel 6.3:

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 2 \\ 0 & 3 & 0 & -1 & 6 \\ 0 & 1 & \frac{1}{4} & 0 & 0 \end{array}\right) \overset{\substack{\mathrm{I}_{23} \\ \mathrm{IV}_{13}(-3)}}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{3}{4} & -1 & 6 \end{array}\right)$$

Eine solche Matrix heißt in *Zeilenstufenform.* Durch Zeilenumformung vom Typ III kann erreicht werden, daß die erste von Null verschiedene Zahl jeder Zeile zu 1 wird:

$$\overset{\mathrm{III}_3(-\frac{4}{3})}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 2 \\ 0 & 1 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 1 & \frac{4}{3} & -8 \end{array}\right)$$

Schließlich kann mit Typ IV erreicht werden, daß alle Zahlen oberhalb der ersten 1 jeder Zeile zu Null werden:

$$\overset{\mathrm{IV}_{32}(-\frac{4}{3}),\ \mathrm{IV}_{31}(-1)}{\longrightarrow} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{4}{3} & 10 \\ 0 & 1 & 0 & -\frac{1}{3} & 2 \\ 0 & 0 & 1 & \frac{4}{3} & -8 \end{array}\right)$$

Die Lösung ist damit (wie zuvor) $x_3 = -8 - \frac{4}{3}x_4$, $x_2 = 2 + \frac{1}{3}x_4$ und $x_1 = 10 + \frac{4}{3}x_4$. ◁

Eine nützliche Definition ist also

Definition 6.7 Eine Matrix $A \in M(m \times n, \mathbb{K})$ heißt in *Zeilenstufenform*, falls gilt:

i) Es gibt eine Zahl $r \leq m$, so daß die Zeilen mit Index $i = 1, \dots, r$ nicht identisch Null sind, während die Zeilen mit Index $i = r+1, \dots, m$ identisch Null sind.

ii) Bezeichne $j_i := \min\{j \; : \; a_{ij} \neq 0\}$ den kleinsten Spaltenindex der Zeilen $i = 1, \ldots, r$ mit Eintrag $\neq 0$, so gilt $j_1 < j_2 < \cdots < j_r$.

Satz 6.8 *Jede Matrix $A \in M(m \times n, \mathbb{K})$ läßt sich durch endlich viele elementare Zeilenumfomungen auf spezielle Zeilenstufenform*

$$\tilde{A} = \begin{pmatrix} 0 & \ldots & 0 & 1 & * & \ldots & * & 0 & * & \ldots & * & 0 & * & \ldots & * & 0 & \ldots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & * & \ldots & * & 0 & * & \ldots & * & 0 & \ldots \\ 0 & & & & & \ldots & & & & & 0 & 1 & * & \ldots & * & 0 & \\ 0 & & & & & & & & & \ldots & & & & & 0 & 1 & \ldots \\ \vdots & & & & & & & & & \ldots & & & & & & & \ddots \end{pmatrix}$$

bringen mit $a_{ij_i} = 1$ und $a_{kj_i} = 0$ für $k \neq i$, wenn $j_i := \min\{j \; : \; a_{ij} \neq 0\}$. Ein $*$ steht für ein beliebiges Element aus $\mathbb{K}$. Jede Zeile wird nach einer Reihe von Nullen mit 1 begonnen, und zwar später als in jeder der vorangehenden Zeilen. Oberhalb und unterhalb der führenden 1 sind alle Einträge Null.

Beweis. Suche die erste Spalte j, die nicht identisch Null ist. Durch Zeilenvertauschung (Typ I) läßt sich erreichen, daß der erste Eintrag a_{1j} dieser Spalte $\neq 0$ ist. Teile die erste Zeile durch diesen Eintrag a_{1j} (Typ III). Diese neue erste Zeile wird nun festgehalten. Addiere zu jeder Zeile $i = 2, \ldots m$ das $(-a_{ij})$-fache der neuen ersten Zeile (Typ IV). Dadurch wird $a_{ij} = 0$ für alle $i > 1$. Wiederhole das Verfahren für die Zeilen 2 bis m, wobei j nun der kleinste Spaltenindex mit einem $a_{ij} \neq 0$ für $i = 2, \ldots, m$ ist, u.s.w. Das Ergebnis ist eine Matrix in Zeilenstufenform, wobei jede Zeile nach einer Reihe von Nullen mit 1 begonnen wird. Durch erneute Umformungen vom Typ IV werden dann (wie in Beispiel 6.6) oberhalb der führenden 1 jeder Zeile die Einträge auf 0 gebracht. □

Definition 6.9 Die Anzahl der Zeilen einer Matrix $A \in M(m \times n, \mathbb{K})$, die nach Überführung in Zeilenstufenform nicht identisch Null sind, heißt der *Rang* von A, geschrieben $\mathrm{rang}(A)$.

Wir werden später sehen, daß der Rang unabhängig von der Wahl der elementaren Zeilenumformungen ist.

Liegt die erweiterte Koeffizientenmatrix in spezieller Zeilenstufenform vor (was nach Satz 6.8 durch elementare Zeilenumformungen immer erreicht werden kann, wobei sich nach Satz 6.2 die Menge der Lösungen des LGS nicht ändert), so läßt sich die Lösung des LGS $A \cdot x = b$ mit $A \in M(m \times n, \mathbb{K})$ und $b \in \mathbb{K}^m$ wie folgt ermitteln:

Satz 6.10 (Gaußsches Eliminationsverfahren) *Es sei $A \cdot x = b$ ein lineares Gleichungssystem, $A \in M(m \times n, \mathbb{K})$ und $b \in \mathbb{K}^m$. Die erweiterterte Koeffizientenmatrix $(A|b)$ liege in spezieller Zeilenstufenform vor.*

- *Ist* $\operatorname{rang}(A|b) \neq \operatorname{rang}(A) = r$, *d.h.* $b_{r+1} \neq 0$, *so gibt es keine Lösung, denn die* $(r+1)$*-te Gleichung* $0 = \sum_{j=1}^{n} a_{r+1,j} x_j = b_{r+1}$ *ist nicht lösbar.*
- *Ist* $\operatorname{rang}(A|b) = \operatorname{rang}(A)$, *so sind die* $n - r$ *Variablen* x_k *mit* $k \notin \{j_1, \dots, j_r\}$ *im Sinne von Definition 6.7.ii), welche also Stufen der Länge* ≥ 2 *entsprechen,* freie Variablen. *Sie können beliebig aus* $\mathbb{K}$ *gewählt werden.*
- *Die verbleibenden* x_{j_i} *mit* $i = 1, \dots, r$ *sind* gebundene Variablen, *deren Wert sich nach Wahl der freien* x_k *bestimmt zu* $x_{j_i} = b_i - \sum_{k \notin \{j_1,\dots,j_r\}} a_{ik} x_k$. □

Offenbar gilt: Für $n = r$ ist die Lösung eindeutig bestimmt, für $m = r$ (keine Nullzeilen) ist das LGS für alle $b \in \mathbb{K}^m$ lösbar. Achtung: die sofortige Lösung des LGS durch $x_{j_i} = b_i - \sum_{k \notin \{j_1,\dots,j_r\}} a_{ik} x_k$ ist nur richtig, wenn $(A|b)$ in spezieller Zeilenstufenform vorliegt.

Mit $y = \begin{pmatrix} x_{j_1} \\ \vdots \\ x_{j_r} \end{pmatrix} \in \mathbb{K}^r$ für die gebundenen Variablen und $w = \begin{pmatrix} x_{k_1} \\ \vdots \\ x_{k_{n-r}} \end{pmatrix} \in \mathbb{K}^{n-r}$ mit $k_1 < \cdots < k_{n-r} \notin \{j_1, \dots, j_r\}$ für die freien Variablen und Zusammenfassung der Matrixelemente $c_{il} := a_{ik_l}$ zu einer Matrix $C = (c_{il}) \in M(r \times (n-r), \mathbb{K})$ ergibt sich die Lösung von $A \cdot x = b$ (in spezieller Zeilenstufenform) zu $y = \tilde{b} - C \cdot w$ mit $w \in \mathbb{K}^{n-r}$ beliebig und $\tilde{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_r \end{pmatrix} \in \mathbb{K}^r$.

Beispiel 6.11

$$\left(\begin{array}{cccc|c} 2 & 1 & 3 & -2 & 1 \\ 5 & 3 & 8 & -4 & -1 \\ 4 & 1 & 5 & -6 & 9 \end{array}\right) \xrightarrow{III_1(\frac{1}{2}),IV_{12}(-5),IV_{13}(-2)} \left(\begin{array}{cccc|c} 1 & \frac{1}{2} & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 & -\frac{7}{2} \\ 0 & -1 & -1 & -2 & 7 \end{array}\right)$$

$$\xrightarrow{III_2(2),IV_{23}(1))} \left(\begin{array}{cccc|c} 1 & \frac{1}{2} & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 1 & 1 & 2 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{IV_{21}(-\frac{1}{2})} \left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 4 \\ 0 & 1 & 1 & 2 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Als Lösung lesen wir ab:

$$\underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{y} = \underbrace{\begin{pmatrix} 4 \\ -7 \end{pmatrix}}_{\tilde{b}} - \underbrace{\begin{pmatrix} 1 & -2 \\ 1 & 2 \end{pmatrix}}_{C} \cdot \underbrace{\begin{pmatrix} x_3 \\ x_4 \end{pmatrix}}_{w}$$

mit $x_3, x_4 \in \mathbb{K}$ beliebig. Wir haben $\operatorname{rang}(A|b) = \operatorname{rang}(A) = 2$. ◁

7 Vektorräume

Vektorräume sind der zentrale Gegenstand der linearen Algebra.

Definition 7.1 Sei K ein Körper. Eine Menge V zusammen mit einer inneren Verknüpfung $+ : V \times V \to V$ (der Addition) und einer äußeren Verknüpfung $\cdot : K \times V \to V$ (Multiplikation mit Skalaren) heißt *Vektorraum über* K, wenn gilt:

(V1) V zusammen mit der Addition ist eine kommutative Gruppe. (Wie üblich wird das neutrale Element wird mit $0 \in V$ bezeichnet und das zu $v \in V$ inverse Element mit $-v$.)

(V2) Für die Multiplikation mit Skalaren gilt

$$\begin{aligned} (\lambda+\mu)\cdot v &= \lambda\cdot v + \mu\cdot v\,, & \lambda\cdot(v+w) &= \lambda\cdot v+\lambda\cdot w\,,\\ \lambda\cdot(\mu\cdot v) &= (\lambda\mu)\cdot v\,, & 1\cdot v &= v\,, \end{aligned}$$

für alle $\lambda, \mu \in K$ und $v, w \in V$.

Ein Element eines Vektorraums V heißt *Vektor*.

Die wichtigsten Fälle sind Vektorräume über $K = \mathbb{R}$ und $K = \mathbb{C}$, und schreiben oft auch $\mathbb{K}$ statt K. Wir sprechen dann von reellen bzw. komplexen Vektorräumen.

Satz 7.2 *In einem Vektorraum über K gilt*

i) $0 \cdot v = 0 \in V \quad \forall v \in V$

ii) $\lambda \cdot 0 = 0 \in V \quad \forall \lambda \in K$

iii) $(-1) \cdot v = -v \quad \forall v \in V$

iv) $\lambda \cdot v = 0 \in V \quad \Leftrightarrow \quad \lambda = 0$ *oder* $v = 0$

Beweis. Beweis von i),ii),iii) analog zu Körpern am Ende von 3.2.
iv): Ist $\lambda \cdot v = 0$, aber $\lambda \neq 0$, so gilt $v = 1 \cdot v = (\lambda^{-1}\lambda) \cdot v = \lambda^{-1} \cdot (\lambda \cdot v) = 0$. □

Beispiel 7.3 (für Vektorräume)

i) $K^n = \{x = (x_1, \ldots, x_n) \ : \ x_i \in K\}$ ist ein Vektorraum mit Addition

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n)$$

und skalarer Multiplikation

$$\lambda \cdot (x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n)\,.$$

Wir schreiben Vektoren aus K^n oft auch als Spalten statt als Zeilen.

ii) Der Vektorraum $M(m\times n, K)$ der $(m\times n)$-Matrizen mit Einträgen aus K ist ein Vektorraum mit komponentenweiser Addition und skalarer Multiplikation: Schreiben wir $A = (a_{ij}), B = (b_{ij}) \in M(m \times n, K)$, so ist

$$A + B := (a_{ij} + b_{ij}) , \qquad \lambda \cdot A := (\lambda a_{ij}) .$$

iii) $\mathbb{C}$ kann als reeller Vektorraum aufgefaßt werden durch $\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$, $(\lambda, x + \mathrm{i}y) \mapsto \lambda x + \mathrm{i}\lambda y$. $\triangleleft$

Im Laufe dieser Vorlesung werden wir weitere wichtige Beispiele für Vektorräume kennenlernen. Hier ein Ausblick ohne Erläuterungen:

- Polynome
- konvergente Zahlenfolgen, Nullfolgen, Potenzreihen
- stetige Funktionen, differenzierbare Funktionen, integrierbare Funktionen
- lineare Abbildungen
- lineare stetige Operatoren auf Hilbert-Räumen

Definition 7.4 Sei V ein Vektorraum. Eine Teilmenge $W \subseteq V$ heißt *Untervektorraum*, wenn

(UV1) $W \neq \emptyset$

(UV2) $v, w \in W \quad \Rightarrow \quad v + w \in W$
(d.h. W ist abgeschlossen bezüglich der Addition)

(UV3) $v \in W\ ,\ \lambda \in K \quad \Rightarrow \quad \lambda \cdot v \in W$
(d.h. W ist abgeschlossen bezüglich der Multiplikation mit Skalaren)

Ein Untervektorraum ist automatisch ein Vektorraum. Beispiele sind

i) der Nullvektor $W = \{0\}$

ii) Vielfache $W = \{\lambda \cdot v \ : \ \lambda \in K\}$ eines ausgewählten Vektors $v \in V$, speziell jede Gerade in der Ebene durch den Nullpunkt, $W = \{(x, y) \in \mathbb{R}^2 \ : \ ax + by = 0\}$ für $a, b \in \mathbb{R}$.

Satz 7.5 *Seien W_i mit $i \in I$ (Indexmenge) jeweils Untervektorräume von V, so ist der Durchschnitt $W := \bigcap_{i\in I} W_i \subseteq V$ wieder ein Untervektorraum von V.*

Beweis. $0 \in W_i\ \forall i \in I$, also $0 \in W$ (damit ist W nicht leer). Seien $v, w \in W$, so sind $v, w \in W_i\ \forall i \in I$. Dann ist auch $v + w \in W_i\ \forall i \in I$ und somit $v + w \in W$. Analog ist $\lambda \cdot v \in W$. $\square$

Dagegen ist die Vereinigung von Untervektorräumen im allgemeinen nicht wieder ein Untervektorraum. Man kann eine solche Vereinigung aber zu einen Vektorraum abschließen.

Definition 7.6 (Linearkombinationen) Seien $v_1, \ldots, v_r \in V$ (nicht notwendig verschiedene) Vektoren aus V. Ein Vektor $v \in V$ heißt *Linearkombination* von $v_1, \ldots, v_r$, wenn es $\lambda_1, \ldots, \lambda_r \in K$ gibt, so daß

$$v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = \sum_{i=1}^{r} \lambda_i \cdot v_i \,.$$

Wir bezeichnen mit

$$\operatorname{span}_K(v_1, \ldots, v_r) = \left\{ \sum_{i=1}^{r} \lambda_i \cdot v_i \;:\; \lambda_i \in K \right\}$$

den Raum aller Linearkombinationen von $v_1, \ldots, v_r$. Offenbar ist $\operatorname{span}_K(v_1, \ldots, v_r)$ ein Untervektorraum von V, er heißt der durch $v_1, \ldots, v_r$ *aufgespannte* (oder *erzeugte*) *Untervektorraum*.

Die Definition läßt sich verallgemeinern auf Untervektorräume, die von einer Familie $(v_i)_{i \in I}$ aus möglicherweise unendlich vielen Vektoren aufgespannt wird. Dann ist $\operatorname{span}_K(v_i)_{i \in I}$ definiert als die Menge aller *endlichen* Linearkombinationen, d.h. zu jedem $v \in \operatorname{span}_K(v_i)_{i \in I}$ gibt es Indizes $i_1, \ldots, i_r \in I$ und Skalare $\lambda_1, \ldots, \lambda_r \in K$, so daß $v = \sum_{j=1}^{r} \lambda_j v_{i_j}$.

Beispiel 7.7 (für Linearkombinationen)

i) Für $V = \mathbb{R}^3$ und $v_1, v_2 \in V$ ist $\operatorname{span}_{\mathbb{R}}(v_1)$ die Gerade durch 0 und v_1, falls $v_1 \neq 0$, und $\operatorname{span}_{\mathbb{R}}(v_1, v_2)$ die Ebene durch $0, v_1, v_2$, falls $v_1 \neq 0$ und $v_2 \notin \operatorname{span}_{\mathbb{R}}(v_1)$.

ii) Im Vektorraum K^n setzen wir $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, wobei die 1 an der i-ten Stelle steht. Dann ist $\operatorname{span}_K(e_i)_{i=1,\ldots,n} = K^n$. $\triangleleft$

Im weiteren wird es wichtig sein, ob sich ein Vektor $v \in V$ in eindeutiger Weise als Linearkombination von vorgegebenen Vektoren $v_1, \ldots, v_r$ darstellen läßt oder nicht.

Definition 7.8 Sei V ein Vektorraum über K. Eine Familie von endlich vielen Vektoren $v_1, \ldots, v_r$ heißt *linear unabhängig* (bzw. die Vektoren $v_1, \ldots, v_r$ heißen linear unabhängig), wenn aus $\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = 0$ für $\lambda_1, \ldots, \lambda_r \in K$ folgt $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$.

Eine beliebige Familie $(v_i)_{i \in I}$ heißt linear unabhängig, wenn jede endliche Teilfamilie linear unabhängig ist. Entsprechend heißt eine Familie $(v_i)_{i \in I}$, die nicht linear unabhängig ist, *linear abhängig*. In diesem Fall gibt es also $v_{i_1}, \ldots, v_{i_r} \neq 0$ mit $i_1, \ldots, i_r \in I$ und $\lambda_1, \ldots, \lambda_r \neq 0$ mit $\sum_{j=1}^{r} \lambda_j \cdot v_{i_j} = 0$.

Satz 7.9 *Eine Familie* $(v_1, \ldots, v_n)$ *von Vektoren* $v_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{mj} \end{pmatrix} \in K^m$ *ist genau dann linear unabhängig, wenn für die Matrix* $A = (v_{ij}) \in M(m \times n, K)$ *gilt* $\mathrm{rang}(A) = n$.

Beweis. Nach Definition des Matrixprodukts gilt

$$\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n = 0 \quad \Leftrightarrow \quad A \cdot x = 0$$

mit $x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \in K^n$. Die Familie $(v_1, \ldots, v_n)$ ist also genau dann linear unabhängig, wenn $A \cdot x = 0$ nur die triviale Lösung $x = 0$ besitzt. Das erfordert $\mathrm{rang}(A) = n$. □

Beispiel 7.10 Sei z.B. $v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $v_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$, $v_3 = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$ so betrachten wir

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \xrightarrow{IV_{12}(-2)\ ,\ IV_{13}(-3)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix} \xrightarrow{III_2(-1), IV_{23}(6)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Damit ist (v_1, v_2, v_3) nicht linear unabhängig. Dagegen ist (v_1, v_2) linear unabhängig, denn Weglassen der 3. Spalte ergibt $\mathrm{rang}(v_1, v_2) = 2$. ◁

Wir werden später sehen, daß in beliebigen endlich erzeugten Vektorräumen die lineare Unabhängigkeit stets durch Lösen linearer Gleichungssysteme ermittelt werden kann.

Satz 7.11 *Für eine Familie* $(v_i)_{i \in I}$ *sind folgende Bedingungen äquivalent:*

i) $(v_i)_{i \in I}$ *ist linear unabhängig*

ii) *Jeder Vektor* $v \in \mathrm{span}_K(v_i)_{i \in I}$ *läßt sich in eindeutiger Weise als Linearkombination von Vektoren aus* $(v_i)_{i \in I}$ *darstellen.*

Beweis. i)⇒ii). Sei $v \in \mathrm{span}_K(v_i)_{i \in I}$ auf zwei verschiedene Arten als Linearkombination darstellbar,

$$v = \sum_{i \in I} \lambda_i \cdot v_i = \sum_{i' \in I} \mu_{i'} \cdot v_{i'} \ ,$$

wobei nur endlich viele Skalare $\lambda_i, \mu_{i'}$ ungleich Null sind. Es gibt also eine endliche Teilmenge $J \subseteq I$ (Vereinigung der Indizes i, i', für die $\lambda_i \neq 0$ oder $\mu_{i'} \neq 0$), so daß

$$\sum_{j \in J} (\lambda_j - \mu_j) \cdot v_j = 0 \ .$$

Da nach Voraussetzung i) jede endliche Teilfamilie von $(v_i)_{i\in I}$ linear unabhängig ist, ist $\lambda_j = \mu_j$ für alle $j \in J$ und weiter für alle $i \in I$ (auf dem Komplement $I \setminus J$ ist $\lambda_i = \mu_i = 0$).

ii)⇒i). Der Nullvektor läßt sich als Linearkombination von $(v_i)_{i\in I}$ darstellen, in der sämtliche Skalare Null sind. Ist die Linearkombination eindeutig, so ist $(v_i)_{i\in I}$ linear unabhängig. □

8 Erzeugendensystem, Basis und Dimension

Definition 8.1 Eine Familie $(v_i)_{i\in I}$ von Vektoren $v_i \in V$ heißt *Erzeugendensystem* von V, wenn $V = \mathrm{span}_K(v_i)_{i\in I}$, wenn also jeder Vektor $v \in V$ eine endliche Linearkombination von $(v_i)_{i\in I}$ ist.

Ein Erzeugendensystem $\mathcal{B} = (v_i)_{i\in I}$ heißt *Basis* von V, wenn $\mathcal{B}$ linear unabhängig ist.

Der Vektorraum V heißt *endlich erzeugt*, falls es ein endliches Erzeugendensystem $\mathcal{B} = (v_1, \ldots, v_r)$ von V gibt.

Beispiel 8.2 (für Erzeugendensysteme und Basen)

i) Im Vektorraum K^n ist $\mathcal{B} = (e_1, \ldots, e_n)$, mit $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, ein Erzeugendensystem, denn jeder Vektor $x = (x_1, \ldots, x_n)$ läßt sich schreiben als $x = \sum_{i=1}^n x_i e_i$ mit $x_i \in K$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardbasis* des K^n.

ii) Im Vektorraum $M(m \times n, K)$ ist $\mathcal{B} = (E_{ij})_{\substack{i=1,\ldots m\\ j=1,\ldots,n}}$ mit

$$E_{ij} = \begin{pmatrix} & & & 0 & & & \\ & \mathrm{O} & & \vdots & & \mathrm{O} & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & \mathrm{O} & & \vdots & & \mathrm{O} & \\ & & & 0 & & & \end{pmatrix},$$

wobei die 1 im Schnittpunkt der i-ten Zeile mit der j-ten Spalte steht, ein Erzeugendensystem. Jede Matrix $A = (a_{ij})$ läßt sich darstellen als $A = \sum_{i=1}^m \sum_{j=1}^n a_{ij} E_{ij}$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardmatrixbasis*.

iii) In $\mathbb{C}$, aufgefaßt als reeller Vektorraum, ist $\mathcal{B} = (1, \mathrm{i})$ eine Basis, denn jede komplexe Zahl kann als $z = x \cdot 1 + y \cdot \mathrm{i}$ mit $x, y \in \mathbb{R}$ geschrieben werden.

Fassen wir $\mathbb{C} = \mathbb{C}^1$ dagegen als komplexen Vektorraum auf, dann ist $\mathcal{B} = (1, \mathrm{i})$ zwar ein Erzeugendensystem, aber keine Basis mehr, denn 1 und i sind nicht mehr linear unabhängig über $\mathbb{C}$: $1 \cdot 1 + \mathrm{i} \cdot \mathrm{i} = 0$. ◁

Ist $\mathcal{B} = (v_1, \ldots, v_n)$ ein endliches Erzeugendensystem bzw. eine endliche Basis, dann ist jede Permutation (Umordnung) der Vektoren v_i wieder ein endliches Erzeugendensystem bzw. eine endliche Basis. Z.B. ist $\mathcal{B} = (v_1, v_2, v_3)$ mit $v_1 = (1, 0, 0)$, $v_2 = (0, 0, 1)$, $v_3 = (0, 1, 0)$ eine Basis von K^3, aber es ist übersichtlicher, statt $\mathcal{B}$ mit der durch Umordnung erhaltenen Standardbasis (v_1, v_3, v_2) zu arbeiten.

Für die Definition einer Basis gibt es mehrere äquivalente Möglichkeiten. Wir beschränken uns zunächst auf endlich erzeugte Vektorräume:

Satz 8.3 *Für eine Familie $\mathcal{B} = (v_1, \ldots, v_n)$ von Vektoren $v_i \in V$, $V \neq \{0\}$, sind folgende Bedingungen äquivalent:*

i) *$\mathcal{B}$ ist ein linear unabhängiges Erzeugendensystem von V (also eine Basis).*

ii) *$\mathcal{B}$ ist ein Erzeugendensystem von V, aber $(v_1, \ldots, v_{r-1}, v_{r+1}, \ldots, v_n)$ ist für jedes $1 \leq r \leq n$ kein Erzeugendensystem mehr.*

iii) *Zu jedem $v \in V$ gibt es eindeutig bestimmte $\lambda_1, \ldots, \lambda_n \in K$ mit $v = \sum_{i=1}^n \lambda_i \cdot v_i$ (Eindeutigkeit der Zerlegung von v nach der Basis).*

iv) *$\mathcal{B}$ ist linear unabhängig, während $(v_1, \ldots, v_n, v)$ linear abhängig ist für alle $v \in V$.*

Beweis. i)⇒ii). Wäre $(v_1, \ldots, v_{r-1}, v_{r+1}, \ldots, v_n)$ ein Erzeugendensystem, dann gibt es $\lambda_1, \ldots, \lambda_{r-1}, \lambda_{r+1}, \ldots, \lambda_n \in K$ mit $v_r = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + \sum_{j=r+1}^n \lambda_j \cdot v_j$. Damit ist $0 = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + (-1) \cdot v_r + \sum_{j=r+1}^n \lambda_j \cdot v_j$, so daß $\mathcal{B}$ nicht linear unabhängig wäre.

ii)⇒iii). Angenommen, es gibt ein $v \in V$ und für dieses zwei verschiedene Darstellungen

$$v = \sum_{i=1}^n \lambda_i \cdot v_i \qquad \text{und} \qquad v = \sum_{i=1}^n \mu_i \cdot v_i$$

mit $\lambda_r \neq \mu_r$ für mindestens ein r. (Es gibt mindestens eine Darstellung, da $\mathcal{B}$ Erzeugendensystem.) Subtraktion beider Gleichungen und Division durch $(\lambda_r - \mu_r)$ ergibt

$$v_r = \sum_{i=1}^{r-1} \frac{\lambda_i - \mu_i}{\mu_r - \lambda_r} \cdot v_i + \sum_{j=r+1}^{n} \frac{\lambda_j - \mu_j}{\mu_r - \lambda_r} \cdot v_j$$

Das würde bedeuten, daß $(v_1, \ldots, v_{r-1}, v_{r+1}, \ldots, v_n)$ ein Erzeugendensystem wäre, denn in einer Linearkombination ließe sich v_r durch die v_i, v_j mit $i, j \neq r$ ausdrücken, im Widerspruch zu ii).

iii)⇒iv). Nach Satz 7.11 ist $\mathcal{B}$ linear unabhängig. Andererseits gibt es für jedes $v \in V$ Skalare $\lambda_1, \ldots, \lambda_n \in K$ mit $0 = \sum_{i=1}^n \lambda_i \cdot v_i + (-1) \cdot v$, so daß $(v_1, \ldots, v_n, v)$ nicht linear unabhängig ist.

iv)⇒i). Da $(v_1,\dots,v_n,v)$ für alle $v \in V$ linear abhängig ist, gibt es $\lambda_1,\dots,\lambda_n,\lambda \in K$, welche nicht alle gleich 0 sind, mit $0=\sum_{i=1}^n \lambda_i\cdot v_i+\lambda\cdot v$. Wäre $\lambda=0$, so auch $\lambda_i=0$ für alle $i=1,\dots,n$, da $\mathcal{B}$ linear unabhängig ist. Also ist $\lambda\neq 0$ und somit

$$v=\sum_{i=1}^n \frac{-\lambda_i}{\lambda}\cdot v_i\ .$$

Damit ist $\mathcal{B}$ ein Erzeugendensystem. □

Die wichtigste Eigenschaft einer Basis ist die Eindeutigkeit der Zerlegung eines gegebenen Vektors nach einer Basis des Vektorraums. Ist eine Basis fixiert, dann kann man an Stelle des Vektors $v\in V$ mit einer Folge von Zahlen $(\lambda_1,\dots,\lambda_n)$ mit $\lambda_i\in K$ arbeiten. Diese Zahlen λ_i heißen die *Koordinaten* von $v=\sum_{i=1}^n\lambda_i\cdot v_i$ bezüglich der Basis $(v_1,\dots,v_n)$.

Beispiel 8.4 Es sei $V=\mathrm{span}_{\mathbb{R}}(v_1,v_2,v_3)$ mit $v_1=\begin{pmatrix}1\\2\\4\end{pmatrix}$, $v_2=\begin{pmatrix}2\\3\\5\end{pmatrix}$, $v_3=\begin{pmatrix}1\\3\\6\end{pmatrix}$. Wir überprüfen, ob (v_1,v_2,v_3) linear unabhängig sind durch Berechnen des Rangs der Matrix $A=(v_{ij})$:

$$A=\begin{pmatrix}1&2&1\\2&3&3\\4&5&6\end{pmatrix}\xrightarrow{IV_{12}(-2)\ ,\ IV_{13}(-4)}\begin{pmatrix}1&2&1\\0&-1&1\\0&-3&2\end{pmatrix}\xrightarrow{IV_{23}(-3)}\begin{pmatrix}1&2&1\\0&-1&1\\0&0&-1\end{pmatrix}$$

Folglich ist (v_1,v_2,v_3) linear unabhängig, außerdem Erzeugendensystem von V und damit Basis. Falls $w=\begin{pmatrix}2\\2\\3\end{pmatrix}\in V$, so muß es eindeutig bestimmte Koeffizienten $\lambda_1,\lambda_2,\lambda_3\in\mathbb{R}$ geben mit $w=\lambda_1v_1+\lambda_2v_2+\lambda_3v_3$. Das ist aber gerade ein lineares Gleichungssystem

$$(A|w)=\left(\begin{array}{ccc|c}1&2&1&2\\2&3&3&2\\4&5&6&3\end{array}\right)\xrightarrow{IV_{12}(-2)\ ,\ IV_{13}(-4)}\left(\begin{array}{ccc|c}1&2&1&2\\0&-1&1&-2\\0&-3&2&-5\end{array}\right)$$

$$\xrightarrow{IV_{23}(-3)}\left(\begin{array}{ccc|c}1&2&1&2\\0&-1&1&-2\\0&0&-1&1\end{array}\right)\xrightarrow{IV_{32}(1),\ IV_{31}(1),\ III_3(-1)}\left(\begin{array}{ccc|c}1&2&0&3\\0&-1&0&-1\\0&0&1&-1\end{array}\right)$$

$$\xrightarrow{IV_{21}(2),\ III_2(-1)}\left(\begin{array}{ccc|c}1&0&0&1\\0&1&0&1\\0&0&1&-1\end{array}\right)$$

welches die Koeffizienten $\lambda_1=1$, $\lambda_2=1$, $\lambda_3=-1$ liefert. Insbesondere ist $w\in V$. ◁

Satz 8.5 *Ist V nicht endlich erzeugt, dann gibt es eine unendliche linear unabhängige Familie von Vektoren.*

Beweis. Durch Induktion nach der Anzahl n linear unabhängiger Vektoren von V: Seien $(v_1, \ldots, v_n)$ linear unabhängig. Wäre $(v_1, \ldots, v_n, v_{n+1})$ linear abhängig für jedes $v_{n+1} \in V$, so wäre $(v_1, \ldots, v_n)$ Erzeugendensystem, Widerspruch. $\square$

Satz 8.6 (Austauschlemma von Steinitz) *Es sei $(v_1, \ldots, v_n)$ Erzeugendensystem (bzw. eine Basis) eines Vektorraums V über K und $w = \sum_{i=1}^n \lambda_i v_i$ mit $\lambda_r \neq 0$ für ein $r \in \{1, \ldots, n\}$. Dann ist auch $(v_1, \ldots, v_{r-1}, w, v_{r+1}, \ldots, v_n)$ ein Erzeugendensystem (bzw. eine Basis) von V.*

Beweis. Unter den Voraussetzungen gilt

$$v_r = \frac{1}{\lambda_r} \cdot w + \sum_{i=1}^{r-1} \frac{-\lambda_i}{\lambda_r} \cdot v_i + \sum_{j=r+1}^{n} \frac{-\lambda_j}{\lambda_r} \cdot v_j \ .$$

Sei $v \in V$ ein beliebiger Vektor, dann gibt es $\mu_1, \ldots, \mu_n \in K$ mit $v = \sum_{i=1}^n \mu_i \cdot v_i$. Einsetzen von v_r liefert

$$v = \sum_{i=1}^{r-1} \Big(\mu_i - \frac{\mu_r \lambda_i}{\lambda_r}\Big) \cdot v_i + \frac{\mu_r}{\lambda_r} \cdot w + \sum_{j=r+1}^{n} \Big(\mu_j - \frac{\mu_r \lambda_j}{\lambda_r}\Big) \cdot v_j \ .$$

Damit ist $(v_1, \ldots, v_{r-1}, w, v_{r+1}, \ldots, v_n)$ ein Erzeugendensystem.

Ist $\mathcal{B} = (v_1, \ldots, v_n)$ Basis, dann betrachten wir zur Untersuchung der linearen Unabhängigkeit

$$0 = \sum_{i=1}^{r-1} \mu_i \cdot v_i + \mu \cdot w + \sum_{j=r+1}^{n} \mu_j \cdot v_j \ .$$

Einsetzen von w liefert

$$0 = \sum_{i=1}^{r-1} (\mu_i + \mu\lambda_i) \cdot v_i + \mu\lambda_r \cdot v_r + \sum_{j=r+1}^{n} (\mu_j + \mu\lambda_j) \cdot v_j \ .$$

Da $\mathcal{B}$ linear unabhängig, folgt $\mu_i + \mu\lambda_i = 0$ für $i \neq r$ und $\mu\lambda_r = 0$, also $\mu = 0$ und dann $\mu_i = 0$. $\square$

Satz 8.7 (Basisauswahlsatz) *Aus jedem endlichen Erzeugendensystem läßt sich eine Basis auswählen. Insbesondere hat jeder endlich erzeugte Vektorraum eine Basis.*

Beweis. Man nehme aus einem endlichen Erzeugendensystem einzelne Vektoren weg, so daß die reduzierte Familie immer noch ein Erzeugendensystem des Vektorraums bleibt. Ist das nicht mehr möglich, so liegt eine Basis vor. $\square$

Wir zeigen nun, daß alle Basen eines endlich erzeugten Vektorraums aus der gleichen Anzahl an Vektoren bestehen.

Satz 8.8 (Austauschsatz) *In einem Vektorraum V über K sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis und $(w_1, \ldots, w_r)$ eine linear unabhängige Familie von Vektoren aus V. Dann ist $r \leq n$, und es gibt paarweise verschiedene Indizes $i_1, \ldots, i_r \in \{1, 2, \ldots, n\}$, so daß man nach Austausch von v_{i_j} durch w_j für alle $1 \leq j \leq r$ wieder eine Basis von V erhält. Nach Permutation der Indizes (Umnumerierung) erreicht man, daß $\mathcal{B}^* = (w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ eine Basis von V ist.*

Beweis. Wir tauschen schrittweise die Basisvektoren aus. Im ersten Schritt finden wir ein $1 \leq s \leq n$, so daß $(v_1, \ldots, v_{s-1}, w_1, v_{s+1}, \ldots, v_n)$ eine Basis ist. Nun nennen wir $v_1 \mapsto v_s$ und ordnen durch Vertauschen von $v_s \leftrightarrow w_1$ die Basis um, so daß $(w_1, v_2, \ldots, v_n)$ eine Basis ist.

Dann gibt es $\mu_1, \lambda_2, \ldots, \lambda_n \in K$ mit $w_2 = \mu_1 \cdot w_1 + \sum_{i=2}^n \lambda_i \cdot v_i$. Mindestens eines der λ_i ist von 0 verschieden, denn sonst wäre $w_2 = \mu_1 \cdot w_1$, und die Familie $(w_1, \ldots, w_r)$ wäre nicht linear unabhängig. Es läßt sich deshalb ein v_s mit $2 \leq s \leq n$ durch w_2 austauschen, so daß nach Umnumerierung $(w_1, w_2, v_3 \ldots, v_n)$ eine Basis von V ist.

Ist $r \leq n$, so führt die wiederholte Anwendung dieses Verfahrens auf eine Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$. Wäre $r > n$, dann erhalten wir im n-ten Schritt eine Basis $(w_1, \ldots, w_n)$ von V. Dann ist w_{n+1} nach dieser Basis zerlegbar, so daß $(w_1, \ldots, w_r)$ für $r > n$ nicht mehr linear unabhängig wäre. □

Satz 8.9 *Sei V ein endlich erzeugter Vektorraum. Dann besteht jede Basis von V aus der gleichen Anzahl von Vektoren.*

Beweis. Seien zwei Basen $\mathcal{B}_1 = (v_1, \ldots, v_n)$ und $\mathcal{B}_2 = (w_1, \ldots, w_r)$ von V gegeben. Ist $n < r$, dann wäre nach Satz 8.8 $\mathcal{B}_2$ nicht linear unabhängig, also keine Basis. Ist $n > r$, dann wäre nach Satz 8.8 $\mathcal{B}_1$ nicht linear unabhängig, also keine Basis. Folglich gilt $n = r$. □

Definition 8.10 In einem Vektorraum V über K heißt die durch

$$\dim_K V := \begin{cases} 0 & \text{falls } V = \{0\}, \\ n & \text{falls } V \text{ endlich erzeugt ist und eine aus } n \text{ Vektoren} \\ & \text{bestehende Basis besitzt,} \\ \infty & \text{falls } V \text{ nicht endlich erzeugt ist} \end{cases}$$

definierte Zahl die *Dimension* von V.

Offenbar ist $\dim_K(K^n) = n$ und $\dim_{\mathbb{R}}(\mathbb{C}) = 2$, aber $\dim_{\mathbb{C}}(\mathbb{C}) = 1$. Wir schreiben auch $\dim(V)$ statt $\dim_K(V)$, wenn der Körper klar ist. Die Dimension ist eine entscheidende Charakterisierung eines Vektorraums.

Satz 8.11 *Ist $W \subseteq V$ Untervektorraum eines endlich erzeugten Vektorraums V, so ist auch W endlich erzeugt, und es gilt*

i) $\dim(W) \leq \dim(V)$,

ii) *aus* $\dim(W) = \dim(V)$ *folgt* $W = V$.

Beweis. i) Wäre W nicht endlich erzeugt, dann gäbe es nach Satz 8.5 eine unendliche linear unabhängige Familie, die auch in V linear unabhängig wäre, Widerspruch. Ebenso kann es nach dem Austauschsatz höchstens $n := \dim_K(V)$ linear unabhängige Vektoren in W geben.

ii) Sei $\dim(W) = \dim(V) = n$ und $(w_1, \ldots, w_n)$ eine Basis von W. Wäre $V \neq W$, so gibt es ein $v \in V \setminus W$, das keine Linearkombination von $(w_1, \ldots, w_n)$ ist. Damit wäre $(w_1, \ldots, w_n, v)$ linear unabhängig, Widerspruch. □

Dieser Satz ist sehr hilfreich. Wenn wir schon wissen, daß ein Vektorraum V die Dimension n hat und n linear unabhängige Vektoren $v_1, \ldots, v_n \in V$ gegeben sind, dann ist $W = \operatorname{span}(v_1, \ldots, v_n)$ ein Untervektorraum, in dem $(v_1, \ldots, v_n)$ eine Basis ist. Aus der Gleichheit der Dimensionen folgt $W = V$, also ist $(v_1, \ldots, v_n)$ ein Erzeugendensystem von V. Es ist wesentlich einfacher, die lineare Unabhängigkeit mit dem Gaußschen Algorithmus für lineare Gleichungssysteme zu überprüfen als die Eigenschaft des Erzeugendensystems. Daraus wird die Bedeutung der Dimension ersichtlich.

Ein weiteres nützliches Hilfsmittel ist:

Satz 8.12 (Basisergänzungssatz) *In einem endlich erzeugten Vektorraum V sei eine Familie $(w_1, \ldots, w_r)$ von linear unabhängigen Vektoren gegeben. Dann läßt sich diese Familie zu einer Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ von V ergänzen.*

Beweis. Man nehme irgendeine Basis von V und wende den Austauschsatz an. □

In nicht endlich erzeugten (also unendlich-dimensionalen) Vektorräumen gibt es einige Besonderheiten. Zwar gilt unter Benutzung des *Auswahlaxioms*, daß auch jeder unendlich-dimensionale Vektorraum eine Basis besitzt, jedoch kann man eine solche nicht angeben. Später führen wir *unitäre und euklidische Vektorräume* ein, welche auch unendlich-dimensional sein können. In diesen gibt es sogenannte Orthonormalbasen, so daß sich jeder Vektor als eindeutige unendliche Linearkombination darstellen läßt, was Methoden der Analysis erfordert. Die Orthonormalbasen sind dann keine Basen im Sinne von Definition 8.1.

Definition 8.13 Sei V ein Vektorraum und $W_1, \ldots, W_r \subseteq V$ Untervektorräume. Dann heißt

$$W = W_1 + \cdots + W_r := \{v \in V \ : \ \text{es gibt } v_j \in W_j \text{ mit } v = v_1 + \cdots + v_r \ \}$$

die *Summe* von $W_1, \ldots, W_r$.

Offenbar ist W wieder ein Untervektorraum von V, und es gilt $\dim(\sum_{i=1}^r W_i) \leq \sum_{i=1}^r \dim(W_i)$. Für $r = 2$ können wir mehr zeigen:

Satz 8.14 *Für endlich-dimensionale Untervektorräume $W_1, W_2 \subseteq V$ gilt $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$.*

Beweis. $W_1 \cap W_2$ ist ein endlich-dimensionaler Untervektorraum von V. Man nehme eine Basis $(v_1, \dots, v_m)$ von $W_1 \cap W_2$ und ergänze sie zu Basen $(v_1, \dots, v_m, w_1, \dots, w_r)$ von W_1 und $(v_1, \dots, v_m, w'_1, \dots, w'_s)$ von W_2. Damit wird $W_1 + W_2$ von $\mathcal{B} = (v_1, \dots, v_m, w_1, \dots, w_r, w'_1, \dots, w'_s)$ erzeugt. Wir zeigen: $\mathcal{B}$ ist linear unabhängig, also Basis. Dazu sei

$$0 = \underbrace{\sum_{i=1}^{m} \lambda_i v_i + \sum_{j=1}^{r} \mu_j w_j}_{=v \in W_1} + \underbrace{\sum_{k=1}^{s} \mu'_k w_k}_{=-v \in W_2} \ .$$

Das bedeutet $v \in W_1 \cap W_2$, also $\mu_j = 0$ und $\mu'_k = 0$ und somit $v = 0$. Aus der linearen Unabhängigkeit von $(v_1, \dots, v_m)$ folgt schließlich auch $\lambda_i = 0$. Damit ist ist $\dim(W_1 + W_2) = m + r + s$ sowie $\dim(W_1) = m + r$ und $\dim(W_2) = m + s$. $\square$

Um eine Basis von $W_1 \cap W_2$ zu erhalten, wählt man zunächst Basen $(w_1, \dots, w_k)$ von W_1 und $(w_{k+1}, \dots, w_l)$ von W_2. Ein Vektor aus $W_1 \cap W_2$ liegt insbesondere in W_1 und W_2 und läßt sich somit darstellen als $v = \sum_{i=1}^{k} \lambda_i v_i = \sum_{j=k+1}^{l} (-\lambda_j) w_j$. Dieses Problem läßt sich als lineare Gleichungssystem $A \cdot x =$ mit $A = (v_1, \dots, v_k, v_{k+1}, \dots, v_l)$ und $x = (\lambda_1, \dots, \lambda_k, \lambda_{k+1}, \dots, \lambda_l)$ auffassen. Ist $\operatorname{rang}(A) = r$, so gibt es $l - r$ freie Variablen unter den λ_i, d.h. es gilt $\dim W_1 \cap W_2 = l - r$. Durch Trennen in $\sum_{i=1}^{k} \lambda_i v_i = \sum_{j=k+1}^{l} (-\lambda_j) w_j$ läßt sich eine Basis von $W_1 \cap W_2$ ablesen.

Beispiel 8.15 Sei $W_1 = \operatorname{span}(v_1, v_2, v_3)$ und $W_2 = \operatorname{span}(v_4, v_5)$ mit $v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $v_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$, $v_3 = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}$ sowie $v_4 = \begin{pmatrix} 1 \\ 5 \\ 0 \end{pmatrix}$, $v_5 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$. Nach Beispiel 7.10 gilt $\dim(W_1) = 2$, insbesondere $W_1 = \operatorname{span}(v_1, v_2)$. Die Dimension $\dim(W_2) = 2$ ist klar. Setzen wir $A = (v_1, v_2, v_4, v_5)$, so suchen wir zur Bestimmung von $W_1 \cap W_2$ die Lösungsmenge $x \in \mathbb{R}^4$ des LGS $Ax = 0$ durch elementare Zeilenumformungen. Die rechte Seite $b = 0$ kann weggelassen werden:

$$\begin{pmatrix} 1 & 4 & 1 & 0 \\ 2 & 5 & 5 & 2 \\ 3 & 6 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 4 & 1 & 0 \\ 0 & -3 & 3 & 2 \\ 0 & -6 & -3 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 4 & 1 & 0 \\ 0 & 1 & -1 & -\frac{2}{3} \\ 0 & 0 & -9 & -3 \end{pmatrix}$$

$$\mapsto \begin{pmatrix} 1 & 4 & 1 & 0 \\ 0 & 1 & -1 & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 4 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} \end{pmatrix}$$

Damit ergibt sich $\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_4 \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \cdot \frac{\lambda_5}{3} = 0$ und $\lambda_5 \in \mathbb{R}$ beliebig. Das bedeutet $\sum_{i=1,2,4,5} \lambda_i v_i = \frac{\lambda_5}{3} \cdot (-3v_1 + v_2 - v_4 + 3v_5) = 0$, d.h. Linearkombinationen von

$$\underbrace{3v_1 - v_2}_{\in W_1} = \underbrace{3v_5 - v_4}_{\in W_2} = \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix}$$

spannen $W_1 \cap W_2$ auf. Somit ist $\dim(W_1 \cap W_2) = 1$ und nach der Dimensionsformel (Satz 8.14) $\dim(W_1 + W_2) = 3$. $\triangleleft$

Aus Satz 8.14 erhält man schrittweise eine allgemeine Dimensionsformel für Summen von Vektorräumen. Dabei entstehen komplizierte Durchschnitte $(W_l \cap (W_1 + \cdots + W_{l-1})$ für $l = 2, \ldots, r$, die sich im allgemeinen nicht besser ausdrücken lassen. Besonders transparent ist folgende Situation:

Definition 8.16 Sei V ein Vektorraum und $W_1, \ldots, W_k$ Untervektorräume von V. Ein Vektorraum W heißt *direkte Summe* der Untervektorräume $W_1, \ldots, W_k$, bezeichnet mit

$$W = W_1 \oplus W_2 \oplus \cdots \oplus W_k ,$$

wenn gilt:

(DS1) $W = W_1 + \cdots + W_k$

(DS2) Von Null verschiedene Vektoren $w_1 \in W_1, \ldots, w_k \in W_k$ sind linear unabhängig in V.

In diesem Fall gilt $\dim(W) = \sum_{i=1}^k \dim(W_i)$.

Zusammenfassung Teil II

Begriffe

- Lineare Gleichungssysteme, Matrix-Schreibweise, erweiterte Koeffizientenmatrix
- Gaußsches Lösungsverfahren: spezielle Zeilenstufenform, Rangbedingung
- Vektorräume, Untervektorräume, lineare Unabhängigkeit, Linearkombinationen
- Erzeugendensystem, Basis, Dimension, Summen und Durchschnitte von Vektorräumen

Methoden

- Lösung linearer Gleichungssysteme
- Bestimmung des Rangs einer Matrix
- Überprüfung einer Familie von Vektoren auf lineare Unabhängigkeit
- Zerlegung eines Vektors nach einer Basis
- Bestimmung von Basen für $W_1 + W_2$ und $W_1 \cap W_2$ für Untervektorräume $W_1, W_2 \subseteq V$.

Teil III
Folgen und Reihen

9 Folgen und Grenzwerte

Unter einer *Folge* $(a_n)_{n\in\mathbb{N}}$ komplexer Zahlen versteht man eine Abbildung $a : \mathbb{N} \to \mathbb{C}$, d.h. jedem $n \in \mathbb{N}$ wird eine komplexe Zahl $a_n \in \mathbb{C}$ zugeordnet. Analog ist eine Folge reeller Zahlen definiert.

Definition 9.1 Eine Folge $(a_n)_{n\in\mathbb{N}}$ heißt *konvergent*, wenn es eine Zahl $a \in \mathbb{C}$ gibt mit folgender Eigenschaft: Zu jedem $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$, so daß $|a_n - a| < \epsilon$ für alle $n \geq N$. Die Zahl a heißt *Grenzwert* oder *Limes* der Folge $(a_n)_{n\in\mathbb{N}}$, geschrieben $a = \lim\limits_{n\to\infty} a_n$. Eine Folge, die gegen 0 konvergiert, heißt *Nullfolge*. Eine Folge, die nicht konvergent ist, heißt *divergent*.

Wichtig ist, daß das $N \in \mathbb{N}$ in der Definition von ϵ abhängt. Verkleinert man ϵ, dann muß im allgemeinen N vergrößert werden.

Im Fall der Konvergenz ist der Grenzwert einer Folge eindeutig: Angenommen, a, b wären Grenzwerte der Folge $(a_n)_{n\in\mathbb{N}}$, dann gibt es zu jedem $\epsilon = \frac{1}{3}|a-b| > 0$ ein $N \in \mathbb{N}$ mit $|a_n - a| < \epsilon$ und $|a_n - b| < \epsilon$. Nach der Dreicksungleichung gilt aber $|a-b| = |(a-a_n)+(a_n-b)| \leq |a_n - a| + |a_n - b| < 2\epsilon$, im Widerspruch zu $2\epsilon = \frac{2}{3}|a-b|$. Geometrisch bedeutet Definition 9.1, daß alle Folgenglieder a_n mit $n \geq N$ in der Kreisscheibe

$$K_\epsilon(a) := \{z \in \mathbb{C} \ : \ |z-a| < \epsilon\}$$

in $\mathbb{C}$ mit Mittelpunkt a und Radius ϵ liegen, bzw. für reelle Folgen im offenen Intervall $I_\epsilon(a) = \{x \in \mathbb{R} \ : \ |x-a| < \epsilon\}$ mit Mittelpunkt a und halber Länge (Radius) ϵ. Die Teilmengen $K_\epsilon(a) \subseteq \mathbb{C}$ bzw. $I_\epsilon(a) \subseteq \mathbb{R}$ heißen auch die *ϵ-Umgebungen* von a.

Beispiel 9.2
i) Die konstante Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_n = a$ für alle $n \in \mathbb{N}$ konvergiert gegen a.

ii) Die Folge $(\frac{1}{n+1})_{n\in\mathbb{N}}$ konvergiert nach Satz 3.11.ii) gegen 0, ist also eine Nullfolge.

iii) Die Folge $((-1)^n)_{n\in\mathbb{N}}$ divergiert: Angenommen, $a \in \mathbb{C}$ wäre Grenzwert dieser Folge, dann gäbe es für $\epsilon = 1$ ein $N \in \mathbb{N}$ mit $|(-1)^n - a| < 1$ für alle $n \geq N$. Damit wäre nach der Dreiecksungleichung

$$\begin{aligned} 2 &= |(-1)^{n+1} - (-1)^n| = |(-1)^{n+1} - a + a - (-1)^n| \\ &\leq |(-1)^{n+1} - a| + |a - (-1)^n| < 2\,, \qquad \text{Widerspruch.} \end{aligned}$$

◁

Ist $(a_n)_{n\in\mathbb{N}}$ konvergent gegen a und gibt es für die Folge $(b_n)_{n\in\mathbb{N}}$ ein $m \in \mathbb{Z}$ mit $a_n = b_{n+m}$ für alle $n \geq N$, so ist auch $(b_n)_{n\in\mathbb{N}}$ konvergent mit $\lim\limits_{n\to\infty} b_n = a$. Damit können in einer konvergenten Folge die Folgenglieder a_n mit $n \leq N$ beliebig abgeändert werden, insbesondere auch weggelassen oder endlich viele Glieder vor a_n eingeschoben werden, ohne Konvergenz und Grenzwert zu ändern.

Satz 9.3 *Es seien $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ Folgen mit $\lim\limits_{n\to\infty} a_n = a$ und $\lim\limits_{n\to\infty} b_n = b$. Dann gilt:*

i) $\lim\limits_{n\to\infty}(a_n + b_n) = a + b$

ii) $\lim\limits_{n\to\infty}(a_n \cdot b_n) = ab$

iii) *Ist $b \neq 0$, dann gibt es ein $N \in \mathbb{N}$ mit $b_n \neq 0$ für alle $n \geq N$, und* $\lim\limits_{n\to\infty} \dfrac{a_{n+N}}{b_{n+N}} = \dfrac{a}{b}$

Beweis. i) Zu gegebenem $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$ mit $|a_n - a| < \frac{\epsilon}{2}$ und $|b_n - b| < \frac{\epsilon}{2}$ für alle $n \geq N$ (man wähle die größere der Schranken N beider Folgen). Die Dreiecksungleichung liefert

$$|a_n + b_n - (a+b)| \leq |a_n - a| + |b_n - b| < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} \qquad \text{für alle } n \geq N\,.$$

ii) Wähle N derart, daß $|a_n - a| < \min(1, \frac{\epsilon}{2(|b|+1)})$ und $|b_n - b| < \frac{\epsilon}{2(|a|+1)}$ für alle $n \geq N$. Dann ist $|a_n| = |a_n - a + a| \leq |a_n - a| + |a| < |a| + 1$ und

$$\begin{aligned}|a_n b_n - ab| &= |a_n(b_n - b) + (a_n - a)b| \leq |a_n(b_n - b)| + |(a_n - a)b| \\ &= |a_n|\,|b_n - b| + |a_n - a|\,|b| < (|a|+1)\frac{\epsilon}{2(|a|+1)} + \frac{\epsilon}{2(|b|+1)}|b| < \epsilon\,.\end{aligned}$$

iii) Wähle N derart, daß $|b_n - b| < \min(\frac{1}{2}|b|, \frac{\epsilon}{2}|b|^2)$ für alle $n \geq N$. Für diese n gilt dann $|b| = |(b - b_n) + b_n| \leq |b - b_n| + |b_n|$, also $|b_n| \geq |b| - |b - b_n| > \frac{1}{2}|b| > 0$. Weiter folgt für $n \geq N$

$$\Big|\frac{1}{b_n} - \frac{1}{b}\Big| = \Big|\frac{b - b_n}{b_n b}\Big| = \frac{|b_n - b|}{|b|\,|b_n|} < \frac{2|b_n - b|}{|b|^2} < \epsilon\,.$$

Somit gilt $\lim\limits_{n\to\infty} \dfrac{1}{b_{N+n}} = \dfrac{1}{b}$, und mit ii) folgt die Behauptung. □

Folgerung 9.4 *Die Menge der konvergenten Folgen bildet einen Vektorraum mit Addition $(a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} := (a_n + b_n)_{n\in\mathbb{N}}$ und skalarer Multiplikation $\lambda \cdot (a_n)_{n\in\mathbb{N}} := (\lambda a_n)_{n\in\mathbb{N}}$. Dieser Vektorraum wird oft mit c bezeichnet. Die Menge der Nullfolgen bildet einen Untervektorraum $c_0 \subseteq c$. Ein weiterer Untervektorraum von c ist gegeben durch die Menge d der Folgen, in denen nur endlich-viele Folgenglieder ungleich Null sind. Insbesondere ist $d \subseteq c_0$ Untervektorraum.*

Für $k \in \mathbb{N}$ seien Folgen $\delta_k = (\delta_{kn})_{n\in\mathbb{N}} \in d$ definiert durch $\delta_{kn} = \begin{cases} 1 & \text{falls } k = n\,, \\ 0 & \text{sonst}\,. \end{cases}$

Die Familie $(\delta_k)_{k\in\mathbb{N}}$ ist Basis von d, da jede Folge aus d sich als eindeutige endliche Linearkombination der $(\delta_k)_{k\in\mathbb{N}}$ schreiben läßt. Insbesondere sind d, und damit erst recht c_0 und c, unendlich-dimensionale Vektorräume. Zu beachten ist, daß in c_0 oder c die Familie $(\delta_k)_{k\in\mathbb{N}}$ zwar linear unabhängig ist, aber *kein* Erzeugendensystem und damit keine Basis ist!

Satz 9.5 (elementare Grenzwerte) *Im folgenden sei $n \in \mathbb{N}^\times$.*

i) $\lim\limits_{n\to\infty} \dfrac{1}{n^s} = 0$ *für alle $s \in \mathbb{Q}_+^\times$. (Dabei ist $n^{\frac{p}{q}} = \sqrt[q]{n^p} \in \mathbb{R}$.)*

ii) $\lim\limits_{n\to\infty} \sqrt[n]{a} = 1$ *für alle $a \in \mathbb{R}_+^\times$.*

iii) $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$.

iv) $\lim\limits_{n\to\infty} z^n = 0$ *für alle $z \in \mathbb{C}$ mit $|z| < 1$.*

v) $\lim\limits_{n\to\infty} \dfrac{n^k}{z^n} = 0$ *für alle $z \in \mathbb{C}$ mit $|z| > 1$ und alle $k \in \mathbb{N}$.*

Beweis. i) Zu gegebenem $\epsilon > 0$ wähle nach dem Archimedischen Axiom $N \in \mathbb{N}$ derart, daß $N > \epsilon^{-\frac{1}{s}}$. Dann gilt für $n \geq N$ zunächst $\frac{1}{n} \leq \frac{1}{N}$, dann $\frac{1}{n^p} \leq \frac{1}{N^p}$ und $\sqrt[q]{\frac{1}{n^p}} \leq \sqrt[q]{\frac{1}{N^p}}$, also $|\frac{1}{n^s}| \leq |\frac{1}{N^s}| < \epsilon$.

ii) Für $a \geq 1$ setze $x_n := \sqrt[n]{a} - 1 > 0$. Nach der Bernoullischen Ungleichung ist $a = (1 + x_n)^n \geq 1 + nx_n$, damit $x_n < \frac{a}{n}$ und $0 < |\sqrt[n]{a} - 1| = x_n < \epsilon$ für alle $n \geq N > \frac{a}{\epsilon}$. Für $0 < a < 1$ gilt $\lim\limits_{n\to\infty} \sqrt[n]{\dfrac{1}{a}} = 1$. Die Behauptung folgt dann aus Satz 9.3.iii).

iii) Sei $n \geq 2$. Nach Binomialentwicklung gilt für $x_n := \sqrt[n]{n} - 1 > 0$

$$n = (1 + x_n)^n \geq 1 + \binom{n}{2} x_n^2 = 1 + \frac{n(n-1)}{2} x_n^2\,,$$

also $n - 1 \geq \frac{n(n-1)}{2} x_n^2$ und schließlich $x_n \leq \sqrt{\frac{2}{n}}$ für $n \geq 2$. Wähle $N > \frac{2}{\epsilon^2}$, so folgt $|\sqrt[n]{n} - 1| = x_n < \epsilon$ für alle $n \geq N$.

iv) Es gilt $|z^n - 0| = |z|^n$. Die Aussage folgt aus Satz 3.11.ii).

v) Betrachte $n > 2k + 1$ und $|z| = 1 + x$ mit $x > 0$. Dann ergibt die Binomialentwicklung

$$(1 + x)^n > \binom{n}{k+1} x^{k+1} = \frac{n(n-1)\cdots(n-k)}{(k+1)!} x^{k+1} > n \cdot \left(\frac{n}{2}\right)^k \frac{x^{k+1}}{(k+1)!}\,,$$

also $\left|\frac{n^k}{z^n}\right| < \frac{2^k(k+1)!}{x^{k+1}} \cdot \frac{1}{n}$. Wähle $N \in \mathbb{N}$ mit $N > \max\left(\frac{2^{k+1}(k+1)!}{x^{k+1}} \cdot \frac{1}{\epsilon}\,,\ 2k+1\right)$, dann folgt $\left|\frac{n^k}{z^n}\right| < \epsilon$ für alle $n \geq N$. $\square$

Beispiel 9.6 Es sei $n \in \mathbb{N}^\times$.

i) $\displaystyle\lim_{n\to\infty} \frac{2n^2+3n}{3n^2+4} = \lim_{n\to\infty} \frac{2+\frac{3}{n}}{3+\frac{4}{n^2}} = \frac{\lim\limits_{n\to\infty} 2+\dfrac{3}{n}}{\lim\limits_{n\to\infty} 3+\dfrac{4}{n^2}} = \frac{2}{3}$

ii) Für $k \in \mathbb{N}$ gilt: $\lim\limits_{n\to\infty} \sqrt[n]{n^k} = \lim\limits_{n\to\infty} (\sqrt[n]{n})^k = (\lim\limits_{n\to\infty} (\sqrt[n]{n}))^k = 1.$ ◁

Einige weitere Rechenregeln für Grenzwerte:

Satz 9.7 i) *Die Folge $(a_n)_{n\in\mathbb{N}}$ konvergiere gegen $a \in \mathbb{C}$. Dann gilt*

$$\lim_{n\to\infty} |a_n| = |a| \;, \quad \lim_{n\to\infty} \overline{a_n} = \bar{a} \;, \quad \lim_{n\to\infty} \mathrm{Re}(a_n) = \mathrm{Re}(a) \;, \quad \lim_{n\to\infty} \mathrm{Im}(a_n) = \mathrm{Im}(a) \;.$$

Insbesondere sind die Grenzwerte reeller Folgen reell, und es gilt

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \mathrm{Re}(a_n) + \mathrm{i} \cdot \lim_{n\to\infty} \mathrm{Im}(a_n) \;.$$

ii) *Es gelte $\lim\limits_{n\to\infty} a_n = a$ und $\lim\limits_{n\to\infty} b_n = b$, ferner gebe es ein $N \in \mathbb{N}$ mit $a_n \leq b_n$ für alle $n \geq N$. Dann gilt $a \leq b$.*

iii) *Zu einer reellen Folge $(c_n)_{n\in\mathbb{N}}$ gebe es konvergente reelle Folgen $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ und ein $N \in \mathbb{N}$ derart, daß $a_n \leq c_n \leq b_n$ für alle $n \geq N$. Gilt $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n = a$, dann konvergiert auch $(c_n)_{n\in\mathbb{N}}$ mit $\lim\limits_{n\to\infty} c_n = a$.*

iv) *Für eine Folge $(a_n)_{n\in\mathbb{N}}$ komplexer Zahlen gebe es ein $N \in \mathbb{N}$ und ein $0 < q < 1$, so daß für alle $n \geq N$ gilt $a_n \neq 0$ und $|\frac{a_{n+1}}{a_n}| \leq q$. Dann gilt $\lim\limits_{n\to\infty} a_n = 0$.*

Beweis. i) Nach Satz 5.4 gilt $|\mathrm{Re}(a_n) - \mathrm{Re}(a)| = |\mathrm{Re}(a_n - a)| \leq |a_n - a|$, damit folgt aus der Konvergenz von $(a_n)_{n\in\mathbb{N}}$ die Konvergenz von $(\mathrm{Re}(a_n))_{n\in\mathbb{N}}$. Analog für $\mathrm{Im}(a_n)$ und $\overline{a_n}$. Für die Beträge folgt die Aussage aus Satz 3.8: $||a_n| - |a|| \leq |a_n - a|$. Der zweite Teil ist klar.

ii) Nach Voraussetzung gibt es zu $\epsilon > 0$ ein $N \in \mathbb{N}$, so daß gleichzeitig $a_n \leq b_n$ sowie $|a_n - a| < \epsilon$ und $|b_n - b| < \epsilon$ gilt für alle $n \geq N$. Nach Fallunterscheidung bedeutet das $-\epsilon < a_n - a < \epsilon$ und $-\epsilon < b_n - b < \epsilon$, somit schließlich $a - \epsilon < a_n \leq b_n < b + \epsilon$. Das ergibt $a - b < 2\epsilon$ für alle $\epsilon > 0$, also $a - b \leq 0$.

iii) Wie in ii) gilt $a - \epsilon < a_n \leq c_n \leq b_n < a + \epsilon$ für alle $n \geq N$, also $|c_n - a| < \epsilon$.

iv) Aus $|\frac{a_{n+1}}{a_n}| \leq q$ folgt $|a_{n+1}| \leq q|a_n|$, damit für $n = N + k$ die Relation

$$|a_n| \leq q|a_{n-1}| \leq q^2|a_{n-2}| \cdots \leq q^k|a_N| = q^n \frac{|a_N|}{q^N} \;.$$

Zu jedem $\epsilon > 0$ gibt es nach Satz 3.11.ii) ein $N' \in \mathbb{N}$ mit $q^n < \epsilon\frac{q^N}{|a_N|}$ für alle $n \geq N'$. Damit gilt $|a_n - 0| < \epsilon$ für alle $n \geq \max(N', N)$. □

Beispiel 9.8 i) Für $s \in \mathbb{Q}_+^*$ gilt $1 \leq \sqrt[n]{n^s} \leq \sqrt[n]{n^k}$ für ein $k \geq s$ und damit $\lim_{n\to\infty} \sqrt[n]{n^s} = 1$.

ii) Für $0 \leq a < b$ gilt $b \leq \sqrt[n]{a^n + b^n} \leq \sqrt[n]{2(b^n)} = b\sqrt[n]{2}$, damit $\lim_{n\to\infty} \sqrt[n]{a^n + b^n} = b$. ◁

Entsprechend der allgemeinen Definition beschränkter Teilmengen heißt eine Folge $(a_n)_{n\in\mathbb{N}}$ beschränkt, wenn es ein $s \in \mathbb{R}$ gibt mit $|a_n| \leq s$ für alle $n \in \mathbb{N}$.

Satz 9.9 *Jede konvergente Folge ist beschränkt.*

Beweis. Es sei a der Grenzwert einer konvergenten Folge $(a_n)_{n\in\mathbb{N}}$, dann gibt es zu $\epsilon = 1$ ein $N \in \mathbb{N}$ mit $|a_n - a| < 1$ für alle $n \geq N$. Dann ist $|a_n| = |a_n - a + a| \leq |a_n - a| + |a| < |a| + 1$ für alle $n \geq N$, und insgesamt gilt $|a_n| \leq s := \max\big(|a_0|, |a_1|, \ldots, |a_N|, |a| + 1\big)$. □

Die Umkehrung wäre falsch: Aus der Beschränktheit einer Folge folgt nicht die Konvergenz, wie das Beispiel $((-1)^n)_{n\in\mathbb{N}}$ zeigt.

Definition 9.10 Eine Folge $(a_n)_{n\in\mathbb{N}}$ *reeller* Zahlen heißt

i) *monoton wachsend* bzw. *streng monoton wachsend*, wenn für alle $n \in \mathbb{N}$ gilt $a_n \leq a_{n+1}$ bzw. $a_n < a_{n+1}$;

ii) *monoton fallend* bzw. *streng monoton fallend*, wenn für alle $n \in \mathbb{N}$ gilt $a_n \geq a_{n+1}$ bzw. $a_n > a_{n+1}$;

iii) *(streng) monoton*, wenn sie (streng) monoton wachsend oder (streng) monoton fallend ist.

Bemerkung: Monotonie kann nicht für Folgen komplexer Zahlen definiert werden.

Satz 9.11 *Jede beschränkte monotone Folge $(a_n)_{n\in\mathbb{N}}$ konvergiert*

i) *gegen* $\sup A$*, falls die Folge monoton wachsend ist,*

ii) *gegen* $\inf A$*, falls die Folge monoton fallend ist,*

wobei $A := \{a_n \ : \ n \in \mathbb{N}\} \subseteq \mathbb{R}$.

Beweis. i) Sei $s := \sup A$, d.h. die kleinste obere Schranke. Dann gibt es zu jedem $\epsilon > 0$ ein $a_N \in A$ mit $s - \epsilon < a_N$. Da $(a_n)_{n\in\mathbb{N}}$ monoton wachsend ist, gilt $s - \epsilon < a_n \leq s$, d.h. $|a_n - s| < \epsilon$, für alle $n \geq N$. ii) ist analog. □

Satz 9.12 (Eulersche Zahl e) *Für $n \in \mathbb{N}^\times$ sei $a_n := (1 + \frac{1}{n})^n$. Dann gilt: Die Folge $(a_n)_{n\geq 1}$ ist konvergent, und für ihren Grenzwert $e := \lim_{n\to\infty}(1 + \frac{1}{n})^n$ gilt* $2 < e \leq 3$.

Genauer gilt: e ist irrational mit $e = 2,71828\ldots$. Wir werden später sehen, daß e in der Analysis eine wichtige Rolle spielt.

Beweis. Wir zeigen: $(a_n)_{n\in\mathbb{N}^\times}$ ist monoton wachsend und beschränkt. Monotonie folgt aus

$$\begin{aligned}\frac{a_n}{a_{n-1}} &= \frac{(\frac{n+1}{n})^n}{(\frac{n}{n-1})^{n-1}} = \frac{n}{n-1}\Big(\frac{\frac{n+1}{n}}{\frac{n}{n-1}}\Big)^n = \frac{n}{n-1}\Big(\frac{n^2-1}{n^2}\Big)^n = \frac{n}{n-1}\Big(1-\frac{1}{n^2}\Big)^n \\ &\overset{\text{Bernoulli}}{>} \frac{n}{n-1}\Big(1-n\frac{1}{n^2}\Big) = 1\;,\end{aligned}$$

d.h. $a_n > a_{n-1}$ für alle $n \geq 2$. Weiter gilt nach der binomischen Formel für $n \geq 2$

$$\begin{aligned}\Big(1+\frac{1}{n}\Big)^n &= \sum_{k=0}^n \binom{n}{k}\frac{1}{n^k} = 1 + \sum_{k=1}^n \frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}\cdot\frac{1}{k!} < 1 + \sum_{k=1}^n \frac{1}{k!} \\ &\leq 1 + \sum_{k=1}^n \frac{1}{2^{k-1}} = 1 + \sum_{k=0}^{n-1}\frac{1}{2^k} = 1 + \frac{1-\frac{1}{2^n}}{1-\frac{1}{2}} < 3\;.\end{aligned}$$

Dabei wurde für $k \geq 1$ die Ungleichung $\frac{1}{1\cdot 2\cdot 3\cdots k} \leq \frac{1}{2^{k-1}}$ sowie im letzten Schritt die geometrische Summenformel benutzt. Also ist a_n beschränkt mit $a_n < 3$, und nach Satz 9.11 konvergiert die Folge. Eine untere Schranke ergibt sich z.B. für $n = 2$ zu $(1+\frac{1}{2})^2 = \frac{9}{4} > 2$. □

Oft sind Folgen rekursiv definiert durch einen Startwert a_0 (oder mehrere Startwerte $a_0, a_1, \ldots, a_N$) und eine Rekursionsformel.

Satz 9.13 (Folge zur Berechnung der Quadratwurzeln) *Es sei $a \in \mathbb{R}^*_+$. Für einen beliebigen Startwert $x_0 > 0$ konvergiert die rekursiv definierte Folge*

$$x_{n+1} := \frac{1}{2}\Big(x_n + \frac{a}{x_n}\Big) \qquad \textit{für } n \in \mathbb{N}$$

gegen $\sqrt{a} \in \mathbb{R}^*_+$, d.h. $\lim\limits_{n\to\infty} x_n = \sqrt{a}$.

Beweis. Nach Induktion ist $x_n > 0$ für alle $n \in \mathbb{N}$. Weiter gilt

$$x_{n+1}^2 - a = \frac{1}{4}\Big(x_n + \frac{a}{x_n}\Big)^2 - a = \frac{1}{4}\Big(x_n - \frac{a}{x_n}\Big)^2 \geq 0\;,$$

d.h. $x_n \geq \sqrt{a}$ für alle $n \geq 1$. Damit ergibt sich, daß die Folge $(x_n)_{n\geq 1}$ monoton fallend ist:

$$x_n - x_{n+1} = \frac{1}{2x_n}(x_n^2 - a) > 0 \qquad n \geq 1\;.$$

Nach Satz 9.11 konvergiert $(x_n)_{n\in\mathbb{N}}$ gegen einen Grenzwert $x \geq \sqrt{a}$. Für diesen Grenzwert gilt die Gleichung $x = \frac{1}{2}(x + \frac{a}{x})$, mit der positiven Lösung $x = \sqrt{a}$. □

Das Rekursionsverfahren hat mehrere interessante Eigenschaften: 1) Es ist stabil, also unabhängig vom Startwert. 2) Man kann zeigen, daß die Folge sehr schnell (quadratisch) gegen den Grenzwert konvergiert, so daß schon wenige Folgenglieder eine brauchbare Näherung liefern.

Allgemeiner gilt (Beweis ist analog):

Satz 9.14 (Folge zur Berechnung der k-ten Wurzeln) *Es sei $a \in \mathbb{R}_+^*$. Für einen beliebigen Startwert $x_0 > 0$ konvergiert die rekursiv definierte Reihe*

$$x_{n+1} := \frac{1}{k}\Big((k-1)x_n + \frac{a}{x_n^{k-1}}\Big) \qquad \text{für } n \in \mathbb{N}$$

gegen $\sqrt[k]{a} \in \mathbb{R}_+^*$, d.h. $\lim\limits_{n\to\infty} x_n = \sqrt[k]{a}$.

10 Der Satz von Bolzano-Weierstraß. Cauchy-Folgen

Definition 10.1 Es sei $(a_n)_{n\in\mathbb{N}}$ eine Folge komplexer Zahlen und $n_0 < n_1 < n_2 < \dots$ eine aufsteigende Folge natürlicher Zahlen. Dann heißt die Folge $(a_{n_k})_{k\in\mathbb{N}} = (a_{n_0}, a_{n_1}, a_{n_2}, \dots)$ eine *Teilfolge* von $(a_n)_{n\in\mathbb{N}}$.

Direkt aus der Definition folgt, daß jede Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ einer konvergenten Folge $(a_n)_{n\in\mathbb{N}}$ konvergent ist gegen den gleichen Grenzwert, d.h. für konvergente Folgen gilt $\lim\limits_{n\to\infty} a_n = \lim\limits_{k\to\infty} a_{n_k}$.

Definition 10.2 Eine Zahl $h \in \mathbb{C}$ heißt *Häufungspunkt* der Folge $(a_n)_{n\in\mathbb{N}}$, wenn es eine Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ gibt, die gegen h konvergiert.

Satz 10.3 *Eine Zahl $h \in \mathbb{C}$ ist genau dann Häufungspunkt der Folge $(a_n)_{n\in\mathbb{N}}$, wenn für alle $\epsilon > 0$ gilt: Die Menge $\{n \in \mathbb{N} \,:\, |h - a_n| < \epsilon\}$ ist unendlich.*

Beweis. ($\Rightarrow$) Es gebe eine gegen h konvergente Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$. Also gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $|a_{n_k} - h| < \epsilon$ für alle $k \geq N$. Die Menge $\{n_k \in \mathbb{N} \,:\, k \geq N\}$ ist unendlich.

($\Leftarrow$) Jede ϵ-Umgebung von h enthalte unendlich viele Folgenglieder. Wir konstruieren induktiv eine Abbildung $k \to n_k$ mit $n_{k+1} > n_k$ und $|a_{n_k} - h| < \frac{1}{k+1}$, also eine gegen h konvergente Teilfolge.
i) Induktionsanfang: Wegen $\{n \in \mathbb{N} \,:\, |a_n - h| < 1\} \neq \varnothing$ gibt es ein $n_0 \in \mathbb{N}$ mit $|a_{n_0} - h| < 1$.
ii) Induktionsschritt $k \to k+1$: Sei a_{n_k} gewählt mit $|a_{n_k} - h| < \frac{1}{k+1}$. Nach Voraussetzung ist die Menge $\{n \in \mathbb{N} \,:\, |a_n - h| < \frac{1}{k+2}\}$ unendlich, insbesondere enthält sie eine Zahl $m > n_k$. Setze $n_{k+1} := m$. □

Beispiel 10.4 i) Die alternierende Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_n := (-1)^n$ besitzt die beiden Häufungspunkte 1 und -1, denn $(a_{2n})_{n\in\mathbb{N}}$ konvergiert gegen 1 und $(a_{2n+1})_{n\in\mathbb{N}}$ gegen -1.

ii) Die durch $a_n := (-1)^n + \frac{1}{n+1}$ gegebenen Folge hat ebenfalls 1 und -1 als Häufungspunkte, denn $\lim\limits_{n\to\infty} a_{2n} = \lim\limits_{n\to\infty}(1 + \dfrac{1}{n+1}) = 1$ und $\lim\limits_{n\to\infty} a_{2n+1} = \lim\limits_{n\to\infty}(-1 + \dfrac{1}{n+1}) = -1$.

iii) Die durch $a_n = n$ definierte Folge hat keine Häufungspunkte: jede Teilfolge enthält unendlich viele Elemente, die paarweise einen Abstand ≥ 1 haben.

iv) Eine konvergente Folge hat ihren Grenzwert als einzigen Häufungspunkt, da jede Teilfolge gegen den gleichen Grenzwert konvergiert.

Satz 10.5 (Bolzano-Weierstraß) *Jede beschränkte Folge komplexer Zahlen besitzt eine konvergente Teilfolge.*

Beweis. Wir zeigen den Satz zunächst für Folgen reeller Zahlen $a_n \in \mathbb{R}$. Wegen der Beschränktheit der Folge gibt es reelle Zahlen $c < d$ mit $a_n \in [c, d]$ für alle $n \in \mathbb{N}$. Wir konstruieren induktiv eine Intervallschachtelung $(I_k)_{k\in\mathbb{N}}$, so daß $I_k = [c_k, d_k]$ unendlich viele Folgenglieder a_n enthält, sowie eine zugehörige Folge $(n_k)_{k\in\mathbb{N}}$ natürlicher Zahlen mit $n_{k+1} > n_k$ und $a_{n_k} \in I_k$.
i) Induktionsanfang: Wähle $I_0 = [c_0, d_0]$ mit $c_0 := c$ und $d_0 = d$ sowie ein $a_{n_0} \in I_0$.
ii) Induktionsschritt $k \to k+1$. Sei $I_k = [c_k, d_k]$ bereits konstruiert und $m_k := \frac{1}{2}(c_k + d_k)$ der Mittelpunkt. Da I_k unendlich viele Folgenglieder enthält, können nicht beide Teilintervalle $[c_k, m_k]$ und $[m_k, d_k]$ nur endlich viele Folgenglieder enthalten. Setze $I_{k+1} := [c_{k+1}, d_{k+1}]$ mit $c_{k+1} := c_k$ und $d_{k+1} := m_k$, falls $[c_k, m_k]$ unendlich viele Folgenglieder enthält, ansonsten $I_{k+1} := [c_{k+1}, d_{k+1}]$ mit $c_{k+1} := m_k$ und $d_{k+1} := d_k$. Offenbar gilt $I_{k+1} \subseteq I_k$ und $|I_{k+1}| = \frac{1}{2}|I_k| = \frac{1}{2^{k+1}}|I_0|$. Da die Menge $\{n \in \mathbb{N} \,:\, a_n \in I_{k+1}\}$ unendlich ist, enthält sie ein Element $m > n_k$. Setze $n_{k+1} := m$, d.h. es gilt $a_{n_{k+1}} \in I_{k+1}$ und $n_{k+1} > n_k$.

Nach dem Intervallschachtelungsprinzip gibt es eine eindeutig bestimmte Zahl $h \in \mathbb{R}$ mit $h \in I_k$ für alle $k \in \mathbb{N}$. Dann gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $I_k \subseteq [h - \epsilon, h + \epsilon]$ für alle $k \geq N$, d.h. $|a_{n_k} - h| \leq \epsilon < 2\epsilon$ für alle $k \geq N$. Das bedeutet $\lim\limits_{k\to\infty} a_{n_k} = h$, und der Satz von Bolzano-Weierstraß ist im reellen Fall bewiesen.

Ist $(a_n)_{n\in\mathbb{N}}$ eine beschränkte Folge komplexer Zahlen, dann bildet $(\mathrm{Re}(a_n))_{n\in\mathbb{N}}$ eine beschränkte Folgen reeller Zahlen. Nach obigem Beweis gibt es eine konvergente Teilfolge $(\mathrm{Re}(a_{n_k}))_{k\in\mathbb{N}}$ mit $\lim_{k\to\infty} \mathrm{Re}(a_{n_k}) = h_1$. Die Folge der Imaginärteile $(\mathrm{Im}(a_{n_k}))_{k\in\mathbb{N}}$ ist ebenfalls eine beschränkte Folge reeller Zahlen und enthält somit eine konvergente Teilfolge $(\mathrm{Im}(a_{(n_k)_l}))_{l\in\mathbb{N}}$ mit $\lim_{l\to\infty} \mathrm{Im}(a_{(n_k)_l}) = h_2$. Dann gilt auch $\lim_{l\to\infty} \mathrm{Re}(a_{(n_k)_l}) = h_1$ als Teilfolge einer konvergenten Folge. Insgesamt ist eine konvergente komplexe Teilfolge $(a_{(n_k)_l})_{l\in\mathbb{N}}$ von $(a_n)_{n\in\mathbb{N}}$ konstruiert mit $\lim\limits_{l\to\infty} a_{(n_k)_l} = h_1 + \mathrm{i}h_2$. □

Nach Definition 10.2 kann man den Satz von Bolzano-Weierstraß auch so formulieren, daß jede beschränkte Folge komplexer Zahlen mindestens einen Häufungspunkt besitzt.

Definition 10.6 Es sei $(a_n)_{n\in\mathbb{N}}$ eine beschränkte Folge *reeller* Zahlen.

- Der größte Häufungspunkt a^* von $(a_n)_{n\in\mathbb{N}}$ heißt *limes superior*, geschrieben $a^* = \limsup a_n$.
- Der kleinste Häufungspunkt a_* von $(a_n)_{n\in\mathbb{N}}$ heißt *limes inferior*, geschrieben $a_* = \liminf a_n$.

Für konvergente Folgen gilt $a_* = a^* = \lim_{n\to\infty} a_n$.

Der Beweis des Satzes von Bolzano-Weierstraß nutzt ganz entscheidend das Intervallschachtelungsprinzip. Unter Verwendung des Satzes von Bolzano-Weierstraß beweisen wir nun das Cauchysche Konvergenzkriterium, das es auch ohne Kenntnis des Grenzwertes erlaubt, die Konvergenz von Folgen zu entscheiden.

Satz 10.7 (Cauchysches Konvergenzkriterium) *Eine Folge $(a_n)_{n\in\mathbb{N}}$ komplexer Zahlen konvergiert genau dann, wenn es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ gibt, so daß $|a_n - a_m| < \epsilon$ gilt für alle $n, m \geq N$.*

Beweis. ($\Rightarrow$) Die Folge $(a_n)_{n\in\mathbb{N}}$ sei konvergent mit Grenzwert a. Dann gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $|a_n - a| < \frac{\epsilon}{2}$ und $|a_m - a| < \frac{\epsilon}{2}$ für alle $n, m \geq N$. Die Dreiecksungleichung liefert $|a_n - a_m| = |a_n - a + (a - a_m)| \leq |a_n - a| + |a_m - a| < \epsilon$.

($\Leftarrow$) Es gelte $|a_n - a_m| < \epsilon$ für alle $n, m \geq N$. Dann ist die Folge $(a_n)_{n\in\mathbb{N}}$ beschränkt, denn zu $\epsilon = 1$ gibt es ein $N \in \mathbb{N}$ mit $|a_n - a_N| < 1$ für alle $n \geq N$; somit gilt $|a_n| \leq \max(|a_0|, |a_1|, \dots, |a_N|, |a_N| + 1)$. Nach dem Satz von Bolzano-Weierstraß besitzt $(a_n)_{n\in\mathbb{N}}$ eine konvergente Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ mit $a := \lim\limits_{k\to\infty} a_{n_k}$. Wir zeigen, daß auch $(a_n)_{n\in\mathbb{N}}$ gegen a konvergiert: Zu beliebigem $\epsilon > 0$ gibt es ein $N' \in \mathbb{N}$, so daß gleichzeitig gilt $|a_n - a_m| < \frac{\epsilon}{2}$ für alle $m, n \geq N'$ und $|a_{n_k} - a| < \frac{\epsilon}{2}$ für ein $n_k \geq N'$. Die Dreiecksungleichung liefert $|a_n - a| = |a_n - a_{n_k} + (a_{n_k} - a)| \leq |a_n - a_{n_k}| + |a_{n_k} - a| < \epsilon$. $\square$

Definition 10.8 Eine Folge $(a_n)_{n\in\mathbb{N}}$ komplexer Zahlen heißt *Cauchy-Folge* (oder *Fundamentalfolge*), wenn zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert, so daß $|a_n - a_m| < \epsilon$ für alle $n, m \geq N$.

Nach Satz 10.7 sind in $\mathbb{C}$ genau die Cauchy-Folgen die konvergenten Folgen. Der Beweis benutzt Bolzano-Weierstraß, wobei der entscheidende Teil für *reelle Folgen* bewiesen wird, wo das Intervallschachtelungsprinzip zur Verfügung steht. Damit wird das Cauchysche Konvergenzkriterium letztendlich auf die Vollständigkeit von $\mathbb{R}$ zurückgeführt. Bemerkenswert ist, daß auch die Umkehrung gilt:

Satz 10.9 *Für einen archimedisch angeordneten Körper folgt das Intervallschachtelungsprinzip aus dem Cauchyschen Konvergenzkriterium.*

Beweis. Es sei $([a_n, b_n])_{n\in\mathbb{N}}$ eine Intervallschachtelung. Zu $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$ mit $|a_N - b_N| < \epsilon$. Wegen $a_n < b_N$ für alle $n \geq N$ gilt $|a_n - a_m| < \epsilon$ für alle $m, n \geq N$, d.h. $(a_n)_{n\in\mathbb{N}}$ ist eine Cauchy-Folge. Sei $s := \lim_{n\to\infty} a_n$ der Grenzwert. Da $(a_n)_{n\in\mathbb{N}}$ monoton wachsend ist, gilt $a_n \leq s$ für alle $n \in \mathbb{N}$. Aus $a_k < b_n$ und Satz 9.7 für die konstante Folge $(b_n)_k = b_n$ folgt $s \leq b_n$ für alle $n \in \mathbb{N}$. Somit gilt $s \in [a_n, b_n]$ für alle $n \in \mathbb{N}$. □

Wir werden sehen, daß sich Folgen in sehr viel allgemeineren Räumen als $\mathbb{R}$ oder $\mathbb{C}$ betrachten lassen. In diesen Fällen ist die Intervallschachtelung nicht mehr sinnvoll, während die Definition der Vollständigkeit über die Konvergenz von Cauchy-Folgen im wesentlichen die gleiche bleibt, solange ein *Abstand* erklärt ist.

11 Reihen

Ist $(a_n)_{n\in\mathbb{N}}$ eine Folge komplexer Zahlen, so werden durch

$$s_0 := a_0\ , \quad s_1 := a_0 + a_1\ , \quad \ldots\ , \quad s_m := \sum_{k=0}^{m} a_k$$

die sogenannten *Partialsummen* definiert. Diese bilden dann die *Folge der Partialsummen* $(s_m)_{m\in\mathbb{N}}$, die wieder auf Konvergenz untersucht werden kann. Eine solche Folge der Partialsummen heißt *(unendliche) Reihe*, und für diese schreibt man $\sum_{n=0}^{\infty} a_n$. Das Symbol hat eine doppelte Bedeutung: Zum einen ist $\sum_{n=0}^{\infty} a_n$ gleichbedeutend mit der Folge der Partialsummen $(s_m)_{m\in\mathbb{N}}$ mit $s_m := \sum_{k=0}^{m} a_k$. Konvergiert diese Folge, dann meint man mit $\sum_{n=0}^{\infty} a_n$ auch den eindeutig bestimmten Grenzwert der Folge der Partialsummen. Über das Cauchysche Konvergenzkriterium für Folgen erhalten wir:

Satz 11.1 (Cauchy-Kriterium für Reihen) *Eine Reihe* $\sum_{n=0}^{\infty} a_n$ *ist genau dann konvergent, wenn es zu jedem* $\epsilon > 0$ *ein* $N \in \mathbb{N}$ *gibt, so daß* $\Big|\sum_{k=n}^{m} a_k\Big| < \epsilon$ *für alle* $m \geq n > N$.

Beweis. Klar wegen $|s_m - s_{n-1}| = \left|\sum_{k=n}^{m} a_k\right|$. $\square$

Daraus ergibt sich, daß aus der Konvergenz von $\sum_{n=0}^{\infty} a_n$ auch die Konvergenz von $\sum_{n=p}^{\infty} a_n$ folgt, und umgekehrt (wenn alle a_n definiert sind).

Satz 11.2 *Die geometrische Reihe* $\sum_{n=0}^{\infty} z^n$ *konvergiert für* $z \in \mathbb{C}$ *mit* $0 < |z| < 1$, *und in diesem Fall gilt* $\sum_{n=0}^{\infty} z^n = \dfrac{1}{1-z}$.

Beweis. Nach Satz: 2.3 (analog zu beweisen für $z \in \mathbb{C} \setminus \{0,1\}$) sind für $z \notin \{0,1\}$ die Partialsummen gegeben durch

$$s_n := \sum_{k=0}^{n} z^k = \frac{1-z^{n+1}}{1-z} \ ,$$

und nach Satz 9.5.iv) ist die Folge $\left(\frac{1-z^{n+1}}{1-z}\right)_{n\in\mathbb{N}}$ für $|z| < 1$ konvergent mit dem angegebenen Grenzwert. $\square$

Beispiel 11.3 (periodische Dezimalbrüche) Für $q = 0.\overline{135} := 135 \sum_{n=1}^{\infty} 10^{-3n}$ gilt $q = \frac{135}{1000} \cdot \sum_{n=0}^{\infty} 10^{-3n} = \frac{135}{1000} \cdot \frac{1}{1-10^{-3}} = \frac{135}{999} = \frac{15}{111} = \frac{5}{37}$. $\triangleleft$

Satz 11.4 *Die harmonische Reihe* $\sum_{n=1}^{\infty} \dfrac{1}{n}$ *ist divergent.*

Beweis. Wir zeigen, daß die Folge der Partialsummen keine Cauchy-Folge ist. Für beliebige $N \in \mathbb{N}^\times$ gilt

$$s_{2N} - s_N = \sum_{k=N+1}^{2N} \frac{1}{k} \geq \sum_{k=N+1}^{2N} \frac{1}{2N} = \frac{1}{2} \ .$$

Wäre $(s_n)_{n\geq 1}$ eine Cauchy-Folge, so gäbe es zu $\epsilon = \frac{1}{2}$ ein $N \in \mathbb{N}$ mit $|s_n - s_m| < \frac{1}{2}$ für alle $n, m \geq N$, Widerspruch. $\square$

Es ist im allgemeinen leichter, die Konvergenz einer Reihe zu beweisen, als den Grenzwert konkret anzugeben. Für den Konvergenzbeweis stehen mehrere Kriterien zur Verfügung.

Lemma 11.5 *Sind* $\sum_{n=0}^{\infty} a_n$ *und* $\sum_{n=0}^{\infty} b_n$ *konvergent, so gilt*

$$\sum_{n=0}^{\infty}(a_n + b_n) = \sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} b_n \qquad \textit{und} \quad \sum_{n=0}^{\infty}(za_n) = z\sum_{n=0}^{\infty} a_n \qquad (z \in \mathbb{C}) .$$

Beweis. Folgt aus Satz 9.3. □

Somit bilden konvergente Reihen einen Vektorraum.

Lemma 11.6 *Ist* $\sum_{n=0}^{\infty} a_n$ *konvergent, so bildet* $(a_n)_{n\in\mathbb{N}}$ *eine Nullfolge.*

Beweis. Es sei s der Grenzwert der Reihe. Zu jedem $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$ mit $|s_n - s| < \frac{\epsilon}{2}$ für alle $n > N$. Aus $a_n = s_n - s_{n-1}$ folgt $|a_n - 0| = |s_n - s_{n-1}| = |s_n - s + (s - s_{n-1})| \leq |s_n - s| + |s_{n-1} - s| < \epsilon$ für alle $n \geq N$. □

Das ist nützlich in negierter Form: Ist $(a_n)_{n\in\mathbb{N}}$ keine Nullfolge, so kann die Reihe $\sum_{n=0}^{\infty} a_n$ nicht konvergent sein. Daraus ergibt sich, daß für $z \in \mathbb{C}$ mit $|z| \geq 1$ die Reihe $\sum_{n=0}^{\infty} z^n$ nicht konvergieren kann, da $(z^n)_{n\in\mathbb{N}}$ dann keine Nullfolge ist.

Die Umkehrung von Lemma 11.6 wäre falsch: Ist $(a_n)_{n\in\mathbb{N}}$ Nullfolge, so kann zunächst nichts über die Konvergenz gesagt werden, wie das Beispiel der harmonischen Reihe zeigt.

Lemma 11.7 *Eine Reihe* $\sum_{n=0}^{\infty} a_n$ *mit positiven reellen Gliedern* $a_n \geq 0$ *konvergiert genau dann, wenn die Folge der Partialsummen beschränkt ist.*

Beweis. Für $a_n \geq 0$ ist die Folge der Partialsummen monoton. □

Beispiel 11.8 *Die folgende Reihe ist konvergent:* $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1.$

Beweis. Wegen $\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$ (Partialbruchzerlegung) gilt

$$s_m := \sum_{k=1}^{m} \frac{1}{k(k+1)} = \sum_{k=1}^{m} \frac{1}{k} - \sum_{k=1}^{m} \frac{1}{(k+1)} = 1 - \frac{1}{m+1} .$$

Die Folge $(1 - \frac{1}{n+1})_{n\geq 1}$ ist konvergent mit dem angegebenen Grenzwert. ◁

Satz 11.9 *Für $s \in \mathbb{Q}_+^\times$ ist*

$$\sum_{n=1}^{\infty} \frac{1}{n^s} \quad \begin{cases} \textit{konvergent} & \textit{für } s > 1 \\ \textit{divergent} & \textit{für } s \leq 1 \end{cases}$$

Beweis. Für $s \leq 1$ gilt für die Partialsummen

$$s_n = \frac{1}{1^s} + \frac{1}{2^s} + \cdots + \frac{1}{n^s} \geq \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} \,,$$

und für die (divergente) harmonische Reihe ist die Folge der Partialsummen unbeschränkt.

Für $s > 1$ betrachten wir die Teilfolge s_{2^k-1} der Partialsummen. Durch Zusammenfassen der Summanden $a_{2^j}, \dots a_{2^{j+1}-1}$ entsteht

$$\begin{aligned} s_{2^k-1} &= 1 + \Big(\frac{1}{2^s} + \frac{1}{3^s}\Big) + \Big(\frac{1}{4^s} + \cdots + \frac{1}{7^s}\Big) + \cdots + \Big(\frac{1}{(2^{k-1})^s} + \cdots + \frac{1}{(2^k-1)^s}\Big) \\ &\leq 1 + 2 \cdot \frac{1}{2^s} + 4 \cdot \frac{1}{4^s} + \cdots + 2^{k-1} \cdot \frac{1}{(2^{k-1})^s} \\ &= \sum_{j=0}^{k-1} \frac{2^j}{(2^j)^s} = \sum_{j=0}^{k-1} \Big(\frac{1}{2^{s-1}}\Big)^j = \frac{1 - \big(\frac{1}{2}\big)^{(s-1)k}}{1 - \big(\frac{1}{2}\big)^{s-1}} < \frac{2^{s-1}}{2^{s-1} - 1} \,. \end{aligned}$$

Damit ist die Teilfolge $(s_{2^k-1})_{k\geq 1}$ beschränkt. Da jedes $n \in \mathbb{N}$ durch ein $2^k - 1 \geq n$ abgeschätzt werden kann und die Folge der Partialsummen monoton ist, ist auch $(s_n)_{n\in\mathbb{N}}$ selbst beschränkt und damit konvergent. □

Die Reihe $\zeta(z) := \sum_{n=1}^{\infty} \frac{1}{n^z}$ heißt *Riemannsche Zeta-Funktion*, wobei zunächst $z \in \mathbb{Q}$ mit $z > 1$ ist. Später wird sich zeigen, daß die Reihe auch für gewisse $z \in \mathbb{C}$ sinnvoll ist. Über die Grenzwerte können wir zunächst nicht viel sagen. Später werden wir z.B. $\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ zeigen.

Satz 11.10 (Konvergenzkriterium von Leibniz) *Es sei $(a_n)_{n\in\mathbb{N}}$ eine monoton fallende Nullfolge (insbesondere $a_n \in \mathbb{R}_+$). Dann ist die Reihe $\sum_{n=0}^{\infty} (-1)^n a_n$ konvergent, und für den Grenzwert s gilt $\Big|s - \sum_{j=0}^{n} (-1)^j a_j\Big| \leq a_{n+1}$.*

Solche Reihen heißen *alternierend.*

Beweis. Für die geraden Partialsummen gilt $s_{2k+2} - s_{2k} = a_{2k+2} - a_{2k+1} < 0$, d.h. die Teilfolge $(s_{2k})_{k\in\mathbb{N}}$ ist monoton fallend. Für die ungeraden Partialsummen gilt

$s_{2k+3} - s_{2k+1} = -a_{2k+3} + a_{2k+2} > 0$, d.h. die Teilfolge $(s_{2k+1})_{k\in\mathbb{N}}$ ist monoton wachsend. Weiter gilt $s_{2k+1} - s_{2k} = -a_{2k+1} < 0$, d.h. $s_{2k} \geq s_{2k+1} \geq s_1$ für alle $k \in \mathbb{N}$. Da also $(s_{2k})_{k\in\mathbb{N}}$ monoton und beschränkt ist, konvergiert die gerade Teilfolge gegen einen Grenzwert s_g, und analog konvergiert $(s_{2k})_{k\in\mathbb{N}}$ gegen einen Grenzwert $s_u \leq s_g$. Nach Satz 9.3 gilt $s_g - s_u = \lim\limits_{k\to\infty}(s_{2k} - s_{2k+1}) = \lim\limits_{k\to\infty} a_{2k+1} = 0$. Damit konvergiert die gesamte Folge der Partialsummen gegen den gemeinsamen Grenzwert $s = s_g = s_u$.

Die Fehlerabschätzung ergibt sich daraus, daß s zwischen s_k und s_{k+1} liegt und $|s_k - s_{k+1}| = a_{k+1}$. □

Beispiel 11.11 Die alternierende harmonische Reihe $\sum_{n=0}^{\infty}(-1)^n\frac{1}{n+1}$ ist konvergent. Der Grenzwert liegt zwischen $s_1 = \frac{1}{2}$ und $s_0 = 1$. Mit später bereitgestellten Methoden kann man $\sum_{n=0}^{\infty}(-1)^n\frac{1}{n+1} = \ln 2$ zeigen.

Die alternierende Reihe $\sum_{n=0}^{\infty}(-1)^n\frac{1}{2n+1}$ ist konvergent. Der Grenzwert liegt zwischen $s_1 = \frac{1}{3}$ und $s_0 = 1$. Mit später bereitgestellten Methoden kann man $\sum_{n=0}^{\infty}(-1)^n\frac{1}{2n+1} = \frac{\pi}{4}$ zeigen. ◁

12 Absolute Konvergenz von Reihen

Definition 12.1 Eine Reihe $\sum\limits_{n=0}^{\infty} a_n$ heißt *absolut konvergent*, wenn auch die Reihe $\sum\limits_{n=0}^{\infty} |a_n|$ konvergent ist.

Satz 12.2 (Majorantenkriterium) *Es seien $\sum\limits_{n=0}^{\infty} a_n$ und $\sum\limits_{n=0}^{\infty} b_n$ Reihen und $p \in \mathbb{N}$ derart, daß $|a_n| \leq |b_n|$ für alle $n \geq p$.*

i) *Ist $\sum\limits_{n=0}^{\infty} |b_n|$ konvergent, so konvergieren auch $\sum\limits_{n=0}^{\infty} a_n$ und $\sum\limits_{n=0}^{\infty} |a_n|$, und es gilt*

$$\Big|\sum_{n=p}^{\infty} a_n\Big| \leq \sum_{n=p}^{\infty} |a_n| \leq \sum_{n=p}^{\infty} |b_n| \,.$$

Insbesondere ist jede absolut konvergente Reihe auch konvergent.

ii) *Ist $\sum\limits_{n=p}^{\infty} a_n$ divergent, so ist auch $\sum\limits_{n=p}^{\infty} |b_n|$ divergent.*

Das Beispiel der alternierenden harmonischen Reihe zeigt, daß aus Konvergenz nicht absolute Konvergenz folgt.

Beweis. ii) folgt aus i) durch Negation.

i) Zu jedem $\epsilon > 0$ gibt es nach dem Cauchyschen Konvergenzkriterium ein $N \in \mathbb{N}$ mit $N \geq p$, so daß $\sum\limits_{k=n}^{m} |b_k| < \epsilon$ für alle $m \geq n > N$. Nach der Dreiecksungleichung gilt für die *endliche Summe*

$$\Big|\sum_{k=n}^{m} a_k\Big| = |a_n + \cdots + a_m| \leq |a_n| + \cdots + |a_m| = \sum_{k=n}^{m} |a_k| \leq \sum_{k=n}^{m} |b_k| < \epsilon\,,$$

so daß $\sum\limits_{n=p}^{\infty} a_n$ und $\sum\limits_{n=p}^{\infty} |a_n|$ nach dem Cauchyschen Konvergenzkriterium konvergent sind. Aus Satz 9.7.i) folgt schließlich

$$\begin{aligned}\Big|\sum_{n=p}^{\infty} a_n\Big| = \Big|\lim_{m\to\infty} \sum_{n=p}^{m} a_n\Big| &= \lim_{m\to\infty} \Big|\sum_{n=p}^{m} a_n\Big| \\ &\leq \lim_{m\to\infty} \sum_{n=p}^{m} |a_n| \quad \begin{cases} = \sum\limits_{n=p}^{\infty} |a_n| \\ \leq \lim\limits_{m\to\infty} \sum\limits_{n=p}^{m} |b_n| = \sum\limits_{n=p}^{\infty} |b_n| \end{cases}\end{aligned} \qquad \square$$

Beispiel 12.3 i) $\sum\limits_{n=1}^{\infty} \frac{n!}{n^n}$ ist konvergent, da für $n > 2$ gilt:

$\frac{n!}{n^n} = \frac{1}{n} \cdot \frac{2}{n} \cdots \frac{n-1}{n} \cdot \frac{n}{n} \leq \frac{2}{n^2}$, und $\sum\limits_{n=1}^{\infty} \frac{2}{n^2}$ ist (absolut) konvergent.

ii) $\sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n(n+1)}}$ ist divergent, da $\frac{1}{\sqrt{n(n+1)}} \geq \frac{1}{n+1}$, und die harmonische Reihe ist divergent. ◁

Satz 12.4 (Quotientenkriterium) *Es sei* $\sum\limits_{n=0}^{\infty} a_n$ *eine Reihe mit* $a_n \neq 0$ *für alle* $n \in \mathbb{N}$, *und es gebe ein* $N \in \mathbb{N}$ *und ein* $0 < q < 1$, *so daß* $\left|\frac{a_{n+1}}{a_n}\right| \leq q$ *für alle* $n \geq N$. *Dann konvergiert die Reihe absolut.*

Beweis. Für $n \geq N$ gilt dann $|a_n| \leq q|a_{n-1}| \leq \ldots (q)^{n-N}|a_N|$, so daß die Reihe $\sum\limits_{n=N}^{\infty} |a_n|$ durch die konvergente Reihe $\sum\limits_{n=N}^{\infty} q^n \frac{|a_N|}{q^N} = \frac{|a_N|}{1-q}$ majorisiert wird. $\square$

Die Eigenschaft $\left|\frac{a_{n+1}}{a_n}\right| \leq q$ für ein $0 < q < 1$ ist automatisch erfüllt, wenn die Folge $(|\frac{a_{n+1}}{a_n}|)_{n\in\mathbb{N}}$ gegen ein $\alpha < 1$ konvergiert: Nach Definition des Grenzwertes gibt es dann zu $\alpha < q < 1$ ein $N \in \mathbb{N}$, so daß $\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{a_{n+1}}{a_n} - \alpha + \alpha\right| \leq$

$\alpha + \left|\frac{a_{n+1}}{a_n} - \alpha\right| \leq q$ für alle $n \geq N$. Gilt $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = q > 1$, dann ist $(a_n)_{n\in\mathbb{N}}$ keine Nullfolge, und die Reihe divergiert.

Wichtig ist, daß das q im Quotientenkriterium unabhängig von n sein muß. Für die Reihe $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s}$ mit $s \in \mathbb{Q}_+^*$ ist zwar $\left|(\frac{n}{n+1})^s\right| < 1$ für alle $n \geq 1$, aber zu jedem $q < 1$ findet man ein $N \in \mathbb{N}$ mit $q < \left|(\frac{n}{n+1})^s\right| < 1$ für alle $n \geq N$. Über die Konvergenz kann das Quotientenkriterium in diesem Fall keine Aussage machen. Die Reihe ist für $s > 1$ absolut konvergent und für $s \leq 1$ divergent.

Die Konvergenz der geometrischen Reihe kann nicht mit dem Quotientenkriterium begründet werden!

Beispiel 12.5 $\displaystyle\sum_{n=0}^{\infty} \frac{1}{n!}\Big(\frac{n}{4}\Big)^n$ ist (absolut) konvergent, denn

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \frac{(n+1)^{n+1}\cdot 4^n \cdot n!}{n^n \cdot 4^{n+1}\cdot (n+1)!} = \lim_{n\to\infty} \frac{1}{4}\Big(1+\frac{1}{n}\Big)^n = \frac{e}{4} < 1\,. \qquad \triangleleft$$

Beispiel 12.6 (komplexe Binomialreihen) Über $\displaystyle\binom{\alpha}{k} := \prod_{j=1}^{k} \frac{\alpha - j + 1}{j}$ für $k \in \mathbb{N}$ und $\alpha \in \mathbb{C}$ lassen sich komplexe Binomialkoeffizienten definieren. Wir betrachten für $z \in \mathbb{C}$ die zugehörige *binomische Reihe*

$$B_\alpha(z) := \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n\,,$$

die für $\alpha \in \mathbb{N}$ bei $k = \alpha$ abbricht und in die binomische Formel übergeht. Das Quotientenkriterium liefert $\displaystyle\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty}\left|\frac{\alpha - n}{n+1}\cdot z\right| = |z|$, so daß die binomische Reihe für $|z| < 1$ absolut konvergent ist und für $|z| > 1$ divergiert. $\triangleleft$

Satz 12.7 (Wurzelkriterium) *Für eine Reihe* $\displaystyle\sum_{n=0}^{\infty} a_n$ *existiere ein* $N \in \mathbb{N}$ *und ein* $0 < q < 1$, *so daß* $\sqrt[n]{|a_n|} \leq q$ *für alle* $n \geq N$. *Dann konvergiert die Reihe* $\displaystyle\sum_{n=0}^{\infty} a_n$ *absolut.*

Ist $\sqrt[n]{|a_n|} \geq 1$ für unendlich viele $n \in \mathbb{N}$, so ist $(a_n)_{n\in\mathbb{N}}$ keine Nullfolge, und die Reihe ist divergent.

Beweis. Die Reihe $\displaystyle\sum_{n=N}^{\infty} |a_n|$ besitzt die konvergente Majorante $\displaystyle\sum_{n=N}^{\infty} q^n$. □

Die Eigenschaft $\sqrt[n]{|a_n|} \leq q$ für ein $0 < q < 1$ ist automatisch erfüllt, wenn $\limsup \sqrt[n]{|a_n|} < 1$, insbesondere wenn $(\sqrt[n]{|a_n|})_{n\in\mathbb{N}}$ gegen $\alpha < 1$ konvergiert.

Beispiel 12.8 Die Reihe $\displaystyle\sum_{n=1}^{\infty} n^k z^n$ für $z \in \mathbb{C}$ mit $|z| < 1$ ist absolut konvergent: Wir betrachten $\sqrt[n]{|n^k z^n|} = \sqrt[n]{n^k}|z|$. Wegen $\lim_{n\to\infty} \sqrt[n]{n^k} = 1$ gibt es ein $N \in \mathbb{N}$, so daß $\sqrt[n]{n^k} < \frac{2}{1+|z|}$. Damit kann im Wurzelkriterium $q = \frac{2|z|}{1+|z|} < 1$ gewählt werden. $\triangleleft$

Es sei $\displaystyle\sum_{n=0}^{\infty} a_n$ eine Reihe und $\tau : \mathbb{N} \to \mathbb{N}$ eine bijektive Abbildung. Dann nennt man die Reihe $\displaystyle\sum_{n=0}^{\infty} a_{\tau(n)}$ eine *Umordnung* der Reihe. Im Gegensatz zu endlichen Summen gilt das Kommutativgesetz "$\displaystyle\sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} a_{\tau(n)}$" nicht immer. Wir sehen uns dazu das Beispiel der alternierenden harmonischen Reihe an und betrachten folgende Umordnung:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{(-1)^{\tau(n)+1}}{\tau(n)} &= 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + \Big(\tfrac{1}{5} + \tfrac{1}{7} - \tfrac{1}{6}\Big) + \Big(\tfrac{1}{9} + \tfrac{1}{11} + \tfrac{1}{13} + \tfrac{1}{15} - \tfrac{1}{8}\Big) + \dots \\ &\quad + \Big(\tfrac{1}{2^n+1} + \tfrac{1}{2^n+3} + \dots + \tfrac{1}{2^{n+1}-1} - \tfrac{1}{2n+2}\Big) + \dots\end{aligned}$$

In dieser Umordnung kommen alle negativ zu zählenden Glieder $\frac{1}{2n+2}$ vor, aber immer mehr verzögert gegenüber den positiven Gliedern. Nun gilt die Abschätzung

$$\begin{aligned}\Big(\tfrac{1}{2^n+1} + \tfrac{1}{2^n+3} + \dots + \tfrac{1}{2^{n+1}-1} - \tfrac{1}{2n+2}\Big) > 2^{n-1}\tfrac{1}{2^{n+1}} - \tfrac{1}{2n+2} &= \tfrac{1}{4} - \tfrac{1}{2n+2} \\ &\geq \tfrac{1}{12} \quad \text{für } n \geq 2\,.\end{aligned}$$

Wie schon im Divergenzbeweis der harmonischen Reihe ist damit für $\epsilon = \frac{1}{12}$ das Cauchysche Konvergenzkriterium verletzt, so daß die umgeordnete Reihe divergiert. Man kann übrigens für die alternierende harmonische Reihe zu jedem $x \in \mathbb{R}$ eine Umordnung derart finden, daß die umgeordnete Reihe gegen x konvergiert.

Wir zeigen nun, daß *absolut* konvergente Reihen beliebig umgeordnet werden dürfen:

Satz 12.9 (Umordnungssatz) *Es sei* $\displaystyle\sum_{n=0}^{\infty} a_n$ *eine absolut konvergente Reihe. Dann ist auch jede Umordnung dieser Reihe absolut konvergent mit dem gleichen Grenzwert.*

Beweis. Es sei $s = \sum_{n=0}^{\infty} a_n$ und $\tau : \mathbb{N} \to \mathbb{N}$ bijektiv. Zu jedem $\epsilon > 0$ gibt es wegen der absoluten Konvergenz ein $N \in \mathbb{N}$, so daß $\sum_{n=N}^{\infty} |a_n| < \frac{\epsilon}{2}$. Dann gilt

$$\Big| s - \sum_{n=0}^{N-1} a_n \Big| = \Big| \lim_{M\to\infty} \sum_{n=0}^{M} a_n - \sum_{n=0}^{N-1} a_n \Big| = \Big| \lim_{M\to\infty} \sum_{n=N}^{M} a_n \Big| = \Big| \sum_{n=N}^{\infty} a_n \Big| \leq \sum_{n=N}^{\infty} |a_n| < \frac{\epsilon}{2} \;.$$

Sei nun ein $N' \in \mathbb{N}$ so gewählt, daß $\{0, 1, \dots, N-1\} \subseteq \{\tau(0), \tau(1), \dots, \tau(N')\}$, d.h. $N' \geq \max(\tau^{-1}(0), \dots \tau^{-1}(N-1)\}$. Dann gilt für alle $m \geq N'$ nach der Dreiecksungleichung

$$\Big| \sum_{n=0}^{m} a_{\tau(n)} - s \Big| \leq \Big| \sum_{n=0}^{m} a_{\tau(n)} - \sum_{n=0}^{N-1} a_n \Big| + \Big| \sum_{n=0}^{N-1} a_n - s \Big| < \sum_{n=N}^{\infty} |a_n| + \frac{\epsilon}{2} < \epsilon \;,$$

denn in $\sum_{n=0}^{m} a_{\tau(n)} - \sum_{n=0}^{N-1} a_n$ heben sich $a_0, a_1, \dots, a_{N-1}$ gegeneinander weg und die Summe über die verbleibenden Terme ist wie angegeben beschränkt. Damit konvergiert $\sum_{n=0}^{\infty} a_{\tau(n)}$ gegen den gleichen Grenzwert s. Die absolute Konvergenz folgt durch Wiederholung des Beweises für die Reihe $\sum_{n=0}^{\infty} |a_{\tau(n)}|$. □

Der Umordnungssatz wird später benötigt, um Identitäten zwischen verschiedenen Reihen zu beweisen.

13 Polynome

Ein Polynom in x mit Koeffizienten in $\mathbb{C}$ ist ein formaler Ausdruck der Form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = \sum_{i=0}^{n} a_i x^i \;, \qquad a_i \in \mathbb{C} \;.$$

Die Menge aller solcher Polynomome wird mit $\mathbb{C}[x]$ bezeichnet, bzw. mit $\mathbb{R}[x]$ wenn $a_i \in \mathbb{R}$. Die Unbestimmte x werden wir zunächst als $x \in \mathbb{C}$ oder $x \in \mathbb{R}$ auffassen; später werden Verallgemeinerungen wichtig. Es gibt eine offensichtliche Addition

$$\Big(\sum_{i=0}^{n} a_i x^i \Big) + \Big(\sum_{i=0}^{m} b_i x^i \Big) := \sum_{i=0}^{\max(m,n)} (a_i + b_i) x^i \;,$$

wobei $a_i = 0$ für $i > n$ und $b_i = 0$ für $i > m$ gesetzt wird. Das Nullpolynom $f(x) = 0$ (alle a_i verschwinden) ist das neutrale Element. Außerdem erhält man durch Ausmultiplizieren und Ordnen nach Potenzen von x eine Multiplikation

$$\Big(\sum_{i=0}^{n} a_i x^i\Big) \cdot \Big(\sum_{j=0}^{m} b_j x^j\Big) =: \sum_{k=0}^{m+n} c_k x^k \;, \qquad c_k = \sum_{j=0}^{k} a_{k-j} b_j \;.$$

(Es wird wieder $a_i = 0$ für $i > n$ und $b_j = 0$ für $j > m$ gesetzt.) Die auftretenden Summanden ergeben sich aus den Diagonalen im folgenden Schema (mit $k = \max(m, n)$):

$$\begin{array}{ccccc}
a_k b_0 & a_k b_1 & & \dots & a_k b_k \\
\vdots & & & & \vdots \\
a_2 b_0 & & & \dots & a_2 b_k \\
& & & & \\
a_1 b_0 & a_1 b_1 & & \dots & a_1 b_k \\
& & & & \\
a_0 b_0 & a_0 b_1 & a_0 b_2 & \dots & a_0 b_k
\end{array}$$

Der maximale Exponent von x, für den der Koeffizient in $f(x)$ ungleich Null ist, heißt der *Grad des Polynoms* und wird mit $\deg(f)$ bezeichnet. Genauer ist für $f(x) = \sum_{i=0}^{n} a_i x^i$

$$\deg(f) = \begin{cases} -\infty & \text{für } f = 0 \\ \max\{i \in \mathbb{N} \;:\; a_i \neq 0\} & \text{sonst} \end{cases}$$

Konstante, aber nichtverschwindende, Polynome haben den Grad 0. Es gilt

$$\deg(f + g) \leq \max(\deg(f), \deg(g)) \;, \qquad \deg(f \cdot g) = \deg(f) + \deg(g)$$

(mit $(-\infty) + n = -\infty$ und $(-\infty) + (-\infty) = -\infty$).

Satz 13.1 (Division mit Rest) *Sind $f, g \in \mathbb{C}[x]$ und ist $g \neq 0$, dann gibt es eindeutig bestimmte Polynome $q, r \in \mathbb{C}[x]$ mit*

$$f = q \cdot g + r \;, \qquad \deg(r) < \deg(g) \;.$$

Beweis. i) Wir beweisen zunächst die Eindeutigkeit. Sei $f = q \cdot g + r = q' \cdot g + r'$ mit $\deg(r) < \deg(g)$ und $\deg(r') < \deg(g)$, dann folgt

$$(q - q') \cdot g = r' - r \;.$$

Wäre $q \neq q'$, dann ist der Grad der linken Seite $\deg((q-q') \cdot g) > \deg(g)$, während der Grad der rechten Seite $\deg(r' - r) < \deg(g)$ ist, Widerspruch. Also ist $q = q'$ und dann $r = r'$.

ii) Existenz der Polynome durch explizite Konstruktion mittels “Division mit Rest”. Sei $f = \sum_{i=0}^{n} a_i x^i$ und $g = \sum_{i=0}^{m} b_i x^i$, mit $a_n, b_m \neq 0$. Für $n < m$ ist $q = 0$ und $r = f$. Sei also $n \geq m$. Wir setzen

$$q_{(1)} := \frac{a_n}{b_m} x^{n-m} , \qquad f_{(1)} := f - q_{(1)} \cdot g .$$

Da der höchste Koeffizient von f durch Subtraktion entfernt wurde, gilt $\deg(f_{(1)}) < \deg(f)$. Auf diese Weise konstruiert man eine Folge von Monomen $q_{(k)} = c_k x^{i(k)}$, so daß für $f_{(k)} := f_{(k-1)} - q_{(k)} \cdot g$ gilt $\deg(f_{(k)}) < \deg(f_{(k-1)})$. Das Verfahren bricht im l-ten Schritt ab, wenn $\deg(f_{(l)}) < \deg(g)$. Dann ist $r := f_{(l)}$ und $q := q_{(1)} + \cdots + q_{(l)}$. $\square$

Beispiel 13.2 Es sei $f = x^3$ und $g = 2x^2 + x$, also $a_3 = 1$ und $b_2 = 2$. Dann ist $q_{(1)} = \frac{a_3}{b_3} x^{3-2} = \frac{1}{2}x$ und $f_{(1)} = f - \frac{1}{2}x \cdot (2x^2 + x) = -\frac{1}{2}x^2$. Im nächsten Schritt ist $q_{(2)} = \frac{-\frac{1}{2}}{2} x^0 = -\frac{1}{4}$ und $f_{(2)} = f_{(1)} + \frac{1}{4} \cdot (2x^2 + x) = \frac{1}{4}x$. Hier bricht das Verfahren ab. Somit gilt $x^3 = (\frac{1}{2}x - \frac{1}{4})(2x^2 + x) + \frac{1}{4}x$. $\triangleleft$

Ist $r = 0$ in Satz 13.1, so heißt g ein *Teiler* von f.

Definition 13.3 Eine Zahl $\alpha \in \mathbb{C}$ heißt ***Nullstelle*** eines Polynoms $f = \sum_{i=0}^{n} a_i x^i$, wenn $f(\alpha) = \sum_{i=0}^{n} a_n \alpha^i = 0$.

Satz 13.4 *Eine Zahl $\alpha \in \mathbb{C}$ ist genau dann* Nullstelle *eines Polynoms $f \in \mathbb{C}[x]$, wenn $x - \alpha$ Teiler von f ist.*

Beweis. Für $\deg(f) = 0$ hat f keine Nullstellen, für $\deg(f) = -\infty$ ist $f = 0$ und $x - \alpha$ ein Teiler für beliebige $\alpha \in \mathbb{C}$. Verbleibt $\deg(f) \geq 1$. Die Division mit Rest liefert eindeutig bestimmte Polynome $q, r \in \mathbb{C}[x]$ mit $f = q \cdot (x - \alpha) + r$ und $\deg(q) = \deg(f) - 1$, $\deg(r) < \deg(x - \alpha) = 1$, also $r \in \mathbb{C}$. Dann ist $f(\alpha) = r$. $\square$

Hat auch q eine Nullstelle α_2, so läßt sich ein weiterer Linearfaktor $x - \alpha_2$ abspalten, usw. Aus der Abbruchbedingung der Division mit Rest folgt, daß ein Polynom vom Grad $n \geq 0$ höchstens n Nullstellen haben kann.

Satz 13.5 (Identitätssatz für Polynome) *Stimmen die Werte der Polynome $f(x) = \sum_{i=0}^{n} a_i x^i$ und $g(x) = \sum_{i=0}^{n} b_i x^i$ an $n+1$ verschiedenen Stellen überein, dann gilt $a_i = b_i$ für alle $i = 0, \ldots, n$ und somit $f(x) = g(x)$ für alle $x \in \mathbb{C}$.*

Beweis. Das Polynom $f - g$ hat $n + 1$ verschiedene Nullstellen und einen Grad $\leq n$. Damit ist $f - g$ das Nullpolynom. $\square$

Folgerung 13.6 *Die Polynome in x vom Grad $\leq n$ bilden einen $n+1$-dimensionalen Vektorraum $P_n[x]$. Durch $\mathcal{B} = \{x^0, x^1, x^2, \ldots, x^n\}$ ist eine Basis in $P_n[x]$ gegeben.*

Beweis. Das Nullpolynom ist das Nullelement. Klar ist, daß $P_n[x]$ durch $\mathcal{B}$ erzeugt wird. Ist $\sum_{i=0}^n a_i x^i = 0$, so folgt $a_i = 0$ nach dem Identitätssatz. □

Je nach Problemstellung können sich andere Basen als sinnvoll erweisen. Besonders nützliche Basen tragen z.T. eigene Namen. Erwähnt seien Polynome von Hermite, Laguerre, Legendre, Tschebyschew, Jacobi, Gegenbauer.

Der Identitätssatz liefert die sehr wichtige Methode des

13.7 Koeffizientenvergleich. *Gibt es für ein Polynom zwei Darstellungen, so sind die entsprechenden Koeffizienten einander gleich.*

Satz 13.8 (Additionstheorem der Binomialkoeffizienten) *Für alle $s, t \in \mathbb{C}$ und $n \in \mathbb{N}$ gilt*

$$\sum_{k=0}^{n} \binom{s}{k}\binom{t}{n-k} = \binom{s+t}{n}.$$

Beweis. i) Sei zunächst $s, t \in \mathbb{N}$. Dann gilt

$$(1+x)^{s+t} = \sum_{n=0}^{s+t} \binom{s+t}{n} x^n$$

$$(1+x)^s \cdot (1+x)^t = \Big(\sum_{k=0}^{s} \binom{s}{k} x^k\Big)\Big(\sum_{l=0}^{t} \binom{t}{l} x^l\Big) = \sum_{n=0}^{s+t} \Big(\sum_{k+l=n} \binom{s}{k}\binom{t}{l}\Big) x^n$$

Die Behauptung folgt aus $\displaystyle\sum_{k+l=n} \binom{s}{k}\binom{t}{l} = \sum_{k=0}^{n} \binom{s}{k}\binom{t}{n-k}$ und Koeffizientenvergleich.

ii) Sei $s \in \mathbb{C}$ und $t \in \mathbb{N}$. Interpretieren wir beide Seiten des Additionstheorems als Polynome in s, dann stimmen diese nach i) in unendlich vielen Stellen überein, nach dem Identitätssatz dann für alle $s \in \mathbb{C}$.

iii) Für $s, t \in \mathbb{C}$ werden schließlich beide Seiten bei festem s als Polynome in t aufgefaßt, die nach ii) an unendlich vielen Stellen übereinstimmen, damit überall. □

Definition 13.9 Seien $f, g \in \mathbb{C}[x]$ Polynome, mit $f \neq 0$, und sei $A \subseteq \mathbb{C}$ eine Teilmenge, die alle Nullstellen von f enthält. Dann heißt der für $x \in \mathbb{C} \setminus A$ erklärte Quotient $R(x) = \frac{g(x)}{f(x)}$ eine *rationale Funktion*.

Durch Polynomdivision läßt sich der Definitionsbereich maximal erweitern. Für rationale Funktionen gilt

Satz 13.10 (Partialbruchzerlegung) *Für* $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{C}$ *und* $\nu_1, \ldots, \nu_n \in \mathbb{N}^\times$ *sei* $f(x) = \alpha_0(x-\alpha_1)^{\nu_1} \cdots (x-\alpha_n)^{\nu_n}$ *ein Polynom vom Grad* $\deg(f) = \nu_1 + \cdots + \nu_n$*. Dann gibt es zu* $g \in \mathbb{C}[x]$ *ein eindeutig bestimmtes Polynom* q *vom Grad* $\deg(g) - \deg(\mathrm{f})$ *(mit* $q = 0$ *falls* $\deg(g) < \deg(\mathrm{f})$*) und eindeutig bestimmte Koeffizienten* $b_{k,j_k} \in \mathbb{C}$ *mit* $k = 1, \ldots, n$ *und* $j_k = 1, \ldots \nu_k$*, so daß für alle* $x \notin \{\alpha_1, \ldots, \alpha_n\}$ *gilt*

$$\begin{aligned}\frac{g(x)}{f(x)} &= q(x) + \sum_{k=1}^{n} \sum_{j_k=1}^{\nu_k} \frac{b_{k,j_k}}{(x-\alpha_k)^{j_k}} \\ &= q(x) + \sum_{k=1}^{n} \Big(\frac{b_{k,1}}{(x-\alpha_k)^1} + \frac{b_{k,2}}{(x-\alpha_k)^2} + \cdots + \frac{b_{k,\nu_k}}{(x-\alpha_k)^{\nu_k}} \Big) .\end{aligned}$$

Diese Darstellung heißt Partialbruchzerlegung.

Beweis. Das Polynom q wird durch Division mit Rest erhalten, $g = q \cdot f + r$. Damit genügt es, den Satz für rationale Funktionen $\frac{g}{f}$ mit $\deg(g) < \deg(f)$ und $q = 0$ zu beweisen. Wir schreiben $f(x) = (x-\alpha_1)^{\nu_1} h(x)$ mit $h(\alpha_1) \neq 0$. Dann gilt für $x \notin \{\alpha_1, \ldots, \alpha_n\}$

$$\begin{aligned}\frac{g(x)}{f(x)} &= \frac{1}{(x-\alpha_1)^{\nu_1}} \Big(\frac{g(x)}{h(x)} - \frac{g(\alpha_1)}{h(\alpha_1)} + \frac{g(\alpha_1)}{h(\alpha_1)} \Big) \\ &= \frac{1}{(x-\alpha_1)^{\nu_1}} \Big(\frac{g(x)h(\alpha_1) - g(\alpha_1)h(x)}{h(x)\, h(\alpha_1)} + \frac{g(\alpha_1)}{h(\alpha_1)} \Big) .\end{aligned}$$

Da $g(x)h(\alpha_1) - g(\alpha_1)h(x)$ die Nullstelle $x = \alpha_1$ hat, gibt es ein Polynom $g'(x)$ vom Grad $\deg(g) - 1$ mit

$$g(x)h(\alpha_1) - g(\alpha_1)h(x) = (x-\alpha_1) \cdot h(\alpha_1) \cdot g'(x) ,$$

und es folgt

$$\frac{g(x)}{f(x)} = \frac{b_{1,\nu_1}}{(x-\alpha_1)^{\nu_1}} + \frac{g'(x)}{f'(x)} , \qquad b_{1,\nu_1} = \frac{g(\alpha_1)}{h(\alpha_1)} , \qquad f'(x) = (x-\alpha_1)^{\nu_1 - 1} h(x)$$

Iteration des Verfahrens liefert die Behauptung. □

Wir betrachten Reihen von rationalen Funktionen $\displaystyle\sum_{n=N}^{\infty} \frac{g(n)}{f(n)}$. Die Reihe konvergiert genau dann (absolut), wenn $\deg(f) \geq \deg(g) + 2$ und f keine Nullstellen $n \in \mathbb{N}$ mit $n \geq N$ hat. Liegen zusätzlich alle Nullstellen von f in $\mathbb{Z}$ und sind $< N$, dann kann die Reihe konkret berechnet werden als Linearkombinationen von 1 und ζ-Funktionen.

Beispiel 13.11 Gesucht sei die Reihe $\sum_{n=0}^{\infty} \frac{1}{n^2(n+1)^2}$. Dazu ist die Partialbruchzerlegung von $\frac{1}{x^2(x+1)^2}$ zu bestimmen. Für die Nullstelle $\alpha_1 = 0$ ist wegen $h(x) = (x+1)^2$ und $g(x) = 1$ zunächst $b_{1,2} = 1$ und $xg'(x) = 1 - (x+1)^2$, also

$$\frac{1}{x^2(x+1)^2} = \frac{1}{x^2} - \frac{2+x}{x(x+1)^2} = \frac{1}{x^2} - \frac{1}{(x+1)^2} - \frac{2}{x(x+1)^2}\,.$$

Im nächsten Schritt folgt $-\frac{2}{x(x+1)^2} = -\frac{2}{x} + \frac{2(1+1+x)}{(x+1)^2}$, insgesamt also

$$\frac{1}{x^2(x+1)^2} = \frac{1}{x^2} + \frac{1}{(x+1)^2} - \frac{2}{x} + \frac{2}{x+1}\,.$$

Es folgt mit $\zeta(2) := \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ und Beispiel 11.8:

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n^2(n+1)^2} &= \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} - 2\sum_{n=1}^{\infty} \frac{1}{n(n+1)} \\ &= 2\sum_{n=1}^{\infty} \frac{1}{n^2} - 1 - 2\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 2\zeta(2) - 3 = \frac{\pi^2}{3} - 3\,. \qquad \triangleleft\end{aligned}$$

14 Potenzreihen

Potenzreihen sind Verallgemeinerungen von Polynomen auf unendliche Summen, also Reihen, $P(z) = \sum_{n=0}^{\infty} a_n z^n = a_0 + \sum_{n=1}^{\infty} a_n z^n$ mit $a_n \in \mathbb{C}$ und einer Unbestimmten $z \in \mathbb{C}$. Beispiele sind die geometrische Reihe $\sum_{n=0}^{\infty} z^n$ und die binomische Reihe $B_s(z) = \sum_{n=0}^{\infty} \binom{s}{n} z^n$. Später werden auch die in einen anderen Ursprung $z_0 \in \mathbb{C}$ verschobenen Potenzreihen $P(z, z_0) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ betrachtet. Im allgemeinen werden diese Reihen nicht für alle $z \in \mathbb{C}$ konvergieren.

Satz 14.1 *Konvergiert eine Potenzreihe $P(z) = \sum_{n=0}^{\infty} a_n z^n$ in einem Punkt $z_0 \in \mathbb{C}$, dann konvergiert sie absolut in jedem Punkt $z \in \mathbb{C}$ mit $|z| < |z_0|$.*

Beweis. Die Folge $(a_n z_0^n)_{n\in\mathbb{N}}$ ist als Nullfolge insbesondere beschränkt, d.h. es gilt $|a_n z_0^n| \leq S$ für alle $n \in \mathbb{N}$. Dann ist $|a_n z^n| = q^n |a_n z_0^n| \leq q^n S$ mit $q := |\frac{z}{z_0}| < 1$.

Damit besitzt $\sum_{n=0}^{\infty} a_n z^n$ die konvergente Majorante $\sum_{n=0}^{\infty} q^n S$, so daß die Reihe für $|z| < |z_0|$ absolut konvergent ist. $\square$

Dieser Satz liefert für jede Potenzreihe $P(z)$ die Existenz eines *Konvergenzkreises* mit Radius

$$R(P) := \sup\{r \in \mathbb{R} \;:\; P(r) \text{ konvergiert } \} \in \mathbb{R}_+ \cup \{\infty\} .$$

Man nennt $R(P)$ den *Konvergenzradius* von P.

Satz 14.2 *Die Potenzreihe P ist für alle $z \in \mathbb{C}$ mit $|z| < R(P)$ absolut konvergent und für alle $z \in \mathbb{C}$ mit $|z| > R(P)$ divergent.*

Beweis. i) Für $|z| < R(P)$ gibt es ein $r \in \mathbb{R}$ mit $|z| < r < R(P)$, so daß $P(r)$ konvergent ist. Dann ist $P(z)$ absolut konvergent nach Satz 14.1.

ii) Sei $|z| > R(P)$. Wäre $P(z)$ konvergent, so wäre nach Satz 14.1 $P(r)$ (sogar absolut) konvergent für alle $R(P) < r < |z|$, im Widerspruch zur Supremumseigenschaft von $R(P)$. $\square$

Der Konvergenzradius einer Potenzreihe $P(z)$ kann unendlich sein; in diesem Fall ist $P(z)$ absolut konvergent für alle $z \in \mathbb{C}$. Der Konvergenzradius $R(P)$ ist 0, wenn $P(z)$ nur für $z = 0$ konvergiert. Die Bestimmung des Konvergenzradius von $P = \sum_{n=0}^{\infty} a_n z^n$ kann über das Wurzelkriterium oder das Quotientenkriterium versucht werden:

i) $R(P) = \frac{1}{L}$, wenn $L = \limsup \sqrt[n]{|a_n|}$ ist (Cauchy-Hadamard),

ii) $R(P) = \frac{1}{q}$, wenn $\{|\frac{a_{n+1}}{a_n}|\}_{n\geq 1}$ gegen q konvergiert (Euler).

Über Konvergenz von $P(z)$ auf dem Rand des Konvergenzkreises, d.h. für $|z| = R(P)$, kann keine Aussage gemacht werden.

Beispiel 14.3 Für $P(z) := \sum_{n=0}^{\infty} z^n$ ist nach Satz 11.2 $R(P) = 1$.

Für $P(z) = B_\alpha(z) := \sum_{n=0}^{\infty} \binom{\alpha}{n} z^n$ ist nach Beispiel 12.6 $R(P) = 1$.

Für den Polylogarithmus $P(z) = \mathrm{Li}_s(z) := \sum_{n=1}^{\infty} \frac{z^n}{n^s}$ ist $R(P) = 1$.

Absolut konvergente Reihen können nach Satz 12.9 beliebig umgeordnet werden. Dadurch wird es möglich, zwei Reihen innerhalb des gemeinsamen Konvergenzkreises zu multiplizieren und nach gemeinsamen Potenzen von z umzuordnen, ähnlich zum Produkt von Polynomen auf Seite 59:

Satz 14.4 *Konvergieren die Potenzreihen* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *und* $g(z) = \sum_{n=0}^{\infty} b_n z^n$ *im Punkt* $z \in \mathbb{C}$ *absolut, so gilt*

$$f(z) \cdot g(z) = \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{n} a_{n-k} b_k \Big) z^n .$$

Beweis. Wegen der absoluten Konvergenz gibt es $A, B \in \mathbb{R}_+^\times$ mit $\sum_{n=0}^{\infty} |a_n z^n| \leq A$ und $\sum_{n=0}^{\infty} |b_n z^n| \leq B$. Nach dem Cauchy-Kriterium gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $\sum_{n=k}^{m} |a_n z^n| < \frac{\epsilon}{3B}$ und gleichzeitig $\sum_{n=k}^{m} |b_n z^n| < \frac{\epsilon}{3A}$ für alle $m \geq k \geq N$. Wir betrachten die Partialsummen $f_N(z) = \sum_{n=0}^{2N} a_n z^n$, $g_N(z) = \sum_{n=0}^{2N} b_n z^n$ und $h_N(z) := \sum_{n=0}^{2N} \sum_{k=0}^{n} a_k b_{n-k} z^n$. Dann ist

$$\begin{aligned}
&|f_N(z) g_N(z) - h_N(z)| \\
&= \Big| \sum_{n=0}^{4N} \sum_{k=\max(0,n-2N)}^{\min(2N,n)} a_k b_{n-k} z^n - \sum_{n=0}^{2N} \sum_{k=0}^{n} a_k b_{n-k} z^n \Big| = \Big| \sum_{n=2N+1}^{4N} \sum_{k=n-2N}^{2N} a_k b_{n-k} z^n \Big| \\
&= \Big| \sum_{n=2N+1}^{3N} \sum_{k=n-2N}^{N} a_k b_{n-k} z^n + \sum_{n=2N+1}^{3N} \sum_{k=N+1}^{2N} a_k b_{n-k} z^n + \sum_{n=3N+1}^{4N} \sum_{k=n-2N}^{2N} a_k b_{n-k} z^n \Big| \\
&\leq \sum_{k=0}^{N} |a_k z^k| \sum_{n=2N+1}^{3N} |b_{n-k} z^{n-k}| + \sum_{k=N+1}^{2N} |a_k z^k| \sum_{n=2N+1}^{3N} |b_{n-k} z^{n-k}| \\
&\qquad + \sum_{k=N+1}^{2N} |a_k z^k| \sum_{n=3N+1}^{4N} |b_{n-k} z^{n-k}| \\
&< \frac{\epsilon}{3A} \sum_{k=0}^{N} |a_k z^k| + B \sum_{k=N+1}^{2N} |a_k z^k| + \frac{\epsilon}{3A} \sum_{k=N+1}^{2N} |a_k z^k| \leq \epsilon .
\end{aligned}$$

Somit gilt $\lim_{m\to\infty} f_m(z) g_m(z) = \lim_{m\to\infty} h_m(z)$. Analog ergibt sich

$$\Big(\sum_{n=0}^{\infty} |a_n z^n| \Big) \Big(\sum_{n=0}^{\infty} |b_n z^n| \Big) = \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{n} |a_{n-k}|\, |b_k| \Big) |z|^n .$$

Aus $\Big| \sum_{k=0}^{n} a_{n-k} b_k z^n \Big| \leq \sum_{k=0}^{n} |a_{n-k}|\, |b_k| |z|^n$ folgt nun die absolute Konvergenz von $f(z) \cdot g(z)$. $\square$

Satz 14.5 (Multiplikation der Binomialreihen) i) *Für alle* $s, t \in \mathbb{C}$ *und* $z \in \mathbb{C}$ *mit* $|z| < 1$ *gilt* $B_s(z) \cdot B_t(z) = B_{s+t}(z)$.

ii) *Für alle* $x \in \mathbb{R}$ *mit* $|x| < 1$ *und alle* $s \in \mathbb{Q}$ *gilt* $B_s(x) = (1+x)^s$

Beweis. i) Nach Satz 13.8 gilt

$$B_s(z) \cdot B_t(z) = \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{n} \binom{s}{k} \binom{t}{n-k} \Big) z^n = \sum_{n=0}^{\infty} \binom{s+t}{n} z^n = B_{s+t}(z) \ .$$

ii) Aus $B_p(x) = (1+x)^p$ für $p \in \mathbb{N}$ folgt mit $s = \frac{p}{q}$ und $q \in \mathbb{N}^\times$

$$\underbrace{B_{\frac{p}{q}}(x) \cdots B_{\frac{p}{q}}(x)}_{q \text{ mal}} = B_p(x) = (1+x)^p \ ,$$

also $B_{\frac{p}{q}}(x) = (1+x)^{\frac{p}{q}}$ wegen der Eindeutigkeit der Wurzel reeller Zahlen. Für negative s benutzt man $B_s(x) \cdot B_{-s}(x) = B_0(x) = 1$. $\square$

Daraus ergibt sich z.B.

$$\begin{aligned}
\sqrt{1+x} &= \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} x^n = 1 + \frac{\frac{1}{2}}{1} x + \frac{\frac{1}{2}}{1} \cdot \frac{(\frac{1}{2}-1)}{2} x^2 + \frac{\frac{1}{2}}{1} \cdot \frac{(\frac{1}{2}-1)}{2} \frac{(\frac{1}{2}-2)}{3} x^3 + \dots \\
&= 1 + \frac{1}{2} x - \frac{1}{2 \cdot 4} x^2 + \frac{1 \cdot 3}{2 \cdot 4 \cdot 6} x^3 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 8} x^4 + \dots \\
\frac{1}{\sqrt{1+x}} &= \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} x^n = 1 + \frac{-\frac{1}{2}}{1} x + \frac{-\frac{1}{2}}{1} \cdot \frac{(-\frac{1}{2}-1)}{2} x^2 + \dots \\
&= 1 - \frac{1}{2} x + \frac{1 \cdot 3}{2 \cdot 4} x^2 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} x^3 + \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8} x^4 + \dots
\end{aligned}$$

Als weiteres Beispiel können wir die (sehr wichtige) Exponentialreihe einführen und untersuchen:

Satz 14.6 i) *Für alle* $z \in \mathbb{C}$ *ist die Exponentialreihe* $\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}$ *absolut konvergent, und es gilt* $|\exp(z)| \leq \exp(|z|)$.

ii) *Für alle* $w, z \in \mathbb{C}$ *gilt* $\exp(w) \cdot \exp(z) = \exp(w+z)$.

iii) *Für* $N \in \mathbb{N}$ *mit* $|z| < 1 + \frac{N}{2}$ *gilt* $\Big| \exp(z) - \sum_{n=0}^{N} \frac{z^n}{n!} \Big| \leq \frac{2|z|^{N+1}}{(N+1)!}$.

iv) *Es gilt* $\exp(z) = \lim_{n \to \infty} (1 + \frac{z}{n})^n$, *insbesondere* $e = \exp(1)$.

Beweis. i) Das Quotientenkriterium liefert $\lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| = \lim_{n\to\infty} \frac{|z|}{n+1} = 0 < 1$, damit ist $\exp(z)$ für alle $z \in \mathbb{C}$ absolut konvergent. Die Ungleichung $|\exp(z)| \leq \exp(|z|)$ folgt aus Satz 12.2.

ii) Für $t \in \mathbb{C}$ ist nach Satz 14.4

$$\begin{aligned}\exp(wt)\cdot\exp(zt) &= \sum_{n=0}^{\infty}\Big(\sum_{k=0}^{n}\frac{w^{n-k}}{(n-k)!}\frac{z^k}{k!}\Big)t^n = \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\sum_{k=0}^{n}\binom{n}{k}w^{n-k}z^k\Big)t^n \\ &= \sum_{n=0}^{\infty}\frac{(w+z)^n t^n}{n!} = \exp((w+z)t)\;.\end{aligned}$$

iii) Für $|z| \leq \frac{N+2}{2}$ gilt

$$\begin{aligned}\Big|\exp(z)-\sum_{n=0}^{N}\frac{z^n}{n!}\Big| &= \Big|\sum_{n=N+1}^{\infty}\frac{z^n}{n!}\Big| \leq \sum_{n=N+1}^{\infty}\frac{|z|^n}{n!} \\ &= \frac{|z|^{N+1}}{(N+1)!}\Big(1+\frac{|z|}{N+2}+\frac{|z|^2}{(N+2)(N+3)}+\dots\Big) \\ &\leq \frac{|z|^{N+1}}{(N+1)!}\sum_{k=0}^{\infty}\Big(\frac{|z|}{N+2}\Big)^k \leq \frac{|z|^{N+1}}{(N+1)!}\sum_{k=0}^{\infty}\frac{1}{2^k} = \frac{2|z|^{N+1}}{(N+1)!}\;.\end{aligned}$$

iv) Aus ii) folgt $\exp(z) = \big(\exp(\frac{z}{n})\big)^n$ für alle $n \in \mathbb{N}^\times$. Damit gilt

$$\begin{aligned}\Big|\exp(z)-\big(1+\tfrac{z}{n}\big)^n\Big| &= \Big|\big(\exp(\tfrac{z}{n})\big)^n-\big(1+\tfrac{z}{n}\big)^n\Big| \\ &= \Big|\big(\exp(\tfrac{z}{n})-(1+\tfrac{z}{n})\big)\sum_{k=0}^{n-1}\big(\exp(\tfrac{z}{n})\big)^k\big(1+\tfrac{z}{n}\big)^{n-k-1}\Big| \\ &\leq \big|\exp(\tfrac{z}{n})-(1+\tfrac{z}{n})\big|\sum_{k=0}^{n-1}\big(\exp(\tfrac{|z|}{n})\big)^k\big(1+\tfrac{|z|}{n}\big)^{n-k-1}\;.\end{aligned}$$

Offenbar gilt $1+\frac{|z|}{n} \leq \exp(\frac{|z|}{n})$ und deshalb

$$\sum_{k=0}^{n-1}\big(\exp(\tfrac{|z|}{n})\big)^k\big(1+\tfrac{|z|}{n}\big)^{n-k-1} \leq n\big(\exp(\tfrac{|z|}{n})\big)^{n-1} \leq n\big(\exp(\tfrac{|z|}{n})\big)^n = n\exp(z)\;.$$

Schließlich gilt nach iii) für $N=1$ und $|\frac{z}{n}| < \frac{3}{2}$, also $n \geq \frac{3}{2}|z|$,

$$\big|\exp(\tfrac{z}{n})-(1+\tfrac{z}{n})\big| \leq 2|\frac{|z|^2}{2n^2}$$

und damit $\Big|\exp(z)-\big(1+\frac{z}{n}\big)^n\Big| \leq \dfrac{|z|^2}{n^2}\cdot n\exp(z)$. Für $n\to\infty$ folgt die Behauptung.

□

Diese Eigenschaften haben mehrere Folgerungen.

i) Wegen $\exp(0) = 1$ existiert für alle $z \in \mathbb{C}$ das Inverse $\big(\exp(z)\big)^{-1} = \exp(-z) \in \mathbb{C}$, damit ist $\exp(z) \neq 0$ für alle $z \in \mathbb{C}$.

ii) Die Exponentialreihe konvergiert sehr schnell; im Schritt $s_N \to s_{N+1}$ der Partialsummen verbessert sich die Konvergenz um den Faktor $N+2$. Damit kann die Eulersche Zahl e numerisch gut berechnet werden. Die Folge $e = \lim\limits_{n\to\infty} (1+\frac{1}{n})^n$ konvergiert viel schlechter.

iii) Es gilt $\exp s = e^s$ für alle $s \in \mathbb{Q}$.

iv) Im Reellen folgt aus $x_1 > x_2 > 0$ die Beziehung $\exp(x_1) > \exp(x_2)$. Über $\exp(-x) = \frac{1}{\exp(x)}$ folgt dann $\exp(x_1) > \exp(x_2)$ für alle $x_1 > x_2$. Damit ist die zugehörige reelle Exponentialfunktion streng monoton wachsend auf $\mathbb{R}$.

Restgliedabschätzungen wie bei der Exponentialfunktion werden oft benötigt.

Satz 14.7 *Eine Potenzreihe* $P(z) = \sum\limits_{n=0}^{\infty} a_n z^n$ *habe einen Konvergenzradius* $R(P) > 0$. *Dann gibt es zu jedem* $0 < r < R(P)$ *und jedem* $p \in \mathbb{N}$ *eine Konstante* $c_p \in \mathbb{R}$, *so daß* $\Big|\sum\limits_{n=p}^{\infty} a_n z^n\Big| < c_p |z|^p$ *für alle* $z \in \mathbb{C}$ *mit* $|z| \leq r$.

Beweis. Setze $c_p := \sum\limits_{k=0}^{\infty} |a_{k+p}| r^k$. Dann gilt $\Big|\sum\limits_{n=p}^{\infty} a_n z^n\Big| \leq \sum\limits_{n=p}^{\infty} |a_n||z|^n = \sum\limits_{k=0}^{\infty} |a_{k+p}||z|^k |z|^p \leq c_p |z|^p$. □

Eine wichtige Anwendung ist folgende Aussage über mögliche Nullstellen einer Potenzreihe:

Satz 14.8 *Eine Potenzreihe* $P(z) = \sum\limits_{n=0}^{\infty} a_n z^n$, *mit* $a_k \neq 0$ *für mindestens ein* k, *habe einen Konvergenzradius* $R(P) > 0$. *Dann gibt es ein* $0 < r < R(P)$, *so daß höchstens endlich viele Nullstellen in der Kreisscheibe* $\overline{K_r(0)} = \{z \in \mathbb{C} \;:\; |z| \leq r\}$ *liegen.*

Beweis. Es sei $N = \min(k \;:\; a_k \neq 0)$. Zu beliebigem $0 < r < R(P)$ gibt es nach Satz 14.7 ein $c > 0$ mit $\Big|P(z) - a_N z^N\Big| \leq c|z|^{N+1}$. Wäre die Aussage falsch, so enthielte jeder Kreis mit Radius $\frac{r}{k}$ um 0, mit $k \in \mathbb{N}^\times$, eine Nullstelle $z_k \neq 0$ von P. Für diese gilt $|a_N z_k^N| \leq c|z_k|^{N+1}$, also $|a_N| \leq c|z_k| \leq \frac{r}{k}$. Für $k \to \infty$ ergibt sich ein Widerspruch zu $a_N \neq 0$. □

Satz 14.9 (Identitätssatz für Potenzreihen) *Für zwei Potenzreihen* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *und* $g(z) = \sum_{n=0}^{\infty} b_n z^n$ *mit positiven Konvergenzradien gebe es eine Nullfolge* $(z_k)_{k\in\mathbb{N}}$ *mit* $f(z_k) = g(z_k)$*. Dann gilt* $a_n = b_n$ *für alle* $n \in \mathbb{N}$.

Beweis. Folgt für $P(z) = f(z) - g(z)$ aus Satz 14.8. □

Gilt z.B. $f(z) = g(z)$ innerhalb einer Kreisscheibe um 0 oder auch nur entlang eines die 0 enthaltenden Kurvenstücks, so ist $f = g$ innerhalb des gemeinsamen Konvergenzkreises.

Wiederholung

- Folgen, Konvergenz, Grenzwert
- elementare Grenzwerte, Eulersche Zahl
- Satz von Bolzano-Weierstraß
- Cauchy-Folgen, Vollständigkeit über Cauchy-Kriterium
- Reihen: Konvergenz, geometrische Reihe, Leibniz-Kriterium
- absolute Konvergenz: Majorantenkriterium, Quotientenkriterium, Wurzelkriterium, Umordnungssatz
- Polynome: Polynomdivision, Nullstellen, Partialbruchzerlegung
- Potenzreihen: Konvergenzradius, Multiplikationssatz
- Binomialreihe, Exponentialreihe
- Identitätssätze für Polynome und Potenzreihen

Teil IV
Metrische Räume und Stetigkeit

15 Euklidische, unitäre und normierte Vektorräume

Definition 15.1 Es sei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ und V ein Vektorraum über $\mathbb{K}$. Ein Skalarprodukt auf V ist eine Abbildung $\langle\ .\ \rangle : V \times V \to \mathbb{K}$ mit

(S1) $\langle\ .\ \rangle$ ist linear in der zweiten Variablen, d.h.

$$\langle w, \lambda_1 v_1 + \lambda_2 v_2 \rangle = \lambda_1 \langle w, v_1 \rangle + \lambda_2 \langle w, v_2 \rangle$$

für alle $v_1, v_2, w \in V$ und $\lambda_1, \lambda_2 \in \mathbb{K}$.

(S2) Es gilt $\langle w, v \rangle = \overline{\langle v, w \rangle}$ (komplexe Konjugation) für alle $v, w \in V$.

(S3) Es gilt $\langle v, v \rangle \geq 0$ für alle $v \in V$ und $\langle v, v \rangle = 0$ genau dann, wenn $v = 0$.

Ein reeller bzw. komplexer Vektorraum mit Skalarprodukt heißt *euklidischer bzw. unitärer Vektorraum* oder (in der Funktionalanalysis) *Prä-Hilbert-Raum*.

Achtung: Wir verwenden hier die Konvention der Physik. In der mathematischen Literatur wird das Skalarprodukt als *linear in der ersten Komponente definiert*. In unserer Konvention folgt aus (S1) und (S2) für die erste Komponente

$$\langle \lambda_1 w_1 + \lambda_2 w_2, v \rangle = \overline{\langle v, \lambda_1 w_1 + \lambda_2 w_2 \rangle} = \overline{\lambda_1 \langle v, w_1 \rangle + \lambda_2 \langle v, w_2 \rangle} = \overline{\lambda_1} \langle w_1, v \rangle + \overline{\lambda_2} \langle w_2, v \rangle .$$

In reellen Vektorräumen ist das Skalarprodukt also auch in der ersten Variablen linear. Außerdem vereinfacht sich (S2) in reellen Vektorräumen zur Symmetrie $\langle v, w \rangle = \langle w, v \rangle$. Es sei bemerkt, daß aus (S2) folgt $\langle v, v \rangle \in \mathbb{R}$, so daß (S3) sinnvoll ist.

Beispiel 15.2 (Standardskalarprodukt in $\mathbb{R}^n$) Durch

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n , \qquad \text{für} \quad \begin{array}{l} x = (x_1, \dots, x_n) \in \mathbb{R}^n , \\ y = (y_1, \dots, y_n) \in \mathbb{R}^n \end{array}$$

wird ein Skalarprodukt $\langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ definiert. Dieses heißt das *kanonische Skalarprodukt* im $\mathbb{R}^n$. ◁

Beispiel 15.3 (Standardskalarprodukt in $\mathbb{C}^n$) Durch

$$\langle w, z \rangle := \overline{w_1} z_1 + \overline{w_2} z_2 + \cdots + \overline{w_n} z_n , \qquad \text{für} \quad \begin{array}{l} w = (w_1, \dots, w_n) \in \mathbb{C}^n , \\ z = (z_1, \dots, z_n) \in \mathbb{C}^n \end{array}$$

wird ein Skalarprodukt $\langle\ ,\ \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ definiert. Dieses heißt das *kanonische Skalarprodukt* im $\mathbb{C}^n$. ◁

Beispiel 15.4 Es sei $V = \mathbb{R}^2$. Dann definiert

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = (2x_1 + x_2)y_1 + (x_1 + 2x_2)y_2 = (2y_1 + y_2)x_1 + (y_1 + 2y_2)x_2$$

ein Skalarprodukt. (S3) folgt aus $\langle x, x\rangle = 2x_1^2 + 2x_1x_2 + 2x_2^2 = (x_1 + x_2)^2 + x_1^2 + x_2^2 \geq 0$ mit Gleichheit genau dann, wenn $x_1 = x_2 = 0$. Dieses Skalarprodukt ist verschieden vom kanonischen. Tatsächlich gibt es unendlich viele verschiedene Skalarprodukte auf $\mathbb{K}^n$. ◁

Beispiel 15.5 Es sei

$$\ell^2 := \Big\{ (a_n)_{n\in\mathbb{N}} \;:\; \sum_{n=0}^{\infty} |a_n|^2 < \infty \Big\}$$

der Vektorraum der *quadrat-summierbaren* komplexen Zahlenfolgen. Dann definiert

$$\langle a, b\rangle := \sum_{n=0}^{\infty} \overline{a_n} b_n \;, \qquad a = (a_n)_{n\in\mathbb{N}},\; b = (b_n)_{n\in\mathbb{N}} \in \ell^2$$

ein Skalarprodukt auf ℓ^2. Dabei wird verwendet, daß wegen $|\overline{a_n}b_n| = |a_n|\,|b_n| \leq \frac{1}{2}(|a_n|^2 + |b_n|^2)$ die unendliche Reihe nach Majorantenkriterium absolut konvergiert. Ist $\langle a, a\rangle = 0$, so folgt $|a_n|^2 = 0$ für alle n, und $(0)_{n\in\mathbb{N}}$ ist das neutrale Element von ℓ^2. ◁

Wir kommen nun zur wichtigsten Ungleichung für Skalarprodukte.

Satz 15.6 (Cauchy-Schwarzsche Ungleichung) *Es sei V ein Vektorraum über $\mathbb{K}$ mit Skalarprodukt $\langle\ ,\ \rangle : V \times V \to \mathbb{K}$. Dann gilt*

$$\big|\langle w, v\rangle\big|^2 \leq \langle v, v\rangle\langle w, w\rangle \qquad \textit{für alle } v, w \in V$$

mit Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis. i) Sei $w \neq 0$ (für $w = 0$ ist die Ungleichung offensichtlich erfüllt), somit ist $\langle w, w\rangle \neq 0$. Dann gilt für *beliebiges* $\lambda \in \mathbb{C}$

$$\begin{aligned} 0 \leq \langle \lambda w - v, \lambda w - v\rangle &= \lambda\overline{\lambda}\langle w, w\rangle - \overline{\lambda}\langle w, v\rangle - \lambda\langle v, w\rangle + \langle v, v\rangle \\ &= \langle w, w\rangle \left|\lambda - \frac{\langle w, v\rangle}{\langle w, w\rangle}\right|^2 + \frac{1}{\langle w, w\rangle}\Big(\langle w, w\rangle\langle v, v\rangle - \big|\langle w, v\rangle\big|^2\Big) \;. \end{aligned}$$

Insbesondere gilt die Ungleichung für $\lambda = \frac{\langle w,v\rangle}{\|w\|^2}$ und wird zur Ungleichung von Cauchy-Schwarz.

ii) Gilt das Gleichheitszeichen in Cauchy-Schwarz, so ist

$$\langle \lambda w - v, \lambda w - v\rangle = \langle w, w\rangle \left|\lambda - \frac{\langle w, v\rangle}{\langle w, w\rangle}\right|^2$$

für beliebiges $\lambda \in \mathbb{C}$. Damit gilt $\langle \lambda w - v, \lambda w - v \rangle = 0$ für $\lambda = \frac{\langle w,v \rangle}{\langle w,w \rangle}$, somit $v = \lambda w$. □

Definition 15.7 Es sei V ein Vektorraum über $\mathbb{K}$. Eine *Norm* auf V ist eine Abbildung $\| \ \| : V \to \mathbb{R}_+$ mit

(N1) $\|v\| = 0 \quad \Leftrightarrow \quad v = 0$,

(N2) $\|\lambda v\| = |\lambda| \, \|v\|$ für alle $v \in V$, $\lambda \in \mathbb{K}$,

(N3) $\|v + w\| \leq \|v\| + \|w\|$ für alle $v, w \in V$ (Dreiecksungleichung)

Ein Vektorraum mit Norm heißt *normierter Vektorraum*.

Satz 15.8 *Jeder euklidische und unitäre Vektorraum ist auch ein normierter Vektorraum mit* $\|v\| := \sqrt{\langle v, v \rangle}$.

Beweis. (N1) und (N2) sind klar, (N3) folgt aus Cauchy-Schwarz:

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w \rangle = \langle v, v \rangle + 2\mathrm{Re}(\langle v, w \rangle) + \langle w, w \rangle \\ &\leq \langle v, v \rangle + 2|\langle v, w \rangle| + \langle w, w \rangle \\ &\leq \langle v, v \rangle + 2\sqrt{\langle v, v \rangle \langle w, w \rangle} + \langle w, w \rangle = (\|v\| + \|w\|)^2 \,. \end{aligned}$$

□

Damit liefern die Beispiele 15.2–15.5 auch Beispiele für normierte Vektorräume. Dabei werden die durch die Standard-Skalarprodukte aus den Beispielen 15.2,15.3 und 15.5 induzierten Normen meist mit $\| \ \|_2$ bezeichnet. Weitere Beispiele sind:

Beispiel 15.9

i) $V = \mathbb{K}^n$ ist mit $\|x\|_1 := \sum_{i=1}^n |x_i|$ für $x = (x_1, \dots, x_n) \in \mathbb{K}^n$ ein normierter Vektorraum $(\mathbb{K}^n, \| \ \|_1)$. (N1) und (N2) sind klar, (N3) folgt aus der Dreiecksungleichung in $\mathbb{R}$ bzw. $\mathbb{C}$ für jede Komponente.

ii) $\|x\|_p := \big(\sum_{i=1}^n |x_i|^p \big)^{\frac{1}{p}}$ ist für jedes $1 \leq p < \infty$ eine Norm auf $\mathbb{K}^n$. Vorerst können wir die Norm nur für $p \in \mathbb{Q}$ definieren. Der typische Beweis erfordert die Minkowskische Ungleichung, die erst später bewiesen wird.◁

Beispiel 15.10 Es sei

$$\ell^1 := \Big\{ (a_n)_{n \in \mathbb{N}} \ : \ \sum_{n=0}^{\infty} |a_n| < \infty \Big\}$$

der Vektorraum der absolut-konvergenten Reihen (oder der summierbaren komplexen Zahlenfolgen). Dann definiert

$$\|a\|_1 := \sum_{n=0}^{\infty} |a_n|$$

eine Norm auf ℓ^1. Tatsächlich gilt

$$\|a+b\|_1 = \sum_{n=0}^{\infty} |a_n + b_n| \leq \sum_{n=0}^{\infty} |a_n| + \sum_{n=0}^{\infty} |b_n| = \|a\|_1 + \|b\|_1 \,.$$

Aus $\|a\|_1 = 0$ folgt $a_n = 0$ für alle n, und $(0)_{n\in\mathbb{N}}$ ist das neutrale Element von ℓ^1. ◁

Beispiel 15.11 Sei M eine beliebige Menge und $\ell^\infty(M)$ die Menge der beschränkten Abbildungen $f : M \to \mathbb{K}$. Mit der punktweisen Addition $(f_1 + f_2)(m) := f_1(m) + f_2(m)$ und $(\lambda f)(m) := \lambda f(m)$ wird $\ell^\infty(M)$ zu einem Vektorraum. Dann wird durch

$$\|f\|_\infty := \sup_{m\in M} |f(m)|$$

eine Norm (die Supremums-Norm) erklärt. Tatsächlich gilt

$$\begin{aligned} \|f+g\|_\infty &:= \sup_{m\in M} |f(m)+g(m)| \leq \sup_{m\in M} (|f(m)| + |g(m)|) \\ &\leq \sup_{m',m''\in M} (|f(m')| + |g(m'')|) = \|f\|_\infty + \|g\|_\infty \,. \end{aligned}$$

Ist $\|f\|_\infty = 0$, so ist $f(m) = 0$ für alle m, und f ist die Nullabbildung. Somit ist $(\ell^\infty(M), \|\;\|_\infty)$ ein normierter Vektorraum.

Ein wichtiger Spezialfall ist $\ell^\infty := \ell^\infty(\mathbb{N})$. Der Vektorraum $\ell^\infty(\{1,2,\dots,n\})$ kann mit $\mathbb{K}^n$ identifiziert werden, und die Supremumsnorm mit $\|(x_1,\dots,x_n)\|_\infty := \max_{i=1,\dots,n} |x_i|$. ◁

Satz 15.12 *Sind $a = (a_n)_{n\in\mathbb{N}} \in \ell^1$ und $b = (b_n)_{n\in\mathbb{N}} \in \ell^\infty$, so gilt für die Folge $ab := (a_n b_n)_{n\in\mathbb{N}}$ der Produkte $\|ab\|_1 \leq \|a\|_1 \|b\|_\infty$. Insbesondere ist $ab \in \ell^1$.*

Beweis. Die Folge ab hat wegen $|b_n| \leq \|b\|_\infty$ die konvergente Majorante $a\|b\|_\infty \in \ell^1$. □

Mit $\|a\|_1$, $\|a_2\|$ und $\|a\|_\infty$ haben wir auf dem unendlich-dimensionalen Vektorraum aller (nicht notwendig konvergenten) Folgen $(a_n)_{n\in\mathbb{N}}$ *verschiedene* Normen erklärt, so daß ℓ^1, ℓ^2 und ℓ^∞ verschiedene normierte Vektorräume werden. Die konstante Folge $a_n = 1$ liegt in ℓ^∞, nicht aber in ℓ^1 und ℓ^2. Die harmonische Reihe $a_n = \frac{1}{n+1}$ liegt in ℓ^2, nicht aber in ℓ^1. In endlich-dimensionalen Vektorräumen kann es diese Unterscheidung nicht geben:

Definition 15.13 Zwei Normen $\|\;\|_a$ und $\|\;\|_b$ auf einem Vektorraum V heißen *äquivalent*, wenn es positive Zahlen $c, C > 0$ gibt, so daß für beliebige $x \in V$ gilt $c\|x\|_a \leq \|x\|_b \leq C\|x\|_a$.

Satz 15.14 *Je zwei Normen auf einem n-dimensionalen Vektorraum V über $\mathbb{K}$ sind äquivalent.*

Beweis. i) Zunächst zeigen wir für $V = \mathbb{R}^n$, daß eine beliebige Norm $\| \ \|$ äquivalent ist zur Standardnorm $\| \ \|_2 = \sqrt{\langle \ , \ \rangle}$. Es sei $(e_1, \dots, e_n)$ die Standardbasis. Dann gilt für $x = \sum_{i=1}^n x_i e_i$ nach Dreiecksungleichung und Cauchy-Schwarz

$$\|x\| \leq \sum_{i=1}^n |x_i| \|e_i\| \leq \sqrt{\sum_{i=1}^n |x_i|^2} \underbrace{\sqrt{\sum_{i=1}^n \|e_i\|^2}}_{C} = C\|x\|_2 \,. \qquad (*)$$

Zur Umkehrung betrachten wir die Einheitssphäre $S^{n-1} := \{y \in \mathbb{R}^n \ : \ \|y\|_2 = 1\} \subseteq \mathbb{R}^n$. Sei $c := \inf_{y \in S^{n-1}} \|y\|$. Wir zeigen im folgenden Lemma: Das Infimum ist Minimum, d.h. es gibt ein $a \in S^{n-1}$ mit $\|a\| = c$. Wegen $0 \notin S^{n-1}$ ist dann $c > 0$, und es folgt $c\|x\|_2 \leq \|x\|$ für alle $x \in \mathbb{R}^n$ (klar für $x = 0$, ansonsten setze $y := \frac{x}{\|x\|_2} \in S^{n-1}$).

Damit ist der Beweis für $V = \mathbb{R}^n$ abgeschlossen. Für $V = \mathbb{C}^n$ entsteht für $x = \sum_{i=1} x_i e_i$ mit $x_i \in \mathbb{C}$ in (*) die Ungleichung $\|x\| \leq C\sqrt{\sum_{i=1}^n ((\mathrm{Re}(x_i))^2 + (\mathrm{Im}(x_i))^2)} = C\|(\mathrm{Re}(x), \mathrm{Im}(x))\|_2$, so daß wir $(\mathrm{Re}(x), \mathrm{Im}(x))$ als Punkt im $\mathbb{R}^{2n}$ auffassen können. Nach analoger Diskussion der Sphäre S^{2n-1} ist dann auch der Fall $V = \mathbb{C}^n$ bewiesen.

ii) Sei nun V ein beliebiger n-dimensionaler Vektorraum und $\| \ \|_a$, $\| \ \|_b$ zwei Normen auf V. Wir wählen eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V und zerlegen einen beliebigen Vektor $v = x_1 v_1 + \dots + x_n v_n \in V$ nach dieser Basis. Dann sind

$$\|(x_1, \dots, x_n)\|_{a,\mathcal{B}} := \|x_1 v_1 + \dots + x_n v_n\|_a \,, \quad \|(x_1, \dots, x_n)\|_{b,\mathcal{B}} := \|x_1 v_1 + \dots + x_n v_n\|_b$$

zwei Normen auf $\mathbb{K}^n$. Somit gibt es $c, C > 0$, so daß für alle $x_1, \dots, x_n \in \mathbb{K}$ gilt

$$c\|(x_1, \dots, x_n)\|_{a,\mathcal{B}} \leq \|(x_1, \dots, x_n)\|_{b,\mathcal{B}} \leq C\|(x_1, \dots, x_n)\|_{a,\mathcal{B}} \,.$$

Folglich ist $c\|v\|_a \leq \|v\|_b \leq C\|v\|_a$ für beliebige $v \in V$. □

Lemma 15.15 *Sei $\| \ \|$ eine beliebige Norm auf $\mathbb{R}^n$ und $S^{n-1} := \{y \in \mathbb{R}^n \ : \ \|y\|_2 = 1\}$ die Einheitssphäre bezüglich der Standard-Norm $\| \ \|_2$. Sei $c := \inf_{y \in S^{n-1}} \|y\|$. Dann gibt es ein $a \in S^{n-1}$ mit $\|a\| = c$.*

Beweis. Nach Cauchy-Schwarz gilt $0 < \|x\| \leq C$ für alle $x \in S^{n-1}$. Damit ist die Menge $T := \{\|x\| \ : \ x \in S^{n-1}\} \subseteq \mathbb{R}$ beschränkt, besitzt also ein Infimum c. Angenommen, c ist nicht Minimum. Dann muß T eine unendliche Menge von Punkten sein, und da $c + \frac{1}{n+1}$ keine untere Schranke von T ist, gibt es eine gegen c konvergente Folge von Punkten aus T. Somit gibt es auch eine Folge $(x_k)_{k \in \mathbb{N}}$ von Punkten $x_k \in S^{n-1}$, für die die Folge $(\|x_k\|)_{k \in \mathbb{N}}$ reeller Zahlen gegen c konvergiert.

Wir betrachten die Komponenten x_{ki} von $x_k = (x_{k1}, \dots, x_{kn})$. Wegen $\sum_{i=1}^n |x_{ki}|^2 = 1$ ist die Folge $(x_{k1})_{k\in\mathbb{N}}$ beschränkt. Damit enthält sie nach dem Satz von Bolzano-Weierstraß eine gegen a_1 konvergente Teilfolge $(x_{k_m 1})_{m\in\mathbb{N}}$. Analog konstruiert man eine gegen a_2 konvergente Teilfolge von $(x_{k_m 2})_{m\in\mathbb{N}}$, usw., bis schließlich eine Teilfolge von (x_k) erhalten ist, die komponentenweise gegen $a = (a_1, \dots, a_n)$ konvergiert. Diese konvergente Teilfolge sei wieder mit $(x_k)_{k\in\mathbb{N}}$ bezeichnet. Dann gilt $a_1^2 + \dots + a_n^2 = \lim_{k\to\infty}(x_{k1}^2 + \dots + x_{kn}^2) = 1$, d.h. $a \in S^{n-1}$. Andererseits gilt nach Dreiecksungleichung und Cauchy-Schwarz

$$c \le \|a\| \le \|a - x_k\| + \|x_k\| \le C\|a - x_k\|_2 + \|x_k\| \quad \text{für alle } k \in \mathbb{N} .$$

Nun ist $\lim_{k\to\infty} \|a - x_k\|_2 = \lim_{k\to\infty} \sqrt{|a_1 - x_{k1}|^2 + \dots |a_n - x_{kn}|^2} = 0$ und $\lim_{k\to\infty} \|x_k\| = c$, also $\|a\| = c$ und damit c Minimum. □

Abschließend sei bemerkt, daß die durch ein Skalarprodukt induzierte Norm die einzige ist, die die Parallelogrammgleichung erfüllt:

Satz 15.16 *Es sei $(V, \|\ \|)$ ein normierter Vektorraum über $\mathbb{K}$. Die Norm $\|\ \|$ geht genau dann aus einem Skalarprodukt hervor, wenn die* Parallelogrammgleichung *gilt,*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2 \qquad \textit{für alle } v, w \in V .$$

Beweisidee. Die Richtung ($\Rightarrow$) ist einfaches Nachrechnen. Für die Umkehrung ($\Leftarrow$) definiert man über die *Polarisationsformeln*

$$\begin{aligned} \langle v, w\rangle &= \tfrac{1}{4}\big(\|v + w\|^2 - \|v - w\|^2 - \mathrm{i}\|v + \mathrm{i}w\|^2 + \mathrm{i}\|v - \mathrm{i}w\|^2\big) && \text{für } \mathbb{K} = \mathbb{C} , \\ \langle v, w\rangle &= \tfrac{1}{4}\big(\|v + w\|^2 - \|v - w\|^2\big) && \text{für } \mathbb{K} = \mathbb{R} \end{aligned}$$

zunächst Abbildungen $\langle\ ,\ \rangle : V \times V \to \mathbb{K}$, für die man dann die Eigenschaften (S1)–(S3) zeigen muß, wobei die Parallelogrammgleichung eingeht. Für die Linearität ist das eine sehr mühsame Rechnung! □

16 Metrische Räume

Für relle oder komplexe Zahlen hatten wir einen Abstand $|z - w|$ eingeführt. Zentral war die Dreiecksungleichung, die in vielen Konvergenzbeweisen benötigt wurde. Dieser Abstandsbegriff wird nun umfassend verallgemeinert:

Definition 16.1 Ein *metrischer Raum* ist eine Menge X zusammen mit einer Abbildung $d : X \times X \to \mathbb{R}_+$, dem *Abstand* oder der *Metrik*, wenn gilt:

(D1) $d(x, y) = 0 \quad \Leftrightarrow \quad x = y$,

(D2) $d(x, y) = d(y, x)$ (Symmetrie),

(D3) $d(x, z) \le d(x, y) + d(y, z)$ (Dreiecksungleichung).

Jeder normierte Vektorraum ist auch metrischer Raum mit Abstand $d(x,y) := \|x-y\|$. Für den $\mathbb{K}^n$ mit der aus dem Skalarprodukt erhaltenen Standardnorm gilt der *Satz des Pythagoras* $d(x,y) = \sqrt{|x_1-y_1|^2+\cdots+|x_n-y_n|^2}$.

Beispiel 16.2 Auf der Sphäre $S^2 = \{x = (x_1,x_2,x_3) \in \mathbb{R}^3 \ : \ \|x\|_2^2 = x_1^2 + x_2^2 + x_3^2 = 1\}$ läßt sich neben einem aus einer Norm auf dem umgebenen $\mathbb{R}^3$ geerbten Abstand folgende Metrik einführen: zu zwei Punkten $x,y \in S^2$, $x \neq y$ gibt es einen eindeutigen Großkreis G durch diese Punkte. Dieser Großkreis ist der Schnitt von S^2 mit der Ebene durch $(0,x,y)$. Dann wird der Abstand erklärt als Winkel $d(x,y) := |\measuredangle(x,0,y)|$ zwischen x,y auf G. ◁

Beispiel 16.3 Riemannsche Mannigfaltigkeiten sind eine (auch für die Physik) wichtige Klasse von Räumen M, in denen für Tangentialvektoren an Kurven durch einen beliebigen Punkt $x \in M$ ein Skalarprodukt erklärt ist (und dieses hängt in regulärer Weise vom Punkt $x \in M$ ab). Unter Verwendung der Integralrechnung läßt sich dann die Länge $L(c)$ einer (genügend regulären) Kurve zwischen Punkten $x,y \in M$ definieren, und

$$d(x,y) = \inf\{L \ : \ \text{es gibt eine Kurve } c \text{ zwischen } x,y \text{ mit } L(c) = L\}$$

definiert eine Metrik auf M. ◁

Über die Metrik führen wir den zentralen Begriff der *offenen Teilmengen* ein:

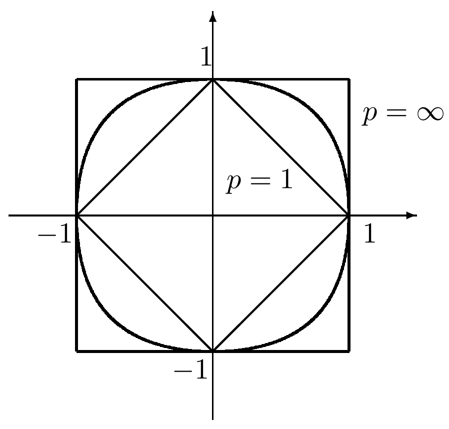
Definition 16.4 Es sei (X,d) ein metrischer Raum, $a \in X$ und $r > 0$. Dann heißt die Teilmenge

$$K_r(a) := \{x \in X \ : \ d(a,x) < r\} \subseteq X$$

die *offene Kugel* in (X,d) mit Mittelpunkt a und Radius r.

Diese Kugeln können je nach Metrik verschiedene Formen haben:

Beispiel 16.5 Betrachtet werde $X = \mathbb{R}^2$ mit den Metriken $d_p(x,y) = \|x-y\|_p$, die aus den Normen $\| \ \|_p$, $1 \leq p \leq \infty$, aus Beispiel 15.9 induziert werden. Dann haben die Einheitskugeln mit $r = 1$ um $a = 0$ folgende Gestalt:



Definition 16.6 Es sei (X, d) ein metrischer Raum.

i) Eine Teilmenge $U \subseteq X$ heißt *Umgebung* eines Punktes $a \in X$, wenn es eine offene Kugel $K_\epsilon(a) \subseteq U$ gibt. Speziell heißt $K_\epsilon(a)$ die *ϵ-Umgebung* von a.

ii) Eine Teilmenge $U \subseteq X$ heißt *offen*, wenn jeder Punkt $x \in U$ eine in U enthaltene ϵ-Umgebung besitzt, d.h. $\forall x \in U\ \exists \epsilon > 0\ :\ K_\epsilon(x) \subseteq U$. Außerdem werden die gesamte Menge $X \subseteq X$ und die leere Menge $\varnothing \subseteq X$ als offen erklärt.

iii) Die Gesamtheit aller offenen Teilmengen von (X, d) heißt die *von d erzeugte Topologie auf X*.

iv) Eine Teilmenge $A \subseteq X$ heißt *abgeschlossen*, wenn $X \setminus A$ offen ist.

v) Sei $Y \subseteq X$ eine Teilmenge. Ein Punkt $y \in X$ heißt *Randpunkt* von Y, wenn es in jeder Umgebung von y sowohl Punkte aus Y als auch aus $X \setminus Y$ gibt. Die Menge aller Randpunkte von Y heißt der *Rand* von Y und wird mit ∂Y bezeichnet.

Damit sind X und $\varnothing$ sowohl offen als auch abgeschlossen in X. Verschiedene Metriken (und Normen) könnten verschiedene Topologien erzeugen. In unendlich-dimensionalen Vektorräumen ist das in der Tat der Fall, im endlich-dimensionalen Fall nach Satz 15.14 aber nicht.

Satz 16.7 *In einem metrischen Raum (X, d) gilt:*

i) *Der Durchschnitt endlich vieler offener Mengen ist offen.*

ii) *Die Vereinigung beliebig vieler offener Mengen ist offen.*

iii) *Der Durchschnitt beliebig vieler abgeschlossener Mengen ist abgeschlossen.*

iv) *Die Vereinigung endlich vieler abgeschlossener Mengen ist abgeschlossen.*

v) *Hausdorffsches Trennungsaxiom: Zu je zwei Punkten $x, y \in X$ mit $x \neq y$ gibt es disjunkte offene Umgebungen U von x und V von y, d.h. $U \cap V = \varnothing$.*

Beweis. i) Seien $U_1, \dots, U_n \subseteq X$ offen und $U := \bigcap_{i=1}^n U_i$. Entweder ist $U = \varnothing$ (und alles klar), oder es gibt ein $x \in U$. Damit ist $x \in U_i$ für alle i, und da U_i offen ist, gibt es $K_{\epsilon_i}(x) \subseteq U_i$. Sei $\epsilon := \min_{i=1,\dots,n}(\epsilon_i)$, so folgt $K_\epsilon(x) \subseteq U_i$ für alle i und damit $K_\epsilon(x) \subseteq U$.

ii) klar.

iii) Für $A_i \subseteq X$ folgt $X \setminus (\bigcap_i A_i) = \bigcup_i (X \setminus A_i)$. Nach ii) ist $\bigcap_i A_i$ abgeschlossen für A_i abgeschlossen.

iv) Für $A_i \subseteq X$ folgt $X \setminus (\bigcup_{i=1}^n A_i) = \bigcap_{i=1}^n (X \setminus A_i)$. Nach i) ist $\bigcup_{i=1}^n A_i$ abgeschlossen für A_i abgeschlossen.

v) Sei $\epsilon = d(x,y)$. Dann sind $K_{\frac{\epsilon}{3}}(x)$ und $K_{\frac{\epsilon}{3}}(y)$ disjunkt nach Dreiecksungleichung. $\square$

Beispiel 16.8 In $\mathbb{R}$ versehen mit $d(x,y) = |x-y|$ sind für $a < b$ die Intervalle $]a,b[$ offen und die Intervalle $[a,b]$ abgeschlossen.
Ist $(I_n)_{n\in\mathbb{N}}$ eine Intervallschachtelung in $\mathbb{R}$ mit $x \in I_n$ für alle n, also $\{x\} = \bigcap_{n=0}^{\infty} I_n$, so ist jeder einzelne Punkt $\{x\}$ abgeschlossen nach Satz 16.7.iii). Das Komplement $\mathbb{R} \setminus \{x\}$ ist somit offen.

Auch offene Intervallschachtelungen können genau einen gemeinsamen Punkt haben, z.B. $J_n =]-\frac{1}{n+1}, \frac{1}{n+1}[$ mit $\bigcap_{n=0}^{\infty} J_n = \{0\}$. Deshalb ist der Durchschnitt beliebig vieler offener Mengen im allgemeinen nicht offen. $\triangleleft$

Satz 16.9 *Sei (X,d) ein metrischer Raum und $Y \subseteq X$. Dann gilt:*

i) $Y^\circ := \{y \in Y \ : \ y \notin \partial Y\}$ *ist offen in X.*

ii) $\bar{Y} := Y \cup \partial Y$ *ist abgeschlossen in X.*

iii) ∂Y *ist abgeschlossen in X.*

Man nennt $\bar{Y}$ den *Abschluß* von Y in X und Y° das *Innere* von Y.

Beweis. i) Für $Y^\circ = \varnothing$ ist alles klar. Ansonsten sei $x \in Y^\circ$ beliebig. Da x kein Randpunkt von Y ist, gibt es eine Umgebung U von x mit entweder $U \subseteq Y$ oder $U \subseteq X \setminus Y$. Wegen $x \in U$ und $x \notin X \setminus Y$ ist der zweite Fall ausgeschlossen. Durch Einschränkung kann $U \subseteq Y$ offen gewählt werden. Dann kann U kein $a \in \partial Y$ enthalten, denn sonst wäre U auch Umgebung von a und enthielte damit Punkte aus $X \setminus Y$. Somit ist $U \cap \partial Y = \varnothing$ und $U \subseteq Y^\circ$.

ii) Für $X = \bar{Y} := Y \cup \partial Y$ ist alles klar. Ansonsten gibt es ein $x \in X \setminus \bar{Y}$. Da X offen ist und $x \notin \partial Y$, gibt es eine offene Umgebung U von x, die Y nicht trifft, d.h. $U \subseteq X \setminus Y$. Wie in i) folgt dann $U \cap \partial Y = \varnothing$, also $U \subseteq X \setminus \bar{Y}$. Damit ist $\bar{Y}$ abgeschlossen.

iii) $X \setminus \partial Y = Y^\circ \cup (X \setminus Y)^\circ$ ist offen. $\square$

Definition 16.10 Für eine Teilmenge $Y \subseteq X$ eines metrischen Raumes (X,d) wird der *Durchmesser* definiert als $\operatorname{diam}(Y) := \sup_{x,y\in Y} d(x,y)$. Eine Teilmenge $Y \subseteq X$ heißt *beschränkt*, wenn $\operatorname{diam}(Y) < \infty$.

Beispiel 16.11 Die n-dimensionale Einheitsvollkugel $B^n := \{x \in \mathbb{R}^n \ : \ \|x\|_2 \leq 1\}$ ist abgeschlossen im $\mathbb{R}^n$. Ihr Rand ist die Einheitssphäre $S^{n-1} := \partial B^n = \{x \in \mathbb{R}^n \ : \ \|x\|_2 = 1\}$, sie ist ebenfalls abgeschlossen im $\mathbb{R}^n$. Es ist dann sinnvoll, offene Teilmengen von S^{n-1} zu betrachten. Jeder Durchschnitt der S^{n-1} mit einer offenen Teilmenge des $\mathbb{R}^n$ definiert eine offene Teilmenge von S^{n-1}. Im Sinne dieser Teilmengen ist S^{n-1} dann offen und abgeschlossen zugleich. Es gilt $\operatorname{diam}(B^n) = 2$. $\triangleleft$

17 Konvergenz und Vollständigkeit in metrischen Räumen

Definition 17.1 Sei (X, d) ein metrischer Raum und $(x_k)_{k\in\mathbb{N}}$ eine Folge von Punkten aus X. Die Folge (x_k) heißt *konvergent*, wenn es ein $a \in X$ gibt mit folgenden äquivalenten Eigenschaften:

i) Zu jedem $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$ mit $d(a, x_k) < \epsilon$ für alle $k \geq N$.

ii) Zu jeder Umgebung U von a gibt es ein $N \in \mathbb{N}$ mit $x_k \in U$ für alle $k \geq N$.

iii) Es gilt $\lim_{k\to\infty} d(x_k, a) = 0$.

Im Konvergenzfall heißt $a \in X$ der *Grenzwert* der Folge $(x_k)_{k\in\mathbb{N}}$, und man schreibt $\lim_{k\to\infty} x_k = a$.

Definition 17.2

i) Eine Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus X heißt *Cauchy-Folge*, wenn zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ existiert, so daß $d(x_n, x_m) < \epsilon$ für alle $m, n \geq N$.

ii) Ein metrischer Raum (X, d) heißt *vollständig*, wenn jede Cauchy-Folge in ihm konvergiert.

iii) Ein (bezüglich des Abstands $d(x, y) = \|x - y\|$) vollständiger normierter Vektorraum heißt *Banach-Raum*.

iv) Ein (bezüglich des Abstands $d(x, y) = \sqrt{\langle x - y, x - y\rangle}$) vollständiger euklidischer oder unitärer Vektorraum heißt *Hilbert-Raum*.

Aus der Dreiecksungleichung folgt:

i) Existiert der Grenzwert, so ist er eindeutig bestimmt.

ii) Jede konvergente Folge in einem beliebigen metrischen Raum (X, d) ist Cauchy-Folge. (Die Umkehrung gilt nur in vollständigen metrischen Räumen.)

Nach Satz 15.14 ist in endlich-dimensionalen normierten Vektorräumen der Konvergenzbegriff unabhängig von der Wahl der Norm. Insbesondere konvergiert im $\mathbb{R}^n$ eine Punktfolge $(x_k)_{k\in\mathbb{N}}$ mit $x_k = (x_{k1}, \dots, x_{kn})$ genau dann gegen $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, wenn für jede Koordinate $1 \leq i \leq n$ die Folge x_{ki} gegen a_i konvergiert.

Satz 17.3 *Jeder endlich-dimensionale normierte Vektorraum ist vollständig.*

Beweis. Nach Satz 15.14 genügt es in endlich-dimensionalen Vektorräumen, die Vollständigkeit bezüglich einer Norm nachzuweisen. Nach Wahl einer Basis können wir uns dann auf $\mathbb{K}^n$ beschränken. Sei also $(x_k)_{k\in\mathbb{N}}$ eine Cauchy-Folge bezüglich der $\| \, \|_1$-Norm von Punkten $x_k = (x_{k1}, \dots, x_{kn}) \in \mathbb{K}^n$. Dann gibt es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$, so daß für alle $k, l \geq N$ gilt

$$\|x_k - x_l\|_1 = |x_{k1} - x_{l1}| + \dots + |x_{kn} - x_{ln}| < \epsilon \;.$$

Also ist $(x_{ki})_{k\in\mathbb{N}}$ eine Cauchy-Folge in $\mathbb{K}$ für jedes $i = 1, \ldots, n$. Wegen der Vollständigkeit von $\mathbb{K}$ konvergiert $(x_{ki})_{k\in\mathbb{N}}$ gegen ein $a_i \in \mathbb{K}$. Damit konvergiert (x_k) gegen $a = (a_1, \ldots, a_n)$. □

Satz 17.4 *Die normierten Vektorräume ℓ^1 und ℓ^2 sind vollständig. Insbesondere ist ℓ^2 ein Hilbert-Raum.*

Beweis. Sei $p = 1$ oder $p = 2$. Sei $(x_{(k)})_{k\in\mathbb{N}}$ eine Cauchy-Folge von Zahlenfolgen $x_{(k)} = (x_{(k)n})_{n\in\mathbb{N}} \in \ell^p$, d.h. $\|x_{(k)} - x_{(l)}\|_p < \epsilon$ für alle $k, l \geq N$. Wegen $|y_n| \leq \|(y_n)_{n\in\mathbb{N}}\|_p$ gilt $|x_{(k)n} - x_{(l)n}| \leq \|x_{(k)} - x_{(l)}\|_p < \epsilon$, d.h. für jedes n ist $(x_{(k)n})_{k\in\mathbb{N}}$ eine Cauchy-Folge in $\mathbb{K}$, die somit gegen eine Grenzfolge $x = (x_n)_{n\in\mathbb{N}}$ konvergiert, $x_n := \lim_{k\to\infty} x_{(k)n}$.

Wir zeigen: $x = (x_n) \in \ell^p$ mit $\|x - x_{(k)}\|_p \to 0$. Aus $\|x_{(k)} - x_{(l)}\|_p < \epsilon$ für alle $k, l \geq N$ folgt insbesondere

$$\sum_{n=0}^{K} |x_{(k)n} - x_{(l)n}|^p < \epsilon^p \qquad \text{für alle } K \in \mathbb{N}\ ,\ k, l \geq N\ .$$

Da für festes $n \in \mathbb{N}$ die Folge $(x_{(l)n})_{l\in\mathbb{N}}$ konvergent ist, ergibt sich im Limes $l \to \infty$

$$s_K := \lim_{l\to\infty} \Big(\sum_{n=0}^{K} |x_{(k)n} - x_{(l)n}|^p \Big) = \sum_{n=0}^{K} |x_{(k)n} - x_n|^p \leq \epsilon^p\ .$$

Für festes $k \geq N$ ist die Folge der Partialsummen (s_K) monoton und beschränkt, so daß im Limes $K \to \infty$ folgt $\|x_{(k)} - x\|_p \leq \epsilon$. Schließlich gilt $\|x\|_p \leq \|x - x_{(k)}\|_p + \|x_{(k)}\|_p < \infty$, also $x \in \ell^p$. □

Es wird sich später zeigen, daß (unter einer technischen Annahme, der Separabilität) *jeder* Hilbert-Raum vermöge einer *Orthonormalbasis* mit ℓ^2 identifiziert werden kann.

Mit ähnlichen Techniken zeigt man: $\ell^\infty(M)$ ist Banach-Raum für eine beliebige Menge M.

Die folgende Eigenschaft ist eine sehr nützliche Charakterierung abgeschlossener Mengen:

Satz 17.5 *Sei (X, d) ein metrischer Raum. Eine Teilmenge $Y \subseteq X$ ist genau dann abgeschlossen in X, wenn für jede in X konvergente Folge $(y_k)_{k\in\mathbb{N}}$ mit $y_k \in Y$ gilt $\lim_{k\to\infty} y_k = x \in Y$.*

Beweis. ($\Rightarrow$) Sei $Y \subseteq X$ abgeschlossen und $x = \lim_{k\to\infty} y_k$ mit $y_k \in Y$. Angenommen, $x \notin Y$, also $x \in X \setminus Y$. Da $X \setminus Y$ offen ist, gibt es in $X \setminus Y$ eine ϵ-Umgebung U_ϵ um x. Da x der Grenzwert der Folge (y_k) ist, gibt es ein $n \in \mathbb{N}$, so daß $y_k \in U_\epsilon \subseteq X \setminus Y$ für alle $k \geq n$, Widerspruch.

($\Leftarrow$) Die Grenzwerte aller konvergenten Folgen von Punkten aus Y liegen in Y. Wir zeigen: $X \setminus Y$ ist offen. Angenommen, es gibt einen Punkt $\tilde{x} \in X \setminus Y$, der keine ϵ-Umgebung in $X \setminus Y$ besitzt. Anders formuliert: Jede ϵ-Umgebung von $\tilde{x}$ enthält einen Punkt aus Y. Somit gibt es zu jedem $k \in \mathbb{N}$ ein $y_k \in Y$ mit $d(\tilde{x}, y_k) < \epsilon_k := \frac{1}{k+1}$. Auf diese Weise wird eine Folge $(y_k)_{k\in\mathbb{N}}$ von Punkten aus Y konstruiert, die gegen $\tilde{x}$ konvergiert. Nach Voraussetzung ist $\tilde{x} \in Y$, Widerspruch. □

18 Stetigkeit

Definition 18.1 Es seien (X, d_X) und (Y, d_Y) metrische Räume und $a \in X$. Eine Abbildung $f : X \to Y$ heißt *stetig im Punkt* $a \in X$, wenn es zu jeder Umgebung $V \subseteq Y$ von $f(a) \in V$ eine Umgebung $U \subseteq X$ von $a \in U$ gibt mit $f(U) \subseteq V$. Die Abbildung $f : X \to Y$ heißt *stetig auf* X, wenn f in jedem Punkt $a \in X$ stetig ist.

Stetigkeit bedeutet also, daß eine genügend kleine Umgebung von a unter f nicht auseinandergerissen wird. Insbesondere kann V als ϵ-Umgebung gewählt werden, und die zugehörige Umgebung U von a enthält zumindest eine offene Kugel um a:

Folgerung 18.2 *Die Abbildung $f : X \to Y$ ist genau dann stetig in a, wenn für alle $\epsilon > 0$ ein $\delta > 0$ existiert, so daß*

$$d_Y\big(f(x), f(a)\big) < \epsilon \qquad \textit{für alle } x \in X \textit{ mit } d_X(x, a) < \delta\ .$$

Zu beachten ist, daß δ von ϵ und im allgemeinen auch von a abhängt: wird ϵ verkleinert, so muß im allgemeinen auch δ verkleinert werden, damit die Relationen richtig bleiben.

Definition 18.3 Sei (X, d) ein metrischer Raum und $D \subseteq X$. Dann nennen wir eine Abbildung $f : D \to \mathbb{K}$ eine (komplexwertige für $\mathbb{K} = \mathbb{C}$, reellwertige für $\mathbb{K} = \mathbb{R}$) *Funktion*. Die Teilmenge $D \subseteq X$ heißt *Definitionsbereich*, das Bild $f(D) \subseteq \mathbb{K}$ *Wertebereich*.

Entsprechend heißt eine Funktion $f : D \to \mathbb{K}$ stetig im Punkt $a \in D$, wenn es zu jedem $\epsilon > 0$ ein $\delta > 0$ gibt, so daß gilt:

$$|f(x) - f(a)| < \epsilon \qquad \text{für alle } x \in D \text{ mit } d(x, y) < \delta\ .$$

Die üblichen Rechenregeln für relle oder komplexe Zahlen übertragen sich punktweise auf Funktionen $f, g : X \to \mathbb{K}$ (mit $c \in \mathbb{K}$):

$$\begin{aligned}
&(f+g)(x) := f(x) + g(x)\ , \quad (f \cdot g)(x) := f(x) \cdot g(x)\ , \quad (cf)(x) := c \cdot f(x)\ ,\\
&\tfrac{f}{g}(x) := \tfrac{f(x)}{g(x)} \qquad \text{falls } g(x) \neq 0 \text{ für alle } x \in X\ ,\\
&\bar{f}(x) := \overline{f(x)}\ , \qquad (\operatorname{Re} f)(x) := \operatorname{Re}(f(x))\ , \quad (\operatorname{Im} f)(x) := \operatorname{Im}(f(x))\ .
\end{aligned}$$

Beispiel 18.4 *Die Funktion $f(z) = z^2$ ist stetig auf ganz $\mathbb{C}$.*

Beweis. Sei $z_0 \in \mathbb{C}$ fest gewählt und $\epsilon > 0$ beliebig. Es gilt $|f(z) - f(z_0)| = |z^2 - z_0^2| = |z - z_0|\,|z + z_0| \leq |z - z_0|(|z - z_0| + 2|z_0|)$. Es genügt deshalb, $\delta(\delta + 2|z_0|) \leq \epsilon$ zu wählen, z.B. $\delta := \min(1, \frac{\epsilon}{1+2|z_0|})$. ◁

Beispiel 18.5 *Die Funktion $f(z) = \exp(z)$ ist stetig in ganz $\mathbb{C}$.*

Beweis. Sei $z_0 \in \mathbb{C}$ fest gewählt und $\epsilon > 0$ beliebig. Es gilt

$$|\exp(z) - \exp(z_0)| = |\exp(z_0) \cdot (\exp(z - z_0) - 1)| \leq \exp(|z_0|)|\exp(z - z_0) - 1|\;.$$

Nach der Abschätzung in Satz 14.6.iii) für $N = 0$ gilt für $|z - z_0| < 1$ die Relation $|\exp(z - z_0) - 1| < 2|z - z_0|$. Somit genügt es, $\delta := \min(1, \frac{\epsilon}{2\exp(|z_0|)}$ zu wählen. ◁

Beispiel 18.6 *Für $k \in \mathbb{N}^\times$ ist die Wurzelfunktion $f : \mathbb{R}_+ \to \mathbb{R}_+$, $f(x) := \sqrt[k]{x}$ stetig.*

Beweis. Sei $x_0 > 0$. Für $y \geq 0$ gilt $y^k \leq (1+y)^k - 1$, also $(\sqrt[k]{\frac{x}{x_0}} - 1)^k \leq \frac{x}{x_0} - 1$ und $\sqrt[k]{x} - \sqrt[k]{x_0} \leq \sqrt[k]{x - x_0}$ für $x \geq x_0$ bzw. (Tausch $x_0 \leftrightarrow x$) $|\sqrt[k]{x} - \sqrt[k]{x_0}| \leq \sqrt[k]{|x - x_0|}$ für alle $x, x_0 \in \mathbb{R}_+$. Dann genügt es, $\delta = \epsilon^k$ zu wählen. ◁

Beispiel 18.7 Unter Verwendung der demnächst einzuführenden Funktion sin sei $f : \mathbb{R} \to \mathbb{R}$ definiert durch $f(x) = \begin{cases} 0 & \text{für } x = 0\;, \\ \sin\frac{1}{x} & \text{für } x \neq 0\;. \end{cases}$ Dann gilt $f(U) = [-1, 1]$ für *jede* Umgebung U von 0, so daß f in 0 nicht stetig ist (wähle z.B. $V =]-\frac{1}{2}, \frac{1}{2}[$). ◁

Ein historisch bedeutsames Beispiel ist die nirgends stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f(x) = \begin{cases} 1 & \text{für } x \in \mathbb{Q} \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases}$

Eine wichtige Klasse stetiger Abbildungen ist:

Definition 18.8 Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen heißt *Lipschitz-stetig*, wenn es eine Konstante $L \geq 0$ gibt, so daß für alle $x, x' \in X$ gilt

$$d_Y(f(x), f(x')) \leq L \cdot d_X(x, x')\;.$$

Jede Lipschitz-stetige Abbildung ist stetig: Man wähle $\delta := \frac{\epsilon}{L}$ für $L \neq 0$ und $\delta = 1$ für $L = 0$.

Satz 18.9 *Es sei $(V, \|\;\|)$ ein normierter Vektorraum. Dann ist die Norm, aufgefaßt als Abbildung $\|\;\| : V \to \mathbb{R}$ zwischen metrischen Räumen, Lipschitz-stetig mit Lipschitz-Konstante 1.*

Beweis. Zu zeigen ist $\big|\|x\| - \|y\|\big| = d(\|x\|, \|y\|) \leq d(x, y) = \|x - y\|$. Das folgt aus den Dreiecksungleichungen $\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$ und $\|y\| = \|y - x + x\| \leq \|x - y\| + \|x\|$. □

Beispiel 18.10 Folgende Funktionen sind Lipschitz-stetig:

i) die konstante Funktion $f(x) = c$ (mit $L = 0$),

ii) lineare Funktionen $f(x) = ax + b$ (mit $L = |a|$),

iii) $f(x) = |x|$, $f(x) = \mathrm{Re}(x)$, $f(x) = \mathrm{Im}(x)$, $f(x) = \overline{x}$ jeweils mit $L = 1$. ◁

Satz 18.11 (Folgenkriterium der Stetigkeit) *Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y ist genau dann stetig im Punkt $a \in X$, wenn für jede Folge $(x_k)_{k\in\mathbb{N}}$ von Punkten aus X mit $\lim_{k\to\infty} x_k = a$ gilt*

$$\lim_{k\to\infty} f(x_k) = f(a)\ .$$

Beweis. ($\Rightarrow$) Sei $f : X \to Y$ stetig in a. Dann gibt es zu jedem $\epsilon > 0$ ein $\delta > 0$, so daß $d\big(f(x), f(a)\big) < \epsilon$ für alle $x \in X$ mit $d(x, a) < \delta$. Ist $(x_n)_{n\in\mathbb{N}}$ eine beliebige gegen a konvergente Folge in X, dann gibt es nach Definition des Grenzwertes ein $N \in \mathbb{N}$, so daß $d(x_n, a) < \delta$ für alle $n \geq N$. Also ist $d\big(f(x_n), f(a)\big) < \epsilon$ für alle $n \geq N$ wegen der Stetigkeit von f in a, und $f(a)$ ist Grenzwert der Folge $(f(x_n))_{n\in\mathbb{N}}$ in Y.

($\Leftarrow$) Die Folgenbedingung sei für alle gegen a konvergenten Folgen erfüllt. Angenommen, f wäre nicht stetig in a. Das bedeutet: Es gibt ein $\epsilon > 0$, so daß für alle $\delta > 0$ gilt: Aus $d(x, a) < \delta$ folgt $d\big(f(x), f(a)\big) \geq \epsilon$. Wähle $\delta = \frac{1}{n+1}$ und für jedes $n \in \mathbb{N}$ einen Punkt $x_n \in X$ mit $d(x_n, a) < \frac{1}{n+1}$. Damit gibt es eine gegen a konvergente Folge $(x_n)_{n\in\mathbb{N}}$, für die gilt $d\big(f(x_n), f(a)\big) \geq \epsilon$ für alle $n \in \mathbb{N}$, im Widerspruch zu $\lim_{n\to\infty} f(x_n) = f(a)$. □

Beispiel 18.12 Eine Abbildung $f : \mathbb{R}^2 \to \mathbb{R}$ sei definiert durch

$$f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2} & \text{für } (x, y) \neq (0, 0) \\ 0 & \text{für } (x, y) = (0, 0) \end{cases}$$

Dann ist f nicht stetig in $a = (0, 0)$, denn für die gegen $(0, 0)$ konvergente Folge $z_k = (\frac{1}{k+1}, \frac{1}{k+1})$ gilt $f(z_k) = 1 \neq 0$. Für die Folge $z_k = (\frac{1}{k+1}, 0)$ gilt dagegen $f(z_k) = 0$. Es reicht also nicht, das Folgenkriterium für genügend viele Beispiel-Folgen zu überprüfen! ◁

Satz 18.13 *Sei X metrischer Raum. Eine Abbildung $f : X \to \mathbb{K}^n$ mit $f(x) = (f_1(x), \dots, f_n(x))$ ist genau dann stetig im Punkt $a \in X$, wenn jede Komponentenfunktion $f_i : X \to \mathbb{K}$ stetig in a ist.*

Beweis. Die Folge $(f(x_k))_{k\in\mathbb{N}}$ ist genau dann in $\mathbb{K}^n$ konvergent gegen $f(a)$, wenn jede Komponente $(f_i(x_k))_{k\in\mathbb{N}}$ gegen $f_i(a)$ konvergiert. (Siehe die Bemerkung im Anschluß an Definition 17.2.) □

Die Rechenregeln für Grenzwerte aus Satz 9.3 übertragen sich auf stetige Funktionen:

Satz 18.14 *Die Funktionen $f, g : X \to \mathbb{C}$ seien stetig in $a \in X$. Dann gilt:*

i) *Die Funktionen $f + g$, $f \cdot g$ sind stetig in a.*

ii) *Ist $g(a) \neq 0$, so ist auch $g(x) \neq 0$ in einer Umgebung $U \subseteq X$ von a, und die Funktion $\frac{f}{g} : U \to \mathbb{C}$ ist stetig in $a \in U$.*

Beweis. i) Ist $(x_n)_{n\in\mathbb{N}}$ eine beliebige Folge in X mit $\lim_{n\to\infty} x_n = a$, so gilt $\lim_{n\to\infty} f(x_n) = f(a)$ und $\lim_{n\to\infty} g(x_n) = g(a)$. Satz 9.3 liefert $\lim_{n\to\infty}(f + g)(x_n) = (f+g)(a)$ und $\lim_{n\to\infty}(f\cdot g)(x_n) = (f\cdot g)(a)$. Nach dem Folgenkriterium sind dann $f + g$ und $f \cdot g$ stetig.

ii) Wir zeigen: Ist g stetig in $a \in X$, dann enthält $U := \{x \in X : |g(x)| \geq \frac{1}{2}|g(a)|\} \subseteq X$ eine offene Kugel um a. Für $g(a) = 0$ ist $U = X$, ansonsten gibt es zu $\epsilon = \frac{1}{2}g(a)$ ein $\delta > 0$, so daß für alle $x \in X$ mit $d(x, a) < \delta$ gilt $|g(x)-g(a)| < \epsilon$. Dann gilt (Lemma 3.8)

$$|g(x)| \geq \big|\,|g(a)| - |g(x) - g(a)|\big| \geq \frac{1}{2}|g(a)| \qquad \text{für } d(x, a) < \delta\ .$$

Somit ist $X \cap K_\delta(a) \subseteq U$, und $X \cap K_\delta(a)$ enthält als offene Teilmenge eine offene Kugel. Damit ist U Umgebung von a.

Sei nun $(x_n)_{n\in\mathbb{N}}$ eine gegen a konvergente Folge von Punkten $x_k \in U$, dann konvergiert $(\frac{f}{g}(x_n))_{n\in\mathbb{N}}$ gegen $\frac{f}{g}(a)$. □

Folgerung 18.15 *Die Menge der stetigen Funktionen auf einem metrischen Raum X bildet einen Vektorraum $\mathcal{C}(X)$. Die Menge $\mathcal{C}_b(X) = \{f \in \mathcal{C}(X) : \sup_{x\in X} |f(x)| < \infty\}$ der stetigen beschränkten Funktionen bildet einen Untervektorraum, der mit der durch $\|f\|_\infty := \sup_{x\in X} |f(x)|$ definierten Supremums-Norm zu einem normierten Vektorraum wird.*

Wir werden später zeigen, daß $(\mathcal{C}_b(X), \|\ \|)$ vollständig ist und daß für kompakte metrische Räume K gilt $\mathcal{C}(K) = \mathcal{C}_b(K)$.

Beispiel 18.16 i) Die reellwertige Funktion $f(x) = x^s$ für $x \in \mathbb{R}^*_+$ und $s \in \mathbb{Q}$ ist stetig, aber nicht beschränkt

ii) Jedes Polynom ist auf ganz $\mathbb{C}$ stetig, aber nur die konstanten Polynome sind beschränkt.

iii) Jedes Polynom in mehreren Variablen

$$f(x_1, \dots, x_m) = \sum_{k_1=0}^{n_1} \cdots \sum_{k_m=0}^{n_n} a_{k_1\dots,k_m} x_1^{k_1} \cdots x_m^{k_m}$$

ist stetig auf $\mathbb{C}^m$ bzw. $\mathbb{K}^m$, aber nicht beschränkt.

iv) Jede rationale Funktion ist stetig auf ihrem Definitionsbereich, aber nicht notwendig beschränkt.

v) Die Projektion

$$p : \mathbb{R}^{n+1} \setminus \{0\} \to S^n \,, \qquad p(x) = \frac{x}{\|x\|_2}$$

ist stetig und beschränkt. ◁

Sind $f : X \to Y$ und $g : Y \to Z$ Abbildungen zwischen metrischen Räumen, dann hatten wir die zusammengesetzte Abbildung $g \circ f : X \to \mathbb{C}$ erklärt durch $(g \circ f)(x) := g(f(x))$.

Satz 18.17 *Seien $f : X \to Y$ stetig in a und $g : Y \to Z$ stetig in $f(a)$, dann ist $f \circ g : X \to Z$ stetig in a.*

Beweis. Konvergiert $(x_n)_{n\in\mathbb{N}}$ gegen a, dann konvergiert $(f(x_n))_{n\in\mathbb{N}}$ gegen $f(a)$ und schließlich $(g(f(x_n)))_{n\in\mathbb{N}}$ gegen $g(f(a))$. □

Beispiel 18.18 i) Mit $f : X \to \mathbb{C}$ sind auch $|f|$, $\overline{f}$, $\operatorname{Re} f$ und $\operatorname{Im} f$ stetig als Komposition mit den stetigen Funktionen $g(y) = |y|$ usw.

ii) Für stetige Funktionen $f, g : X \to \mathbb{C}$ sind auch $\max(f, g) = \frac{1}{2}(f + g + |f - g|)$ und $\min(f, g) = \frac{1}{2}(f + g - |f - g|)$ stetig. ◁

Satz 18.19 *Jede Potenzreihe $f(z) = \sum_{n=0}^{\infty} a_n z^n$ ist im Inneren ihres Konvergenzkreises stetig.*

Beweis. Es sei $R(f)$ der Konvergenzradius von $f(z)$ und $|z_0| < r < R(f)$. Wegen der absoluten Konvergenz der Reihe gibt es zu $\epsilon > 0$ ein $N \in \mathbb{N}$ mit $\sum_{n=N}^{\infty} |a_n| r^n < \frac{\epsilon}{3}$. Da das Polynom $\sum_{n=0}^{N-1} a_n z^n$ stetig ist, gibt es ein $r - |z_0| > \delta > 0$, so daß $\Big| \sum_{n=0}^{N-1} a_n z^n - \sum_{n=0}^{N-1} a_n z_0^n \Big| < \frac{\epsilon}{3}$ für alle $z \in \mathbb{C}$ mit $|z - z_0| < \delta$. Dann ist $|z| \leq |z - z_0| + |z_0| < \delta + |z_0| < r$, so daß

$$|f(z) - f(z_0)| \leq \Big| \sum_{n=0}^{N-1} a_n z^n - \sum_{n=0}^{N-1} a_n z_0^n \Big| + \sum_{n=N}^{\infty} |a_n z| + \sum_{n=N}^{\infty} |a_n z_0| < \epsilon \,.$$

□

19 Grenzwerte von Funktionen

Stetigkeit ist eng verbunden mit dem Begriff des Grenzwertes für Funktionen.

Definition 19.1 Es sei (X, d) ein metrischer Raum. Ein Punkt $a \in X$ heißt *Häufungspunkt einer Teilmenge* $D \subseteq X$, wenn jede ϵ-Umgebung $K_\epsilon(a)$ von a unendlich viele Punkte aus D enthält.

Ein Häufungspunkt von D ist nicht notwendigerweise in D enthalten. Für die offene Kreisscheibe $D = K_r(a) := \{z \in \mathbb{C} \ : \ |a - z| < r\}$ sind alle Punkte des abgeschlossenen Kreises $\overline{K_r(a)} := \{z \in \mathbb{C} \ : \ |a - z| \leq r\}$ Häufungspunkte, aber die Randpunkte $|z - a| = r$ gehören nicht zu D.

Definition 19.2 Es sei $a \in X$ ein Häufungspunkt von $D \subseteq X$. Eine Funktion $f : D \to \mathbb{C}$ hat in a den *Grenzwert* λ, wenn die fortgesetzte Funktion $F : D \cup \{a\} \to \mathbb{C}$ mit $F(x) = \begin{cases} f(x) & \text{für } x \neq a \\ \lambda & \text{für } x = a \end{cases}$ in a stetig ist. In diesem Fall schreibt man $\lambda = \lim_{x\to a} f(x)$.

Die Stetigkeit von F in a kann wieder über ϵ-δ-Techniken oder Folgen überprüft werden. Gehört a zum Definitionsbereich von f, dann ist f nach Satz 18.11 in a genau dann stetig, wenn $\lim_{x\to a} f(x) = f(a)$. Damit gibt es folgende Möglichkeiten für Unstetigkeiten einer Funktion f im Punkt a:

i) f hat in a keinen Grenzwert,

ii) f hat in a einen Grenzwert ungleich $f(a)$.

Die Bedeutung des Grenzwertes von Funktionen besteht darin, daß man Funktionen unter Umständen stetig über den Definitionsbereich hinaus *fortsetzen* kann.

Beispiel 19.3 Für $z \in D := \mathbb{C} \setminus \{0\}$ sei $f(z) = \frac{\exp(cz)-1}{z}$ für $c \in \mathbb{C}$. Wir zeigen: $\lim_{z\to 0} f(z) = c$. Auf D gilt

$$\frac{\exp(cz) - 1}{z} = \sum_{n=1}^{\infty} \frac{(cz)^n}{zn!} = c + \sum_{k=1}^{\infty} \frac{c(cz)^k}{(k+1)!} \ .$$

Nach Satz 14.7 gibt es zu $r = 1$ ein $c_1 \in \mathbb{R}$ mit

$$\Big|\frac{\exp(cz) - 1}{z} - c\Big| \leq \Big|\sum_{k=1}^{\infty} \frac{c(cz)^k}{(k+1)!}\Big| < c_1|z| \ .$$

für alle $|z| < 1$. Wähle $\delta = \min(1, \frac{\epsilon}{c_1})$. Damit ist die Funktion $F(z) = \begin{cases} \dfrac{\exp(cz) - 1}{z} & \text{für } z \neq 0 \\ c & \text{für } z = 0 \end{cases}$ stetig auf ganz $\mathbb{C}$.

Man kann übrigens beweisen, daß $f(z) = \exp(cz)$ die einzige Lösung der beiden Bedingungen $f(w)f(z) = f(w+z)$ für alle $z, w \in \mathbb{C}$ und $\lim_{z\to 0} \frac{f(z)-1}{z} = c$ ist (siehe Königsberger, *Analysis 1*, §8.1). ◁

Für des Rechnen mit Grenzwerten gelten die üblichen Regeln (der Beweis ist ähnlich wie in Satz 9.3 und Satz 18.14.

Satz 19.4 *Gilt* $\lim_{x\to a} f(x) = \lambda$ *und* $\lim_{x\to a} g(x) = \mu$, *so folgt* $\lim_{x\to a}(f+g)(x) = \lambda + \mu$, $\lim_{x\to a}(f\cdot g)(x) = \lambda \cdot \mu$ *und, falls* $\mu \neq 0$, $\lim_{x\to a} \frac{f}{g}(x) = \frac{\lambda}{\mu}$.

Ist die Komposition $g \circ f$ *definiert und ist* $\lim_{x\to a} f(x) = y$ *und* g *stetig in* y, *dann folgt* $\lim_{x\to a}(g\circ f)(x) = g(y)$.

Im Reellen kann man zusätzlich einseitige Grenzwerte definieren:

Definition 19.5 Es sei a ein Häufungspunkt von $D \subseteq \mathbb{R}$ und $D_- := D\cap\,]-\infty, a[$ und $D_+ := D\cap\,]a,\infty[$. Eine Funktion $f : D \to \mathbb{C}$ hat in a *linksseitig* bzw. *rechtsseitig* den Grenzwert λ, falls die Einschränkung von f auf D_- bzw. D_+ den Grenzwert λ hat. In diesem Fall schreibt man

$$\lambda = \lim_{x\nearrow a} f(x) = f(a-) \qquad \text{bzw.} \qquad \lambda = \lim_{x\searrow a} f(x) = f(a+)\ .$$

Ist $a \in D$ und $f(a) = f(a-)$ bzw. $f(a) = f(a+)$, dann heißt f in a *linksseitig* bzw. *rechtsseitig* stetig.

Beispiel 19.6 Es sei $f(x) = [x] \in \mathbb{Z}$ der *ganze Teil* einer reellen Zahl x, d.h. $[x] = \sup\{g \in \mathbb{Z} \ :\ g \leq x\}$. Dann hat f in $g \in \mathbb{Z}$ linksseitig den Grenzwert $g-1$ und rechtsseitig den Grenzwert g und ist rechtsseitig stetig. ◁

Über die einseitigen Grenzwerte kann man den Grenzwert einer im Reellen definierten Funktion in $\pm\infty$ definieren:

Definition 19.7 Der Definitionsbereich $D \subseteq \mathbb{R}$ einer Funktion $f : D \to \mathbb{C}$ sei nach oben unbeschränkt. Eine Zahl $\lambda \in \mathbb{C}$ heißt Grenzwert von f in ∞, wenn es zu jedem $\epsilon > 0$ ein N gibt mit $|f(x)-\lambda| < \epsilon$ für alle $x \in D$ mit $x > N$. In diesem Fall schreibt man $\lim_{x\to\infty} f(x) = \lambda$. Analog ist ein Grenzwert in $-\infty$ definiert.

Beispiel 19.8 $\lim\limits_{x\to\infty} (\sqrt{x^2+x} - x) = \frac{1}{2}$ für $x \geq 0$.

Beweis. Für $x > 1$ gilt unter Verwendung der Binomialreihe

$$\sqrt{x^2+x} - x = x\Big(\sqrt{1+\frac{1}{x}} - 1\Big) = x\sum_{n=1}^{\infty}\binom{\frac{1}{2}}{n}\frac{1}{x^n} = \frac{1}{2} + \sum_{k=1}^{\infty}\binom{\frac{1}{2}}{k+1}\frac{1}{x^k}\ .$$

Da der Konvergenzradius gleich 1 ist, gibt es nach Satz 14.7 ein $c_1 \in \mathbb{R}$ mit $\Big|\sum\limits_{k=1}^{\infty}\binom{\frac{1}{2}}{k+1}\frac{1}{x^k}\Big| < c_1\dfrac{1}{x}$ für alle $x > 1$. ◁

Dieses Beispiel demonstriert bereits das allgemeine Prinzip der Zurückführung von Grenzwerten in ∞ auf Grenzwerte in 0:

Lemma 19.9 *Setzt man $g(\xi) = f(\frac{1}{\xi})$ für $\frac{1}{\xi} \in D$, dann gilt: f besitzt genau dann in ∞ einen Grenzwert, wenn g rechtsseitig in 0 einen Grenzwert besitzt, und dann gilt $\lim\limits_{x\to\infty} f(x) = g(0+)$.*

Analog gilt gegebenenfalls $\lim\limits_{x\to-\infty} f(x) = g(0-)$.

Schließlich führen wir $\pm\infty$ als uneigentliche Grenzwerte reellwertiger Funktionen ein:

Definition 19.10 Eine Funktion $f : D \to \mathbb{R}$ hat in $x_0 \in \mathbb{R} \cup \{-\infty, \infty\}$ den *uneigentlichen Grenzwert* ∞ bzw. $-\infty$, wenn es zu jedem $M \in \mathbb{R}$ eine Umgebung $V \subseteq D \setminus \{x_0\}$ gibt, so daß für alle $x \in V$ gilt $f(x) > M$ bzw. $f(x) < M$.

Beispiel 19.11 Für Polynome $f(x) = \sum_{i=0}^n a_i x^i$ mit $n \geq 1$ gilt

$$\lim_{x\to\infty} f(x) = \begin{cases} \infty & \text{falls } a_n > 0 \\ -\infty & \text{falls } a_n < 0 \end{cases}$$

Ist $a_n > 0$, so gilt

$$\lim_{x\to-\infty} f(x) = \begin{cases} \infty & \text{falls } n \text{ gerade} \\ -\infty & \text{falls } n \text{ ungerade} \end{cases}$$

Für $a_n < 0$ tauschen sich die Vorzeichen. ◁

20 Der Zwischenwertsatz

Wir geben zunächst eine äquivalente Charakterisierung der globalen Stetigkeit:

Satz 20.1 *Eine Abbildung $f : X \to Y$ ist genau dann stetig auf ganz X (d.h. in jedem Punkt $a \in X$), wenn das Urbild $f^{-1}(V)$ jeder offenen Menge $V \subseteq Y$ offen in X ist.*

Beweis. ($\Rightarrow$) Ist $f^{-1}(V) = \varnothing$, so ist diese Menge offen. Ansonsten gibt es zu jedem Punkt $a \in f^{-1}(V)$, also $f(a) \in V$, wegen der Offenheit von V eine ϵ-Umgebung $K_\epsilon(f(a)) \subseteq V$. Wegen der Stetigkeit von f existiert eine δ-Umgebung $K_\delta(a) \subseteq X$ mit $f(K_\delta(a)) \subseteq K_\epsilon(f(a)) \subseteq V$. Also ist $K_\delta(a) \subseteq f^{-1}(V) \subseteq X$, d.h. $f^{-1}(V)$ ist offen.

($\Leftarrow$) Sei $a \in X$ beliebig. Eine beliebige offene Umgebung $V \subseteq Y$ von $f(a)$ enthält eine offene Kugel $K_\epsilon(f(a)) \subseteq Y$. Deren Urbild $f^{-1}(K_\epsilon(f(a))$ ist nach Voraussetzung offen, enthält also eine δ-Umgebung von a. Damit ist f stetig. □

Durch Bildung der Komplemente folgt: Eine Abbildung $f : X \to Y$ ist genau dann stetig auf ganz X, wenn das Urbild $f^{-1}(A)$ jeder abgeschlossenen Menge $A \subseteq Y$ abgeschlossen in X ist.

Insbesondere gilt:

Satz 20.2 *Für eine stetige Funktion $f : X \to \mathbb{R}$ gilt:*

i) $U := \{x \in X \,:\, f(x) < c\}$ *ist offen, ebenso* $U' := \{x \in X \,:\, f(x) > c\}$.

ii) $A := \{x \in X \,:\, f(x) \leq c\}$ *ist abgeschlossen, ebenso* $A' := \{x \in X \,:\, f(x) \geq c\}$.

iii) $N := \{x \in X \,:\, f(x) = c\}$ *ist abgeschlossen.* □

Beispiel 20.3 In einem beliebigen normierten Vektorraum $(V, \| \, \|)$ sind der Einheitsball $B_V = \{v \in V \,:\, \|v\| \leq 1\}$ und die Einheitsphäre $S_V = \{v \in V \,:\, \|v\| = 1\}$ abgeschlossen als Urbild von $[0, 1]$ bzw. $\{1\}$ unter der (nach Satz 18.9) stetigen Abbildung $\| \, \| : V \to \mathbb{R}$. ◁

Definition 20.4 Ein metrischer Raum X heißt *zusammenhängend*, wenn X und $\varnothing$ die einzigen Teilmengen von X sind, die sowohl offen als auch abgeschlossen sind.

Damit besitzen nichtzusammenhängende metrische Räume X eine Zerlegung $X = U \cup V$ mit U, V offen und $U, V \neq \varnothing$ und $U \cap V = \varnothing$. Die (Nicht-)Existenz solcher Zerlegungen ist abhängig von der Wahl der Metrik.

In $\mathbb{R}$ führt diese Definition auf Intervalle:

Satz 20.5 *Eine Teilmenge $X \subseteq \mathbb{R}$ mit mindestens zwei verschiedenen Punkten ist genau dann zusammenhängend, wenn X ein Intervall ist.*

Beweis. ($\Rightarrow$) Sei $X = I$ ein Intervall. Angenommen, $I = U \cup V$ mit U, V offen und $U, V \neq \varnothing$ und $U \cap V = \varnothing$. Dann gibt es Punkte $u \in U \subseteq I$ und $v \in V \subseteq I$ mit $u < v$ oder $u > v$. Sei $u < v$, dann ist $[u, v] \subseteq I$. Sei $s := \sup\{[u, v] \cap U\}$. Dann gibt es eine gegen s konvergente Folge von Punkten aus $[u, v] \cap U$. Da $U = I \setminus V$ in I abgeschlossen ist, ist $s \in U$, also ist $]s, v] \subseteq V$. Andererseits ist U offen in I, enthält also auch eine Kugel $K_\epsilon(s)$, Widerspruch.

Sei umgekehrt X kein Intervall. Dann gibt es $u < s < v \in \mathbb{R}$ mit $u, v \in X$ und $s \notin X$. Also sind $U = X \cap]-\infty, s[$ und $V = X \cap]s, \infty[$ offen, disjunkt und nichtleer, außerdem ist $X = U \cup V$. Somit ist X nicht zusammenhängend. □

Theorem 20.6 *Es sei $f : X \to Y$ eine stetige Abbildung zwischen metrischen Räumen X, Y. Ist X zusammenhängend, so ist auch $f(X)$ zusammenhängend.*

Beweis. Wäre $f(X)$ nicht zusammenhängend, so gäbe es disjunkte nichtleere offene Mengen U, V mit $f(X) = U \cup V$. Wegen der Stetigkeit von f sind $f^{-1}(U)$ und $f^{-1}(V)$ offen, nichtleer und disjunkt, denn $f(x)$ liegt entweder in U oder in V. Somit wäre $X = f^{-1}(U) \cup f^{-1}(V)$, Widerspruch. □

Satz 20.7 (Zwischenwertsatz) *Sei X ein zusammenhängender metrischer Raum, $a, b \in X$ und $f : X \to \mathbb{R}$ stetig. Dann nimmt f jeden Wert zwischen $f(a)$ und $f(b)$ an.*

Beweis. Nach Satz 20.6 ist $f(X)$ zusammenhängend. Ist $f(a) \neq f(b)$, dann ist $f(X)$ nach Satz 20.5 ein Intervall. Für $f(a) = f(b)$ ist nichts zu zeigen. □

Für Funktionen auf Intervallen gilt insbesondere:

Folgerung 20.8 (Zwischenwertsatz für Intervalle) *Es seien $a < b$ reelle Zahlen und $f : [a, b] \to \mathbb{R}$ stetig. Dann existiert zu jedem $\gamma \in [f(a), f(b)]$ bzw. $\gamma \in [f(b), f(a)]$ ein $c \in [a, b]$ mit $\gamma = f(c)$.* □

Der Zwischenwertsatz beschreibt die anschauliche Tatsache, daß man stetige Funktionen lückenlos durchzeichnen kann. Das ist eine weitere Version der Vollständigkeit von $\mathbb{R}$. Der Zwischenwertsatz ist nützlich in vielen Existenzbeweisen:

Satz 20.9 (Existenz n-ter Wurzeln) *Jedes reelle Polynom $P(x) = x^n - \alpha$ mit $\alpha > 0$ hat eine positive Nullstelle.*

Beweis. Es ist $P(0) = -\alpha < 0$ und $P(1 + \alpha) = (1 + \alpha)^n - 1 > 0$ (binomische Formel). Da Polynome stetig sind, hat P in $[0, 1 + \alpha]$ mindestens eine Nullstelle y mit $y^n = \alpha$. □

Satz 20.10 *Jedes reelle Polynom $P(x) = \sum_{i=0}^{2n+1} a_i x^i$ ungeraden Grades ($a_{2n+1} \neq 0$) besitzt in $\mathbb{R}$ mindestens eine Nullstelle.*

Beweis. (für $a_{2n+1} > 0$) Wegen $\lim_{x\to\infty} P(x) = \infty$ und $\lim_{x\to-\infty} P(x) = -\infty$ gibt es reelle Zahlen $a < b$ mit $P(a) < 0$ und $P(b) > 0$. Nach dem Zwischenwertsatz gibt es mindestens eine Punkt $c \in [a, b]$ mit $P(c) = 0$. □

Manche Aufgaben, die intuitiv schwer zugänglich sind, finden ihre Lösung im Zwischenwertsatz:

Beispiel 20.11 Zwei Punkte der Erdoberfläche (die Meeresoberfläche eingeschlossen) heißen *antipodal*, wenn Ihre Verbindungsgerade den Erdmittelpunkt enthält. Es gilt: Zu jedem Zeitpunkt gibt es unendlich viele Paare antipodaler Punkte der Erdoberfläche, die die gleiche Lufttemperatur haben. (Die Lufttemperatur wird dabei als stetige Funktion T : Erdoberfläche $\to \mathbb{R}$ aufgefaßt, und die Erdoberfläche als metrischer Raum.)

Beweis: Angenommen, es gibt zumindest zwei antipodale Punkte mit verschiedener Lufttemperatur (sonst ist nichts zu zeigen). Diese seien z.B. Nordpol N und Südpol S. Wir laufen entlang eines beliebigen Längenkreises von N nach S und registrieren die Differenz $f(x) := T(x) - T(x')$, wobei x auf dem Längenkreis liegt und x' sein antipodaler Punkt auf dem gleichen Längenkreis ist. Wegen $N' = S$ und $S' = N$ gilt $f(N) = -f(S) \neq 0$. Somit hat f eine Nullstelle. ▷

Definition 20.12 Sei $X \subseteq \mathbb{R}$. Eine Funktion $f : X \to \mathbb{R}$ heißt

i) *monoton wachsend* bzw. *streng monoton wachsend*, wenn für alle $x, y \in X$ mit $x < y$ gilt $f(x) \leq f(y)$ bzw. $f(x) < f(y)$;

ii) *monoton fallend* bzw. *streng monoton fallend*, wenn für alle $x, y \in X$ mit $x < y$ gilt $f(x) \geq f(y)$ bzw. $f(x) > f(y)$;

iii) *(streng) monoton*, wenn sie (streng) monoton wachsend oder (streng) monoton fallend ist.

Jede streng monotone (reellwertige) Funktion auf $X \subseteq \mathbb{R}$ ist injektiv und besitzt damit eine Umkehrfunktion $g : f(X) \to \mathbb{R}$, die im gleichen Sinn streng monoton ist. In diesem Fall gehen die Graphen von f und g durch Spiegelung an der Diagonalen $y = x$ auseinander hervor:

$$G(f) = \{(x, y) \;:\; y = f(x)\;,\; x \in X\} \quad \Leftrightarrow \quad G(g) = \{(y, x) \;:\; y = f(x)\;,\; x \in X\}\;.$$

Satz 20.13 *Es sei $I \subseteq \mathbb{R}$ ein Intervall und $f : I \to \mathbb{R}$ eine stetige streng monotone Funktion. Dann bildet f das Intervall I bijektiv auf das Intervall $I' := f(I)$ ab, und die Umkehrfunktion $f^{-1} : I' \to I$ ist ebenfalls streng monoton (im gleichen Sinn) und stetig.*

Beweis. Nach Satz 20.6 ist $I' := f(I)$ zusammenhängend, nach Satz 20.5 ein Intervall. Damit bildet f als streng monotone Funktion I bijektiv auf I' ab. Entsprechend wird jedes Teilintervall von I bijektiv auf ein Teilintervall abgebildet, und zwar (wegen der strengen Monotonie) offene Intervalle auf offene und abgeschlossene Intervalle auf abgeschlossene. Insbesondere ist das Urbild in I' unter f^{-1} jeder offenen Umgebung in I wieder offen. Damit ist f^{-1} stetig. □

Die Voraussetzung, daß $I \subseteq \mathbb{R}$ ein Intervall ist, ist wesentlich. Es gibt bijektive stetige Abbildungen zwischen metrischen Räumen, deren Umkehrung nicht stetig ist. Wir werden später sehen, daß $x \mapsto \exp(\mathrm{i}x)$ eine bijektive stetige Abbildung $f : [0, 2\pi[\to S^1$ definiert, deren Umkehrabbildung nicht stetig ist.

Definition 20.14 Eine Abbildung $f : X \to Y$ zwischen metrischen Räumen X, Y heißt *Homöomorphismus*, falls f bijektiv ist und sowohl $f : X \to Y$ als auch $f^{-1} : Y \to X$ stetig ist. Die metrischen Räume X, Y heißen *homöomorph*, falls es einen Homöomorphismus $f : X \to Y$ gibt.

Beispiel 20.15 $X = \mathbb{R}_+$ ist homöomorph zum Intervall $I = [0, 1[$ vermöge der Abbildung $f(x) = \frac{x}{1+x} =: y$. Die Umkehrung ist $f^{-1}(y) = \frac{y}{1-y}$. Als rationale Funktionen sind f, f^{-1} stetig. ◁

21 Die Exponentialfunktion

21.1 Logarithmus und komplexe Potenzen

Die Exponentialfunktion $\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}$ ist nach Beispiel 18.5 in jedem Punkt $z \in \mathbb{C}$ stetig. Insbesondere ist auch die reelle Exponentialfunktion $\exp : \mathbb{R} \to \mathbb{R}$ stetig und nach Bemerkung iv) im Anschluß an Satz 14.6 streng monoton wachsend. Wegen $\exp(0) = 1$ und $\lim_{x\to\infty} \exp(x) = \infty$ nimmt exp auf $\mathbb{R}_+$ jeden Wert $y \geq 1$ genau einmal an. Dann folgt aus $\exp(-x) = \frac{1}{\exp(x)}$, daß exp auf $\mathbb{R}_-$ jeden Wert $0 < y \leq 1$ genau einmal annimmt, d.h. $\exp : \mathbb{R} \to \mathbb{R}_+^\times$ ist bijektiv. Somit existiert die Umkehrfunktion, der (natürliche) *Logarithmus*:

$$\ln : \mathbb{R}_+^\times \to \mathbb{R} \qquad \text{mit} \quad \ln(x) := \{y \in \mathbb{R} \;:\; \exp(y) = x\} \;,$$

und nach Satz 20.13 ist $\ln : \mathbb{R}_+^\times \to \mathbb{R}$ stetig und streng monoton wachsend. Speziell ist $\ln(1) = 0$.

Satz 21.1 *Es gilt die Funktionalgleichung* $\ln(xy) = \ln(x) + \ln(y)$ *für alle* $x, y \in \mathbb{R}_+^\times$ *sowie* $\lim_{x\to 0} \frac{\ln(1+x)}{x} = 1$.

Beweis. i) Setze $\xi := \ln(x)$ und $\eta := \ln y$, dann ist $\exp(\xi + \eta) = \exp(\xi) \cdot \exp(\eta) = x \cdot y$. Einsetzen in ln liefert die Behauptung.

ii) Sei $(x_n)_{n\in\mathbb{N}}$ eine Nullfolge mit $x_n \neq 0$ für alle n und $y_n := \ln(1+x_n)$. Wegen $\ln(1) = 0$ ist auch $(y_n)_{n\in\mathbb{N}}$ eine Nullfolge. Dann gilt $\frac{\ln(1+x_n)}{x_n} = \frac{y_n}{\exp(y_n)-1}$, und nach Beispiel 19.3 konvergiert $(\frac{y_n}{\exp(y_n)-1})_{n\in\mathbb{N}}$ gegen 1. □

Die Funktionalgleichung des Logarithmus führt auf folgende Definition allgemeiner komplexer Potenzen:

$$x^z := \exp(z \ln x) \;, \qquad x \in \mathbb{R}_+^\times \;, \quad z \in \mathbb{C} \;.$$

Insbesondere gilt $e^z = \exp(z)$ für alle $z \in \mathbb{C}$ und $x^0 = 1$ für alle $x \in \mathbb{R}_+^\times$. Die so definierten komplexen Potenzen haben folgende Eigenschaften:

i) Als Komposition stetiger Funktionen ist $x \mapsto x^z$ stetig in jedem Punkt $x \in \mathbb{R}_+^\times$ (dabei ist z festgehalten), und $z \mapsto x^z$ ist stetig in jedem Punkt $z \in \mathbb{C}$ (dabei ist x festgehalten).

ii) Für reelle Exponenten $a \in \mathbb{R}$ ist $x \mapsto x^a$ streng monoton wachsend falls $a > 0$ und streng monoton fallend falls $a < 0$ (für $a = -b$ mit $b > 0$ gilt $x^{-b} = \exp(-b \ln x) = \frac{1}{\exp(b \ln x)} = \frac{1}{x^b}$). Folglich bildet die Funktion $x \mapsto x^a$ den Definitionsbereich $\mathbb{R}_+^\times$ bijektiv auf $\mathbb{R}_+^\times$ ab.

iii) Für $x, y \in \mathbb{R}_+^\times$, $a, b \in \mathbb{R}$ und $z, w \in \mathbb{C}$ gelten die Identitäten

$$(x^a)^b = x^{a \cdot b} , \qquad x^z \cdot y^z = (xy)^z .$$

1) $(x^a)^b := \exp(b \ln(x^a)) = \exp(b \ln(\exp(a \ln x))) = \exp(ba \ln x) = x^{ab}$, dabei ist $x^a \in \mathbb{R}_+^\times$ entscheidend.
2) $x^z \cdot y^z = \exp(z \ln x) \cdot \exp(z \ln y) = \exp(z(\ln x + \ln x)) = \exp(z \ln(xy)) = (xy)^z$.

iv) Es gelten folgende Grenzwerte:

$$\lim_{x\to\infty} x^a = \begin{cases} \infty & \text{für } a > 0 \\ 0 & \text{für } a < 0 \end{cases} \qquad \lim_{x\to 0} x^a = \begin{cases} 0 & \text{für } a > 0 \\ \infty & \text{für } a < 0 \end{cases}$$

$$\lim_{x\to\infty} \frac{\ln x}{x^a} = 0 \quad \text{für } a > 0 , \qquad \lim_{x\to 0} x^a \ln x = 0 \quad \text{für } a > 0 .$$

Die erste Zeile folgt aus der Tatsache, daß $x \mapsto x^a$ streng monoton ist und $\mathbb{R}_+^\times$ bijektiv auf $\mathbb{R}_+^\times$ abbildet. In der ersten Gleichung der zweiten Zeile verwende $\lim_{y\to\infty} \frac{y}{\exp(ay)} = 0$ für $a > 0$ und setze $y = \ln x$.

v) Als Konsequenz aus $\lim_{x\to 0} x^a = 0$ für $a > 0$ kann $x \mapsto x^a$ stetig nach $x = 0$ fortgesetzt werden, so daß $f_a(x) = \begin{cases} x^a & \text{für } x > 0 \\ 0 & \text{für } x = 0 \end{cases}$ eine stetige Funktion $f_a : \mathbb{R}_+ \to \mathbb{R}_+$ ist (für $a > 0$).

Satz 21.2 *Für alle $s \in \mathbb{C}$ und $x \in]-1, 1[$ gilt*

$$(1+x)^s = B_s(x) = \sum_{n=0}^{\infty} \binom{s}{n} x^n , \qquad \ln(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n .$$

Beweis. Für $s \in \mathbb{Q}$ hatten wir $(1+x)^s = B_s(x)$ bereits in Satz 14.5.ii) gezeigt. Für $z \in \mathbb{C}$ mit $|z| < 1$ und $s \in \mathbb{C}^\times$ schreiben wir

$$\frac{B_s(z) - 1}{s} = \sum_{n=1}^{\infty} f_n(s, z) , \qquad f_n(s, z) := \frac{1}{s}\binom{s}{n} z^n = \frac{(s-1)\cdots(s-n+1)}{n!} z^n .$$

Es gilt $\left|\dfrac{f_{n+1}(s,z)}{f_n(s,z)}\right| = \dfrac{|s-n|}{n+1}|z|$, so daß für $|z| < 1$ und $|s| < M$ die Reihe $\sum_{n=1}^{\infty} f_n(s, z)$ nach dem Quotientenkriterium absolut konvergent ist. Sie definiert deshalb für festes s eine stetige Funktion in z, andererseits kann sie für festes z umgeordnet werden in eine Potenzreihe in s, die in jedem Punkt $s \in \mathbb{C}$ mit $|s| < M$, insbesondere in $s = 0$, stetig ist. Wegen $f_n(0, z) = \frac{(-1)^{n-1}}{n} z^n$ gilt somit

$$L(z) := \lim_{s\to 0} \frac{B_s(z) - 1}{s} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n , \qquad |z| < 1 .$$

Zusammen mit dem Additionstheorem $B_{s+t}(z) = B_s(z) \cdot B_t(z)$ aus Satz 14.5.i) folgt aus Beispiel 19.3, daß für festes z gilt $B_s(z) = \exp(sL(z))$. Für $s = 1$ ergibt sich $B_1(z) = (1+z) = \exp(L(z))$ für alle $z \in \mathbb{C}$ mit $|z| < 1$. Für $z = x \in]-1, 1[$ wird durch Logarithmieren $\ln(1+x) = L(x)$ und dann $B_s(x) = \exp(sL(x)) = (1+x)^s$ erhalten. □

Die Logarithmus-Reihe $\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k$ ist zunächst nur für $|x| < 1$ erklärt. Nach dem Leibniz-Kriterium konvergiert sie auch für $x = 1$ (alternierende harmonische Reihe). Da die Logarithmus-Funktion $\ln(1+x)$ für $x > -1$ erklärt ist, ist zu vermuten, daß die alternierende harmonische Reihe gegen $\ln 2$ konvergiert. Nach dem Leibniz-Kriterium gilt für $0 \leq x < 1$

$$\Big| \ln(1+x) - \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} x^k \Big| \leq \frac{x^{n+1}}{n+1} .$$

Alle Funktionen in dieser Ungleichung sind stetig in $x = 1$, so daß wir im Limes $x \to 1$ erhalten

$$\Big| \ln(2) - \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \Big| \leq \frac{1}{n+1} \qquad \Rightarrow \qquad \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = \ln 2 .$$

21.2 Trigonometrische Funktionen

Für die Exponentialfunktion gilt $\overline{\exp(z)} = \exp(\bar{z})$, deshalb für rein imaginäre Zahlen $z = \mathrm{i}x$ mit $x \in \mathbb{R}$ zunächst $\overline{\exp(\mathrm{i}x)} = \exp(-\mathrm{i}x) = \frac{1}{\exp(\mathrm{i}x)}$, also

$$\left| e^{\mathrm{i}x} \right|^2 = \exp(\mathrm{i}x) \exp(-\mathrm{i}x) = 1 .$$

Damit liegt jeder Punkt $e^{\mathrm{i}x}$ auf dem Einheitskreis $S^1 = \{z \in \mathbb{C} \; ; \; |z| = 1\}$. Wir definieren

$$\cos x := \mathrm{Re}(e^{\mathrm{i}x}) = \frac{1}{2}(e^{\mathrm{i}x} + e^{-\mathrm{i}x}) , \qquad \sin x := \mathrm{Im}(e^{\mathrm{i}x}) = \frac{1}{2\mathrm{i}}(e^{\mathrm{i}x} - e^{-\mathrm{i}x}) .$$

Damit gilt für alle $x \in \mathbb{R}$

$$e^{\mathrm{i}x} = \cos x + \mathrm{i} \sin x , \qquad \cos^2 x + \sin^2 x = 1 .$$

Insbesondere sind $\sin, \cos : \mathbb{R} \to [-1, 1]$ stetige Funktionen.

Satz 21.3 *Es gelten die Additionstheoreme*

$$\cos(x+y) = \cos x \cos y - \sin x \sin y , \quad \sin(x+y) = \sin x \cos y + \cos x \sin y ,$$

die Potenzreihendarstellungen

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}\,, \qquad \sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

sowie der Grenzwert $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = 1$.

Beweis. Zum Beweis der Additionstheoreme zerlegt man $e^{\mathrm{i}(x+y)} = e^{\mathrm{i}x} \cdot e^{\mathrm{i}y}$ nach Real- und Imaginärteil. Die Potenzreihen ergeben sich aus der Exponentialfunktion unter Beachtung, daß i^n reell ist für n gerade und rein imaginär für n ungerade. Der Grenzwert $\lim_{x\to 0} \frac{\sin x}{x} = 1$ folgt aus der Reihendarstellung des Sinus und Satz 14.7. □

Wir kommen nun zur Definition der Zahl π:

Satz 21.4 *Der Cosinus hat im Intervall* $[0,2]$ *genau eine Nullstelle, die mit* $\frac{\pi}{2}$ *bezeichnet wird. Es gilt* $\cos\frac{\pi}{2} = 0$ *und* $\sin\frac{\pi}{2} = 1$.

Beweis. Wir zeigen, daß $\cos : [0,2] \to [-1,1]$ streng monoton fallend ist mit $\cos 0 = 1 > 0$ und $\cos 2 < 0$. Der Zwischenwertsatz liefert dann die Existenz einer Nullstelle, aus der Monotonie folgt ihre Eindeutigkeit. Insbesondere is $0 < \pi < 4$.

Die Reihen für cos und sin sind alternierend. Für $x \in \,]0,2]$ bilden die Beträge der Summanden a_k in der Cosinusreihe bzw. der Sinusreihe eine streng monoton fallende Nullfolge ab $k=1$ bzw. $k=0$. Nach der Fehlerabschätzung im Leibniz-Kriterium (Satz 11.10) gilt damit für $x \in \,]0,2]$

$$\underbrace{1 - \frac{x^2}{2}}_{s_1} < \cos x < \underbrace{1 - \frac{x^2}{2} + \frac{x^4}{24}}_{s_3}\,, \qquad \underbrace{x - \frac{x^3}{6}}_{s_1} < \sin x < \underbrace{x}_{s_0}\,,$$

insbesondere ist $\cos 2 < -\frac{1}{3}$ und $\sin x > 0$ für $x \in \,]0,2]$. Die Monotonie des Cosinus folgt aus der Differenzengleichung

$$\cos x - \cos y = \cos(\tfrac{x+y}{2} + \tfrac{x-y}{2}) - \cos(\tfrac{x+y}{2} - \tfrac{x-y}{2}) = -2 \sin\tfrac{x-y}{2} \sin\tfrac{x+y}{2}\,.$$

Ist $0 \le y < x \le 2$, so ist die rechte Seite negativ. □

Somit gelten $e^{\frac{\mathrm{i}\pi}{2}} = \mathrm{i}$ und damit $e^{\mathrm{i}\pi} = -1$, $e^{\frac{3\mathrm{i}\pi}{2}} = -\mathrm{i}$ und $e^{2\pi\mathrm{i}} = 1$ und damit

$$e^{\mathrm{i}(x+\frac{\pi}{2})} = \mathrm{i}e^{\mathrm{i}x}\,, \qquad e^{\mathrm{i}(x+\pi)} = -e^{\mathrm{i}x}\,, \qquad e^{\mathrm{i}(x+\frac{3\pi}{2})} = -\mathrm{i}e^{\mathrm{i}x}\,, \qquad e^{\mathrm{i}(x+2\pi)} = e^{\mathrm{i}x}\,.$$

Zerlegung nach Real- und Imaginärteil liefert die folgenden Periodizitäten für Cosinus und Sinus:

$$\begin{array}{llll}
\cos(x+\frac{\pi}{2}) = -\sin x\,, & \cos(x+\pi) = -\cos x\,, & \cos(x+\frac{3\pi}{2}) = \sin x\,, & \cos(x+2\pi) = \cos x\,, \\
\sin(x+\frac{\pi}{2}) = \cos x\,, & \sin(x+\pi) = -\sin x\,, & \sin(x+\frac{3\pi}{2}) = -\cos x\,, & \sin(x+2\pi) = \sin x\,.
\end{array}$$

Wegen $\cos x = \cos(-x)$ sind $\pm\frac{\pi}{2}$ die einzigen Nullstellen des Cosinus im Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$ Aus $\cos(x+\pi) = -\cos x$ folgt dann, daß $x_k = \frac{(2k+1)\pi}{2}$ mit $k \in \mathbb{Z}$ die einzigen Nullstellen des Cosinus sind, und entsprechend sind $x_k = k\pi$ die einzigen Nullstellen des Sinus.

Satz 21.5 $e^z = 1 \qquad \Leftrightarrow \qquad z = 2\mathrm{i}\pi k$ *für* $k \in \mathbb{Z}$.

Beweis. ($\Leftarrow$) ist klar. Umgekehrt sei $z = x + \mathrm{i}y$ mit $x, y \in \mathbb{R}$, dann gilt $|e^z| = e^x|e^{\mathrm{i}y}| = e^x$, also $x = 0$ und dann $1 = e^{\mathrm{i}y} = \cos y + \mathrm{i}\sin y$. Folglich ist $\cos y = 1$ und $\sin y = 0$, der Sinus liefert zunächst $y = k\pi$ mit $k \in \mathbb{Z}$, wegen $\cos(0+\pi) = -\cos(0) = -1$ und der 2π-Periodizität sind aber nur $y = 2\pi k$ Lösungen. $\square$

Folglich besitzt jede komplexe Zahl $z \neq 0$ die Darstellung $z = re^{\mathrm{i}\phi}$ mit $r = |z| \in \mathbb{R}_+^\times$ und $\phi \in \mathbb{R}$, wobei ϕ nur bis auf Addition eines Vielfachen von 2π bestimmt ist.

21.3 Weitere trigonometrische und hyperbolische Funktionen

Außerhalb der Nullstellen von cos bzw. sin werden Tangens und Cotangens definiert als

$$\tan x := \frac{\sin x}{\cos x}\,, \qquad x \in \mathbb{R} \setminus \{(k+\tfrac{1}{2})\pi\,,\ k \in \mathbb{Z}\}$$
$$\cot x := \frac{\cos x}{\sin x}\,, \qquad x \in \mathbb{R} \setminus \{k\pi\,,\ k \in \mathbb{Z}\}$$

Diese sind stetig und π-periodisch, und als Hauptzweig wählt man das Intervall $]-\frac{\pi}{2}, \frac{\pi}{2}[$ für den Tangens und das Intervall $]0, \pi[$ für den Cotangens.

Satz 21.6 i) cos *ist im Intervall* $[0, \pi]$ *streng monoton fallend und bildet somit* $[0, \pi]$ *bijektiv auf* $[-1, 1]$ *ab. Die damit existierende stetige Umkehrfunktion ist* $\arccos : [-1, 1] \to [0, \pi]$.

ii) sin *ist im Intervall* $[-\frac{\pi}{2}, \frac{\pi}{2}]$ *streng monoton wachsend und bildet somit* $[-\frac{\pi}{2}, \frac{\pi}{2}]$ *bijektiv auf* $[-1, 1]$ *ab. Die damit existierende stetige Umkehrfunktion ist* $\arcsin : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$.

iii) tan *ist im Intervall* $]-\frac{\pi}{2}, \frac{\pi}{2}[$ *streng monoton wachsend und bildet* $]-\frac{\pi}{2}, \frac{\pi}{2}[$ *bijektiv auf* $\mathbb{R}$ *ab. Die damit existierende stetige Umkehrfunktion ist* $\arctan : \mathbb{R} \to]-\frac{\pi}{2}, \frac{\pi}{2}[$.

Beweis. i) cos ist nach dem Beweis von Satz 21.4 streng monoton fallend in $[0, \frac{\pi}{2}]$, und wegen $\cos(\pi - x) = -\cos(-x) = -\cos x$ ist cos auch in $[\frac{\pi}{2}, \pi]$ streng monoton fallend.

ii) folgt aus i) mit $\sin x = \cos(\frac{\pi}{2} - x)$.

iii) Da sin in $[0, \frac{\pi}{2}[$ streng monoton wächst und cos dort streng monoton fällt und beide nichtnegativ sind, ist tan streng monoton wachsend in $[0, \frac{\pi}{2}[$. Wegen

$\tan(-x) = -\tan x$ ist tan streng monoton wachsend in $[-\frac{\pi}{2}, \frac{\pi}{2}[$. Wegen $\sin\frac{\pi}{2} = 1$ und $\cos\frac{\pi}{2} = 0$ ist tan nach oben unbeschränkt in $[0, \frac{\pi}{2}[$, so daß tan das Intervall $]-\frac{\pi}{2}, \frac{\pi}{2}[$ bijektiv auf $\mathbb{R}$ abbildet. $\square$

Satz 21.7 *Es gilt* $\arctan x = \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}$ *für* $x \in\,]-1, 1[$.

Beweis. Zunächst ist $\tan x = \frac{1}{\mathrm{i}} \frac{e^{\mathrm{i}x} - e^{-\mathrm{i}x}}{e^{\mathrm{i}x} + e^{-\mathrm{i}x}} = \frac{1}{\mathrm{i}} \frac{e^{2\mathrm{i}x} - 1}{e^{2\mathrm{i}x} + 1}$. Auflösen nach $e^{2\mathrm{i}x}$ ergibt

$$e^{2\mathrm{i}x} = \frac{1 + \mathrm{i}\tan x}{1 - \mathrm{i}\tan x} \qquad \overset{x=\arctan y}{\Rightarrow} \qquad e^{2\mathrm{i}\arctan y} = \frac{1 + \mathrm{i}y}{1 - \mathrm{i}y} \;.$$

Im Beweis von Satz 21.2 hatten wir $(1 + z) = \exp(L(z))$ für $z \in \mathbb{C}$ mit $|z| < 1$ gezeigt, wobei $L(z) = \displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n$. Somit gilt $\dfrac{1 + \mathrm{i}y}{1 - \mathrm{i}y} = \exp(L(\mathrm{i}y) - L(-\mathrm{i}y))$. Für gerades $n = 2k$ ist $(\mathrm{i}y)^{2k} = (-\mathrm{i}y)^{2k}$, deshalb heben sich in $L(\mathrm{i}y) - L(-\mathrm{i}y)$ die geraden Potenzen von y auf und die ungeraden verdoppeln sich:

$$L(\mathrm{i}y) - L(-\mathrm{i}y) = 2\sum_{n=0}^{\infty} \frac{1}{2n+1} (\mathrm{i}y)^{2n+1} = 2\mathrm{i} \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} y^{2n+1} \;.$$

Aus $\exp(2\mathrm{i}\arctan y) = \exp(L(\mathrm{i}y) - L(-\mathrm{i}y))$ und Satz 21.5 folgt

$$\arctan y + 2k\pi = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} y^{2n+1} \;, \qquad k \in \mathbb{Z}$$

für alle $y \in \mathbb{R}$. Einsetzen von $y = 0$ liefert $k = 0$. $\square$

Interessant ist der Punkt $x = 1$, denn $\tan\frac{\pi}{4} = \frac{\sin(\frac{\pi}{2} - \frac{\pi}{4})}{\cos\frac{\pi}{4}} = \frac{\cos\frac{\pi}{4}}{\cos\frac{\pi}{4}} = 1$, also $\frac{\pi}{4} = \arctan 1$. Die Arcustangensreihe ist für $x = 1$ nicht mehr absolut konvergent, jedoch konvergent nach dem Leibniz-Kriterium. Da $\left|\tan x - \sum_{n=0}^{N} \frac{(-1)^n}{2n+1} x^{2n+1}\right|$ stetig ist und durch den nachfolgenden Term $\frac{|x|^{2N+3}}{2N+3}$ abgeschätzt werden kann, konvergiert die Reihe für $x = 1$ gegen $\arctan 1$, also $\dfrac{\pi}{4} = \displaystyle\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}$. Über das Additionstheorem des Tangens kann $\frac{\pi}{4} = 4\arctan\frac{1}{5} - \arctan\frac{1}{239}$ gezeigt werden, was eine sehr viel schneller konvergierende Reihendarstellung liefert.

Aus der Exponentialfunktion werden auch die folgenden hyperbolischen Funktionen Cosinus hyperbolicus, Sinus hyperbolicus, Tangens hyperbolicus, Cotangens hyperbolicus erhalten:

$$\cosh x := \frac{1}{2}(e^x + e^{-x}) \;, \qquad \sinh x = \frac{1}{2}(e^x - e^{-x}) \;, \qquad \tanh x := \frac{\sinh x}{\cosh x} \;.$$

Alle diese Funktionen sind stetig in $\mathbb{R}$, und $\coth x := \frac{\cosh x}{\sinh x}$ ist stetig in $\mathbb{R}^\times$. Es gelten $\cosh^2 x - \sinh^2 x = 1$, die Additionstheoreme

$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y \;,$$
$$\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y \;,$$

die Potenzreihendarstellungen

$$\cosh x = \sum_{k=0}^\infty \frac{x^{2k}}{(2k)!} \;, \qquad \sinh x = \sum_{k=0}^\infty \frac{x^{2k+1}}{(2k+1)!}$$

sowie der Grenzwert $\lim\limits_{x\to 0} \dfrac{\sinh x}{x} = 1$. Weiter gilt:

i) cosh wächst streng monoton auf $\mathbb{R}_+$ mit Bild $\cosh(\mathbb{R}) = \{y \in \mathbb{R} \;:\; y \geq 1\}$, die Umkehrfunktion ist $\operatorname{arcosh} : \{y \in \mathbb{R} \;:\; y \geq 1\} \to \mathbb{R}_+$

ii) sinh wächst streng monoton auf $\mathbb{R}$ mit Bild $\sinh(\mathbb{R}) = \mathbb{R}$, die Umkehrfunktion ist $\operatorname{arsinh} : \mathbb{R} \to \mathbb{R}$,

iii) tanh wächst streng monoton auf $\mathbb{R}$ mit Bild $\tanh(\mathbb{R}) = \,]-1,1[$, die Umkehrfunktion ist $\operatorname{artanh} : \,]-1,1[\, \to \mathbb{R}$.

Diese Umkehrfunktionen lassen sich durch den Logarithmus darstellen:

$$\operatorname{arcosh}(x) = \ln(x + \sqrt{x^2-1}) \;, \qquad \operatorname{arsinh}(x) = \ln(x + \sqrt{x^2+1}) \;,$$
$$\operatorname{artanh}(x) = \frac{1}{2} \ln \frac{1+x}{1-x} \;.$$

22 Kompakte Mengen

Definition 22.1 Sei A Teilmenge eines metrischen Raumes X. Unter einer *offenen Überdeckung* von A versteht man eine Familie $(U_i)_{i\in I}$ von offenen Teilmengen $U_i \subseteq X$ mit $A \subseteq \bigcup_{i\in I} U_i$, d.h. zu jedem Punkt $x \in A$ gibt es ein $i \in I$ mit $x \in U_i$.

i) Eine Teilmenge A eines metrischen Raumes X heißt *überdeckungskompakt* oder kurz *kompakt*, wenn es zu jeder offenen Überdeckung $(U_i)_{i\in I}$ von A endlich viele Indizes $i_1, \dots, i_n \in I$ gibt, so daß $A \subseteq U_{i_1} \cup \dots \cup U_{i_n}$. (Heine-Borelsche Überdeckungseigenschaft)

ii) Eine Teilmenge K eines metrischen Raums X heißt *folgenkompakt*, wenn jede Folge von Punkten aus K eine konvergente Teilfolge besitzt, deren Grenzwert in K liegt. (Bolzano-Weierstraß-Eigenschaft)

Es wird in i) also nicht gefordert, daß man A durch endlich viele offene Teilmengen von X überdecken kann. Das geht immer, denn A läßt sich durch X selbst überdecken, und X ist offen. Die Forderung ist, daß man jede exotische unendliche Überdeckung von A auf eine endliche Überdeckung reduzieren kann. Insbesondere wird durch Definition 22.1 für $A = X$ die Kompaktheit und Folgenkompaktheit metrischer Räume erklärt.

Satz 22.2 *Es sei X ein metrischer Raum.*

i) *Jede kompakte Teilmenge $A \subseteq X$ ist auch folgenkompakt.* *Somit gilt der Satz von Bolzano-Weierstraß: Zu jeder Folge $\{y_k\}_{k\in\mathbb{N}}$ von Punkten einer kompakten Teilmenge $A \subseteq X$ gibt es ein $a \in A$ sowie eine gegen a konvergente Teilfolge $\{y_{k_l}\}_{l\in\mathbb{N}}$.*

ii) *Jede folgenkompakte Teilmenge $K \subseteq X$ ist beschränkt und abgeschlossen.*

Beweis. i) Sei $(a_k)_{k\in\mathbb{N}}$ eine Folge von Punkten aus A und $K = \{a_k \;:\; k \in \mathbb{N}\}$. Ist K endlich, so hat (a_k) eine konstante Teilfolge. Sei K also unendlich. Angenommen, K hat keinen Häufungspunkt in A. Dann besitzt jeder Punkt $x \in A$ eine offene Umgebung $U(x) \subseteq X$, die nur endlich viele Punkte aus K enthält. Die offenen Umgebungen $U(x)$ bilden eine offene Überdeckung von A. Da A kompakt ist, genügen bereits endlich viele $U(x_1), \dots, U(x_n)$ zur Überdeckung, und A enthielte nur endlich viele Punkte aus K, Widerspruch. Sei $a \in A$ Häufungspunkt von K. Dann enthält jede offene Kugel um a mit Radius $\frac{1}{l+1}$ unendlich viele Punkte aus K. Setze $k_l := \min(k \;:\; d(a_k, a) < \frac{1}{l+1})$. Dann ist $(a_{k_l})_{l\in\mathbb{N}}$ eine Teilfolge von (a_k), die gegen a konvergiert.

ii) Angenommen, K wäre unbeschränkt. Dann gäbe es zu beliebigem $y \in K$ eine Folge $(x_k)_{k\in\mathbb{N}}$ mit $d(x_k, y) > k$, welche keine konvergente Teilfolge besitzt, Widerspruch. Wäre K nicht abgeschlossen, dann gäbe es eine konvergente Folge von Punkten aus K mit Grenzwert außerhalb K, Widerspruch. □

Es läßt sich (mit größerem Schreibaufwand) zeigen, daß in beliebigen metrischen Räumen kompakt und folgenkompakt äquivalente Eigenschaften sind. In allgemeinen topologischen Räumen fallen beide Arten der Kompaktheit jedoch auseinander. Wir beschränken uns hier auf den Beweis, daß in endlich-dimensionalen normierten Vektorräumen die Umkehrung gilt, indem wir zeigen:

Satz 22.3 *Für eine Teilmenge K eines endlich-dimensionalen normierten Vektorraums V sind folgende Aussagen äquivalent:*

i) *K ist beschränkt und abgeschlossen.*

ii) *K ist kompakt (Heine-Borel).*

iii) *K ist folgenkompakt (Bolzano-Weierstraß).*

Wegen Satz 22.2 ist nur i)⇒ii) zu zeigen. Wir formulieren zunächst einen Zwischenschritt als eigenen Satz:

Satz 22.4 *Sei X ein metrischer Raum, $Y \subseteq X$ eine kompakte Teilmenge und $A \subseteq Y$ eine abgeschlossene Teilmenge. Dann ist A kompakt.*

Beweis. $X \setminus A$ ist offen. Ist $(U_i)_{i\in I}$ eine offene Überdeckung von A, dann ist $Y \subseteq X = (X \setminus A) \cup \bigcup_{i\in I} U_i$. Da Y kompakt, genügen endlich viele U_i mit $X \setminus A$ zur Überdeckung von Y und damit auch von A. □

Beweis von Satz 22.2. Sei zunächst $V = \mathbb{R}^n$. Wegen Satz 22.4 genügt es zu zeigen, daß der abgeschlossene Quader

$$Q = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \ : \ a_\nu \leq x_\nu \leq b_\nu\} \subseteq \mathbb{R}^n$$

kompakt ist, denn jede beschränkte abgeschlossene Teilmenge von $\mathbb{R}^n$ liegt in einem abgeschlossenen Quader.

Sei $(U_i)_{i\in I}$ eine unendliche offene Überdeckung von $Q_0 = Q$, die nicht auf eine endliche reduziert werden kann. Durch Halbierung aller Kanten zerlegen wir Q in 2^n gleich große Teilquader der halben Größe. Es gibt dann mindestens einen abgeschlossenen Teilquader, den wir mit Q_1 bezeichnen, der nicht durch endlich viele U_i überdeckt werden kann. Durch Wiederholung des Verfahrens finden wir eine Folge

$$Q = Q_0 \supset Q_1 \supset Q_2 \supset \ldots$$

von abgeschlossenen Quadern mit $\operatorname{diam}(Q_k) = \frac{1}{2^k} \operatorname{diam}(Q)$, so daß jeder von ihnen nicht durch endlich viele U_i überdeckt werden kann.

Nach dem Intervallschachtelungsprinzip gibt es einen Punkt $x \in Q_k$ für alle k. Dieser Punkt x liegt in irgendeiner Umgebung U_j mit $j \in I$. Da U_j offen, gibt es ein $\epsilon > 0$, so daß $K_\epsilon(x) \subseteq U_i$. Dann finden wir aber auch ein $p \in \mathbb{N}$ mit $\operatorname{diam}(Q_p) = \frac{1}{2^p} \operatorname{diam}(Q) < \epsilon$. Somit gilt $Q_p \subseteq U_j$, d.h. die Quader lassen sich im Widerspruch zur Annahme durch endlich viele U_i überdecken. Also ist Q kompakt.

Sei nun V beliebiger n-dimensionaler reeller normierter Vektorraum[3]. Wähle eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von Vektoren $v_j \in V$ und setze

$$K' = \{(\lambda_1, \ldots, \lambda_n) \ : \ \sum_{j=1}^{n} \lambda_j v_j \in K\} \subseteq \mathbb{R}^n \,.$$

Ist K beschränkt und abgeschlossen, so auch K', und K' ist kompakt. Sei $(U_i)_{i\in I}$ eine Überdeckung von $K \subseteq V$. Jede Teilmenge $U_i \subseteq V$ definiert eine Teilmenge

$$U_i' = \{(\lambda_{i1}, \ldots, \lambda_{in}) \ : \ \sum_{j=1}^{n} \lambda_{ij} v_j \in U_i\} \subseteq \mathbb{R}^n \,.$$

Dann ist $(U_i')_{i\in I}$ eine Überdeckung von K', und es gibt endlich viele Indizes $i_1, \ldots, i_p$ mit $K' \subseteq U_{i_1}' \cup \cdots \cup U_{i_p}'$. Dann ist auch $K \subseteq U_{i_1} \cup \cdots \cup U_{i_p}$, d.h. K ist kompakt. □

Daraus ergibt sich folgende Formulierung des Satzes von Bolzano-Weierstraß:

Folgerung 22.5 *Jede beschränkte Folge in einem endlich-dimensionalen normierten Vektorraum V besitzt eine konvergente Teilfolge.* □

[3]Für komplexe Vektorräume identifiziere $\mathbb{C}^n$ mit $\mathbb{R}^{2n}$.

Satz 22.6 *Es sei* $(V^n, \| \ \|)$ *ein* n*-dimensionaler normierter Vektorraum. Dann gilt: Die Vollkugel* $B^n := \{x \in V^n \ : \ \|x-a\| \leq r\} \subseteq V^n$ *um* $a \in V^n$ *mit Radius* r *und die Sphäre* $S^{n-1} := \{x \in V^n \ : \ \|x-a\| = r\} \subseteq V^n$ *sind kompakt.*

Beweis. Die Norm ist nach Satz 18.9 Lipschitz-stetig mit Lipschitz-Konstante 1, also sind B^n und S^{n-1} abgeschlossen nach Satz 20.2, außerdem beschränkt (klar) und damit kompakt nach Satz 22.3. □

Die Umkehrung von Satz 22.2.ii) gilt nicht in beliebigen metrischen Räumen. Es läßt sich z.B. zeigen, daß die Einheitsvollkugel in einem normierten Vektorraum V genau dann kompakt ist, wenn V endlich-dimensional ist.

Nach diesen Vorbereitungen können wir eine wichtige Eigenschaft stetiger Abbildungen beweisen:

Theorem 22.7 *Sei* $f : X \to Y$ *eine stetige Abbildung zwischen metrischen Räumen* X, Y*. Ist* $A \subseteq X$ *kompakt, dann ist auch* $f(A) \subseteq Y$ *kompakt.*

Beweis. Sei $(U_i)_{i \in \mathbb{N}}$ eine offene Überdeckung von $f(A)$. Aus der Stetigkeit von f folgt, daß $V_i := f^{-1}(U_i)$ offen ist. Dann ist $A \subseteq \bigcup_{i \in I} V_i$, aber tatsächlich genügen endlich viele V_i zur Überdeckung: $A \subseteq V_{i_1} \cup \dots \cup V_{i_n}$, also $f(A) \subseteq U_{i_1} \cup \dots \cup U_{i_n}$. □

Beispiel 22.8 Die Abbildung $[0, 2\pi[\ \ni t \mapsto e^{\mathrm{i}t} \in S^1$ kann kein Hömöomorphismus sein, da ansonsten $f^{-1}(S^1)$ nach Theorem 22.7 kompakt wäre. ◁

Aus Theorem 22.7 ergibt sich der Extremwertsatz, den wir an vielen Stellen brauchen werden:

Satz 22.9 (Extremwertsatz) *Sei* $A \subseteq X$ *eine kompakte Teilmenge eines metrischen Raumes* X *und* $f : X \to \mathbb{R}$ *eine stetige Funktion. Dann ist die Einschränkung* $f|_A$ *auf* A *beschränkt (d.h.* $|f(y)| < \infty$ *für alle* $y \in A$*) und nimmt ihr Supremum und Infimum auf* A *an, d.h. es gibt* $p, q \in A$ *mit*

$$f(p) = \sup\{f(y) \ : \ y \in A\} \qquad \textit{und} \qquad f(q) = \inf\{f(y) \ : \ y \in A\} \ .$$

Anders gesagt: Das Supremum ist sogar das globale Maximum, das Infimum sogar das globale Minimum der stetigen Funktion f *auf* A*.*

Beweis. Nach Satz 22.7 ist $f(A) \subseteq \mathbb{R}$ kompakt und nach Satz 22.2 beschränkt (und abgeschlossen), besitzt also ein Supremum M und Infimum m. Nach Definition von Supremum und Infimum gibt es gegen m bzw. M konvergente Folge von Punkten aus $f(A)$. Wegen der Abgeschlossenheit von $f(A)$ liegen die Grenzwerte in $f(A)$ selbst, d.h. $m, M \in f(A)$. □

Der Extremwertsatz ist ein mächtiges Hilfsmittel in Beweisen. Wir können hier einen Beweis angeben für den

Theorem 22.10 (Fundamentalsatz der Algebra) *Jedes Polynom vom Grad ≥ 1 mit komplexen Koeffizienten besitzt in $\mathbb{C}$ mindestens eine Nullstelle.*

Beweis. Es genügt, das Polynom $P(z) = z^n + a_{n-1}z^{n-1} + \dots a_1 z + a_0$ zu betrachten.

i) Wir zeigen: $|P|$ *nimmt auf* $\mathbb{C}$ *ein Minimum an.* Für $|z| \geq 1$ gilt:

$$P(z) = z^n(1 + r(z)) , \qquad r(z) := \frac{a_{n-1}}{z} + \dots + \frac{a_0}{z^n} ,$$
$$|r(z)| \leq \frac{A}{|z|} \text{ mit } A := |a_{n-1}| + \dots + |a_0| .$$

Damit gilt $|r(z)| \leq \frac{1}{2}$ für $|z| \geq R := \max(1, 2A)$ und weiter $|P(z)| \geq \frac{|z|}{2} \geq A$ für alle $z \in \mathbb{C}$ mit $|z| \geq R$. Auf der kompakten Kreisscheibe $\overline{K_0(R)}$ um 0 mit Radius R nimmt die stetige Funktion $|P(z)|$ ein Minimum an, wegen $|P(0)| = |a_0| \leq A$ ist dieses dann das Minimum in ganz $\mathbb{C}$.

ii) Wir zeigen: *Ist* $P(z_0) \neq 0$*, dann hat* $|P(z)|$ *in* z_0 *kein Minimum.* (Also wird das Minimum $|P(z_0)| = 0$ in einem Punkt $z_0 \in \overline{K_0(R)}$ angenommen.)
Dazu betrachten wir das Polynom $Q(w) := \frac{P(w+z_0)}{P(z_0)}$. Wegen $Q(0) = 1$ gilt $Q(w) = 1 + b_1 w^1 + \dots + b_n w^n$. Da P nicht konstant ist, verschwinden nicht alle b_i. Sei $1 \leq k \leq n$ der kleinste Index mit $b_k \neq 0$. Durch Skalieren[4] $w \mapsto \beta w$ mit $\beta^k = -\frac{1}{b_k}$ erreicht man $Q(\beta w) = 1 - w^k + w^{k+1}Q'(w)$. Hier geht entscheidend ein, daß wir in $\mathbb{C}$ arbeiten; in $\mathbb{R}$ wird $\beta^k = -\frac{1}{b_k}$ nicht immer lösbar sein! Das so konstruierte Polynom $Q'(w)$ ist auf $\overline{K_0(R)}$ beschränkt, $|Q'(w)| \leq c$ mit $c > 0$. Somit gilt $|w^{k+1}Q'(w)| < |w|^k$ für alle $w \in \mathbb{C}$ mit $0 < |w| < \min(1, \frac{1}{c}, R)$. Wählt man $w_0 \in \mathbb{R}$ mit $0 < w_0 < \min(1, \frac{1}{c}, R)$, so ergibt sich

$$|Q(\beta w_0)| \leq 1 - w_0^k + |w_0^{k+1}Q'(w_0)| < 1 ,$$

also $|P(z + \beta w_0)| < |P(z_0)|$. $\square$

Beispiel 22.11 Für reelle Zahlen $a, b, c \in \mathbb{R}$ mit $a, b, ab - c^2 > 0$ sei

$$E := \{(x, y) \in \mathbb{R}^2 \ : \ ax^2 + by^2 + 2cxy = 1\} \qquad \text{(Ellipse)} .$$

Wir zeigen: E ist beschränkt und abgeschlossen, damit kompakt. Denn $a(x + \frac{cy}{a})^2 + \frac{ba-c^2}{a}y^2 = 1$ und $b(y + \frac{cy}{b})^2 + \frac{ba-c^2}{b}x^2 = 1$, somit sind $|y| < \sqrt{\frac{a}{ba-c^2}}$ und $|x| < \sqrt{\frac{b}{ba-c^2}}$ beschränkt. Weiter ist $E = f^{-1}(1)$ abgeschlossen als Urbild der abgeschlossen Menge $\{1\}$ unter der stetigen Abbildung $f : \mathbb{R}^2 \to \mathbb{R}$ mit $f(x, y) = ax^2 + by^2 + 2cxy$.

Folglich besitzt jede stetige Funktion $f : E \to \mathbb{R}$ auf E ein lokales Minimum und ein lokales Maximum. Ein Beispiel ist die Abstandsfunktion $d_{(u,v)}(x, y) =$

[4]Eine komplexe Zahl $b \neq 0$ schreibt sich als $b = |b|(\cos\alpha + \mathrm{i}\sin\alpha)$, dann gilt $\beta^k = b$ für $\beta := \sqrt[k]{|b|}(\cos\frac{\alpha}{k} + \mathrm{i}\sin\frac{\alpha}{k})$.

$\sqrt{(x-u)^2+(y-v)^2}$ eines festen Punkt $(u,v) \in \mathbb{R}^2$ zu $(x,y) \in E$. Somit gilt: Es gibt einen Punkt $(x_0,y_0) \in E$, in dem der Abstand $d(E,(u,v)) := \inf_{(x,y)\in E} \sqrt{(x-u)^2+(y-v)^2}$ von (u,v) zu E minimal wird, $d(E,(u,v)) = d_{(u,v)}(x_0,y_0)$. $\triangleleft$

Definition 22.12 Seien (X,d_X) und (Y,d_Y) metrische Räume. Eine Abbildung $f : X \to Y$ heißt ***gleichmäßig stetig***, wenn zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so daß

$$d_Y\big(f(x_1),f(x_2)\big) < \epsilon \quad \text{für alle } x_1,x_2 \in X \text{ mit } d_X(x_1,x_2) < \delta\ .$$

Der Unterschied zur bisher betrachteten Stetigkeit (Definition 18.2) ist, daß δ nur von ϵ, nicht aber von einem Punkt aus X abhängt. In den Beispielen 18.4 und 18.5 hatten wir jeweils eine z_0-Abhängigkeit von δ erhalten. Die Beispiele 18.10 und 18.6 sind gleichmäßig stetig.

Satz 22.13 *Seien X,Y metrische Räume und sei X kompakt. Dann ist jede stetige Abbildung $f : X \to Y$ auch gleichmäßig stetig.*

Beweis. Sei $\epsilon > 0$ global vorgegeben. Wegen der Stetigkeit von f gibt es zu jedem Punkt $a \in X$ ein $\delta(a) > 0$, so daß $d_Y\big(f(x),f(a)\big) < \frac{\epsilon}{2}$ für alle $x \in K_{2\delta(a)}(a)$. Zunächst ist $X = \bigcup_{a\in X} K_{\delta(a)}(a)$. Da X aber kompakt ist, gibt es endlich viele Punkte $a_1,\dots,a_k$ mit $X = K_{\delta(a_1)}(a_1) \cup \dots \cup K_{\delta(a_k)}(a_k)$. Wir setzen $\delta := \min(\delta(a_1),\dots,\delta(a_k))$.

Seien nun zwei Punkte $x_1,x_2 \in X$ mit $d_X(x_1,x_2) < \delta$ gegeben. Der Punkt x_1 liege in der j-ten Umgebung, d.h. $x_1 \in K_{\delta(a_j)}(a_j)$. Dann ist $d_X(x_1,a_j) < \delta(a_j)$ und $d_X(x_2,a_j) \le d_X(x_2,x_1) + d_X(x_1,a_j) < \delta + \delta(a_j) < 2\delta(a_j)$. Aus der Stetigkeit im Punkt a_j folgt:

$$d_Y\big(f(x_1),f(x_2)\big) \le d_Y\big(f(x_1),f(a_j)\big) + d_Y\big(f(x_2),f(a_j)\big) < \epsilon$$

für alle x_1,x_2 mit $d_X(x_1,x_2) < \delta$. $\square$

Beispiel 22.14 Die Funktion $f(x) = \frac{1}{x}$ ist stetig in jedem Punkt $x \in]0,1]$, aber nicht gleichmäßig stetig: Zu $\epsilon = 1$ und $\delta_n = \frac{1}{n+1}$ gilt für $x_n = \frac{\delta}{2}$ und $x'_n = \delta$ einerseits $|x - x'| < \frac{1}{n+1}$ und andererseits $|f(x_n) - f(x'_n)| = n+1 \ge 1$. $\triangleleft$

Wiederholung

- Skalarprodukt, Ungleichung von Cauchy-Schwarz, Beispiel ℓ^2; Norm
- Metrik, metrische Räume, offene und abgeschlossene Teilmengen, Umgebungen
- Konvergenz von Folgen, Cauchy-Folgen, Vollständigkeit, Banach-Räume, Hilbert-Räume

- Definition 18.1 stetiger Abbildungen, äquivalente Charakterisierungen: ϵ-δ und Folgenkriterium.
- Eigenschaften stetiger Abbildungen:
  - Bild zusammenhängender Mengen ist zusammenhängend: Zwischenwertsatz
  - Urbild offener Mengen ist offen (äquivalent zur Stetigkeit)
  - Urbild abgeschlossener Mengen ist abgeschlossen (äquivalent zur Stetigkeit)
  - Bild kompakter Mengen ist kompakt
  - stetige Abbildungen auf kompakten Mengen sind gleichmäßig stetig
- Grenzwerte von Funktionen
- Exponentialfunktion, Logarithmus, komplexe Potenzen, trigonometrische Funktionen, hyperbolische Funktionen
- endlich-dimensionale normierte Vektorräume:
  - alle Normen sind äquivalent
  - sind vollständig
  - beschränkte und abgeschlossene Teilmengen sind kompakt

Teil V
Differentialrechnung

23 Die Ableitung

Wir behandeln hier die Differentiation von Funktionen, die auf Intervallen $I \subseteq \mathbb{R}$ definiert sind. Die Funktionen dürfen aber komplexwertig sein, z.B. $f(x) = x^s$ für $s \in \mathbb{C}$ und $f(x) = e^{\mathrm{i}x}$, jeweils mit $x \in I \subseteq \mathbb{R}$.

Definition 23.1 Eine Funktion $f : I \to \mathbb{C}$ heißt *differenzierbar* im Punkt $x_0 \in I$, wenn der Grenzwert $f'(x_0) = \lim\limits_{\substack{x \to x_0 \\ x \neq x_0}} \dfrac{f(x) - f(x_0)}{x - x_0}$ existiert, d.h. wenn die Funktion $\phi(x) = \begin{cases} \frac{f(x)-f(x_0)}{x-x_0} & \text{für } x \neq x_0 \\ f'(x_0) & \text{für } x = x_0 \end{cases}$ in einer Umgebung von x_0 stetig ist. Dieser Grenzwert $f'(x_0)$ heißt *Ableitung* oder *Differentialquotient* von f in x_0, manchmal auch bezeichnet mit $(Df)(x_0)$ oder $\frac{df}{dx}(x_0)$. Die Funktion f heißt *differenzierbar in* I, wenn sie in jedem Punkt $x_0 \in I$ differenzierar ist.

Äquivalent dazu ist $f'(x_0) = \lim\limits_{\substack{h \to 0 \\ h \neq 0}} \dfrac{1}{h}(f(x_0 + h) - f(x_0))$. Dabei ist h so zu wählen, daß $x_0 + h \in I$ gilt. Zur Vereinfachung der Schreibweise werde vereinbart, daß $\lim_{h\to 0}$ und $\lim_{x\to x_0}$ stets $h \neq 0$ und $x \neq x_0$ bedeuten. Für reellwertige Funktionen ist $\frac{f(x)-f(x_0)}{x-x_0}$ die Steigung der Sekante des Graphen von f durch die beiden Punkte $(x, f(x))$ und $(x_0, f(x_0))$. Im Grenzübergang $x \to x_0$ geht die Sekante in die Tangente an den Graphen im Punkt x_0 über.

Beispiel 23.2 Es sei $a < 0$ und $b > 0$. Die Funktion $f(x) = |x|$ ist nicht differenzierbar in $x_0 = 0 \in [a, b]$. Denn $\frac{f(x)-f(x_0)}{x-x_0} = \frac{|x|}{x} = \begin{cases} 1 & \text{für } x > 0 \\ -1 & \text{für } x < 0 \end{cases}$
Zu jedem $c \in \mathbb{C}$ und $\epsilon = \frac{1}{2}$ gibt es für alle $\delta > 0$ ein $x \in [a, b]$ mit $|x - 0| < \delta$ und $|\frac{|x|}{x} - c| \geq \frac{1}{2}$. Dagegen wäre $f(x) = |x|$ differenzierbar in $[a, 0]$ und in $[0, b]$ mit Ableitung $f'(x) = -1$ bzw. $f'(x) = 1$. ◁

Satz 23.3 i) *$f(x) = x^n$ für $x \in \mathbb{R}$ und $n \in \mathbb{N}$ $\quad\Rightarrow\quad f'(x) = nx^{n-1}$ (und $f'(x) = 0$ für $n = 0$)*

ii) *$f(x) = e^{cx}$ für $x \in \mathbb{R}$ und $c \in \mathbb{C}$ $\quad\Rightarrow\quad f'(x) = ce^{cx}$ (insbesondere $f(x) = a^x \quad\Rightarrow\quad f'(x) = a^x \ln a$)*

iii) *$f(x) = \ln x$ für $x \in \mathbb{R}_+^\times$ $\quad\Rightarrow\quad f'(x) = \frac{1}{x}$*

Beweis. i) $\frac{\xi^n - x^n}{\xi - x} = \sum_{k=0}^{n-1} x^k \xi^{n-1-k}$. Die rechte Seite konvergiert für $\xi \to x$ gegen nx^{n-1}.

ii) $\frac{e^{c(x+h)} - e^{cx}}{h} = e^{cx} \frac{e^{ch}-1}{h}$. Nach Beispiel 19.3 ist $\lim_{h\to 0} \frac{e^{ch}-1}{h} = c$.

iii) $\frac{\ln(x+h)-\ln x}{h} = \frac{1}{x}\frac{\ln(1+\frac{h}{x})}{\frac{h}{x}}$. Nach Satz 21.1 ist $\lim_{y\to 0}\frac{\ln(1+y)}{y} = 1$. □

Satz 23.4 (Lineare Approximierbarkeit) *Eine Funktion $f : I \to \mathbb{C}$ ist genau dann differenzierbar in $x_0 \in I$, wenn es eine Konstante $c \in \mathbb{C}$ gibt, so daß für die Funktion $\psi : I \to \mathbb{C}$ mit $f(x) = f(x_0)+c(x-x_0)+\psi(x)$ gilt $\lim_{x\to x_0}\frac{\psi(x)}{x-x_0} = 0$. In diesem Fall ist $f'(x_0) = c$.*

Beweis. ($\Rightarrow$) Sei f differenzierbar in x_0 mit $f'(x_0) = c$. Für $x \neq x_0$ gilt $\frac{\psi(x)}{x-x_0} = \frac{f(x)-f(x_0)}{x-x_0} - f'(x_0)$. Der Grenzwert der rechten Seite für $x \to x_0$ ist 0.

($\Leftarrow$) Es gelte $f(x) = f(x_0) + c(x - x_0) + \psi(x)$ mit $\lim_{x\to x_0}\frac{\psi(x)}{x-x_0} = 0$. Dann ist $0 = \lim_{x\to x_0}\big(\frac{f(x)-f(x_0)}{x-x_0} - c\big)$, d.h. der Grenzwert $c = \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$ existiert. □

Für reellwertige Funktionen $f : I \to \mathbb{R}$ ist der Graph der Funktion $g(x) = f(x_0) + f'(x_0)(x - x_0)$ die Tangente an den Graphen von f im Punkt x_0.

Satz 23.5 *Ist $f : I \to \mathbb{C}$ differenzierbar in $x_0 \in I$, dann ist f in x_0 auch stetig.*

Beweis. Nach Satz 23.4 gilt insbesondere $\lim_{x\to x_0}\psi(x) = 0$, also $\lim_{x\to x_0} f(x) = \lim_{x\to x_0}\big(f(x_0) + c(x - x_0) + \psi(x)\big) = f(x_0)$. □

Satz 23.6 *Die Funktionen $f, g : I \to \mathbb{C}$ seien in $x \in I$ differenzierbar. Dann sind auch $f + g$, $f \cdot g$ und für $g(x) \neq 0$ auch $\frac{f}{g}$ differenzierbar in x, und es gilt*

$$\begin{aligned}
(f + g)'(x) &= f'(x) + g'(x) \\
(f \cdot g)'(x) &= f'(x)g(x) + f(x)g'(x) && \textit{(Produktregel bzw. Leibniz-Regel)} \\
(\tfrac{f}{g})'(x) &= \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2} && \textit{(Quotientenregel)}
\end{aligned}$$

Beweis. Man schreibt die Differenzenquotienten wie folgt:

$$\begin{aligned}
\frac{(f + g)(x+h) - (f + g)(x)}{h} &= \frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h}, \\
\frac{(f \cdot g)(x + h) - (f \cdot g)(x)}{h} &= \frac{f(x+h) - f(x)}{h}g(x+h) + f(x)\frac{g(x+h) - g(x)}{h}, \\
\frac{(\frac{f}{g})(x+h) - (\frac{f}{g})(x)}{h} &= \frac{1}{g(x)g(x+h)}\Big(\frac{f(x+h) - f(x)}{h}g(x) - f(x)\frac{g(x+h) - g(x)}{h}\Big).
\end{aligned}$$

Da g nach Satz 23.5 in x stetig ist, ist in der letzten Gleichung nach Satz 18.14.ii) $g(x + h) \neq 0$ für $\delta > h > 0$. Nach Satz 19.4 haben die Grenzwerte für $h \to 0$ die behaupteten Eigenschaften. □

Beispiel 23.7 i) Polynome $f(x) = \sum_{k=0}^{n} a_k x^k$ sind differenzierbar in jedem Punkt $x \in \mathbb{R}$, und es gilt $f'(x) = \sum_{k=1}^{n} k a_k x^{k-1}$.

ii) Rationale Funktionen $f(x) = \frac{\sum_{k=0}^n a_k x^k}{\sum_{l=0}^m b_l x^l}$ sind differenzierbar außerhalb der Nullstellen des Nennerpolynoms, und es gilt $f'(x) = \frac{\sum_{k=0}^n \sum_{l=0}^m (k-l) a_k b_l x^{k+l-1}}{\left(\sum_{l=0}^m b_l x^l\right)^2}$. Insbesondere gilt für die in der Partialbruchzerlegung (Satz 13.10) entstehenden elementaren Funktionen $g(x) = \frac{1}{(x-a)^k}$, daß $g'(x) = \frac{-k}{(x-a)^{k+1}}$.

iii) cos, sin sind als Summen von Exponentialfunktionen differenzierbar in jedem Punkt $x \in \mathbb{R}$, und es gilt $\cos'(x) = -\sin(x)$ und $\sin'(x) = \cos(x)$. (Verwende $\frac{1}{2}(e^{\mathrm{i}x} + e^{-\mathrm{i}x})' = \frac{1}{2}(\mathrm{i}e^{\mathrm{i}x} - \mathrm{i}e^{-\mathrm{i}x})$ und analog für sin).

iv) tan und cot sind als Quotienten der differenzierbaren Funktionen sin und cos differenzierbar in jedem Punkt des Definitionsbereiches, und es gilt $\tan'(x) = \frac{\sin'(x)\cos(x) - \sin(x)\cos'(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}$ und $\cot'(x) = \frac{\cos'(x)\sin(x) - \cos(x)\sin'(x)}{\sin^2(x)} = -\frac{1}{\sin^2(x)}$. ◁

Satz 23.8 (Kettenregel) *Die Funktion $f : I \to J \subseteq \mathbb{R}$ sei differenzierbar in $x_0 \in I$, und die Funktion $g : J \to \mathbb{C}$ sei differenzierbar in $f(x_0) \in J$. Dann ist auch die zusammengesetzte Funktion $g \circ f : I \to \mathbb{C}$ differenzierbar in x_0, und es gilt $(g \circ f)'(x_0) = g'(f(x_0)) \cdot f'(x_0)$.*

Beweis. Sei $y_0 := f(x_0)$ und $\gamma(y) := \begin{cases} \frac{g(y)-g(y_0)}{y-y_0} & \text{für } y \neq y_0 \\ g'(y_0) & \text{für } y = y_0 \end{cases}$. Nach Voraussetzung ist $\gamma : J \to \mathbb{C}$ stetig in y_0, und es gilt $g(y) - g(y_0) = (y - y_0)\gamma(y)$ für alle $y \in J$. Dann erhalten wir

$$\begin{aligned}(g \circ f)'(x_0) &= \lim_{x \to x_0} \frac{g(f(x)) - g(f(x_0))}{x - x_0} = \lim_{x \to x_0} \frac{\gamma(f(x)) \cdot (f(x) - f(x_0))}{x - x_0} \\ &= \lim_{x \to x_0} \gamma(f(x)) \cdot \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\ &= \gamma(f(x_0)) \cdot f'(x_0) = g'(f(x_0) \cdot f'(x_0)\ .\end{aligned}$$

□

Beispiel 23.9 i) Für $f(x) = x^z$ mit $z \in \mathbb{C}$ und $x \in \mathbb{R}_+^\times$ gilt $f'(x) = zx^{z-1}$: Setze $g(y) = e^{zy}$ und $\tilde{g}(x) = \ln x$, dann ist $f(x) = (g \circ \tilde{g})(x) = e^{z \ln x}$ und $f'(x) = ze^{z \ln x} \cdot \frac{1}{x}$.

ii) Für $f(x) = \sqrt{1 + x}$ mit $x > -1$ gilt $f'(x) = \frac{1}{2\sqrt{1+x}}$: Setze $g(y) = y^{\frac{1}{2}}$ mit $g'(y) = \frac{1}{2}y^{\frac{1}{2}-1}$ und $\tilde{g}(x) = 1 + x$. ◁

Satz 23.10 (Differentiation der Umkehrfunktion) *Es sei $g = f^{-1}$ die Umkehrfunktion einer streng monotonen Funktion $f : I \to \mathbb{R}$. Ist f differenzierbar in $y_0 \in I$ mit $f'(y_0) \neq 0$, dann ist g differenzierbar in $x_0 := f(y_0)$, und es gilt $g'(x_0) = \dfrac{1}{f'(y_0)} = \dfrac{1}{f'(g(x_0))}$.*

Beweis. Wie im Beweis von Satz 23.8 ist $\phi(y) := \begin{cases} \frac{f(y)-f(y_0)}{y-y_0} & \text{für } y \neq y_0 \\ f'(y_0) & \text{für } y = y_0 \end{cases}$ eine in y_0 stetige Funktion $\phi : I \to \mathbb{R}$ mit $f(y) - f(y_0) = (y - y_0)\phi(y)$ für alle $y \in I$ und $f'(y_0) = \phi(y_0)$. Wegen der strengen Monotonie von f und $\phi(y_0) \neq 0$ ist $\phi(y) \neq 0$ für alle $y \in I$. Dann folgt mit $y := g(x)$ und $x = f(y)$

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

Für $x \to x_0$ folgt aus der Stetigkeit von g in x_0 und der Stetigkeit von ϕ in $y_0 = g(x_0)$ die Behauptung. $\square$

Beispiel 23.11 i) $\arctan'(x) = \frac{1}{1+x^2}$. Sei $y := \arctan x$, so gilt $\arctan'(x) = \frac{1}{\tan'(y)} = \cos^2 y = \frac{\cos^2 y}{\cos^2 y + \sin^2 y} = \frac{1}{1+\tan^2 y} = \frac{1}{1+x^2}$.

ii) $\arcsin'(x) = \frac{1}{\sqrt{1-x^2}}$ für $x \in \,]-1, 1[$. Mit $y = \arcsin x$ und $y \in \,]-\frac{\pi}{2}, \frac{\pi}{2}[$ gilt $\arcsin'(x) = \frac{1}{\sin'(y)} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-\sin^2 y}} = \frac{1}{\sqrt{1-x^2}}$. $\triangleleft$

Satz 23.12 *Es seien $f_n : I \to \mathbb{C}$ differenzierbare und Lipschitz-stetige Funktionen mit $0 \leq \frac{|f_n(x)-f_n(y)|}{|x-y|} \leq L_n < \infty$ für alle $n \in \mathbb{N}$ und alle $x, y \in I$. Wenn*

i) $\displaystyle\sum_{n=0}^{\infty} f_n(x)$ *konvergent ist in jedem Punkt $x \in I$ und*

ii) $\displaystyle\sum_{n=0}^{\infty} f_n'(x_0)$ *und* $\displaystyle\sum_{n=0}^{\infty} L_n$ *konvergent sind,*

dann ist die Funktion $f = \sum_{n=0}^{\infty} f_n$ in x_0 differenzierbar, und es gilt $f'(x) = \sum_{n=0}^{\infty} f_n'(x)$.

Insbesondere ist jede Potenzreihe $f(x) = \sum_{n=0}^{\infty} a_n x^n$ mit Konvergenzradius $R > 0$ in jedem Punkt $x \in \,]-R, R[$ differenzierbar, und es gilt $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$.

Beweis. Zu $\epsilon > 0$ gibt es ein $N \in \mathbb{N}$, so daß gleichzeitig gilt $\sum_{n=N+1}^{\infty} L_n < \frac{\epsilon}{3}$ und $\big|\sum_{n=N+1}^{\infty} f'(x_0)\big| < \frac{\epsilon}{3}$. Dann gilt für beliebige $x \in I \setminus \{x_0\}$

$$\begin{aligned} \Big|\frac{f(x) - f(x_0)}{x - x_0} - \sum_{n=0}^{\infty} f_n'(x_0)\Big| &= \Big|\sum_{n=0}^{\infty} \Big(\frac{f_n(x) - f_n(x_0)}{x - x_0} - f_n'(x_0)\Big)\Big| \\ &\leq \sum_{n=0}^{N} \Big|\frac{f_n(x) - f_n(x_0)}{x - x_0} - f_n'(x_0)\Big| + \sum_{n=N+1}^{\infty} L_n + \Big|\sum_{n=N+1}^{\infty} f'(x_0)\Big| \ . \end{aligned}$$

Wegen der Differenzierbarkeit der f_n mit $0 \leq n \leq N$ gibt es ein gemeinsames $\delta > 0$, so daß $\sum_{n=0}^{N} \Big|\frac{f_n(x)-f_n(x_0)}{x-x_0} - f_n'(x_0)\Big| < \frac{\epsilon}{3}$ für alle $x \in I \setminus \{x_0\}$ mit $|x - x_0| < \delta$. Somit ist $\Big|\frac{f(x)-f(x_0)}{x-x_0} - \sum_{n=0}^{\infty} f_n'(x_0)\Big| < \epsilon$ für alle $x \in I$ mit $|x - x_0| < \delta$.

Für Potenzreihen ist $f_n = a_n x^n$, und für $x, y \in]-r, r[$ mit $0 < r < R$ gilt $\big|\frac{a_n x^n - a_n y^n}{x-y}\big| = \big|\sum_{k=0}^{n-1} a_n x^k y^{n-1-k}\big| \leq n|a_n|r^{n-1} =: L_n$. Nach dem Wurzelkriterium und $\lim_{n\to\infty} \sqrt[n]{n} = 1$ besitzt die Reihe $\sum_{n=1}^{\infty} n a_n x^{n-1}$ den gleichen Konvergenzradius R, so daß die Reihen $\sum_{n=1}^{\infty} f_n'(x_0) = \sum_{n=1}^{\infty} n a_n x_0^{n-1}$ mit $|x_0| < r$ und $\sum_{n=1}^{\infty} L_n = \sum_{n=1}^{\infty} n a_n r^{n-1}$ konvergieren. Dann folgt die Behauptung aus dem allgemeinen Teil mit $I =]-r, r[$, denn jeder Punkt $-R < x_0 < R$ liegt in I für $r = \frac{|x_0|+R}{2} < R$. □

Beispiel 23.13 Die Funktionen $f(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^s}$ und $g(x) = \sum_{n=1}^{\infty} \frac{\sin(nx)}{n^s}$ sind für alle $s > 2$ differenzierbar in $\mathbb{R}$ mit $f'(x) = -\sum_{n=1}^{\infty} \frac{\sin(nx)}{n^{s-1}}$ und $g'(x) = \sum_{n=1}^{\infty} \frac{\cos(nx)}{n^{s-1}}$.

Zum Beweis der Lipschitz-Bedingung verwende man $\big|\frac{\cos(nx)-\cos(ny)}{x-y}\big| = \frac{2}{|x-y|}\big|\sin\frac{n(x+y)}{2}\big|\big|\sin\frac{n(x-y)}{2}\big| \leq n \sup_{r\in\mathbb{R}_+^\times} \big|\frac{\sin r}{r}\big| \leq n \sup_{r\in]0,\frac{\pi}{2}]} \big|\frac{\sin r}{r}\big| = n$. Der letzte Schritt folgt aus $\lim_{r\to 0} \frac{\sin r}{r} = 1$ und $1 > \frac{\sin r}{r} > 1 - \frac{r^2}{6}$ für $r \in]0, \frac{\pi}{2}]$ nach dem Leibniz-Kriterium. Analog für g. ◁

24 Lokale Extrema, Mittelwertsatz

Extrema von Funktionen lassen sich zunächst für beliebige Definitionsbereiche definieren; für Intervalle liefert dann die Differenzierbarkeit ein wichtiges Kriterium.

Definition 24.1 Sei X metrischer Raum. Eine Funktion $f : X \to \mathbb{R}$ hat in $x_0 \in X$

i) ein *globales Maximum* bzw. *globales Minimum*, wenn $f(x) \leq f(x_0)$ bzw. $f(x) \geq f(x_0)$ für alle $x \in X$,

ii) ein *lokales Maximum* bzw. *lokales Minimum*, wenn es eine Umgebung U von x_0 gibt mit $f(x) \leq f(x_0)$ bzw. $f(x) \geq f(x_0)$ für alle $x \in U \cap X$.

Ein lokales/globales Extremum ist ein lokales/globales Minimum oder Maximum. Gilt in i) und ii) das Gleichheitszeichen nur für den Punkt $x = x_0$, so hat f in x_0 ein *strenges* Extremum.

Satz 24.2 *Eine Funktion $f :]a, b[\to \mathbb{R}$ besitze in $x_0 \in]a, b[$ ein lokales Extremum und sei differenzierbar in x_0. Dann gilt $f'(x_0) = 0$.*

Beweis. (für lokales Maximum) Es gibt eine Umgebung U von x mit $f(x) - f(x_0) \leq 0$ für alle $x \in U$, und U enthält Punkte $x > x_0$ und Punkte $y < x_0$. Somit gilt

$$\frac{f(x) - f(x_0)}{x - x_0} \leq 0 \qquad \text{für alle } x \in U \text{ mit } x > x_0$$
$$\frac{f(y) - f(x_0)}{y - x_0} \geq 0 \qquad \text{für alle } y \in U \text{ mit } y < x_0$$

Da f in x_0 differenzierbar ist, folgt aus beiden Ungleichungen $f'(x_0) = 0$. □

Bemerkungen:

i) Stetige Funktionen $f : [a, b] \to \mathbb{R}$ nehmen ihr globales Maximum und Minimum an. Als Kandidaten dieser Extremalpunkte kommen die Randpunkte a, b und, falls f differenzierbar ist, die Lösungen von $f'(x) = 0$ in Frage. Für ein globales Extremum an Randpunkten a oder b muß $f'(a) = 0$ bzw. $f'(b) = 0$ nicht gelten.

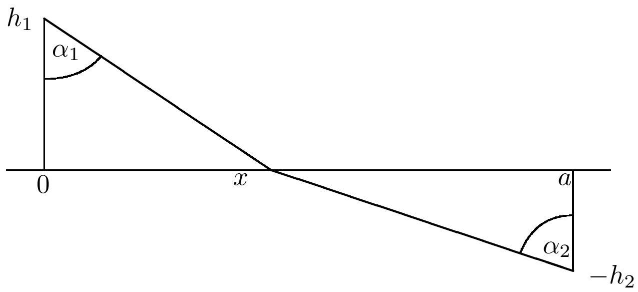
ii) $f'(x) = 0$ ist notwendig, aber nicht hinreichend für das Vorliegen eines lokalen Extremums im Punkt x. Z.B. gilt für $f(x) = x^3$ zwar $f'(0) = 0$, aber f besitzt als streng monoton wachsende Funktion kein lokales Extremum.

Beispiel 24.3 (Fermatsches Prinzip und Brechungsgesetz) Gesucht ist der schnellste Weg zwischen einem Punkt $(0, h_1)$ und einem Punkt $(a, -h_2)$, wenn der Betrag der Geschwindigkeit v_1 ist in $M_1 := \mathbb{R} \times \mathbb{R}^\times_+$ und v_2 in $M_2 := \mathbb{R} \times \mathbb{R}^\times_-$, und die Bewegung in M_1 und M_2 geradlinig verläuft.

Die Zeit für einen Weg durch $(x, 0)$ ist $t(x) = \frac{1}{v_1}\sqrt{h_1^2 + x^2} + \frac{1}{v_2}\sqrt{h_1^2 + (a - x)^2}$. Diese Funktion ist in $]0, a[$ differenzierbar. Liegt ein lokales Extremum in $x \in]0, a[$ vor, so gilt dort

$$0 = t'(x) = \frac{1}{v_1}\frac{x}{\sqrt{h_1^2 + x^2}} - \frac{1}{v_2}\frac{a - x}{\sqrt{h_1^2 + (a - x)^2}} ,$$

also die geometrische Bedingung $\frac{v_1}{v_2} = \frac{\sin \alpha_1}{\sin \alpha_2}$.



◁

Satz 24.4 (Mittelwertsatz der Differentialrechnung) *Die Funktion $f : [a, b] \to \mathbb{R}$ sei stetig auf dem kompakten Intervall $[a, b]$ und differenzierbar auf dem offenen Intervall $]a, b[$. Dann gibt es ein $\xi \in]a, b[$ mit $\dfrac{f(b) - f(a)}{b - a} = f'(\xi)$.*

Ein wichtiger Spezialfall ist der **Satz von Rolle***: Gilt zusätzlich $f(a) = f(b)$, so gibt es ein $\xi \in]a, b[$ mit $f'(\xi) = 0$.*

Beweis. i) für den Satz von Rolle. Ist f konstant auf $[a, b]$, so ist $f'(\xi) = 0$ für alle $\xi \in]a, b[$. Ansonsten nimmt f als stetige Funktion nach dem Extremwertsatz ein Maximum und ein Minimum an, und zumindest eines ist von $f(a) = f(b)$ verschieden. Damit wird dieses Extremum in einem Punkt $\xi \in]a, b[$ angenommen, und wegen der Differenzierbarkeit von f in ξ gilt $f'(\xi) = 0$.

ii) Man betrachte die Funktion $F(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a)$. Diese ist wie f stetig auf $[a, b]$ und differenzierbar auf $]a, b[$, außerdem gilt $F(a) = F(b) = f(a)$. Nach dem Satz von Rolle gibt es $\xi \in]a, b[$ mit $0 = F'(\xi) = f'(\xi) - \frac{f(b)-f(a)}{b-a}$. □

Satz 24.5 (Schrankensatz) *Es sei $f : [a, b] \to \mathbb{C}$ stetig und differenzierbar auf $]a, b[$ mit beschränkter Ableitung $|f'(x)| \leq L$ für alle $x \in]a, b[$. Dann ist f Lipschitz-stetig mit Lipschitz-Konstanten L, d.h. $|f(x) - f(y)| \leq L|x - y|$ für alle $x, y \in [a, b]$.*

Für reellwertige differenzierbare Funktionen $f : [a, b] \to \mathbb{R}$ gilt unter sonst gleichen Voraussetzungen: Ist $m \leq f'(x) \leq M$ für alle $x \in]a, b[$, so gilt $m(x_2 - x_1) \leq f(x_2) - f(x_1) \leq M(x_2 - x_1)$ für alle $x_1, x_2 \in [a, b]$ mit $x_1 \leq x_2$.

Beweis. ii) Die Aussage für reelle Funktionen ist eine Umformulierung des Mittelwertsatzes, insbesondere werden für m, M nur innere Punkte $\xi \in]a, b[$ benötigt.

i) Sei $f(x) \neq f(y)$, insbesondere $x \neq y$, sonst ist nichts zu zeigen. Setze $c := \frac{|f(x)-f(y)|}{f(x)-f(y)} \in \mathbb{C}$ mit $|c| = 1$ und $\phi := \mathrm{Re}(cf)$. Nach dem Mittelwertsatz gibt es zu $x, y \in [a, b]$ ein $\xi \in]x, y[$ mit $\phi(x) - \phi(y) = (x - y)\phi'(\xi)$. Somit gilt

$$\begin{aligned} |f(x) - f(y)| &= cf(x) - cf(y) = \phi(x) - \phi(y) = (x - y)\phi'(\xi) \\ &\leq |x - y| \sup_{\xi \in]x,y[} |\phi'(x)| \leq L|x - y| \end{aligned}$$

wegen $|\mathrm{Re}(cf')(x)| \leq |f'(x)| \leq L$. □

Satz 24.6 *Sei $f : [a, b] \to \mathbb{C}$ stetig und differenzierbar in $]a, b[$ mit $f'(x) = 0$ für alle $x \in]a, b[$. Dann ist f konstant.*

Insbesondere gilt: Zwei differenzierbare Funktionen $f, g : [a, b] \to \mathbb{C}$ mit $f'(x) = g'(x)$ für alle $x \in]a, b[$ unterscheiden sich nur um eine Konstante, $f - g = const.$

Beweis. Nach dem Schrankensatz ist $f(x) = f(y)$ für alle $x, y \in [a, b]$. Der zweite Teil folgt aus dem ersten für die Funktion $f - g$. □

Als Anwendung geben wir eine weitere Charakterisierung der Exponentialfunktion als eindeutige Lösung einer *Differentialgleichung* zu gegebener *Anfangsbedingung*:

Satz 24.7 *Es sei $f : \mathbb{R} \to \mathbb{C}$ eine differenzierbare Funktion, und es gelte $f'(x) = cf(x)$ für alle $x \in \mathbb{R}$ und ein $c \in \mathbb{C}$. Ist $f(0) =: A \in \mathbb{C}$, so gilt $f(x) = A\, e^{cx}$.*

Beweis. Für $F(x) := f(x)e^{-cx}$ gilt nach Produktregel $F'(x) = f'(x)e^{-cx} - cf(x)e^{-cx} = 0$ für alle $x \in \mathbb{R}$. Dann ist F nach Satz 24.6 eine konstante Funktion, also $F(x) = F(0) = f(0) = A$ für alle $x \in \mathbb{R}$. Also gilt $f(x) = A\,e^{cx}$ für alle $x \in \mathbb{R}$. $\square$

Wir können nun eine Lücke im Beweis von Satz 21.2 schließen. Für $x \in \mathbb{R}$ sei $f(x) = B_{sx}(z)$ für $s, z \in \mathbb{C}$ mit $|z| < 1$ und $s \neq 0$. Dann gilt nach Multiplikationssatz (14.5)

$$f'(x) = \lim_{h\to 0} \frac{B_{s(x+h)}(z) - B_{sx}(z)}{h} = sB_{sx}(z) \lim_{h\to 0} \frac{B_{sh}(z) - 1}{sh}$$

In Satz 21.2 war gezeigt: $\lim_{s\to 0} \frac{B_s(z)-1}{s} = L(z)$. Insbesondere folgt auch $\lim_{h\to 0} \frac{B_{sh}(z)-1}{sh} = L(z)$, also $f'(x) = sL(z)f(x)$ und damit $B_{sx}(z) = \exp(sxL(z))$ für alle $x \in \mathbb{R}$.

Satz 24.8 *Für alle $x \in \,]-1,1[$ gilt*

$$\arcsin x = \sum_{k=0}^{\infty} \binom{-\frac{1}{2}}{k} (-1)^k \frac{x^{2k+1}}{2k+1} \;.$$

Beweis: Die Funktionen $f(x) = \arcsin x$ und $g(x) = \binom{-\frac{1}{2}}{k} (-1)^k \frac{x^{2k+1}}{2k+1}$ sind auf $]-1,1[$ differenzierbar, und es gilt

$$f'(x) = \frac{1}{\sqrt{1-x^2}} = (1-x^2)^{-\frac{1}{2}} = \sum_{k=0}^{\infty} \binom{-\frac{1}{2}}{k} (-x^2)^k = g'(x) \;.$$

Damit ist $f(x) - g(x) = f(0) - g(0) = 0$ eine Konstante. $\square$

Satz 24.9 (Verallgemeinerter Mittelwertsatz) *Die Funktionen $f, g : [a,b] \to \mathbb{R}$ seien stetig auf $[a,b]$ und differenzierbar auf $]a,b[$, und es gelte $g'(x) \neq 0$ für alle $x \in \,]a,b[$. Dann ist $g(a) \neq g(b)$, und es gibt ein $\xi \in \,]a,b[$ mit $\dfrac{f(b)-f(a)}{g(b)-g(a)} = \dfrac{f'(\xi)}{g'(\xi)}$.*

Beweis. Wäre $g(b) = g(a)$, so gäbe es nach dem Satz von Rolle ein $\xi \in \,]a,b[$ mit $g'(\xi) = 0$, Widerspruch. Wir können deshalb den Satz von Rolle auf die Funktion $F(x) = f(x) - \dfrac{f(b)-f(a)}{g(b)-g(a)}(g(x) - g(a))$ anwenden. Es gilt $F(a) = F(b) = f(a)$, also gibt es ein $\xi \in \,]a,b[$ mit $0 = F'(\xi) = f'(\xi) - \dfrac{f(b)-f(a)}{g(b)-g(a)} g'(\xi)$. $\square$

Die folgende Rechenregel erlaubt in vielen Fällen eine einfache Berechnung von Grenzwerten:

Satz 24.10 (Regel von de l'Hospital) *Die Funktionen $f, g : \,]a,b[\,\to \mathbb{R}$ seien differenzierbar, und es gelte $g'(x) \neq 0$ für alle $x \in \,]a,b[$. In beiden Fällen*

i) $\lim_{x\searrow a} f(x) = 0$ *und* $\lim_{x\searrow a} g(x) = 0$

ii) $\lim_{x\searrow a} f(x) = \infty$ *und* $\lim_{x\searrow a} g(x) = \infty$

gilt: Existiert der Grenzwert $\displaystyle\lim_{x\searrow a} \frac{f'(x)}{g'(x)}$, *so existiert auch* $\displaystyle\lim_{x\searrow a} \frac{f(x)}{g(x)}$, *und es gilt* $\displaystyle\lim_{x\searrow a} \frac{f(x)}{g(x)} = \lim_{x\searrow a} \frac{f'(x)}{g'(x)}$. *Analog für* $x \nearrow a$ *sowie* $x \to \infty$ *und* $x \to -\infty$.

Beweis. i) Definieren wir $f(a) := 0$ und $g(a) := 0$, so sind die Funktionen f, g stetig in a. Damit sind zu jedem $x \in\,]a, b[$ die Voraussetzungen des verallgemeinerten Mittelwertsatzes für $f, g : [a, x] \to \mathbb{R}$ erfüllt, und es existiert ein $\xi \in\,]a, x[$ mit $\frac{f(x)}{g(x)} = \frac{f'(\xi)}{g'(\xi)}$. Nach Voraussetzung existiert der Limes $x \searrow a$ und hat die behauptete Eigenschaft.

ii) Nach Existenz des rechtsseitigen Grenzwertes $A := \lim_{x\searrow a} \frac{f'(x)}{g'(x)}$ gibt es zu $\epsilon > 0$ ein $\delta > 0$, so daß für alle $x \in\,]a, a+\delta[$ gilt $\left|\frac{f'(x)}{g'(x)} - A\right| < \epsilon$. Nach dem verallgemeinerten Mittelwertsatz gilt dann auch $\left|\frac{f(x)-f(y)}{g(x)-g(y)} - A\right| < \epsilon$ für alle $x, y \in\,]a, a+\delta[$ mit $x \neq y$. Wir betrachten

$$\frac{f(x)}{g(x)} - \frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f(x)-f(y)}{g(x)-g(y)} \left(\frac{1 - \frac{g(y)}{g(x)}}{1 - \frac{f(y)}{f(x)}} - 1 \right) .$$

Wir halten y fest und lassen x gegen a gehen. Wegen $\lim_{x\searrow a} f(x) = \infty$ und $\lim_{x\searrow a} g(x) = \infty$ gibt es ein δ' mit $0 < \delta' < y - a < \delta$, so daß $\left|\frac{1-\frac{g(y)}{g(x)}}{1-\frac{f(y)}{f(x)}} - 1\right| < \frac{\epsilon}{|A|+\epsilon}$ für alle $x \in\,]a, a+\delta'[$. Damit ergibt sich

$$\begin{aligned}\left|\frac{f(x)}{g(x)} - A\right| &\leq \left|\frac{f(x)}{g(x)} - \frac{f(x)-f(y)}{g(x)-g(y)}\right| + \left|\frac{f(x)-f(y)}{g(x)-g(y)} - A\right| \\ &< \left|\frac{f(x)-f(y)}{g(x)-g(y)}\right| \left|\frac{1 - \frac{g(y)}{g(x)}}{1 - \frac{f(y)}{f(x)}} - 1\right| + \epsilon < 2\epsilon\end{aligned}$$

für alle $x \in\,]a, a+\delta'[$. □

Beispiel 24.11 Gesucht sei $\lim_{x\to 2} \frac{\sqrt{x^2+x+3}-3}{\sin(\frac{\pi x}{2})}$. Zähler und Nenner sind differenzierbar in offenen Intervallen $]2-\epsilon, 2[$ und $]2, 2+\epsilon[$ oberhalb und unterhalb des Häufungspunkts 2, und $(\sin(\frac{\pi x}{2}))' = \frac{\pi}{2}\cos(\frac{\pi x}{2}) \neq 0$ für alle $x \in\,]2-\epsilon, 2[$ oder $x \in\,]2, 2+\epsilon[$. Wir betrachten

$$\frac{(\sqrt{x^2+x+3}-3)'}{(\sin(\frac{\pi x}{2}))'} = \frac{\frac{2x+1}{2\sqrt{x^2+x+3}}}{\frac{\pi}{2}\cos(\frac{\pi x}{2})} =: q(x) .$$

Die Funktion $q(x)$ ist stetig in $]2-\epsilon, 2+\epsilon[$ mit $q(2) = \frac{5}{2\sqrt{9}\cdot\frac{\pi}{2}\cdot\cos\pi} = -\frac{5}{3\pi}$, somit $\lim_{x\searrow 2} q(x) = \lim_{x\nearrow 2} q(x) = -\frac{5}{3\pi}$. Insgesamt sind alle Voraussetzungen der Regel von de l'Hospital erfüllt, so daß

$$q(2) = \lim_{x \searrow 2} \frac{\sqrt{x^2+x+3}-3}{\sin(\frac{\pi x}{2})} = \lim_{x \nearrow 2} \frac{\sqrt{x^2+x+3}-3}{\sin(\frac{\pi x}{2})} = \lim_{x \to 2} \frac{\sqrt{x^2+x+3}-3}{\sin(\frac{\pi x}{2})}$$
$$= -\frac{5}{3\pi} \,. \qquad \triangleleft$$

25 Monotonie, höhere Ableitungen, Konvexität

Satz 25.1 (Monotoniekriterium) *Es sei $f : [a,b] \to \mathbb{R}$ stetig und in $]a,b[$ differenzierbar. Dann gilt:*

$$f' > 0 \text{ in }]a,b[\Rightarrow f \text{ wächst in }]a,b[\text{ streng monoton,}$$
$$f' < 0 \text{ in }]a,b[\Rightarrow f \text{ fällt in }]a,b[\text{ streng monoton,}$$
$$f' \geq 0 \text{ in }]a,b[\Leftrightarrow f \text{ wächst in }]a,b[\text{ monoton,}$$
$$f' \leq 0 \text{ in }]a,b[\Leftrightarrow f \text{ fällt in }]a,b[\text{ monoton.}$$

Beweis. ($\Rightarrow$) folgt jeweils aus dem Mittelwertsatz bzw. Schrankensatz. ($\Leftarrow$) folgt aus Definition des Grenzwertes. □

Das Beispiel der streng monoton wachsenden Funktion $f(x) = x^3$ mit $f'(0) = 0$ zeigt, daß in den ersten beiden Implikationen in Satz 25.1 die Umkehrungen ($\Leftarrow$) nicht gelten.

Die Ableitung $f' : I \to \mathbb{C}$ einer auf $I \subseteq \mathbb{R}$ differenzierbaren Funktion $f : I \to \mathbb{C}$ kann erneut auf Differenzierbarkeit in $x_0 \in I$ untersucht werden, und gegebenenfalls heißt die Ableitung von f' in x_0 die *zweite Ableitung* von f in x_0. Man schreibt $f''(x_0) = (f')'(x_0)$ oder auch $(DDf)(x_0)$ oder $\frac{d^2 f}{dx}(x_0)$. Die n-te Ableitung $f^{(n)}$ von f wird dann rekursiv definiert als $f^{(n)}(x_0) := (f^{(n-1)})'(x_0)$, falls $f^{(n-1)}$ in x_0 differenzierbar ist. Dabei ist $f^{(0)}(x_0) := f(x_0)$ und $f^{(1)}(x_0) := f'(x_0)$. Ist die n-te Ableitung $f^{(n)}$ stetig, so heißt f n-mal stetig differenzierbar.

Existiert $f^{(n)}$ für alle $n \in \mathbb{N}$, so heißt f *beliebig oft differenzierbar* (Stetigkeit ist automatisch). Nach Satz 23.12 ist jede Potenzreihe $f(x) = \sum_{n=0}^{\infty} a_n x^n$ beliebig oft differenzierbar im Inneren ihres Konvergenzkreises, und dort gilt $f^{(k)}(x) = \displaystyle\sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k}$. Die Menge der beliebig oft differenzierbaren Funktionen ist "größer" als die Menge der durch Potenzreihen darstellbaren Funktionen. Eine nützliche beliebig oft auf $\mathbb{R}$ differenzierbare Funktion ist die Einschaltfunktion

$$f(x) := \begin{cases} e^{-\frac{1}{x}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Satz 25.2 *Eine in $]a,b[$ differenzierbare Funktion $f :]a,b[\to \mathbb{R}$ sei in $x_0 \in]a,b[$ zweimal differenzierbar, und es gelte $f'(x_0) = 0$ sowie $f''(x_0) \neq 0$. Dann besitzt f*

in x_0 ein strenges lokales Extremum, und zwar ein strenges lokales Minimum für $f''(x_0) > 0$ bzw. ein strenges lokales Maximum für $f''(x_0) < 0$.

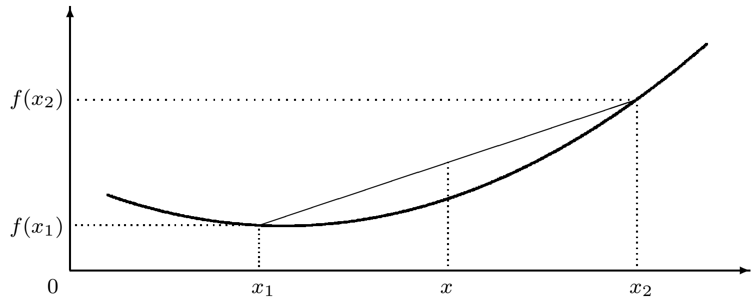
Beweis. (für $f''(x_0) > 0$). Wegen $f''(x_0) = \lim_{x\to x_0} \dfrac{f'(x) - f'(x_0)}{x - x_0} > 0$ gibt es ein $\epsilon > 0$, so daß unter Verwendung von $f'(x_0) = 0$ gilt $\dfrac{f'(x)}{x - x_0} > 0$ für alle $x \in]a, b[$ mit $0 < |x - x_0| < \epsilon$. Nach Satz 25.1 ist f' in $]x_0 - \epsilon, x_0[$ streng monoton fallend und in $]x_0, x_0 + \epsilon[$ streng monoton wachsend, besitzt also in x_0 ein strenges lokales Minimum. □

Definition 25.3 Eine Funktion $f : I \to \mathbb{R}$ heißt *konvex* auf dem Intervall $I \subseteq \mathbb{R}$, wenn für alle $x_1 < x_2 \in I$ und alle $\lambda \in]0, 1[$ gilt

$$f(\lambda x_1 + (1 - \lambda)x_2) \le \lambda f(x_1) + (1 - \lambda)f(x_2) \; .$$

Gilt sogar die Relation '$<$', so heißt f *streng konvex*. Eine Funktion $f : I \to \mathbb{R}$ heißt *(streng) konkav*, wenn $-f$ (streng) konvex ist.

Setzen wir $x = \lambda x_1 + (1 - \lambda)x_2$, dann ist $x_1 < x < x_2$ und $\lambda = \frac{x-x_2}{x_1-x_2}$. Somit ist $x \mapsto \lambda f(x_1) + (1 - \lambda)f(x_2) = f(x_2) + \frac{f(x_1)-f(x_2)}{x_1-x_2}(x - x_2)$ die Sekante durch $(x_1, f(x_1))$ und $(x_2, f(x_2))$. Folglich bedeutet Konvexität, daß $(x, f(x))$ unterhalb der Sekante durch $(x_1, f(x_1))$ und $(x_2, f(x_2))$ liegt, und zwar für beliebige $x_1, x_2 \in I$ mit $x_1 < x < x_2$.



Die Definition von Konvexität erfordert keine Differenzierbarkeit: Z.B: ist $f(x) = |x|$ in jedem Punkt $x \in \mathbb{R}$ konvex. Für differenzierbare Funktionen haben wir

Satz 25.4 (Konvexitätskriterium) *Eine in $[a, b]$ stetige und in $]a, b[$ differenzierbare Funktion f ist genau dann konvex in $[a, b]$, wenn f' in $]a, b[$ monoton wächst.*

Beweis. ($\Rightarrow$) Sei f konvex und $x_1 < x_2$ Punkte aus $]a, b[$. Dann gilt $f(x) \le f(x_2) + \frac{f(x_1)-f(x_2)}{x_1-x_2}(x - x_2) = f(x_1) + \frac{f(x_1)-f(x_2)}{x_1-x_2}(x - x_1)$ für alle $x \in]x_1, x_2[$, also

$$\frac{f(x) - f(x_1)}{x - x_1} \le \frac{f(x_2) - f(x_1)}{x_2 - x_1} \le \frac{f(x_2) - f(x)}{x_2 - x} \; .$$

Für $x \searrow x_1$ einerseits und $x \nearrow x_2$ andererseits folgt aus der Differenzierbarkeit in x_1, x_2 die Relation $f'(x_1) \leq \dfrac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f'(x_2)$. Da $x_1 < x_2$ beliebig sind, wächst f' monoton.

($\Leftarrow$) Sei f' monoton wachsend und $x_1 < x < x_2$ Punkte aus $[a, b]$. Nach dem Mittelwertsatz gibt es Punkte $\xi_1 \in \,]x_1, x[$ und $\xi_2 \in \,]x, x_2[$ mit

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(\xi_1) \qquad \text{und} \qquad \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2) \,.$$

Es ist $\xi_1 < \xi_2$ und damit $f'(\xi_1) \leq f'(\xi_2)$, also $\frac{f(x)-f(x_1)}{x-x_1} \leq \frac{f(x_2)-f(x)}{x_2-x}$ für alle $x \in \,]x_1, x_2[$. Dann ist $x - x_1 = (1-\lambda)(x_2 - x_1)$ und $x_2 - x = \lambda(x_2 - x_1)$ für $\lambda \in \,]0, 1[$, also $\lambda(f(x) - f(x_1)) \leq (1-\lambda)(f(x_2) - f(x))$ und dann $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$. $\square$

Satz 25.5 *Es sei $f : [a, b] \to \mathbb{R}$ stetig und zweimal differenzierbar in $]a, b[$. Dann gilt:*

i) *$f'' \geq 0$ in $]a, b[$ $\quad\Leftrightarrow\quad$ f ist konvex in $[a, b]$.*

ii) *$f'' > 0$ in $]a, b[$ $\quad\Rightarrow\quad$ f ist streng konvex in $[a, b]$.*

Beweis. i) ist Satz 25.1 für f' zusammen mit Satz 25.4.

ii) f ist zumindest konvex. Wäre f nicht streng konvex, so gibt es $x_1 < x < x_2 \in [a, b]$ mit $\frac{f(x)-f(x_1)}{x-x_1} = \frac{f(x_2)-f(x_1)}{x_2-x_1} = \frac{f(x_2)-f(x)}{x_2-x}$. Nach dem Mittelwertsatz gibt es dann Punkte $\xi_1 \in \,]x_1, x[$ und $\xi_2 \in \,]x, x_2[$ mit

$$f'(\xi_1) = \frac{f(x) - f(x_1)}{x - x_1} = \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2) \,,$$

im Widerspruch zur strengen Monotonie von f' nach Satz 25.1 für f'. $\square$

Beispiel 25.6 Die Funktionen $f(x) = x^{2n}$ mit $n \in \mathbb{N}^\times$ sind streng konvex auf $\mathbb{R}$ wegen $f''(x) = 2n(2n-1)x^{2n-2} > 0$. Ebenso ist $f(x) = e^x$ streng konvex auf $\mathbb{R}$ wegen $f''(x) = e^x > 0$. Schließlich ist $f(x) = \ln x$ streng konkav auf $\mathbb{R}^\times_+$, da $(-\ln x)'' = \frac{1}{x^2} > 0$. $\triangleleft$

Satz 25.7 (Höldersche Ungleichung) *Für relle Zahlen $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$ seien Folgen $x = (x_k)_{k\in\mathbb{N}}$ und $y = (y_k)_{k\in\mathbb{N}}$ reeller oder komplexer Zahlen gegeben mit*

$$\|x\|_p := \Big(\sum_{k=0}^{\infty} |x_k|^p\Big)^{\frac{1}{p}} < \infty \,, \qquad \|y\|_q := \Big(\sum_{k=0}^{\infty} |y_k|^q\Big)^{\frac{1}{q}} < \infty \,.$$

Dann konvergiert die Reihe $\sum_{k=0}^{\infty} x_k y_k$, und es gilt die Höldersche Ungleichung

$$\Big|\sum_{k=0}^{\infty} x_k y_k\Big| \leq \sum_{k=0}^{\infty} |x_k y_k| \leq \|x\|_p \, \|y\|_q \,.$$

Speziell gilt für endliche Folgen

$$\Big|\sum_{k=1}^{n} x_k y_k\Big| \le \sum_{k=1}^{n} |x_k y_k| \le \Big(\sum_{k=1}^{n} |x_k|^p\Big)^{\frac{1}{p}} \Big(\sum_{k=1}^{n} |y_k|^q\Big)^{\frac{1}{q}} \qquad \textit{für } \frac{1}{p}+\frac{1}{q}=1\;.$$

Beweis. i) Wir zeigen zunächst: *Für beliebige $a,b \in \mathbb{R}_+^\times$ und p,q wie oben gilt*

$$a^{\frac{1}{p}} \cdot b^{\frac{1}{q}} \le \frac{a}{p} + \frac{b}{q}\;. \tag{*}$$

Die Funktion $f(x) = -\ln x$ ist konvex auf $\mathbb{R}_+^\times$, also gilt mit $\lambda = \frac{1}{p}$ und $(1-\lambda) = \frac{1}{q}$

$$-\ln\Big(\frac{a}{p}+\frac{b}{q}\Big) \le -\frac{1}{p}\ln a - \frac{1}{q}\ln b = -\ln\Big(a^{\frac{1}{p}} \cdot b^{\frac{1}{q}}\Big)\;.$$

Durch Exponentieren folgt die Behauptung.

ii) Für $x_k = 0$ für alle k oder $y_k = 0$ für alle k ist im Satz nichts zu zeigen. Ansonsten sei $\xi_k := \dfrac{|x_k|^p}{\|x\|_p^p}$ und $\eta_k := \frac{|y_k|^q}{\|y\|_q^q}$. Es gilt

$$\xi_k^{\frac{1}{p}} \cdot \eta_k^{\frac{1}{q}} \le \frac{\xi_k}{p} + \frac{\eta_k}{q}$$

nach (*) für $\xi_k, \eta_k \neq 0$ bzw. trivialerweise für $\xi_k = 0$ oder $\eta_k = 0$. Wegen $\sum_{k=0}^\infty \xi_k = 1$ und $\sum_{k=0}^\infty \eta_k = 1$ hat die Reihe $\sum_{k=0}^\infty \frac{|x_k|\cdot|y_k|}{\|x\|_p \cdot \|y\|_q}$ die konvergente Majorante $\sum_{k=0}^\infty \big(\frac{\xi_k}{p} + \frac{\eta_k}{q}\big)$, und es gilt

$$\sum_{k=0}^\infty \frac{|x_k|\cdot|y_k|}{\|x\|_p \cdot \|y\|_q} = \sum_{k=0}^\infty \xi_k^{\frac{1}{p}} \cdot \eta_k^{\frac{1}{q}} \le \sum_{k=0}^\infty \Big(\frac{\xi_k}{p} + \frac{\eta_k}{q}\Big) = \frac{1}{p} + \frac{1}{q} = 1\;,$$

also $\sum_{k=0}^\infty |x_k y_k| \le \|x\|_p \cdot \|y\|_q$. Schließlich hat die Reihe $\sum_{k=0}^\infty x_k y_k$ die konvergente Majorante $\sum_{k=0}^\infty |x_k y_k|$, ist damit selbst konvergent und erfüllt die Höldersche Ungleichung. □

Für $p = q = 2$ ist das die Ungleichung von Cauchy-Schwarz. Allgemein ergibt sich

Satz 25.8 (Minkowskische Ungleichung) *Sei $p > 1$. Dann gilt für beliebige Folgen $x = (x_k)_{k\in\mathbb{N}}$ und $y = (y_k)_{k\in\mathbb{N}}$ mit $\|x\|_p < \infty$ und $\|y\|_p < \infty$ die Minkowskische Ungleichung*

$$\|x+y\|_p \le \|x\|_p + \|y\|_p\;.$$

Damit ist $\ell^p := \{x = (x_k)_{k\in\mathbb{N}} \;:\; \|x\|_p < \infty\}$ für jedes $p \ge 1$ ein normierter Vektorraum bezüglich $\|\;\|_p$, und insbesondere definiert $\|(x_1,\dots,x_n)\|_p := \big(\sum_{k=1}^n |x_k|^p\big)^{\frac{1}{p}}$ eine Norm auf $\mathbb{R}^n$ oder $\mathbb{C}^n$.

Beweis. Setze $z_k := |x_k + y_k|^{p-1}$, dann gilt $|z_k|^q = |x_k + y_k|^{pq-q} = |x_k + y_k|^p$. Nach Dreiecksungleichung und Hölderscher Ungleichung gilt für jedes $N \in \mathbb{N}$

$$\begin{aligned}
\sum_{k=0}^{N} |x_k + y_k|^p = \sum_{k=0}^{N} |x_k + y_k||z_k| &\leq \sum_{k=0}^{N} |x_k z_k| + \sum_{k=0}^{N} |y_k z_k| \\
&\leq \Big(\Big(\sum_{k=0}^{N} |x_k|^p \Big)^{\frac{1}{p}} + \Big(\sum_{k=0}^{N} |y_k|^p \Big)^{\frac{1}{p}} \Big) \Big(\sum_{k=0}^{N} |z_k|^q \Big)^{\frac{1}{q}} \\
&\leq \big(\|x\|_p + \|y\|_p \big) \Big(\sum_{k=0}^{N} |x_k + y_k|^p \Big)^{\frac{1}{q}} .
\end{aligned}$$

Für $x_k + y_k = 0$ für alle k ist nichts zu zeigen, ansonsten folgt mit $1 - \frac{1}{q} = \frac{1}{p}$

$$\Big(\sum_{k=0}^{N} |x_k + y_k|^p \Big)^{\frac{1}{p}} \leq \|x\|_p + \|y\|_p .$$

Somit ist die monoton wachsende Folge $\big\{ \big(\sum_{k=0}^{N} |x_k + y_k|^p \big)^{\frac{1}{p}} \big\}_{N \in \mathbb{N}}$ durch $\|x\|_p + \|y\|_p$ beschränkt, so daß der Limes $N \to \infty$ existiert und die Minkowskische Unleichung erfüllt. □

Eine weitere wichtige Anwendung der Konvexität ist das Newtonsche Iterationsverfahren zur numerischen Berechnung von Nullstellen.

Satz 25.9 (Newtonsches Iterationsverfahren) *Eine stetige und konvexe Funktion $f : [a, b] \to \mathbb{R}$ mit $f(a) < 0$ und $f(b) > 0$ sei zweimal stetig differenzierbar in $]a, b[$. Dann gilt:*

i) *Es gibt genau ein $\xi \in]a, b[$ mit $f(\xi) = 0$.*

ii) *Ist $x_0 \in]a, b[$ ein beliebiger Punkt mit $f(x_0) \geq 0$, dann ist die durch*

$$x_{n+1} := x_n - \frac{f(x_n)}{f'(x_n)}$$

rekursiv definierte Folge $(x_n)_{n \in \mathbb{N}}$ wohldefiniert und konvergiert monoton fallend gegen ξ.

Bemerkung: Analoge Aussagen gelten für konkave Funktionen und/oder für $f(a) > 0$ und $f(b) < 0$. Man kann zeigen, daß das Verfahren quadratisch konvergiert, d.h. x_n approximiert ξ für große n auf $2n$ Dezimalstellen genau.

Beweis. i) Nach dem Zwischenwertsatz gibt es zumindest eine Nullstelle ξ von f in $]a, b[$. Außerdem gibt es nach dem Extremwertsatz in $q \in [a, b]$ ein globales Minimum mit $f(q) \leq f(x)$ für alle $x \in [a, b]$. Damit ist $f(q) < 0$. Ist $q \neq a$, so hat

f in q ein lokales Minimum, d.h. es gilt $f'(q) = 0$. Nach Satz 25.4 und Satz 25.5 ist f' monoton wachsend in $]a, b[$, also gilt $f'(x) \leq 0$ für alle $x \in [a, q]$. Somit liegen alle Nullstellen von f in $]q, b[$.

Angenommen, es gäbe zwei Nullstellen $\xi_1 < \xi_2$. Nach dem Mittelwertsatz gibt es ein $t \in]q, \xi_1[$ mit $f'(t) = \frac{f(\xi_1)-f(q)}{\xi_1 - q} = \frac{-f(q)}{\xi_1 - q} > 0$. Daraus folgt $f'(x) > 0$ für alle $x \in [t, b[$, insbesondere ist f in $[\xi_1, b]$ streng monoton wachsend und kann keine zweite Nullstelle besitzen.

ii) Für den Anfangspunkt der Folge gilt $x_0 \geq \xi$. Wir zeigen durch vollständige Induktion $f(x_n) \geq 0$ und $\xi \leq x_n \leq x_{n-1}$ für alle $n \geq 1$.

a) Aus $x_n \geq \xi$ folgt $f'(x_n) \geq f'(\xi) > 0$, also $\frac{f(x_n)}{f'(x_n)} \geq 0$ und somit $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \leq x_n$.

b) Die Funktion $\phi_n(x) := f(x) - f(x_n) - f'(x_n)(x - x_n)$ beschreibt die Differenz von $f(x)$ zur ihrer Tangenten in x_n. Es gilt $\phi_n'(x) = f'(x) - f'(x_n)$, also $\phi_n'(x) \leq 0$ für $x \leq x_n$. Aus $\phi_n(x_n) = 0$ folgt dann $\phi_n(x) \geq 0$ für $x \leq x_n$, insbesondere

$$0 \leq \phi_n(x_{n+1}) = f(x_{n+1}) - f(x_n) - f'(x_n)(x_{n+1} - x_n) = f(x_{n+1}) \; .$$

Damit ist $x_{n+1} \geq \xi$, so daß $(x_n)_{n \in \mathbb{N}}$ eine monoton fallende und durch ξ nach unten beschränkte Folge ist. Sei $x^* := \lim_{n\to\infty} x_n$ ihr Grenzwert, dann folgt aus der Stetigkeit von f und f' in $]a, b[$ die Gleichung $x^* = x^* - \dfrac{f(x^*)}{f'(x^*)}$, also $f(x^*) = 0$ und damit $x^* = \xi$. □

Beispiel 25.10 Für $k \in \mathbb{N}$ mit $k \geq 2$ und $a \in \mathbb{R}_+^\times$ betrachten wir die durch $f(x) = x^k - a$ definierte Funktion $f : \mathbb{R}_+ \to \mathbb{R}$. Diese ist konvex auf $\mathbb{R}_+$ mit $f(0) = -a < 0$ und $f(1 + a) > 0$. Das Newtonsche Iterationsverfahren zur Berechnung der Nullstellen ist damit anwendbar und liefert die Folge

$$x_{n+1} := x_n - \frac{x_n^k - a}{k x_n^{k-1}} = \frac{1}{k}\Big((k-1)x_n + \frac{a}{x_n^{k-1}}\Big) \; ,$$

die für einen beliebigen Startwert x_0 mit $x_0^k \geq a$ monoton fallend gegen $\sqrt[k]{a}$ konvergiert. (Wählt man $0 < x_0^k < a$, dann ist $x_1^k > a$, und das Verfahren konvergiert ebenfalls. Siehe Satz 9.14. ◁

26 Taylor-Polynome und Taylor-Reihen

Nach Satz 23.4 liefern Funktionswert $f(x_0)$ und Ableitung $f'(x_0)$ einer differenzierbaren Funktion $f : I \to \mathbb{C}$ eine lineare Approximation der Funktion in der Nähe des Punktes x_0 mit Kontrolle des Fehlers. Die Taylorsche Formel beschreibt eine polynomiale Approximation der Funktion:

Satz 26.1 (Taylorsche Formel) *Es sei $f : I \to \mathbb{C}$ eine $(n+1)$-mal stetig differenzierbare Funktion und $a \in I$. Dann gibt es zu jedem $x \in I$ ein $\xi \in]a, x[$ bzw. $\xi \in]x, a[$, so daß*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n+1}(x) ,$$

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1} .$$

Beweis. Nach Zerlegung in Real- und Imaginärteil können wir uns auf reellwertige Funktionen beschränken. Für $n = 0$ handelt es sich um den Mittelwertsatz der Differentialrechnung. Sei also $n \geq 1$ und z.B. $x > a$ fest gewählt. Wir definieren eine Funktion $g : I \to \mathbb{R}$ durch

$$g(t) = f(t) - f(a) - \sum_{k=1}^{n} \frac{f^{(k)}(a)}{k!}(t-a)^k - M(t-a)^{n+1} .$$

Wählen wir

$$M = \frac{1}{(x-a)^{n+1}}\Big(f(x) - f(a) - \sum_{k=1}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k\Big) ,$$

dann gilt $g(x) = 0$. Nach Voraussetzung ist g eine $(n+1)$-mal stetig differenzierbare Funktion auf I, und es gilt $g(a) = 0$ und $g^{(k)}(a) = 0$ für alle $k = 1, \dots, n$.

Nach dem Satz von Rolle gibt es ein $\xi_1 \in]x, a[$ mit $g'(\xi_1) = 0$. Im nächsten Schritt gibt es wieder nach Rolle wegen $g'(a) = 0$ ein $\xi_2 \in]\xi_1, a[$ mit $g''(\xi_2) = 0$. Schließlich gibt es ein $\xi = \xi_{n+1} \in]\xi_n, a[\subseteq]x, a[$ mit

$$0 = g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n+1)!M$$

Also ist $M = \frac{f^{(n+1)}(\xi)}{(n+1)!}$ für ein $\xi \in]x, a[$. □

Für eine mindestens $(n+1)$-mal stetig differenzierbare Funktion f heißt

$$(T_n f)(x; a) := f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

das Taylor-Polynom n-ter Ordnung von f mit Entwicklungspunkt a, und $R_{n+1}(x)$ heißt das Restglied $(n+1)$-ter Ordnung. Für Potenzreihen $f(x) = \sum_{n=0}^{\infty} a_n x^n$ mit Konvergenzradius $R > 0$ stimmt wegen $f^{(k)}(0) = k! a_k$ das Taylor-Polynom N-ter Ordnung von f mit der bei N abgebrochenen Potenzreihe überein, d.h. $(T_N f)(x; 0) = \sum_{n=0}^{N} a_n x^n$ für $|x| < R$.

Damit liefert die Taylorsche Formel Fehlerabschätzungen der folgenden Art:

Beispiel 26.2 Für $f(x) = \sin x$ gilt $f^{(2k)}(x) = (-1)^k \sin x$ und $f^{(2k+1)}(x) = (-1)^k \cos x$, d.h. $|f^{(n)}(\xi)| \leq 1$ für alle $\xi \in \mathbb{R}$ und alle $n \in \mathbb{N}$. Unter Verwendung der Sinus-Reihe gilt somit für alle $x \in \mathbb{R}$ und $n \in \mathbb{N}$

$$\Big|\sin x - \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!}\Big| \leq \frac{|x|^{2n+3}}{(2n+3)!} \; . \qquad \triangleleft$$

Satz 26.3 (hinreichendes Kriterium für lokale Extrema) *Es sei $f : I \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion. In einem inneren Punkt $a \in I$ gelte $f'(a) = \cdots = f^{(n)}(a) = 0$, aber $f^{(n+1)}(a) \neq 0$. Dann hat f im Punkt a*

i) *ein strenges lokales Minimum, falls n ungerade ist und $f^{(n+1)}(a) > 0$;*

ii) *ein strenges lokales Maximum, falls n ungerade ist und $f^{(n+1)}(a) < 0$;*

iii) *kein lokales Extremum, falls n gerade ist.*

Beweis. Einsetzen der Voraussetzungen in die Taylorsche Formel mit ergibt $f(x) = f(a) + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$ für ein ξ zwischen x und a. Wegen der Stetigkeit von $f^{(n+1)}$ gibt es nach Satz 18.14.ii) eine Umgebung $U \subseteq I$ von a, in der $f^{(n+1)}(\xi)$ das gleiche Vorzeichen wie $f^{(n+1)}(a)$ hat. Ist $n+1$ gerade und $f^{(n+1)}(a) > 0$, so folgt $f(x) > f(a)$ für alle $x \in U \setminus \{a\}$, d.h. f hat in a ein strenges lokales Minimum. Analog für ii). Ist dagegen $n+1$ ungerade und z.B. $f^{(n+1)}(a) > 0$, so folgt $f(x) < f(a)$ für $x < a$ und $f(x) > f(a)$ für $x > a$, d.h. f hat in a kein lokales Extremum. □

Satz 26.4 *Es sei $f : I \to \mathbb{C}$ eine n-mal stetig differenzierbare Funktion und $a \in I$. Dann gibt es eine stetige Funktion $r : I \to \mathbb{C}$ mit $r(a) = 0$ und $f(x) = (T_n f)(x;a) + (x-a)^n r(x)$.*

Beweis. Stetigkeit von r in $x \neq a$ folgt aus der Definition, so daß nur $\lim_{x\to a} r(x) = 0$ zu zeigen ist. Nach Zerlegung in Real- und Imaginärteil genügt der Beweis für reellwertige Funktionen. Mit der Taylorschen Formel gilt

$$\big(f(x) - (T_{n-1}f)(x;a)\big) - \frac{f^{(n)}(a)}{n!}(x-a)^n = \frac{f^{(n)}(\xi)}{n!}(x-a)^n - \frac{f^{(n)}(a)}{n!}(x-a)^n$$

für ein ξ zwischen x und a, d.h. $r(x) = \dfrac{f^{(n)}(\xi(x)) - f^n(a)}{n!}$ für $x \neq a$. Mit $x \to a$ geht auch $\xi(x) \to a$, so daß aus der Stetigkeit von $f^{(n)}$ folgt $\lim_{x\to a} r(x) = 0$. □

Für diese Eigenschaft ist das Landau-Symbol "o" gebräuchlich, $f(x) = (T_n f)(x;a) + o\big((x-a)^n\big)$. Darunter versteht man

$$f(x) = o(g(x)) \quad \text{für } x \to a \qquad \Leftrightarrow \qquad \lim_{x\to a} \frac{f(x)}{g(x)} = 0 \; .$$

Definition 26.5 Es sei $f : I \to \mathbb{C}$ eine beliebig oft differenzierbare Funktion. Die Potenzreihe $(Tf)(x;a) := \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$ heißt die *Taylor-Reihe* von f im Punkt $a \in I$. Konvergiert $(Tf)(x;a)$ gegen $f(x)$ für alle x aus einer Umgebung $U \subseteq I$ von a, so sagt man: f besitzt in U eine *Taylor-Entwicklung* um a.

Warnung: Die Taylor-Reihe ist ein formaler Ausdruck! Während die Taylorsche Formel mit Restglied exakt gilt, muß die Taylor-Reihe für $x \neq a$ nicht konvergieren (der Konvergenzradius ist dann 0), und selbst wenn die Taylor-Reihe konvergiert, muß $(Tf)(x;a)$ nicht gleich $f(x)$ sein!

Beispiel 26.6 Wir erinnern an die schon in Abschnitt 25 eingeführte Einschaltfunktion

$$f(x) := \begin{cases} e^{-\frac{1}{x}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Die Ableitungen in $x > 0$ sind von der Form $f^{(n)}(x) = P_n(\frac{1}{x})e^{-\frac{1}{x}})$ für ein Polynom P_n vom Grad $2n$ in $\frac{1}{x}$, denn nach Produkt- und Kettenregel ist

$$f^{(n+1)}(x) = \Big(P_n(\tfrac{1}{x})e^{-\frac{1}{x}}\Big)' = \Big(-P_n'(\tfrac{1}{x})\frac{1}{x^2} + P_n(\tfrac{1}{x})\frac{1}{x^2}\Big)e^{-\frac{1}{x}} \;.$$

Dabei ist die Ableitung P' des Polynoms ein Polynom $(2n-1)$-ter Ordnung. Somit ist $\lim_{x\searrow 0} f^{(n+1)}(x) = 0$ für alle $n \in \mathbb{N}$ und folglich $(Tf)(x;0) = 0 \neq f(x)$. $\triangleleft$

Ist eine Funktion f durch eine Potenzreihe darstellbar, so stimmt die Taylor-Reihe (innerhalb des Konvergenzkreises) mit f überein. Insbesondere lassen sich die im Laufe des Semesters hergeleiteten Potenzreihen reproduzieren:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}\;, \qquad x \in \mathbb{R}\;,$$

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}\;, \qquad x \in \mathbb{R}\;,$$

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}\;, \qquad x \in \mathbb{R}\;,$$

$$\cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}\;, \qquad x \in \mathbb{R}\;,$$

$$\sinh x = \sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!}\;, \qquad x \in \mathbb{R}\;,$$

$$
\begin{aligned}
(1+x)^\alpha &= \sum_{k=0}^\infty \binom{\alpha}{k} x^k , \qquad x \in \mathbb{R} \text{ für } \alpha \in \mathbb{N} , \quad x \in \,]-1,1[\text{ für } \alpha \in \mathbb{C} ,\\
\ln(1+x) &= \sum_{k=1}^\infty \frac{(-1)^{k-1} x^k}{k} , \qquad x \in \,]-1,1] ,\\
\arctan x &= \sum_{k=0}^\infty \frac{(-1)^{k-1} x^{2k+1}}{2k+1} , \qquad x \in [-1,1] ,\\
\arcsin x &= \sum_{k=0}^\infty \binom{-\frac{1}{2}}{k} (-1)^k \frac{x^{2k+1}}{2k+1} , \qquad x \in \,]-1,1[\,.
\end{aligned}
$$

Zu beachten ist, daß einige dieser Potenzreihen sogar für komplexe Zahlen $x \in \mathbb{C} \setminus \mathbb{R}$ gelten, während wir die Differentiation nur im Reellen eingeführt hatten.

Beispiel 26.7 Man kann zeigen, daß die Funktion $\frac{z}{e^z-1}$ in einer Umgebung von 0 in eine Potenzreihe entwickelbar ist,

$$
\frac{z}{e^z-1} = \frac{1}{1+\sum_{n=1}^\infty \frac{z^n}{(n+1)!}} = \sum_{k=0}^\infty \frac{B_k}{k!} z^k .
$$

Die Entwicklungskoeffizienten B_k heißen *Bernoulli-Zahlen*. Offenbar ist $B_0 = 1$ und $B_1 = -\frac{1}{2}$. Ferner gilt

$$
\frac{z}{e^z-1} - B_1 z = \frac{z}{2} \frac{e^z+1}{e^z-1} = \frac{z}{2} \coth \frac{z}{2} ,
$$

d.h. $\frac{z}{e^z-1} - B_1 z$ ist eine gerade Funktion, so daß alle ungeraden Koefizienten $B_{2k+1} = 0$ mit $k \geq 1$ verschwinden. Ersetzt man $z \mapsto 2ix$, so folgt die *Cotangens-Reihe*

$$
\cot x = \frac{1}{x} + \sum_{k=1}^\infty (-1)^k \frac{4^k}{(2k)!} B_{2k} x^{2k-1} .
$$

Mittels $\tan x - \cot x = \frac{\sin^2 x - \cos^2 x}{\sin x \cos x} = -2\cot(2x)$ folgt daraus die *Tangens-Reihe*

$$
\tan x = \sum_{k=1}^\infty (-1)^{k-1} \frac{4^k(4^k-1)}{(2k)!} B_{2k} x^{2k-1} .
$$

◁

27 Partielle Ableitungen

Wir betrachten Funktionen mehrerer Veränderlicher, also Abbildungen $f : U \to \mathbb{R}$ mit $U \subseteq \mathbb{R}^n$. Partielle Ableitungen von f sind gewöhnliche Ableitungen, die man erhält, wenn alle Komponenten von $x = (x_1, \ldots, x_n) \in U$ bis auf eine festgehalten werden.

Definition 27.1 Sei $U \subseteq \mathbb{R}^n$ offen. Eine Funktion $f : U \to \mathbb{C}$ heißt im Punkt $x \in U$ *partiell differenzierbar* in der j-ten Koordinatenrichtung, falls der Grenzwert

$$(\partial_j f)(x) := \lim_{\substack{h\to 0\\ h\neq 0}} \frac{1}{h}\big(f(x+he_j) - f(x)\big)$$

existiert. Dabei ist $e_j \in \mathbb{R}^n$ der j-te Einheitsvektor und h ist so zu wählen, daß $x + he_j \in U$. Der Grenzwert $(\partial_j f)(x)$ heißt die *j-te partielle Ableitung* von f in x.

Eine Funktion $f : U \to \mathbb{C}$ heißt *partiell differenzierbar*, falls $(\partial_j f)(x)$ für alle $x \in U$ und alle $1 \leq j \leq n$ existiert, und *stetig partiell differenzierbar*, falls alle Funktionen $\partial_j f : U \to \mathbb{C}$ stetig sind.

Oft schreibt man auch $\frac{\partial f}{\partial x_j}$ an Stelle von $\partial_j f$. Zur Existenz des Grenzwertes muß U nicht notwendig offen sein.

Die partielle Ableitung erfüllt die Leibniz-Regel: Seien $f, g : U \to \mathbb{C}$ partiell differenzierbare Funktionen auf U, dann gilt

$$(\partial_j(f\cdot g))(x) = (\partial_j f)(x)\cdot g(x) + f(x)\cdot(\partial_j g)(x)\ , \qquad x \in U\ .$$

Beispiel 27.2 Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch $f(x_1, x_2) = e^{x_1^2+x_2^2}\sin x_1$. Dann ist $(\partial_1 f)(x_1, x_2) = (2x_1 \sin x_1 + \cos x_1)e^{x_1^2+x_2^2}$ und $(\partial_2 f)(x_1, x_2) = 2x_2 \sin x_1 e^{x_1^2+x_2^2}$. ◁

Beispiel 27.3 Die partielle Ableitung des Radius $r(x) := \|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$ in der j-ten Koordinatenrichtung ist nach der Kettenregel

$$\frac{\partial r}{\partial x_j}(x_1, \dots, x_n) = \frac{1}{2\sqrt{x_1^2+\cdots+x_n^2}}\cdot 2x_j = \frac{x_j}{r}\ .$$

Damit ist $r : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ partiell differenzierbar mit $(\partial_j r)(x) = \frac{x_j}{r}$.

Entsprechend ist nach der Kettenregel jede differenzierbare Funktion $f(r)$ des Radius, aufgefaßt als Funktion $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$, partiell differenzierbar mit $(\partial_i f)(x) = x_j \frac{f'(r)}{r}$. Zum Beispiel sind für $f(r) = \mathrm{e}^{-ar^2}$ die partiellen Ableitungen gegeben durch $(\partial_j f)(x) = -2ax_j f(r)$. ◁

Das folgende Beispiel zeigt, daß aus der partiellen Differenzierbarkeit einer Funktion *nicht* die Stetigkeit folgt.

Beispiel 27.4 Es sei $f : \mathbb{R}^2 \to \mathbb{R}$ gegeben durch

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2+x_2^2} & \text{für } (x_1, x_2) \neq (0,0) \\ 0 & \text{für } (x_1, x_2) = (0,0) \end{cases}$$

Dann ist f auf $\mathbb{R}^2 \setminus \{(0,0)\}$ partiell differenzierbar mit

$$(\partial_1 f)(x_1, x_2) = \frac{x_2}{x_1^2+x_2^2} - \frac{2x_1^2 x_2}{(x_1^2+x_2^2)^2} = \frac{x_2(x_2^2 - x_1^2)}{(x_1^2+x_2^2)^2}$$

und analog

$$(\partial_2 f)(x_1,x_2) = \frac{x_1(x_1^2-x_2^2)}{(x_1^2+x_2^2)^2} \;.$$

Im Nullpunkt haben wir

$$(\partial_1 f)(0,0) = \lim_{h\to 0} \frac{f(h,0){-}f(0,0)}{h} = 0 \;, \quad (\partial_2 f)(0,0) = \lim_{h\to 0} \frac{f(0,h){-}f(0,0)}{h} = 0 \;,$$

so daß f auf dem gesamten $\mathbb{R}^2$ partiell differenzierbar ist. Jedoch ist f nicht stetig in $(0,0)$. Die Folge $(y_k)_{k\in\mathbb{N}}$ mit $y_k = (\frac{1}{k+1}, \frac{1}{k+1})$ konvergiert gegen $(0,0)$, aber

$$f(y_k) = \frac{\frac{1}{k+1}\cdot\frac{1}{k+1}}{(\frac{1}{k+1})^2+(\frac{1}{k+1})^2} = \frac{1}{2}$$

konvergiert nicht gegen $f(0)=0$. ◁

Wir werden im 3. Semester sehen, daß stetig partiell differenzierbare Funktionen gute Eigenschaften haben.

Die partielle Ableitung einer partiell differenzierbaren Funktion kann nochmals partiell differenziert werden, usw. Induktiv definieren wir:

Definition 27.5 Sei $U\subseteq\mathbb{R}^n$ offen und $k\in\mathbb{N}^\times$. Eine Funktion $f:U\to\mathbb{C}$ heißt $(k{+}1)$-*mal partiell differenzierbar*, wenn sie k-mal partiell differenzierbar ist und alle partiellen Ableitungen $\partial_{i_k}\dots\partial_{i_1}f$ partiell differenzierbar sind. Sind die partiellen Ableitungen $\partial_{i_{k+1}}\partial_{i_k}\dots\partial_{i_1}f$ stetig, so heißt f eine $(k{+}1)$-*mal stetig partiell differenzierbare Funktion*.

Satz 27.6 (Schwarz) *Sei $U\subseteq\mathbb{R}^n$ offen und $f:U\to\mathbb{C}$ eine zweimal stetig partiell differenzierbare Funktion. Dann vertauschen die zweiten partiellen Ableitungen, d.h. für alle $a\in U$ und alle $i,j=1,\dots,n$ gilt*

$$(\partial_i\partial_j f)(a) = (\partial_j\partial_i f)(a) \;.$$

Beweis. Nach Zerlegung in Real- und Imaginärteil kann $f:U\to\mathbb{R}$ angenommen werden. Der Übersichtlichkeit wegen sei $i=1, j=2$ (kann durch Umnumerieren der Koordinaten immer erreicht werden) und dann $n=2$ (die weiteren Komponenten sind festgehalten und spielen keine Rolle).

Für gegebenes $\delta>0$ sei $W_\delta(a)\subseteq U\subseteq\mathbb{R}^2$ der offene Würfel mit Kantenlänge 2δ und Mittelpunkt $a=(x_0,y_0)$. Es sei (x,y) ein beliebiger Punkt von W_δ, d.h. $|x-x_0|<\delta$ und $|y-y_0|<\delta$. Für festgehaltenes y sei $F_y(x)=f(x,y)-f(x,y_0)$. Nach dem Mittelwersatz der Differentialrechnung gibt es einen Punkt $\xi\in]x_0-\delta,x_0+\delta[$, so daß

$$F_y(x)-F_y(x_0) = F_y'(\xi)\cdot(x-x_0)$$

$$\Rightarrow \quad f(x,y)-f(x,y_0)-f(x_0,y)+f(x_0,y_0) = \big((\partial_1 f)(\xi,y) - (\partial_1 f)(\xi,y_0)\big)\cdot(x-x_0)\;.$$

Wir nutzen den Mittelwertsatz nochmals für die Funktion $G_\xi(y) := (\partial_1 f)(\xi,y)$. Es gibt also ein $\eta \in]y_0-\delta, y_0+\delta[$, so daß

$$\begin{aligned} & G_\xi(y) - G_\xi(y_0) = G'_\xi(\eta)\cdot(y-y_0) \\ \Rightarrow \quad & (\partial_1 f)(\xi,y) - (\partial_1 f)(\xi,y_0) = (\partial_2\partial_1 f)(\xi,\eta)\cdot(y-y_0)\;. \end{aligned}$$

Insgesamt gibt es somit ein $(\xi,\eta) \in W_\delta$ mit

$$f(x,y)-f(x,y_0)-f(x_0,y)+f(x_0,y_0) = (\partial_2\partial_1 f)(\xi,\eta)\cdot(x-x_0)(y-y_0)\;.$$

Wir können aber auch erst x festhalten und den Mittelwertsatz in y anwenden, und als letztes den Mittelwertwsatz in x. Im Ergebnis gibt es einen neuen Punkt $(\tilde\xi,\tilde\eta) \in W_\delta$ mit

$$f(x,y)-f(x,y_0)-f(x_0,y)+f(x_0,y_0) = (\partial_1\partial_2 f)(\tilde\xi,\tilde\eta)\cdot(x-x_0)(y-y_0)\;.$$

Somit gilt $(\partial_2\partial_1 f)(\xi,\eta) = (\partial_1\partial_2 f)(\tilde\xi,\tilde\eta)$. Lassen wir δ gegen 0 streben, so konvergieren (ξ,η) und $(\tilde\xi,\tilde\eta)$ gegen (x,y), und aus der vorausgesetzten Stetigkeit der zweiten partiellen Ableitungen folgt $(\partial_2\partial_1 f)(x,y) = (\partial_1\partial_2 f)(x,y)$. □

Entsprechend können bei k-mal stetig partiell differenzierbaren Funktionen die partiellen Ableitungen in beliebiger Reihenfolge geschrieben werden:

$$\partial_{i_1}\dots\partial_{i_k} f = \partial_{\pi(i_1)}\dots\partial_{\pi(i_k)} f$$

für eine beliebige Permutation π der Indizes $i_1,\dots,i_k$, denn jede Permutation läßt sich durch Vertauschen benachbarter Elemente darstellen. Es ist deshalb auch üblich, die mehrfachen partiellen Ableitungen zu schreiben als

$$\partial_{i_1}\dots\partial_{i_k} f = \frac{\partial^k f}{\partial x_{i_1}\dots\partial x_{i_k}}\;.$$

Wiederholung

- Produkt- und Kettenregel, Differentiation der Umkehrfunktion, Differentiation von Potenzreihen
- Mittelwertsatz der Differentialrechnung, Schrankensatz
- Regel von de l'Hospital
- Konvexität: Höldersche Ungleichung
- Taylorsche Formel: Approximation mit Fehlerabschätzung
- Kurvendiskussion: lokale Extrema, notwendige und hinreichende Kriterien dafür; Newton-Verfahren zur Bestimmung der Nullstellen
- Überblick über wichtige Potenzreihen