

# A N A L Y S I S I

Wintersemester 2025/2026

**Ernst Kuwert**

Mathematisches Institut  
Universität Freiburg



# Inhaltsverzeichnis

1	Die Sprache der Mengenlehre . . . . .	1
2	Arithmetik und Anordnung in $\mathbb{R}$ . . . . .	5
3	Vollständige Induktion . . . . .	13
4	Grenzwerte von Folgen . . . . .	21
5	Vollständigkeit der reellen Zahlen . . . . .	29
6	Teilmengen von $\mathbb{R}$ und von $\mathbb{R}^n$ . . . . .	41
7	Stetigkeit . . . . .	51
8	Zwischenwertsatz und Monotonie . . . . .	61
9	Die Ableitung . . . . .	65
10	Mittelwertsatz . . . . .	71
11	Die reelle Exponentialfunktion . . . . .	77
12	Die trigonometrischen Funktionen . . . . .	85
13	Das Riemannsches Integral . . . . .	95
14	Ableitung und Integral . . . . .	107
15	Reihen . . . . .	115
16	Potenzreihen . . . . .	123
17	Taylorentwicklung . . . . .	131



# 1 Die Sprache der Mengenlehre

Wenn es um die Kommunikation mathematischer Sachverhalte geht, ist die Umgangssprache nur teilweise geeignet, die Sprache der Mengenlehre ist oft nützlich. Keine Angst, wir wollen uns nicht wissenschaftlich mit der Mengenlehre befassen, es geht nur um ein paar Vokabeln: wie kann man Mengen beschreiben, was bedeuten die Begriffe Teilmenge und Obermenge, Vereinigung, Durchschnitt, Differenz, Komplement, leere Menge? Mit der Sprache der Mengenlehre lassen sich auch Funktionen definieren, einschließlich der Konzepte Bild und Urbild, Einschränkung, Verkettung und Graph, sowie der Begriffe injektiv, surjektiv, bijektiv und Umkehrfunktion.

Der Begriff der Menge wurde 1895 von G. Cantor eingeführt, hier eine Kurzversion:

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

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

- (a) Das griechische Alphabet ist die Menge

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

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

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

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

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

- (c) Eine Variante der aufzählenden Beschreibung haben wir bei

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

$M$  wird durch die Menge  $\mathbb{N} = \{1, 2, 3, \dots\}$  der natürlichen Zahlen "parametrisiert".

Eine Menge  $M$  heißt Teilmenge der Menge  $N$ , wenn jedes Element von  $M$  auch Element von  $N$  ist (Notation:  $M \subset N$ ). Gibt es außerdem ein Element von  $N$  das nicht zu  $M$  gehört, so ist  $M$  echte Teilmenge von  $N$ .<sup>1</sup> Beispiel: unter den Studierenden der Mathematik bilden die weiblichen eine echte Teilmenge, denn es gibt ja auch Studenten. Statt  $M \subset N$  schreiben wir manchmal  $N \supset M$ , also  $N$  ist Obermenge von  $M$ .<sup>2</sup>

---

<sup>1</sup>soll das betont werden schreibe ich  $M \subsetneq N$ .

<sup>2</sup>manche Autoren schreiben für Teilmenge  $M \subseteq N$  und  $M \subset N$  für echte Teilmenge, alles Geschmackssache.

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

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

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

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

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

Oben haben wir die Worte *und* sowie *oder* benutzt, die Aussagen miteinander logisch verknüpfen. Unter einer *Aussage* verstehen wir einen Satz, bei dem wir sinnvoll nach dem Wahrheitswert fragen können, das heißt ist der Satz *wahr* oder *falsch*? Zum Beispiel ist *Der Mars besteht aus grünem Käse* eine Aussage, während *Geh nach Hause!* oder *Was gibts heute in der Mensa?* keine Aussagen sind. Aus gegebenen Aussagen  $P, Q$  können wir durch logische Verknüpfung neue Aussagen herstellen:<sup>3</sup>

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

Beachte dass das *oder* nicht ausschließend zu verstehen ist, es bedeutet nicht entweder/oder. Eine wichtige logische Verknüpfung ist die Implikation:

$P \Rightarrow Q$	<i>aus P folgt Q</i>	falsch genau wenn $P$ wahr und $Q$ falsch, wahr immer wenn $P$ falsch, oder wenn $P, Q$ beide wahr.
-------------------	----------------------	--

Ich habe mal in einem Beweis einen Fehler gemacht, dann konnte ich auf einmal überraschende Dinge herleiten. Aus einer falschen Aussage kann man allen Unsinn schließen, etwa: *wenn der Mars aus grünem Käse ist dann bin ich der Kaiser von China*. Das Äquivalenzzeichen  $P \Leftrightarrow Q$  bedeutet  $P \Rightarrow Q$  und  $Q \Rightarrow P$ . Es ist also wahr genau wenn  $P, Q$  beide wahr oder beide falsch sind. Die Implikation  $P \Rightarrow Q$  ist gleichbedeutend mit  $\neg Q \Rightarrow \neg P$ . Dazu folgende Tabelle:

---

<sup>3</sup>Negation, Konjunktion und Disjunktion

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

Wir sehen dass  $Q \Rightarrow P$  nicht gleichbedeutend ist mit  $P \Rightarrow Q$ , das wird oft nicht beachtet.

Wir kommen nun zu den Abbildungen oder gleichbedeutend Funktionen. Seien  $X, Y$  Mengen. Eine Abbildung von  $X$  nach  $Y$  (gleichbedeutend: eine Funktion auf  $X$  mit Werten in  $Y$ ) schreiben wir in der Form

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

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

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

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

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

Beachte: dies definiert nur die Menge  $f^{-1}(B)$  und nicht die Abbildung  $f^{-1}$ . Die Abbildung  $f$  muss hier nicht bijektiv sein, die Abbildung  $f^{-1}$  gibt es also eventuell gar nicht. Verkleinern des Definitionsbereichs ergibt eine Einschränkung von  $f$ . Für  $A \subset X$  ist

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

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

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

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

Eine aus der Schule bekannte Möglichkeit, sich Funktionen bzw. Abbildungen vorzustellen, bietet der Graph. Dazu bilden wir das kartesische Produkt<sup>4</sup>

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

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

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

---

<sup>4</sup>René Descartes, 1596–1650

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 2 Arithmetik und Anordnung in $\mathbb{R}$

Die reellen Zahlen  $\mathbb{R}$  sind fundamental für die Analysis. Basierend auf Ideen von Archimedes<sup>5</sup> wurde ihr Konzept im 19. Jahrhundert entwickelt, beteiligt waren unter anderem Cauchy, Bolzano und Weierstraß. Konstruktionen der reellen Zahlen haben schließlich Cantor und Dedekind um 1875 angegeben. Man könnte denken dass es mit der Analysis erst dann losging, aber das wäre ganz falsch. Viele Resultate wurden lange vorher gezeigt, zum Beispiel um 1680 der Hauptsatz der Differential- und Integralrechnung. Die reellen Zahlen wurden dabei intuitiv benutzt, nicht rigoros aber sehr erfolgreich. In dieser Vorlesung beschreiben wir  $\mathbb{R}$  durch Axiome. Wir beginnen mit den Rechenregeln (Arithmetik) und den Ungleichungen (Anordnung), in Kapitel 5 kommt die Vollständigkeit hinzu. In real life werden nur die Axiome benutzt, nicht die konkreten Modelle von Cantor und Dedekind, so ist auch unser Zugang. Allerdings zeigen die Modelle dass unsere Axiome widerspruchsfrei sind und dass es die reellen Zahlen wirklich gibt. Für mehr Informationen siehe Kapitel 2 des Buchs *Zahlen*<sup>6</sup>.

Im folgenden setzen wir voraus dass die Zahlbereiche  $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$  aus der Schule (oder der Vorlesung Lineare Algebra) bekannt sind:

$$\mathbb{N} = \{1, 2, 3, \dots\} \quad \text{natürliche Zahlen}$$

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\} \quad \text{ganze Zahlen}$$

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid q \in \mathbb{N}, p \in \mathbb{Z} \right\} \quad \text{rationale Zahlen}$$

Die reellen Zahlen sind eine Menge  $\mathbb{R}$  auf der folgende Strukturen gegeben sind:

*eine Verknüpfung  $+$ , die je zwei  $a, b \in \mathbb{R}$  ein  $a + b \in \mathbb{R}$  zuordnet (Addition),  
eine Verknüpfung  $\cdot$ , die je zwei  $a, b \in \mathbb{R}$  ein  $a \cdot b \in \mathbb{R}$  zuordnet (Multiplikation),  
eine Relation  $a > b$ , die für  $a, b \in \mathbb{R}$  zutrifft oder nicht (Anordnung).*

Wir schreiben oft  $ab$  statt  $a \cdot b$ . Es sollen eine Reihe von Axiomen gelten, welche die reellen Zahlen charakterisieren. Diese sind in drei Gruppen unterteilt:

*Körperaxiome (K)  
Anordnungsaxiome (A)  
Vollständigkeitsaxiom (V).*

Die Körperaxiome (K) regeln die arithmetischen Eigenschaften von  $\mathbb{R}$ , also Addition und Multiplikation. Sie lauten wie folgt:

---

<sup>5</sup>Archimedes von Syrakus, 287-212 v.C.

<sup>6</sup>H.-D. Ebbinghaus (Hrsg.): Zahlen, Springer 1992

	+	•
<i>Assoziativgesetz:</i>	$(a + b) + c = a + (b + c)$	$(a \cdot b) \cdot c = a \cdot (b \cdot c)$
<i>Kommutativgesetz:</i>	$a + b = b + a$	$a \cdot b = b \cdot a$
<i>Neutrales Element:</i>	<i>Es gibt Zahlen <math>0 \in \mathbb{R}</math> und <math>1 \in \mathbb{R}</math> mit <math>1 \neq 0</math>, so dass für alle <math>a \in \mathbb{R}</math> gilt:</i>	
	$a + 0 = a$	$a \cdot 1 = a$
<i>Inverses Element:</i>	<i>Zu jedem <math>a \in \mathbb{R}</math> gibt es Lösungen <math>x, y \in \mathbb{R}</math> der Gleichungen</i>	
	$a + x = 0$	$a \cdot y = 1$ falls $a \neq 0$
<i>Distributivgesetz:</i>	$a \cdot (b + c) = a \cdot b + a \cdot c.$	

Eine Menge mit Verknüpfungen  $+$  und  $\cdot$ , so dass die Axiome (K) erfüllt sind, heißt Körper (englisch: *field*). Die reellen Zahlen sind also ein Körper, aber keineswegs der einzige. Die ganzen Zahlen  $\mathbb{Z}$  sind kein Körper (welche Gesetze sind erfüllt, welche nicht?), die rationalen Zahlen  $\mathbb{Q}$  sind aber ein Körper (check!). Wir werden sehen dass  $\mathbb{Q}$  als dichte Teilmenge in  $\mathbb{R}$  enthalten ist. In der Algebra treten viele verschiedene Körper auf, ein extremes Beispiel ist  $\mathbb{K} = \{0, 1\}$  mit  $1 + 1 = 0$  und den sonst üblichen Rechenregeln.

**Satz 2.1 (Lösen von Gleichungen)** *Aus den Körpergesetzen folgt:*

- (1) *Die neutralen Elemente 0 bzw. 1 sind eindeutig bestimmt.*
- (2) *Das zu  $a$  inverse Element ist eindeutig bestimmt; es wird mit  $-a$  bezeichnet (Addition) bzw. mit  $\frac{1}{a}$  (Multiplikation).*
- (3) *Zu  $a, b \in \mathbb{R}$  hat die Gleichung  $x + a = b$  genau eine Lösung, sie wird mit  $b - a$  bezeichnet. Ist  $a \neq 0$ , so hat die Gleichung  $x \cdot a = b$  auch genau eine Lösung, diese heißt  $\frac{b}{a}$ .*

BEWEIS: Wir zeigen die Aussagen für die Addition, der Beweis für die Multiplikation ist analog. Sind  $0_1 \in \mathbb{R}$  und  $0_2 \in \mathbb{R}$  neutrale Elemente, so folgt mit dem Kommutativgesetz

$$0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2.$$

Damit ist (1) bewiesen. Um (2) zu zeigen seien  $x_{1,2}$  Lösungen der Gleichung  $a + x = 0$ . Dann folgt mit dem Assoziativ- und Kommutativgesetz

$$x_1 = x_1 + 0 = x_1 + (a + x_2) = (x_1 + a) + x_2 = (a + x_1) + x_2 = 0 + x_2 = x_2.$$

Betrachte nun die Gleichung  $x + a = b$ . Ist  $x$  eine Lösung so folgt

$$x = x + (a + (-a)) = (x + a) + (-a) = b + (-a).$$

Und  $x = b + (-a)$  ist wirklich eine bzw. die eindeutige Lösung, denn

$$x + a = (b + (-a)) + a = b + ((-a) + a) = b.$$

□

Nach Beweis von Schritt (3) ist  $b - a = b + (-a)$ . Für  $b = 0$  ergibt sich  $0 - a = -a$ .

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

$$\left\{ \begin{array}{ll} -(-a) = a & -(a+b) = (-a) + (-b), \\ (a^{-1})^{-1} = a \quad (a \neq 0) & (ab)^{-1} = a^{-1}b^{-1} \quad (a, b \neq 0), \\ a \cdot 0 = 0 & a(-b) = -(ab), \\ (-a)(-b) = ab & a(b-c) = ab - ac. \end{array} \right. \quad (2.1)$$

$$ab = 0 \quad \Rightarrow \quad a = 0 \quad \text{oder} \quad b = 0 \quad (\text{Nullteilerfreiheit}). \quad (2.2)$$

BEWEIS: Die Zahl  $-b$  ist das eindeutig bestimmte inverse Element zu  $b$  bezüglich der Addition. Die erste Zeile in (2.1) folgt damit aus

$$(-a) + a = a + (-a) = 0 \quad (a+b) + ((-a) + (-b)) = (a + (-a)) + (b + (-b)) = 0 + 0 = 0.$$

Die zweite Zeile ist die entsprechende Aussage für die Multiplikation, sie folgt analog. Für die erste Behauptung in der dritten Zeile zeigen wir, dass  $a \cdot 0$  eine Lösung der Gleichung  $a \cdot 0 + x = a \cdot 0$  ist, und zwar gilt

$$a \cdot 0 + a \cdot 0 = a \cdot (0 + 0) = a \cdot 0.$$

Die Gleichung wird aber auch durch  $x = 0$  gelöst, mit der Eindeutigkeit aus Satz 2.1(3) sehen wir  $a \cdot 0 = 0$ . Nun folgt weiter

$$ab + a(-b) = a(b + (-b)) = a \cdot 0 = 0,$$

also  $a(-b) = -(ab)$ , und dann durch zweimalige Anwendung

$$(-a)(-b) = -((-a)b) = -(b(-a)) = -(-ba) = ba = ab.$$

Die letzte Aussage von (2.1) gilt mit

$$a(b-c) = a(b + (-c)) = ab + a(-c) = ab + (-ac) = ab - ac.$$

Ist in (2.2)  $a \neq 0$ , so folgt schließlich

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

Damit sind alle Aussagen des Satzes gezeigt. □

Auch die folgenden Regeln der Bruchrechnung lassen sich aus den Körperaxiomen herleiten, dies sei jedoch den Lesern überlassen.

**Satz 2.3 (Bruchrechnung)** Für  $a, b, c, d \in \mathbb{R}$  mit  $c, d \neq 0$  gilt:

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

Nun zur Anordnung der reellen Zahlen. Dabei geht es um das Konzept von positiv und negativ (Symbole  $a > 0$  und  $a < 0$ ), und die daraus abgeleiteten Begriffe größer und kleiner. Wir formulieren zwei Anordnungsaxiome, ein drittes folgt weiter unten.

(A1) Für jedes  $a \in \mathbb{R}$  gilt genau eine der drei Aussagen  $a > 0$ ,  $a = 0$  oder  $-a > 0$ .

(A2) Aus  $a, b > 0$  folgt  $a + b > 0$  und  $ab > 0$ .

Um einige Rechenregeln abzuleiten, führen wir noch folgende Notation für  $a, b \in \mathbb{R}$  ein:

$$a > b \Leftrightarrow a - b > 0 \quad \text{sowie} \quad a < b \Leftrightarrow b > a.$$

Damit ist  $a < 0$  genau wenn  $-a > 0$ , die Alternativen in A1 sind  $a$  positiv,  $a$  Null oder  $a$  negativ. Und die Ungleichung  $a < b$  bedeutet dass  $a - b$  negativ ist.

### Satz 2.4 (Rechnen mit Ungleichungen)

(1) Für  $a, b \in \mathbb{R}$  gilt genau eine der Relationen  $a > b$ ,  $a = b$  oder  $a < b$ .

(2) Aus  $a > b$ ,  $b > c$  folgt  $a > c$  (Transitivität).

(3) Aus  $a > b$  folgt

$$\begin{cases} a + c > b + c \\ ac > bc, \text{ wenn } c > 0 \\ ac < bc, \text{ wenn } c < 0 \end{cases}$$

(4) Aus  $a > b$  und  $c > d$  folgt

$$\begin{cases} a + c > b + d, \\ ac > bd, \text{ falls } b, d > 0 \end{cases}$$

(5) Für  $a \neq 0$  ist  $a^2 > 0$ .

(6) Aus  $a > 0$  folgt  $1/a > 0$ .

(7) Aus  $a > b$ ,  $b > 0$  folgt  $1/a < 1/b$ .

(8) Ist  $a \in \mathbb{R}$  mit  $a < \varepsilon$  für alle  $\varepsilon > 0$ , so ist  $a \leq 0$ .

BEWEIS: Wie oben diskutiert sind  $a > b$ ,  $a = b$  und  $a < b$  äquivalent zu  $a - b > 0$ ,  $a - b = 0$  und  $a - b < 0$ , Aussage (1) folgt also mit (A1). Für (2) rechnen wir

$$a - c = (a - b) + (b - c) > 0 \quad \text{nach (A2)}.$$

Ähnlich ergeben sich (3) und (4):

$$\begin{aligned} (a + c) - (b + c) &= a - b > 0, \\ ac - bc &= (a - b)c > 0 \quad \text{im Fall } c > 0 \text{ nach (A2)}, \\ bc - ac &= (a - b)(-c) > 0 \quad \text{im Fall } c < 0 \text{ nach (A2)}, \\ (a + c) - (b + d) &= (a - b) + (c - d) > 0 \quad \text{nach (A2)}, \\ ac - bd &= ac - bc + bc - bd \\ &= (a - b)c + b(c - d) > 0 \quad \text{nach (2) und (A2)}. \end{aligned}$$

Die Positivität von Quadraten folgt aus der Fallunterscheidung

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

Dabei haben wir die Regel  $(-a)(-b) = ab$  aus (2.1) benutzt. Nach (A1) ist  $a^2 > 0$  für  $a \neq 0$  bewiesen. Aussage (6) ergibt sich nun mit

$$\frac{1}{a} = \left(\frac{1}{a}\right)^2 \cdot a > 0 \text{ nach (5) und (A2).}$$

Zu guter Letzt haben wir für (7)

$$\frac{1}{b} - \frac{1}{a} = \frac{1}{ab}(a - b) > 0 \text{ mit (6) und (A2).}$$

Wäre nicht  $a \leq 0$  in (8), so folgt  $a > 0$  aus (1). Dann können wir  $\varepsilon = a$  wählen und erhalten  $a < a$ , Widerspruch zu (1).  $\square$

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

$$|a| = \begin{cases} a & \text{falls } a \geq 0 \\ -a & \text{falls } a \leq 0. \end{cases} \quad (2.3)$$

Es ist praktisch bei den Fällen den overlap  $a = 0$  zu lassen, wegen  $-0 = 0$  stimmen die beiden Definitionen ja überein.

Ein anschauliches Modell der reellen Zahlen ist die Zahlengerade. Ist  $a < b$ , so liegt  $b$  rechts von  $a$  im Abstand  $|a - b|$ . Insbesondere ist  $|a|$  der Abstand zum Nullpunkt.

**Satz 2.5 (Rechnen mit Beträgen)** *Für  $a, b \in \mathbb{R}$  gelten folgende Aussagen:*

- (1)  $|-a| = |a|$ .
- (2)  $|a| = \max(a, -a)$ .
- (3)  $|a| \geq 0$ , aus Gleichheit folgt  $a = 0$ .
- (4)  $|ab| = |a| \cdot |b|$ .
- (5)  $|a + b| \leq |a| + |b|$ .
- (6)  $|a - b| \geq ||a| - |b||$ .

BEWEIS: Definition 2.1 liefert direkt

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

Ebenso klar ist Aussage (2), denn es gilt

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

Hieraus folgt leicht Aussage (3).

In (4) bleiben die linke und rechte Seite gleich, wenn wir  $a$  durch  $-a$  ersetzen, dasselbe gilt bezüglich  $b$ . Also können wir  $a, b \geq 0$  annehmen, und erhalten  $|ab| = ab = |a| \cdot |b|$  wie verlangt. Ersetzen wir in (5)  $a$  durch  $-a$  und gleichzeitig  $b$  durch  $-b$ , so bleiben wieder beide Seiten gleich. Damit können wir  $a + b \geq 0$  annehmen, und mit (2) wie folgt abschätzen:

$$|a + b| = a + b \leq |a| + |b|.$$

Schließlich gilt  $|a| = |a - b + b| \leq |a - b| + |b|$  nach (5), also  $|a - b| \geq |a| - |b|$ . Hier können wir aber  $a$  und  $b$  vertauschen, es folgt

$$||a| - |b|| = \max(|a| - |b|, |b| - |a|) \leq |a - b|.$$

□

Wie wir später sehen werden, spielen Ungleichungen in der Analysis eine große Rolle. Wir kommen nun zu dem noch fehlenden, dritten Anordnungsaxiom.

**A3** Zu jedem  $\varepsilon \in \mathbb{R}$ ,  $\varepsilon > 0$ , existiert ein  $n \in \mathbb{N}$  mit  $1/n < \varepsilon$  (*Archimedisches Axiom*).

*Bemerkung.* Es gibt dann auch zu jedem  $K \in \mathbb{R}$  ein  $n \in \mathbb{N}$  mit  $n > K$ . Für  $K \leq 0$  ist das sowieso klar. Im Fall  $K > 0$  wähle  $\varepsilon = 1/K > 0$  in A3, und erhalte ein  $n \in \mathbb{N}$  mit  $1/n < 1/K$  bzw.  $n > K$ .

Eine wichtige Konsequenz von A3 ist, dass die rationalen Zahlen auf der Zahlengerade  $\mathbb{R}$  dicht verteilt sind. Um das zu formulieren führen wir folgende Bezeichnungen für Intervalle mit Grenzen  $a, b \in \mathbb{R}$  ein:

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

**Satz 2.6 ( $\mathbb{Q}$  ist dicht in  $\mathbb{R}$ )** Für alle  $a, b \in \mathbb{R}$ ,  $a < b$ , gibt es ein  $q \in \mathbb{Q}$  mit  $q \in (a, b)$ .

BEWEIS: Nach A3 gibt es ein  $n \in \mathbb{N}$  mit  $1/n < b - a$ . Die Zahlen  $k/n$ ,  $k \in \mathbb{Z}$ , sind rational und bilden ein Gitter, mit Gitterlänge kleiner als  $b - a$ . Wir zeigen dass im Intervall  $(a, b)$  ein Gitterpunkt liegt. Betrachte dazu

$$M = \{k \in \mathbb{Z} : k/n < b\}.$$

$M$  ist nicht leer, denn nach Anmerkung A3 gibt es ein  $k' \in \mathbb{N}$  mit  $k' > -nb$ , also  $(-k')/n < b$ . Weiter ist  $M$  nach oben beschränkt, denn es gilt  $k < nb$  für  $k \in M$ . Sei nun  $m \in \mathbb{Z}$  die größte Zahl in  $M$ , also  $m/n < b$  aber  $(m + 1)/n \geq b$ . Es folgt  $m/n \in (a, b)$ , denn

$$a < b - \frac{1}{n} \leq \frac{m + 1}{n} - \frac{1}{n} = \frac{m}{n} < b.$$

□

Im Beweis wurde folgende Tatsache benutzt.

**Satz 2.7 (Prinzip des größten Sünders)** Sei  $M \subset \mathbb{Z}$  nichtleer und nach oben beschränkt. Dann hat  $M$  ein größtes Element.

BEWEIS: Da  $M \neq \emptyset$  gibt es ein  $k_0 \in M$ . Sei  $b \in \mathbb{R}$  obere Schranke von  $M$ . Nach Axiom A3 gibt es ein  $k_1 \in \mathbb{N}$  mit  $k_1 > b$ . In der Menge  $\{k_0, \dots, k_1\}$  sei nun  $k$  maximal mit  $k \in M$ . Dann ist  $k$  ein größtes Element von ganz  $M$ . Denn ist  $k' \in M$  mit  $k' > k$ , so ist  $k'$  kein Element von  $\{k_0, \dots, k_1\}$  und es folgt  $k' > k_1 > b$ , Widerspruch zu  $b$  obere Schranke von  $M$ .  $\square$



### 3 Vollständige Induktion

Am Amazonas gibt es ein sehr kleines indigenes Volk, die Pirahã, die nur die Zählbegriffe *eins*, *wenige* und *vielen* kennen. Im Gegensatz dazu haben wir die natürlichen Zahlen  $1, 2 = 1 + 1, 3 = 2 + 1, \dots$  zur Verfügung. Laut Kronecker<sup>7</sup> sind diese von Gott gemacht, alles andere ist Menschenwerk. Die Menge der natürlichen Zahlen wird mit  $\mathbb{N}$  bezeichnet. Manche Autoren nehmen die Null hinzu, ich finde das aber nicht so gut.

Die natürlichen Zahlen ergeben sich sukzessive durch Addition der 1. Auf dieser Tatsache beruht die Idee der vollständigen Induktion.

**Induktionsprinzip:** Sei  $M \subset \mathbb{N}$  eine Menge mit den beiden Eigenschaften

- (1)  $1 \in M$ ,
- (2)  $n \in M \Rightarrow n + 1 \in M$ .

Dann gilt schon  $M = \mathbb{N}$ .

Die Aussage in (2) sagt nicht, *dass*  $n$  ein Element von  $M$  ist. Es geht um die Implikation: *wenn*  $n$  ein Element von  $M$  ist, so ist auch  $n + 1$  in  $M$ .

Mit der Notation  $1, 2, 3, \dots$  sind die natürlichen Zahlen nicht definiert, da die Bedeutung der Punkte nicht erklärt wird, es wird nur an die Anschauung appelliert. Das Induktionsprinzip lässt sich so nicht beweisen. Wir könnten  $\mathbb{N}$  als geeignete Teilmenge von  $\mathbb{R}$  definieren und das Induktionsprinzip dann aus den Axiomen von Kapitel 1 herleiten, so wird es im Buch von Hildebrandt<sup>8</sup> gemacht. Aber nach meinem Empfinden wäre das kein schöner Weg, der Begriff des Zählens ist grundlegender als das Konzept der reellen Zahlen, die natürlichen Zahlen sollten unabhängig von  $\mathbb{R}$  erklärt werden. Eine solche *Axiomatik des Zählens* wurde von G. Peano 1889 entwickelt, sie gehört wie die Konstruktionen von Cantor und Dedekind zum abstrakten Fundament der Analysis und wird in der Vorlesung Lineare Algebra genauer diskutiert. Diese Grundlagen sind beruhigend, sie sichern die Existenz der Objekte ab, aber aktiv gebraucht werden sie von uns nicht. Wir verwenden hier das Induktionsprinzip als Ausgangspunkt, es hat folgende Umformulierung.

**Satz 3.1 (Beweisverfahren der vollständigen Induktion)** *Gegeben sei eine Folge von Aussagen  $A(n)$  für  $n \in \mathbb{N}$ . Es möge gelten:*

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

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

BEWEIS: Wir betrachten die Menge  $M = \{n \in \mathbb{N} : A(n) \text{ ist wahr}\}$ . Nach Voraussetzung gilt  $1 \in M$ , und mit  $n \in M$  ist auch  $n + 1 \in M$ . Das Induktionsprinzip ergibt  $M = \mathbb{N}$ , das heißt alle Aussagen  $A(n)$  sind wahr.  $\square$

---

<sup>7</sup>Leopold Kronecker, 1823-1891

<sup>8</sup>Stefan Hildebrandt: Analysis 1, Springer-Verlag 2002

Man bezeichnet (1) als Induktionsanfang und (2) als Induktionsschluss. In den Naturwissenschaften bedeutet *Induktion* den Schluss vom Einzelfall auf ein allgemeines Gesetz. Das kann richtig oder falsch sein, anders als bei unserem Beweisverfahren.

**Beispiel 3.1 (arithmetische Summe)** Wir zeigen die Summenformel

$$A(n) : \quad 1 + 2 + \dots + n = \frac{n(n+1)}{2}.$$

Für  $n = 1$  sind linke und rechte Seite gleich Eins, also gilt der Induktionsanfang. Für den Induktionsschluss starten wir mit der linken Seite von  $A(n+1)$ , und verwenden dann  $A(n)$ :

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

Wir sind wie gewünscht bei der rechten Seite von  $A(n+1)$  gelandet, so dass nach Induktionsprinzip  $A(n)$  für alle  $n \in \mathbb{N}$  bewiesen ist. Übrigens läuft diese Formel unter dem Stichwort *der kleine Gauß*. Die Schüler in der Klasse von Gauß waren mal wieder unartig, als Strafe sollten sie die Zahlen von Eins bis Hundert addieren. Der Lehrer hatte wohl gehofft eine Zeitlang seine Ruhe zu haben. Doch Gauß fand die Antwort 5050 sofort. Sein Trick war, die Zahlen nochmal umgekehrt aufzuschreiben und zu addieren:

$$\begin{array}{cccccc} 1 & 2 & \dots & \dots & n-1 & n \\ n & n-1 & \dots & \dots & 2 & 1 \\ n+1 & n+1 & \dots & \dots & n+1 & n+1 \end{array}$$

Die Stärke des Arguments von Gauß liegt darin, dass es die richtige Formel liefert, während diese für den Induktionsbeweis schon bekannt sein oder geraten werden muss. Auf der anderen Seite sind Induktionsbeweise, wenn die Behauptung vorliegt, oft relativ einfach.

Bevor wir weitere Beispiele betrachten, wollen wir für Summen und Produkte eine bessere Notation einführen. Für durchnummerierte Zahlen  $a_m, \dots, a_n$  setzen wir

$$\sum_{i=m}^n a_i = a_m + \dots + a_n.$$

Der Index  $i$  durchläuft dabei aufsteigend die ganzen Zahlen von der unteren Grenze  $m$  bis zur oberen Grenze  $n$ . Es ist manchmal praktisch den Fall  $n < m$  zuzulassen, das heißt es gibt keine Summanden. Eine solche *leere Summe* bekommt per Definition den Wert Null. Im Fall eines Summanden, also  $n = m$ , ist

$$\sum_{i=m}^m a_i = a_m.$$

Dies ist der Start  $n = m$  einer induktiven Definition der Summe, für  $n > m$  setzt man

$$\sum_{i=m}^n a_i = \left( \sum_{i=m}^{n-1} a_i \right) + a_n.$$

Der Laufindex  $i$  ist nur ein Parameter, die Zahlen können auch durch einen anderen Index nummeriert werden. Dann sind aber die Grenzen anzupassen. Wollen wir zum Beispiel  $i = j + 1$  substituieren und durchläuft  $i$  die Werte  $m, \dots, n$ , so muss  $j$  die Werte  $m - 1, \dots, n - 1$  annehmen, es gilt also

$$\sum_{i=m}^n a_i = \sum_{j=m-1}^{n-1} a_{j+1}.$$

Das Produkt kann ganz analog induktiv erklärt werden, es gilt mit  $n > m$

$$\prod_{i=m}^m a_i = a_m \quad \text{und} \quad \prod_{i=m}^n a_i = \left( \prod_{i=m}^{n-1} a_i \right) \cdot a_n.$$

Das leere Produkt wird gleich Eins gesetzt. Die untere Grenze von Summen und Produkten ist oft  $m = 1$  oder  $m = 0$ . Um die Notation zu üben, könnten Sie den Beweis der Formel für die arithmetische Summe mit dem Summenzeichen aufschreiben.

**Beispiel 3.2 (geometrische Summe)** Sei  $x \in \mathbb{R}$ ,  $x \neq 1$ . Dann gilt für alle  $n \in \mathbb{N}_0$

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

Dabei ist  $x^k = x \cdot \dots \cdot x$  das Produkt mit  $k$  Faktoren  $x$ . Für  $k = 0$  ist es das Produkt mit null Faktoren, das leere Produkt, es gilt also  $x^0 = 1$ . Das trifft auch im Extremfall  $x = 0$  zu. Für  $x \neq 0$  lässt sich diese Definition mit den Potenzgesetzen motivieren, man rechnet  $1 = x^1 \cdot x^{-1} = x^{1-1} = x^0$ . Ein Beweis ist das auch nicht, wir nehmen  $x^0 = 1$  als Definition. Wir werden uns später mit Potenzen noch ausführlich beschäftigen. Nun zeigen wir die Behauptung oben mit Induktion. Für  $n = 0$  gilt

$$\sum_{k=0}^0 x^k = x^0 = 1 = \frac{1 - x^{0+1}}{1 - x}.$$

Jetzt gelte die Formel für ein  $n \in \mathbb{N}_0$ . Dann folgt

$$\sum_{k=0}^{n+1} x^k = \left( \sum_{k=0}^n x^k \right) + x^{n+1} = \frac{1 - x^{n+1}}{1 - x} + x^{n+1} = \frac{1 - x^{(n+1)+1}}{1 - x}.$$

Damit ist der Induktionsschluss verifiziert, die Formel ist bewiesen. Der Beweis ist wieder relativ einfach, die Sache läuft halt nach einem festen Schema ab. Wie kann man aber die Formel finden? Dazu gibt es einen schönen Trick, die *Teleskopsumme*:

$$(1 - x) \sum_{k=0}^n x^k = \sum_{k=0}^n x^k - \sum_{k=0}^n x^{k+1} = \sum_{k=0}^n x^k - \sum_{k=1}^{n+1} x^k = 1 - x^{n+1}.$$

Im dritten Beispiel wollen wir die Potenzen  $(1 + x)^n$ ,  $n \in \mathbb{N}$ , nach unten abschätzen. Für kleine  $n$  sieht das so aus:

$$\begin{aligned} (1 + x)^1 &= 1 + x \\ (1 + x)^2 &= 1 + 2x + x^2 \\ (1 + x)^3 &= 1 + 3x + 3x^2 + x^3. \end{aligned}$$

Um allgemein eine Formel zu finden, multiplizieren wir in Gedanken aus. Nehmen wir in jeder Klammer die 1 als Faktor, so ergibt das eine 1. Wählen wir in einer Klammer  $x$  und sonst immer die 1, so liefert das ein  $x$ . Da es  $n$  Klammern gibt, bekommen wir so insgesamt  $nx$ . Es gilt also

$$(1+x)^n = 1 + nx + \text{Terme mit höheren } x\text{-Potenzen.}$$

Im Fall  $x \geq 0$  sind die restlichen Terme alle nichtnegativ, also ist  $(1+x)^n \geq 1 + nx$ . Für  $x < 0$  ist das Vorzeichen dieser Terme dagegen unklar. Wir können aber folgendes zeigen.

**Satz 3.2 (Bernoullische Ungleichung)** <sup>9</sup> Für  $x \in \mathbb{R}$ ,  $x \geq -1$ , und  $n \in \mathbb{N}$  gilt

$$(1+x)^n \geq 1 + nx.$$

BEWEIS: Für  $n = 1$  gilt  $(1+x)^1 = 1 + 1x$ . Wegen  $1+x \geq 0$  folgt weiter

$$\begin{aligned} (1+x)^{n+1} &= (1+x) \cdot (1+x)^n \\ &\geq (1+x) \cdot (1+nx) \quad (\text{nach Induktionsannahme}) \\ &= 1 + (n+1)x + nx^2 \\ &\geq 1 + (n+1)x. \end{aligned}$$

□

Als nächstes geht es um die Zahl der Elemente gewisser Mengen. Das Abzählen gehört zu den Dingen die jedem klar sind, nur die Mathematiker machen es kompliziert <sup>10</sup>.

**Definition 3.1 (Zahl der Elemente)** Eine Menge  $M$  hat  $n \in \mathbb{N}$  Elemente (Notation:  $\#M = n$ ) wenn eine bijektive Abbildung  $\{1, \dots, n\} \xrightarrow{\sim} M$  existiert.  $M$  heißt endlich, wenn entweder  $\#M = n$  für ein  $n \in \mathbb{N}$  oder  $M = \emptyset$ , wir setzen  $\#\emptyset = 0$ . Eine Menge heißt unendlich, wenn sie nicht endlich ist.

Bei Nachdenken tut sich noch folgende Frage auf: ist die Anzahl eindeutig definiert, oder könnte sowohl  $\#M = m$  als auch  $\#M = n$  gelten mit  $n \neq m$ ? Dann gäbe es Bijektionen sowohl zu  $\{1, \dots, m\}$  als auch zu  $\{1, \dots, n\}$ . Wir hätten dann eine Bijektion

$$\{1, \dots, m\} \xrightarrow{\sim} M \xrightarrow{\sim} \{1, \dots, n\} \quad \text{mit } n \neq m.$$

Die Abbildung ist in beiden Richtungen injektiv. Der folgende Satz liefert  $m \leq n$  und  $n \leq m$ , und damit  $n = m$ . Die Anzahl ist also eindeutig definiert.

**Satz 3.3 (Schubfachprinzip)** <sup>11</sup> Für alle  $n \in \mathbb{N}$  gilt: ist  $f : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$  injektiv für ein  $m \in \mathbb{N}$ , so ist  $m \leq n$ .

BEWEIS: Durch Induktion über  $n \in \mathbb{N}$ , jeweils für alle  $m \in \mathbb{N}$ . Für  $n = 1$  haben wir  $f(m) = 1 = f(1)$ , also  $m = 1$  wegen  $f$  injektiv. Für den Induktionsschluss sei nun

$$f : \{1, \dots, m\} \rightarrow \{1, \dots, n+1\} \quad \text{injektiv.}$$

<sup>9</sup>Jakob Bernoulli, Basel 1689

<sup>10</sup>Mathematiker sind eine Art Franzosen: redet man zu ihnen, so übersetzen sie es in ihre Sprache, und dann ist es alsobald etwas ganz anderes (Goethe).

<sup>11</sup>Peter Gustav Lejeune Dirichlet, 1834

Zu zeigen ist  $m \leq n + 1$ , das ist klar für  $m = 1$ . Für  $m \geq 2$  suchen wir eine Injektion

$$\tilde{f} : \{1, \dots, m-1\} \rightarrow \{1, \dots, n\}, k \mapsto \tilde{f}(k),$$

dann folgt nach Induktion  $m-1 \leq n$ , also  $m \leq n+1$ , der Satz ist bewiesen. Eine naheliegende Wahl ist  $\tilde{f}(k) = f(k)$  für  $k = 1, \dots, m-1$ ; das geht gut falls  $f(k) \in \{1, \dots, n\}$  für alle  $k = 1, \dots, m-1$ . Andernfalls ist  $f(i) = n+1$  für genau ein  $i \in \{1, \dots, m-1\}$ . Aber dann ist  $f(m) \neq n+1$  da  $f$  injektiv, als Bildpunkt von  $i$  können wir also  $f(m) \in \{1, \dots, n\}$  nehmen:

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

Die  $f(1), \dots, f(m)$  sind paarweise verschieden, somit ist  $\tilde{f}$  injektiv. □

Das Schubfachprinzip macht eine Aussage über unsere Wäschekommode: ist in jeder Schublade höchstens eine Socke so haben wir nicht mehr Socken als Schubladen. Auch richtig ist: ist in jeder Schublade mindestens eine Socke, so haben wir mindestens soviel Socken wie Schubladen. Wir drücken das noch mathematisch aus.

**Folgerung 3.1** *Ist  $f : \{1, \dots, m\} \rightarrow \{1, \dots, n\}$  surjektiv für  $m, n \in \mathbb{N}$ , so folgt  $m \geq n$ .*

BEWEIS: Wähle zu jedem  $k \in \{1, \dots, n\}$  ein Urbild  $i_k \in \{1, \dots, m\}$ , also  $f(i_k) = k$ . Zum Beispiel kann man das kleinste Urbild wählen. Die Abbildung

$$g : \{1, \dots, n\} \rightarrow \{1, \dots, m\}, g(k) = i_k,$$

erfüllt  $f(g(k)) = k$ , also ist  $g$  injektiv. Aus Satz 3.3 folgt  $n \leq m$ . □

Wir wollen nun für Mengen von Interesse die Anzahl der Elemente bestimmen. Als erstes geht es um die Menge  $S_n$  aller bijektiven Abbildungen  $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ . Die Ziffern  $\sigma_1 = \sigma(1), \dots, \sigma_n = \sigma(n)$  sind eine Umordnung (oder Permutation) von  $1, \dots, n$ , wir identifizieren jedes  $\sigma$  mit dem Tupel  $(\sigma_1, \dots, \sigma_n)$ .

**Satz 3.4 (Zahl der Permutationen)** *Es gilt  $\#S_n = n! = 1 \cdot 2 \cdot \dots \cdot n$ .*

BEWEIS: Die Behauptung ist klar für  $n = 1$ . In  $S_{n+1}$  betrachten wir die Teilmengen  $S_{n+1,k}$ , bei denen  $n+1$  an der Stelle  $k$  steht:

$$S_{n+1,k} = \{(\tau_1, \dots, \tau_{n+1}) \in S_{n+1} : \tau_k = n+1\} \quad \text{für } k = 1, \dots, n+1.$$

Das ist wie in der Familie, Papa sitzt an der Stelle  $k$  und die anderen müssen sich drumherum sortieren. Für jedes  $k$  haben wir die bijektive Abbildung

$$S_n \rightarrow S_{n+1,k}, \quad (\sigma_1, \dots, \sigma_n) \mapsto (\sigma_1, \dots, \sigma_{k-1}, n+1, \sigma_k, \dots, \sigma_n).$$

Zum Beispiel ist  $S_2$  bijektiv zu  $S_{3,2}$ , mit  $(1, 2) \mapsto (1, 3, 2)$  und  $(2, 1) \mapsto (2, 3, 1)$ . Allgemein ist die Umkehrabbildung gegeben durch

$$S_{n+1,k} \rightarrow S_n, \quad (\tau_1, \dots, \tau_{n+1}) \mapsto (\tau_1, \dots, \tau_{k-1}, \tau_{k+1}, \dots, \tau_{n+1}).$$

Es folgt  $\#S_{n+1,k} = \#S_n$ . Aber  $S_{n+1}$  ist die Vereinigung der  $S_{n+1,k}$ , und je zwei dieser Mengen haben keine gemeinsamen Elemente. Also folgt durch Induktion

$$\#S_{n+1} = \sum_{k=1}^{n+1} \#S_{n+1,k} = \sum_{k=1}^{n+1} \#S_n = \sum_{k=1}^{n+1} n! = (n+1) \cdot n! = (n+1)!.$$



**Satz 3.5 (Zahl der Kombinationen)** Sei  $n \in \mathbb{N}_0$  und  $k \in \{0, 1, \dots, n\}$ . Dann ist die Anzahl  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$  gleich  $\binom{n}{k}$ .

BEWEIS: Wir erledigen zunächst den Fall  $k = 0$ . Es gibt genau eine null-elementige Teilmenge von  $\{1, \dots, n\}$ , nämlich die leere Menge. Mit  $\binom{n}{0} = 1$  stimmt die Behauptung für  $k = 0$ , auch im Sonderfall  $n = 0$ . Jetzt führen wir Induktion über folgende Aussage  $A(n)$ :

Für alle  $k = 0, 1, \dots, n$  ist die Zahl der  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$  gleich  $\binom{n}{k}$ .

Wir oben festgestellt gilt  $A(0)$ . Für den Induktionsschluss müssen wir die Anzahl der  $k$ -elementigen Teilmengen von  $\{1, \dots, n+1\}$  bestimmen, wobei  $n \in \mathbb{N}_0$  und  $k \in \{1, \dots, n+1\}$ . Diese zerfallen in zwei disjunkte Klassen:

**Klasse 1:** Die Menge enthält die Nummer  $n+1$  nicht.

**Klasse 2:** Die Menge enthält die Nummer  $n+1$ .

Klasse 1 besteht aus den  $k$ -elementigen Teilmengen von  $\{1, \dots, n\}$ . Klasse 2 ergibt sich, indem wir zu jeder  $(k-1)$ -elementigen Teilmenge von  $\{1, \dots, n\}$  das Element  $n+1$  dazunehmen. Induktiv ist die Zahl der  $k$ -elementigen Teilmengen von  $\{1, \dots, n+1\}$  also gleich

$$\# \text{Klasse 1} + \# \text{Klasse 2} = \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

Zum Schluss haben wir Lemma 3.1 benutzt. Das Argument gilt auch im Extremfall  $k = n+1$ : es gibt keine  $(n+1)$ -elementigen Teilmengen von  $\{1, \dots, n\}$ , und nach Definition ist  $\binom{n}{n+1} = 0$ . Der Satz bewiesen ist.  $\square$

Als Anwendung erhalten wir die allgemeine Binomische Formel.

**Satz 3.6 (Binomische Formel)** Für  $a, b \in \mathbb{R}$  und  $n \in \mathbb{N}$  gilt

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (3.3)$$

Das Produkt  $(a+b)^n$  besteht aus  $n$  Faktoren. Beim Ausmultiplizieren ergeben sich Terme der Form  $a^k b^{n-k}$  mit  $k = 0, 1, \dots, n$ . Um die Potenz  $a^k$  zu erhalten, müssen wir in  $k$  Klammern den Faktor  $a$  wählen, in den übrigen Klammern  $b$ . Die Zahl dieser Terme ist also gleich der Anzahl der Möglichkeiten, aus den insgesamt  $n$  Klammern  $k$  Stück für  $a$  zu reservieren, also gleich dem Binomialkoeffizienten  $\binom{n}{k}$ . Diese Begründung ist sehr schlüssig, allerdings haben wir das Ausmultiplizieren nicht wirklich durchgeführt. Wir geben alternativ einen Induktionsbeweis.

BEWEIS: Wir zeigen als erstes durch Induktion, dass eine Darstellung

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

mit geeigneten Koeffizienten  $c(n, k) \in \mathbb{R}$  gilt. Für  $n = 1$  ist das klar, wir wählen  $c(1, 0) = c(1, 1) = 1$ . Sei nun (3.4) gezeigt für ein  $n \in \mathbb{N}$ . Dann folgt

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n \\
&= \sum_{k=0}^n c(n, k) a^{k+1} b^{n-k} + \sum_{k=0}^n c(n, k) a^k b^{n-k+1} \\
&= c(n, n) a^{n+1} + \sum_{k=0}^{n-1} c(n, k) a^{k+1} b^{n-k} + \sum_{k=1}^n c(n, k) a^k b^{n+1-k} + c(n, 0) b^{n+1} \\
&= c(n, n) a^{n+1} + \sum_{k=1}^n (c(n, k) + c(n, k-1)) a^k b^{n+1-k} + c(n, 0) b^{n+1}.
\end{aligned}$$

Die rechte Seite hat die Form  $\sum_{k=0}^{n+1} c(n+1, k) a^k b^{n+1-k}$ , und zwar mit

$$c(n+1, k) = \begin{cases} c(n, k) + c(n, k-1) & \text{für } k = 1, \dots, n \\ c(n, n) & \text{für } k = n+1, \\ c(n, 0) & \text{für } k = 0. \end{cases}$$

Jetzt zeigen wir induktiv  $c(n, k) = \binom{n}{k}$ . Für  $n = 1$  und  $k = 0, 1$  stimmt das, und induktiv folgt aus der Rekursionsformel, siehe Lemma 3.1,

$$c(n+1, k) = \begin{cases} \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} & \text{für } k = 1, \dots, n \\ \binom{n}{n} = 1 = \binom{n+1}{n+1} & \text{für } k = n+1, \\ \binom{n}{0} = 1 = \binom{n+1}{0} & \text{für } k = 0. \end{cases}$$

□

## 4 Grenzwerte von Folgen

Wir definieren hier den Begriff des Grenzwerts im Fall von Folgen reeller Zahlen. Die zugrundeliegende Idee kann man auf Archimedes zurückführen, der unter anderem den Flächeninhalt des Kreises berechnet hat. Er hat auch das Volumen der Kugel bestimmt, darauf war er so stolz dass er es auf seinem Grabstein hat verewigen lassen; leider ist das Grab aber verschollen. Sein Ansatz zum Konzept der Konvergenz wurde unter anderem von Weierstraß<sup>12</sup> aufgegriffen und präzisiert.

Eine Folge ist genau gesagt eine Abbildung

$$\mathbb{N} \longrightarrow \mathbb{R}, \quad n \longmapsto a_n.$$

Die Zahlen sind also nummeriert,  $n$  heißt die Nummer und  $a_n$  das  $n$ -te Glied der Folge. Oft werden Folgen durch ein Bildungsgesetz oder durch Aufzählen der ersten Glieder angegeben. Zum Beispiel beschreibt  $a_n = n^2$  oder  $a_n = 1, 4, 9, \dots$  die Folge der Quadratzahlen. Die Folge  $a_n = 1, 1, \dots$  ist die konstante Folge mit Wert 1. Wenn es um eine Folge als Ganzes geht, ist die Notation  $(a_n)_{n \in \mathbb{N}}$  üblich, manchmal auch kurz  $(a_n)$ .

**Definition 4.1 (Konvergenz)**  $(a_n)_{n \in \mathbb{N}}$  konvergiert gegen  $a \in \mathbb{R}$  mit  $n \rightarrow \infty$ , falls gilt:

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

Die Zahl  $a$  heißt Grenzwert der Folge und wir schreiben

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

Eine Folge  $(a_n)_{n \in \mathbb{N}}$  heißt konvergent, wenn es ein  $a \in \mathbb{R}$  gibt das Grenzwert der Folge ist; andernfalls heißt die Folge divergent.

Oft wird eine verbale Fassung der Konvergenz gegeben, etwa: *die  $a_n$  weichen von  $a$  beliebig wenig ab, wenn  $n$  hinreichend groß ist.* Auch wird eine dynamische Vorstellung des Grenzwerts impliziert, die Folge  $a_n$  *strebt gegen  $a$*  oder *ist asymptotisch zu  $a$ .* Ich persönlich finde unsere Fassung klarer. Die Zahl  $\varepsilon > 0$  gibt vor, um wieviel  $a_n$  von  $a$  höchstens abweichen soll. Für die Konvergenz muss man ein  $K$  finden, so dass dies erfüllt ist für alle  $n > K$ . Und das muss man schaffen egal wie klein  $\varepsilon > 0$  gewählt wird. Der Vorteil dieser quantitativen Beschreibung liegt darin, dass wir sie in Zukunft in mathematischen Argumenten benutzen können. Mit einer verbalen Umschreibung würden wir ins Schwimmen kommen.

**Beispiel 4.1 (Harmonische Folge)** Die Folge  $a_n = 1/n$  konvergiert gegen  $a = 0$ . Denn zu gegebenem  $\varepsilon > 0$  wählen wir  $K = 1/\varepsilon$ , und es folgt für alle  $n > K$

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

Eine geringfügige Modifikation von Definition 4.1 lautet wie folgt:

$$\text{zu jedem } \varepsilon > 0 \text{ gibt es ein } N \in \mathbb{N}, \text{ so dass für alle } n \geq N \text{ gilt: } |a_n - a| < \varepsilon. \quad (4.1)$$

Diese Version folgt aus Definition 4.1: angenommen wir haben dort zu  $\varepsilon > 0$  ein  $K$  gefunden. Dann wählen wir  $N > K$  als die nächst größere natürliche Zahl. Für  $n \geq N$  ist  $n > K$  und es

---

<sup>12</sup>Karl Weierstraß, 1815-1897

folgt  $|a_n - a| < \varepsilon$ . Umgekehrt: haben wir zu  $\varepsilon > 0$  ein  $N$  wie in (4.1) gefunden, so ist offenbar  $|a_n - a| < \varepsilon$  für  $n > N$ , das heißt Definition 4.1 gilt mit  $K = N$ . Was ist der Vorteil unserer Definition? Um Konvergenz zu beweisen muss man die Schranke  $K$  bzw.  $N$  berechnen. Die Rechnung führt in der Regel auf eine reelle Zahl, es ist dann praktisch dass  $K$  in  $\mathbb{R}$  gewählt werden kann. Für (4.1) müssten wir jedes Mal zu der natürlichen Zahl  $N$  übergehen, das bleibt uns erspart. Zum Beispiel konnten wir für die harmonische Folge  $K = 1/\varepsilon$  nehmen.

**Beispiel 4.2 (Konstante Folge)** Ist  $a_n = a$  für alle  $n \in \mathbb{N}$ , so folgt  $\lim_{n \rightarrow \infty} a_n = a$ . Denn für  $\varepsilon > 0$  gilt  $|a_n - a| = 0 < \varepsilon$  für alle  $n > 0$ , also können wir immer  $K = 0$  wählen.

Hier kann zu jedem  $\varepsilon > 0$  das gleiche  $K$  gewählt werden, nämlich  $K = 0$ ; aber das ist ein Sonderfall. Im allgemeinen gilt: je kleiner die Fehlerschranke  $\varepsilon > 0$  ist, desto später wird sie in der Regel erfüllt, desto größer muss somit  $K$  gewählt werden. Wir betrachten das für die Folge  $a_n = 1/n$ , und zwar im Fall  $\varepsilon = 1/N$  mit  $N \in \mathbb{N}$ . Wie oben gezeigt kann  $K = 1/\varepsilon$  gewählt werden. Ein kleineres  $K$  tut es nicht, denn dafür hat man

$$N = 1/\varepsilon > K \quad \text{aber} \quad |a_N - 0| = 1/N = \varepsilon.$$

Somit ist hier  $K = 1/\varepsilon$  die optimale, sprich kleinste mögliche Wahl, und dieses  $K$  wird groß für  $\varepsilon > 0$  klein. Manchmal wird das durch die Notation  $K = K(\varepsilon)$  betont. Im allgemeinen wäre es kompliziert, das optimale  $K$  zu berechnen. Aber das ist in der Definition auch gar nicht verlangt, wir brauchen nur irgendein  $K$ . Es reicht daher aus, die Differenz  $|a_n - a|$  so abzuschätzen, dass eine geeignete Wahl von  $K$  ersichtlich wird. Das vereinfacht die Sache erheblich, hier ein Beispiel.

**Beispiel 4.3 (Geometrische Folge)** Für  $q \in \mathbb{R}$  mit  $|q| < 1$  gilt  $\lim_{n \rightarrow \infty} q^n = 0$ . Das ist klar wenn  $q = 0$ . Für  $q \neq 0$  wollen wir die Bernoulli-Ungleichung, Satz 3.2, verwenden. Wir schreiben  $|q| = 1/(1+x)$  mit  $x > 0$ , und erhalten

$$|q^n - 0| = |q|^n = \frac{1}{(1+x)^n} \leq \frac{1}{1+nx} \leq \frac{1}{nx} < \varepsilon \quad \text{für alle } n > 1/(\varepsilon x).$$

Wir können also  $K = 1/(\varepsilon x)$  wählen.

Hier mal ein Beispiel einer Folge, die nicht konvergiert.

**Beispiel 4.4** Die Folge  $a_n = (-1)^n$  ist nicht konvergent. Und zwar hat sie abwechselnd die Werte  $\pm 1$ , insbesondere ist  $a_n - a_{n+1} = (-1)^n(1 - (-1)) = \pm 2$ . Angenommen es gilt  $\lim_{n \rightarrow \infty} a_n = a$  für ein  $a \in \mathbb{R}$ . Dann gibt es zu jedem  $\varepsilon > 0$  ein  $K \in \mathbb{R}$  mit  $|a_n - a| < \varepsilon$  für alle  $n > K$ . Es folgt dann für  $n > K$

$$2 = |a_n - a_{n+1}| = |a_n - a + a - a_{n+1}| \leq |a_n - a| + |a - a_{n+1}| < 2\varepsilon.$$

Das ist ein Widerspruch zum Beispiel für  $\varepsilon = 1$ .

**Satz 4.1 (Eindeutigkeit des Grenzwerts)** Falls die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergent ist, so ist ihr Grenzwert eindeutig bestimmt.

BEWEIS: Seien  $a, a' \in \mathbb{R}$  Grenzwerte der Folge. Durch Einschieben von  $a_n$  folgt

$$|a - a'| = |a - a_n + a_n - a'| \leq |a_n - a| + |a_n - a'|.$$

Nach Definition der Konvergenz gibt es zu jedem  $\varepsilon > 0$  Schranken  $K, K' \in \mathbb{R}$  mit

$$|a_n - a| < \frac{\varepsilon}{2} \text{ für alle } n > K \quad \text{ sowie } \quad |a_n - a'| < \frac{\varepsilon}{2} \text{ für alle } n > K'.$$

Für  $n > \max(K, K')$  folgt  $|a - a'| < \varepsilon$ . Da  $\varepsilon > 0$  beliebig bedeutet das  $a = a'$ .  $\square$

Hier wurde eine banale, aber wichtige Tatsache benutzt: ist  $K' \geq K$  und gilt die Abschätzung  $|a_n - a| < \varepsilon$  für alle  $n > K$ , so gilt sie erst recht für alle  $n > K'$ . Mit anderen Worten: ist für ein  $\varepsilon > 0$  ein geeignetes  $K$  gefunden, so kann dieses immer vergrößert werden, es bleibt dabei geeignet. Dadurch ist es zum Beispiel hier möglich, zu  $\max(K, K')$  überzugehen.

Die Definition der Konvergenz ist auch in folgendem Punkt flexibel: angenommen wir haben für  $\varepsilon > 0$  nur eine Abschätzung  $|a_n - a| < 3\varepsilon$  für  $n > K(\varepsilon)$  geschafft. Das reicht schon für die Konvergenz, denn für  $\varepsilon' > 0$  wenden wir die Information an mit  $\varepsilon = \varepsilon'/3$  und erhalten  $|a_n - a| < 3\varepsilon = \varepsilon'$  für  $n > K(\varepsilon) = K(\varepsilon'/3)$ . Natürlich geht das statt mit  $3\varepsilon$  auch mit  $10\varepsilon$  oder mit  $C\varepsilon$  für eine Konstante  $C$ , auch die Abschätzung  $|a_n - a| \leq \varepsilon$  wäre ausreichend.

Durch den Begriff der  $\varepsilon$ -Umgebung gewinnt der Grenzwertbegriff an Anschaulichkeit.

**Definition 4.2 ( $\varepsilon$ -Umgebung)** Die  $\varepsilon$ -Umgebung von  $a \in \mathbb{R}$  ist die Menge

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

Mit diesem Begriff können wir die Konvergenz einer Folge auch so ausdrücken: zu jedem  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{R}$  so dass  $a_n \in B_\varepsilon(a)$  für alle  $n > K$ . Das ist schlicht und einfach äquivalent, aber eventuell doch anschaulicher. Die Bezeichnung  $B_\varepsilon(a)$  kommt hier aus dem Englischen: im  $\mathbb{R}^3$  ist die Menge der Punkte, die zu einem festen Punkt  $a$  Abstand kleiner als  $\varepsilon > 0$  haben, eine Vollkugel (ball) mit Radius  $\varepsilon$  um  $a$ .

Wir wollen die Eindeutigkeit des Grenzwerts nochmal mithilfe des Umgebungsbegriffs anschauen. Ist  $\varepsilon > 0$  mit  $B_\varepsilon(a) \cap B_\varepsilon(a') \neq \emptyset$ , so folgt mit  $x \in B_\varepsilon(a) \cap B_\varepsilon(a')$

$$|a - a'| = |a - x + x - a'| \leq |x - a| + |x - a'| < 2\varepsilon.$$

Das impliziert im Umkehrschluss

$$B_\varepsilon(a) \cap B_\varepsilon(a') = \emptyset \quad \text{für } \varepsilon \leq \frac{1}{2}|a - a'|.$$

Hieraus folgt aber die Eindeutigkeit des Grenzwerts, denn für  $\varepsilon \leq \frac{1}{2}|a - a'|$  kann  $a_n$  nicht in beiden Umgebungen  $B_\varepsilon(a)$  und  $B_\varepsilon(a')$  liegen. Die Zahl  $\frac{1}{2}|a - a'|$  ist übrigens der Abstand von  $a, a'$  zum Mittelpunkt  $\frac{1}{2}(a + a')$  der Strecke von  $a$  nach  $a'$ .

**Definition 4.3 (Beschränktheit von Folgen)** Eine Folge  $(a_n)_{n \in \mathbb{N}}$  heißt

beschränkt                      wenn es ein  $C \geq 0$  gibt mit  $|a_n| \leq C$  für alle  $n \in \mathbb{N}$ ,  
nach oben beschränkt      wenn es ein  $M \in \mathbb{R}$  gibt mit  $a_n \leq M$  für alle  $n \in \mathbb{N}$ ,  
nach unten beschränkt      wenn es ein  $m \in \mathbb{R}$  gibt mit  $a_n \geq m$  für alle  $n \in \mathbb{N}$ .

Verwenden Sie  $|a_n| = \max(a_n, -a_n)$  um zu zeigen:  $(a_n)$  ist genau dann beschränkt, wenn  $(a_n)$  nach oben und unten beschränkt ist.

**Beispiel 4.5** Die Folge  $a_n = n$ ,  $n \in \mathbb{N}$ , ist nach unten beschränkt, zum Beispiel gilt  $a_n \geq 0$  für alle  $n$ . Sie ist aber nicht nach oben beschränkt: nach Archimedes gibt es zu jedem  $M \in \mathbb{R}$  ein  $n \in \mathbb{N}$  mit  $a_n = n > M$ .

**Satz 4.2 (konvergent  $\Rightarrow$  beschränkt)** Jede konvergente Folge ist beschränkt.

BEWEIS: Es gelte  $a_n \rightarrow a$  mit  $n \rightarrow \infty$ . Zu  $\varepsilon > 0$  gibt es dann ein  $N \in \mathbb{N}$  mit

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

Wir wählen  $\varepsilon = 1$ , also gilt  $|a_n| \leq |a| + 1$  für  $n > N$ . Mit  $C = \max(|a_1|, \dots, |a_N|, |a| + 1)$  folgt

$$|a_n| \leq C \quad \text{für alle } n \in \mathbb{N}.$$

□

Man kann dieses Argument so sehen: die Konvergenz ist eine asymptotische Eigenschaft der Folge. An ihr ändert sich nichts wenn wir endlich viele Folgenglieder manipulieren, hinzufügen oder weglassen. Und dasselbe gilt auch für die Beschränktheit der Folge. Zwar muss die Schranke eventuell angepasst werden, aber die Beschränktheit an sich ist auch eine asymptotische Eigenschaft.

**Satz 4.3 (Rechenregeln für Grenzwerte)** Es gelte  $a_n \rightarrow a, b_n \rightarrow b$  mit  $n \rightarrow \infty$ .

- a) Für alle  $\lambda, \mu \in \mathbb{R}$  ist  $(\lambda a_n + \mu b_n)_{n \in \mathbb{N}}$  konvergent mit Grenzwert  $\lim_{n \rightarrow \infty} (\lambda a_n + \mu b_n) = \lambda a + \mu b$ .
- b) Die Folge  $(a_n \cdot b_n)_{n \in \mathbb{N}}$  ist konvergent mit Grenzwert  $\lim_{n \rightarrow \infty} (a_n \cdot b_n) = a \cdot b$ .
- c) Falls  $b \neq 0$ , so gibt es ein  $K_0 \in \mathbb{R}$  mit  $b_n \neq 0$  für  $n > K_0$  und die Folge  $(a_n/b_n)_{n > K_0}$  ist konvergent mit Grenzwert  $\lim_{n \rightarrow \infty} a_n/b_n = a/b$ .

BEWEIS: Wir zeigen erst a) im Fall  $\lambda = \mu = 1$ . Zu  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{R}$  mit  $|a_n - a| < \varepsilon/2$  und  $|b_n - b| < \varepsilon/2$  für  $n > K$ . Es folgt für  $n > K$

$$|(a_n + b_n) - (a + b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Die volle Aussage in a) folgt aus diesem Spezialfall und b).

Um b) zu zeigen, schätzen wir erst für  $n \in \mathbb{N}$  ab:

$$|a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab| \leq |a_n| \cdot |b_n - b| + |a_n - a| \cdot |b|.$$

Nach Satz 4.2 gibt es ein  $C > 0$  mit  $|a_n| \leq C$  für alle  $n \in \mathbb{N}$ , und so dass auch  $|b| \leq C$ . Also

$$|a_n b_n - ab| \leq C(|a_n - a| + |b_n - b|) \quad \text{für alle } n \in \mathbb{N}.$$

Zu  $\varepsilon > 0$  wähle  $K \in \mathbb{R}$  mit  $|a_n - a|, |b_n - b| < \varepsilon/(2C)$  für  $n > K$ . Es folgt

$$|a_n b_n - ab| < C \left( \frac{\varepsilon}{2C} + \frac{\varepsilon}{2C} \right) = \varepsilon \quad \text{für } n > K.$$

In c) können wir  $a_n = 1$  annehmen, der allgemeine Fall folgt dann aus b) mit

$$\frac{a_n}{b_n} = a_n \cdot \frac{1}{b_n}.$$

Wir müssen erst das Problem loswerden, dass der Nenner Null sein könnte. Es gilt

$$|b_n| = |b - (b - b_n)| \geq |b| - |b_n - b|.$$

Zu  $\varepsilon = |b|/2 > 0$  gibt es ein  $K_0$  mit  $|b_n - b| < \varepsilon$  für  $n > K_0$ . Daraus folgt

$$|b_n| \geq |b| - |b_n - b| > |b| - \varepsilon = |b|/2 > 0.$$

Somit ist  $b_n \neq 0$  für  $n > K_0$ , und es gilt die Abschätzung

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b| |b_n|} \leq C |b_n - b| \quad \text{mit } C = \frac{2}{|b|^2}.$$

Zu jedem  $\varepsilon > 0$  gibt es ein  $K$  mit  $|b_n - b| < \varepsilon/C$  für  $n > K$ . Wir erhalten

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| \leq C |b_n - b| < \varepsilon \quad \text{für alle } n > \max(K_0, K).$$

□

An dieser Stelle gibt es eine Verbindung zur Linearen Algebra. Die Menge  $\mathcal{F}$  aller reellen Zahlenfolgen bildet nämlich einen Vektorraum. Die Vektoraddition und Skalarmultiplikation ist dabei komponentenweise erklärt, das heißt

$$\begin{aligned} (a_n)_{n \in \mathbb{N}} \oplus (b_n)_{n \in \mathbb{N}} &= (a_n + b_n)_{n \in \mathbb{N}}, \\ \lambda \odot (a_n)_{n \in \mathbb{N}} &= (\lambda a_n)_{n \in \mathbb{N}}. \end{aligned}$$

Das ist ähnlich wie die Addition und Skalarmultiplikation von Vektoren im  $\mathbb{R}^3$ , nur hat die Folge unendlich viele Komponenten. Es ist nicht schwierig, die Vektorraumaxiome nachzuprüfen. Die Unterräume des  $\mathbb{R}^3$  sind der Nullvektor, Geraden und Ebenen durch den Nullpunkt, und der  $\mathbb{R}^3$  selbst. Allgemein sind Unterräume  $U$  gekennzeichnet durch die Eigenschaften

$$a, b \in U, \lambda \in \mathbb{R} \quad \Rightarrow \quad a \oplus b, \lambda \odot a \in U.$$

Im Raum  $\mathcal{F}$  aller reellen Folgen haben wir nun folgende Teilmengen; dabei wird  $(a_n)$  als Nullfolge bezeichnet wenn  $\lim_{n \rightarrow \infty} a_n = 0$ .

$$\{\text{Nullfolgen}\} \subset \{\text{Konvergente Folgen}\} \subset \{\text{Beschränkte Folgen}\} \subset \{\text{Reelle Folgen}\}.$$

Mit Regel a) aus Satz 4.3 und Satz 4.2 sieht man leicht, dass es sich auf jeder Stufe um Untervektorräume handelt. Man kann sich überlegen, dass an keiner Stelle Gleichheit gilt, und dass keiner der Räume endlichdimensional ist. So weit dieser Ausflug.

Hier eine Anwendung der Rechenregeln für Grenzwerte.

**Beispiel 4.6 (geometrische Reihe)** Für  $-1 < q < 1$  betrachten wir die Folge (bzw. Reihe)  $a_n = 1 + q + \dots + q^n = \sum_{k=0}^n q^k$ . Dann ergibt sich aus Beispiel 3.2, Beispiel 4.3 und Satz 4.3

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

Wir schreiben hierfür auch

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1 - q} \quad \text{für } q \in (-1, 1). \quad (4.2)$$

Folgen die wie in diesem Fall additiv gegeben sind heißen Reihen. Sie spielen eine große Rolle in der Analysis und werden noch ausführlicher untersucht.

**Satz 4.4 (Grenzwerte und Ungleichungen)** Seien  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  konvergent, mit Grenzwerten  $\lim_{n \rightarrow \infty} a_n = a$  und  $\lim_{n \rightarrow \infty} b_n = b$ . Dann gelten folgende Aussagen:

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

BEWEIS: Nach Voraussetzung gibt es zu jedem  $\varepsilon > 0$  ein  $K \in \mathbb{R}$  mit  $a_n > a - \varepsilon$  und  $b_n < b + \varepsilon$  für alle  $n > K$ . Die Voraussetzung in a) liefert dann  $a - \varepsilon < b + \varepsilon$  beziehungsweise  $(a - b)/2 < \varepsilon$  für jedes  $\varepsilon > 0$ , also  $a \leq b$ . Aussage b) folgt unmittelbar aus a), indem wir  $c, d$  als konstante Folgen auffassen. Unter den Voraussetzungen in c) folgt für  $n > K$  die Ungleichungskette

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

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

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

**Definition 4.4 (Uneigentliche Konvergenz)** Die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergiert uneigentlich (oder divergiert bestimmt) gegen  $+\infty$ , falls gilt:

Zu jedem  $C > 0$  gibt es ein  $K \in \mathbb{R}$ , so dass  $a_n > C$  für alle  $n > K$ .

Wir schreiben  $\lim_{n \rightarrow \infty} a_n = +\infty$  oder  $a_n \rightarrow +\infty$  mit  $n \rightarrow \infty$ . Uneigentliche Konvergenz gegen  $-\infty$  ist analog definiert.

Die Definition der Intervalle aus Kapitel 1 wird ausgedehnt, indem  $\pm\infty$  als offene Intervallgrenzen zugelassen werden, zum Beispiel

$$\begin{aligned} [a, \infty) &= \{x \in \mathbb{R} : a \leq x < \infty\}, \\ (-\infty, b) &= \{x \in \mathbb{R} : -\infty < x < b\}, \\ (-\infty, \infty) &= \{x \in \mathbb{R} : -\infty < x < \infty\} = \mathbb{R}. \end{aligned}$$

Mit dieser Notation kann die Konvergenz  $a_n \rightarrow +\infty$  wie folgt ausgedrückt werden:

Zu jedem  $C > 0$  gibt es ein  $K \in \mathbb{R}$  mit  $a_n \in (C, \infty)$  für alle  $n > K$ .

Bei Konvergenz gegen  $a \in \mathbb{R}$  liegen für  $n > K$  alle Folgenglieder in  $B_\varepsilon(a)$ . Bei Konvergenz gegen  $+\infty$  tritt anstelle der  $\varepsilon$ -Umgebung das rechtsseitig unendliche Intervall  $(C, \infty)$ . Die Konzepte *Konvergenz* und *uneigentliche Konvergenz* lassen sich unter einen Hut bringen, indem die Menge  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$  betrachtet wird. Aber wir verzichten darauf, die Sache würde dadurch auch nicht einfacher. Insbesondere kommen in dieser Vorlesung  $\pm\infty$  als abgeschlossene Intervallgrenzen nicht vor.

**Beispiel 4.7** Für  $q > 1$  gilt  $\lim_{n \rightarrow \infty} q^n = +\infty$ . Denn zu gegebenem  $C > 0$  gibt es nach Beispiel 4.3 ein  $K \in \mathbb{R}$  mit  $(1/q)^n < 1/C$  für  $n > K$ , also  $q^n > C$  für  $n > K$ . Insgesamt haben wir für das Verhalten der Folge  $q^n$  mit  $n \rightarrow \infty$  folgende Tabelle:

$$\begin{aligned} q > 1 &\Rightarrow \lim_{n \rightarrow \infty} q^n = +\infty, \\ q = 1 &\Rightarrow \lim_{n \rightarrow \infty} q^n = 1, \\ -1 < q < 1 &\Rightarrow \lim_{n \rightarrow \infty} q^n = 0, \\ q \leq -1 &\Rightarrow (q^n) \text{ nicht konvergent.} \end{aligned}$$

Der Fall  $-1 < q < 1$  wurde in Beispiel 4.3 behandelt, und der Fall  $q \leq -1$  folgt mit etwas Überlegung aus Beispiel 4.4.

**Beispiel 4.8** Betrachte die rekursiv definierte Folge

$$a_0 = 0, \quad a_{n+1} = a_n^2 + c \quad (c \in \mathbb{R} \text{ Parameter}).$$

Wir behaupten, dass die Folge für  $c > \frac{1}{4}$  gegen  $+\infty$  konvergiert. Und zwar gilt

$$a_{n+1} - a_n = a_n^2 - a_n + c = \left(a_n - \frac{1}{2}\right)^2 + c - \frac{1}{4}.$$

Also folgt induktiv  $a_n \geq n(c - \frac{1}{4}) \rightarrow \infty$ . Die Frage, für welche  $c \in \mathbb{R}$  die Folge divergiert bzw. konvergiert, ist nicht einfach zu beantworten (Stichwort Mandelbrotmenge).

**Satz 4.5 (Konvergenz von Kehrwerten)** Für eine Folge  $(a_n)_{n \in \mathbb{N}}$ ,  $a_n \neq 0$ , gilt:

- (1) Aus  $|a_n| \rightarrow +\infty$  folgt  $1/a_n \rightarrow 0$ .
- (2) Aus  $a_n \rightarrow 0$  und  $a_n > 0$  (bzw.  $a_n < 0$ ) folgt  $1/a_n \rightarrow +\infty$  (bzw.  $1/a_n \rightarrow -\infty$ ).

BEWEIS: Um (1) zu zeigen sei  $\varepsilon > 0$  gegeben. Wegen  $|a_n| \rightarrow +\infty$  gibt es dann ein  $K \in \mathbb{R}$  mit  $|a_n| > 1/\varepsilon$  für alle  $n > K$ , also  $|1/a_n - 0| = 1/|a_n| < \varepsilon$  für  $n > K$ .

Um (2) zu zeigen sei  $C > 0$  gegeben. Wegen  $a_n \rightarrow 0$ ,  $a_n > 0$ , gibt es ein  $K \in \mathbb{R}$  mit  $a_n = |a_n - 0| < 1/C$  für alle  $n > K$ . Es folgt  $a_n > C$  für  $n > K$ .  $\square$

Die Aussagen (1) und (2) sind nicht ganz symmetrisch, bei (1) gibt es keine Voraussetzung ans Vorzeichen während bei (2) explizit  $a_n > 0$  verlangt werden muss.



## 5 Vollständigkeit der reellen Zahlen

Thema dieses Kapitels ist die Vollständigkeit der reellen Zahlen. Eine populäre aber reichlich vage Formulierung des Axioms lautet: *in der Zahlengerade gibt es keine Löcher*. Damit kann man natürlich nicht viel anfangen, unsere Version verwendet das Konzept der Konvergenz von Cauchyfolgen. Wir werden im Laufe des Kapitels mehrere alternative Fassungen kennen lernen, sowie eine Reihe von Anwendungen.

Die arithmetischen Axiome (K) und Anordnungsaxiome (A) sind in den rationalen Zahlen gültig, dennoch ist  $\mathbb{Q}$  für die Analysis ungeeignet. Um das zu begründen beginnen wir mit einer Beobachtung der Pythagoräer<sup>13</sup>. Vorab folgende Tatsache: ist  $p \in \mathbb{N}$  mit  $p^2$  gerade, so ist  $p$  auch gerade. Denn wäre  $p$  ungerade, also  $p = 2m + 1$  mit  $m \in \mathbb{N}_0$ , so hätten wir  $p^2 = 4m^2 + 4m + 1$  und  $p^2$  wäre ungerade, Widerspruch.

**Satz 5.1 (Irrationalität von  $\sqrt{2}$ )** Die Gleichung  $x^2 = 2$  ist in  $\mathbb{Q}$  nicht lösbar.

BEWEIS: (durch Widerspruch) Angenommen es gibt eine rationale Zahl  $x$  mit  $x^2 = 2$ . Indem wir wenn nötig  $x$  durch  $-x$  ersetzen haben wir  $x > 0$ . also  $x = p/q$  mit  $p, q \in \mathbb{N}$ . Die folgende Tatsache ist nun wesentlich: *wir können annehmen dass höchstens eine der Zahlen  $p, q$  gerade ist*. Dazu kürzen wir so lange durch 2 bis es stimmt, das ist spätestens der Fall wenn Zähler oder Nenner gleich 1 sind, insbesondere nach endlich vielen Schritten. Aus der Gleichung  $p^2/q^2 = x^2 = 2$  folgt

$$p^2 = 2q^2 \quad \Rightarrow \quad p^2 \text{ gerade} \quad \Rightarrow \quad p \text{ gerade.}$$

Somit ist  $p = 2m$  mit  $m \in \mathbb{N}$ , es folgt weiter

$$2q^2 = p^2 = 4m^2 \quad \Rightarrow \quad q^2 \text{ gerade} \quad \Rightarrow \quad q \text{ gerade.}$$

Also sind doch  $p, q$  beide gerade, ein Widerspruch. □

In  $\mathbb{R}$  ist die Gleichung  $x^2 = 2$  lösbar, wie wir in Kürze sehen werden. Man könnte das als Grund für die Erweiterung von  $\mathbb{Q}$  zu  $\mathbb{R}$  anführen, ähnlich wie die Erweiterungen  $\mathbb{N} \subset \mathbb{Z}$  und  $\mathbb{Z} \subset \mathbb{Q}$  durch die Lösbarkeit der Gleichungen  $a + x = b$  beziehungsweise  $a \cdot x = b$  motiviert sind. Das ist nicht ganz falsch, wir werden für positive Zahlen Quadratwurzeln und auch  $n$ -te Wurzeln definieren. Andererseits bleibt die Gleichung  $x^2 = -1$  auch in  $\mathbb{R}$  ungelöst. Und die Menge aller Nullstellen von Polynomen mit rationalen Koeffizienten, die sogenannten reellen algebraischen Zahlen, bilden nur eine relativ kleine Teilmenge von  $\mathbb{R}$ , was wir noch präzisieren und zeigen werden. Die meisten reellen Zahlen haben mit der Lösung von Polynomgleichungen mit Koeffizienten in  $\mathbb{Q}$  nichts zu tun.

Im Zentrum steht vielmehr das Ziel der Analysis, neue Objekte – Zahlen, Funktionen, Operationen – durch Grenzprozesse zu konstruieren. Unsere Definition der Konvergenz setzt voraus, dass wir den Grenzwert der Folge bereits kennen. Das Vollständigkeitsaxiom muss die Existenz von Grenzwerten in Situationen garantieren, in denen der Grenzwert a priori nicht bekannt ist. Wir betrachten hierzu drei Beispiele.

---

<sup>13</sup>Hippasos von Metapont, um 450 v. Chr.

**Beispiel 5.1 (Verfahren von Heron)** Sei  $a > 0$ . Wir wollen zeigen dass folgende Iteration gegen eine Lösung der Gleichung  $x^2 = a$  konvergiert:

$$x_{n+1} = \frac{1}{2} \left( x_n + \frac{a}{x_n} \right), \quad \text{Startwert } x_0 > 0. \quad (5.1)$$

Wir können  $x$  und  $a/x$  als Seiten eines Rechtecks mit Flächeninhalt  $a$  auffassen. Im Fall eines Quadrats wäre  $x^2 = a$ , die Lösung wäre gefunden. Im allgemeinen sind die Seiten aber relativ zum Quadrat zu lang bzw. zu kurz, der Übergang von  $x_n$  zum arithmetischen Mittel  $x_{n+1}$  soll das in jedem Schritt verbessern. Was können wir über die Iteration aussagen? Es gilt

$$x_{n+1}^2 = \frac{1}{4} \left( x_n + \frac{a}{x_n} \right)^2 = \frac{1}{4} \left( x_n - \frac{a}{x_n} \right)^2 + a.$$

Für  $n \geq 1$  ist also  $x_n^2 \geq a$ , das heißt  $x_n$  ist die längere Seite. Weiter folgt

$$x_n - x_{n+1} = \frac{1}{2} \left( x_n - \frac{a}{x_n} \right) = \frac{x_n^2 - a}{2x_n} \geq 0 \quad \text{für } n \geq 1, \quad \text{also } x_1 \geq x_2 \geq \dots \geq 0.$$

Somit ist die Folge  $x_n$  für  $n \geq 1$  monoton fallend und durch  $x_n \geq 0$  nach unten beschränkt. Wenn sie gegen ein  $x \in \mathbb{R}$  konvergiert, so folgt  $x \geq 0$  und weiter  $x^2 \geq a > 0$ , und damit

$$x = \lim_{n \rightarrow \infty} x_{n+1} = \lim_{n \rightarrow \infty} \frac{1}{2} \left( x_n + \frac{a}{x_n} \right) = \frac{1}{2} \left( x + \frac{a}{x} \right).$$

Umstellen ergibt  $x^2 = a$ , die Gleichung ist gelöst. Die Konvergenz der Folge  $x_n$  können wir aber nicht mit Definition 4.1 zeigen, denn dafür müssten wir den Grenzwert  $\sqrt{a}$  schon kennen, und den wollen wir ja gerade konstruieren.

**Beispiel 5.2 (Zinseszinsrechnung)** Wird ein Kapital von  $K$  Euro mit Zinssatz  $p\%$  per annum linear verzinst, so ist der Kontostand nach  $x$  Jahren gleich  $K(1 + \frac{p}{100}x)$ . Wir starten im folgenden mit 1 Euro und nehmen als Zinssatz 100%. Nach  $x$  Jahren haben wir dann  $f_1(x) := 1 + x$ , zum Beispiel nach einem Jahr 2 Euro, das Kapital hat sich verdoppelt. Die Idee des Zinseszinses ist, den Zeitraum  $x$  in kürzere Intervalle zu unterteilen und den Zins anteilig anzurechnen mit dem Effekt, dass die schon erworbenen Zinsen ihrerseits wieder Zinsen liefern. Wird der Zeitraum  $[0, x]$  in  $n$  gleich lange Intervalle unterteilt, so ergibt das nach dem ersten Intervall  $1 + \frac{x}{n}$ , nach dem zweiten  $(1 + \frac{x}{n})^2$ , und schließlich nach dem  $n$ -ten Intervall, also insgesamt nach der Zeit  $x$ ,

$$f_n(x) = \left( 1 + \frac{x}{n} \right)^n \quad \text{für } x \in \mathbb{R}, n \in \mathbb{N}. \quad (5.2)$$

Was liefert kontinuierliche Verzinsung, existiert der Grenzwert  $\lim_{n \rightarrow \infty} f_n(x)$ ?

**Beispiel 5.3 (Dezimalbruchdarstellung)** Jede Zahl  $a \in \mathbb{R}$  hat eine Darstellung der Form  $a = a_0, k_1 k_2 \dots$  mit  $a_0 \in \mathbb{Z}$  und  $k_j \in \{0, 1, \dots, 9\}$  für  $j = 1, 2, \dots$ . Genauer bedeutet das

$$a = \lim_{n \rightarrow \infty} a_n \quad \text{mit} \quad a_n = a_0 + \sum_{j=1}^n k_j \cdot 10^{-j}.$$

Als erstes wählen wir  $a_0 \in \mathbb{Z}$  mit  $a_0 \leq a < a_0 + 1$ . Wir bestimmen dann  $a_n$  induktiv mit  $a_n \leq a < a_n + 10^{-n}$ , das ist für  $n = 0$  erfüllt. Sei  $a_{n-1}$  schon bestimmt, es gilt also

$$a_{n-1} + 0 \cdot 10^{-n} = a_{n-1} \leq a < a_{n-1} + 10^{-(n-1)} = a_{n-1} + 10 \cdot 10^{-n}.$$

Wir definieren nun  $a_n = a_{n-1} + k_n \cdot 10^{-n}$  wobei

$$k_n = \max\{k \in \mathbb{Z} : a_{n-1} + k \cdot 10^{-n} \leq a\} \in \{0, 1, \dots, 9\}.$$

In diesem Maximum ist  $k = k_n$  zulässig aber nicht  $k = k_n + 1$ . Das bedeutet

$$a_n = a_{n-1} + k_n \cdot 10^{-n} \leq a < a_{n-1} + (k_n + 1) \cdot 10^{-n} = a_n + 10^{-n}.$$

Damit sind die  $a_n$  wie behauptet definiert. Sie approximieren die Zahl  $a$ , denn es gilt

$$|a - a_n| \leq 10^{-n} \rightarrow 0 \quad \text{mit } n \rightarrow \infty.$$

Eine gegebene Zahl  $a$  hat somit eine Darstellung als Dezimalbruch. Umgekehrt stellt sich aber die Frage: seien  $a_0 \in \mathbb{Z}$  und  $k_1, k_2, \dots$  beliebig mit  $k_j \in \{0, 1, \dots, 9\}$ . Konvergiert dann die Dezimalbruchfolge  $a_n = a_0, k_1 k_2 \dots k_n$  gegen eine gewisse, reelle Zahl?

In diesen Beispielen brauchen wir wie gesagt eine Charakterisierung konvergenter Folgen, die ohne die Kenntnis des Grenzwerts auskommt. Die Idee von Cauchy<sup>14</sup> besteht darin, die Glieder der Folge nicht mit dem unbekanntem Grenzwert, sondern *untereinander* zu vergleichen.

**Definition 5.1 (Cauchyfolge)** Eine Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *Cauchyfolge*, wenn gilt:

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

Beim Nachweis dieser Eigenschaft reicht es aus, die Zahlen  $n, m$  mit  $n < m$  zu betrachten, denn die Definition ist symmetrisch in  $n$  und  $m$  und für  $n = m$  ist nichts zu tun.

(V) Vollständigkeitsaxiom: Jede Cauchyfolge ist konvergent.

Damit sind die Axiome (KAV) der reellen Zahlen komplett. Es gibt auch andere Möglichkeiten die Vollständigkeit zu charakterisieren, diese werden wir als Folgerungen herleiten.

**Satz 5.2** Jede konvergente Folge ist eine Cauchyfolge.

BEWEIS: Sei  $\lim_{n \rightarrow \infty} a_n = a$ . Zu  $\varepsilon > 0$  gibt es dann ein  $K \in \mathbb{R}$  mit  $|a_n - a| < \varepsilon/2$  für  $n > K$ , und für  $n, m > K$  folgt

$$|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a_m - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Die Umkehrung dieser Aussage ist genau das Vollständigkeitsaxiom, wir können also auch sagen: eine Folge ist genau dann konvergent wenn sie eine Cauchyfolge ist.

---

<sup>14</sup>Augustin Louis Cauchy, 1789–1857

**Beispiel 5.4 (Konvergenz von Dezimalbrüchen)** Seien  $a_0 \in \mathbb{Z}$ ,  $k_j \in \{0, 1, \dots, 9\}$  beliebig gegeben. Dann konvergiert die Dezimalbruchfolge

$$a_n = a_0 + \sum_{j=1}^n k_j \cdot 10^{-j}$$

gegen eine reelle Zahl. Für  $n < m$  schätzen wir wie folgt ab, wobei wir die Formel für die geometrische Reihe, Beispiel 4.6, verwenden:

$$|a_m - a_n| = \sum_{j=n+1}^m k_j \cdot 10^{-j} \leq 9 \cdot 10^{-(n+1)} \sum_{k=0}^{m-(n+1)} 10^{-k} \leq 9 \cdot 10^{-(n+1)} \frac{1}{1 - \frac{1}{10}} = 10^{-n}.$$

Nach Beispiel 4.3 ist  $10^{-n}$  eine Nullfolge, also ist  $10^{-n} < \varepsilon$  für  $n > K$ . Die Dezimalbruchfolge ist damit eine Cauchyfolge, sie konvergiert nach dem Vollständigkeitsaxiom. Nach Beispiel 5.3 hat jede reelle Zahl eine Darstellung als Dezimalbruch. Die Abbildung, die jedem Dezimalbruch seinen Grenzwert zuordnet, ist also definiert und surjektiv. Sie ist nicht injektiv, zum Beispiel gilt  $0,99\dots = 1,00\dots$ . Tatsächlich sind die Gegenbeispiele zur Eindeutigkeit der Dezimaldarstellung aber nur von diesem Typ (Übungsaufgabe).

**Definition 5.2 (Monotone Folge)** Eine Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *monoton wachsend*, wenn

$$a_{n+1} \geq a_n \quad \text{für alle } n \in \mathbb{N}.$$

Manche Autoren bezeichnen dies als nichtfallend, und reservieren den Begriff wachsend für eine Folge mit  $a_{n+1} > a_n$ ; bei uns heißt das *streng monoton wachsend*.

**Satz 5.3 (Konvergenzkriterium der Monotonie und Beschränktheit)** Jede *monoton wachsende, nach oben beschränkte Folge* ist eine Cauchyfolge und damit konvergent.

BEWEIS: Sei  $(a_n)_{n \in \mathbb{N}}$  monoton wachsend und  $a_n \leq b$  für alle  $n \in \mathbb{N}$ . Betrachte für  $\varepsilon > 0$  die Intervalle  $I_k = [a_1 + k\varepsilon, a_1 + (k+1)\varepsilon]$  mit  $k \in \mathbb{N}_0$ . Wenn  $I_k$  ein Folgenglied  $a_n$  enthält so folgt

$$a_1 + k\varepsilon \leq a_n, \text{ also } k \leq \frac{a_n - a_1}{\varepsilon} \leq \frac{b - a_1}{\varepsilon}.$$

Das Intervall  $I_0$  enthält  $a_1$ . Sei  $k$  maximal so dass  $I_k$  irgendein Element der Folge enthält, und sei dann  $N$  minimal mit  $a_N \in I_k$ . Aus der Monotonie und der Maximalität von  $k$  folgt

$$n \geq N \Rightarrow a_1 + k\varepsilon \leq a_N \leq a_n \leq a_1 + (k+1)\varepsilon.$$

Also gilt  $|a_n - a_m| \leq \varepsilon$  für  $n, m \geq N$ , die Folge  $(a_n)$  ist eine Cauchyfolge. □

Natürlich gibt es auch konvergente Folgen die nicht monoton sind, zum Beispiel  $a_n = (-1)^n/n$ . Dennoch ist das Kriterium von zentraler Bedeutung, es hat sehr viele Anwendungen.

**Beispiel 5.5 (Konvergenz des Heronverfahrens)** Die in Beispiel 5.1 konstruierte Folge  $x_n$  erfüllt  $x_1 \geq x_2 \geq \dots \geq 0$ . Nach Satz 5.3 konvergiert die Folge gegen ein  $x \geq 0$ . Wir hatten bereits gesehen dass dann die Gleichung  $x^2 = a$  gelöst wird, also haben wir eine Quadratwurzel von  $a$  gefunden.

**Beispiel 5.6 (Die Zahl  $e$  und Zinsrechnung)** Wir betrachten die Verzinsung von einem Euro Kapital für ein Jahr, siehe Beispiel 5.2. Bei Zinseszins mit  $n$  Teilintervallen ergibt sich

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

Wir behaupten, dass die Folge  $f_n$  nach oben beschränkt ist. Berechne dazu mit der binomischen Formel, siehe Satz 3.6,

$$f_n = \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k} = \sum_{k=0}^n \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \sum_{k=0}^n \frac{1}{k!} =: e_n.$$

Wir schätzen  $e_n$  ab: es ist  $k! \geq 2^{k-1}$  für  $k \geq 1$ , also

$$e_n \leq 1 + \sum_{k=1}^n 2^{1-k} \leq 1 + \sum_{k=1}^{\infty} 2^{1-k} = 3.$$

Die Folge  $e_n$  ist offensichtlich monoton wachsend. Für  $f_n$  verwende die Bernoulli-Ungleichung:

$$\begin{aligned} \frac{f_n}{f_{n-1}} &= \frac{(n+1)^n}{n^n} \cdot \frac{(n-1)^{n-1}}{n^{n-1}} \\ &= \frac{n}{n-1} \left( \frac{(n+1)(n-1)}{n^2} \right)^n \\ &= \frac{n}{n-1} \left( 1 - \frac{1}{n^2} \right)^n \\ &\geq \frac{n}{n-1} \left( 1 - n \cdot \frac{1}{n^2} \right) = 1. \end{aligned}$$

Also ist auch die Folge  $f_n$  monoton wachsend. Im Zins-Modell bedeutet das, kürzere Intervalle sind besser. Beide Folgen haben die Zahl 3 als obere Schranke, also sind sie konvergent.

**Definition 5.3 (Eulersche Zahl)** Wir setzen

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{1}{k!}. \quad (5.3)$$

Wie oben gezeigt ist  $2 \leq f_n \leq e_n \leq 3$  (beachte  $f_1 = 2$ ), also gilt auch  $2 \leq e \leq 3$ . Der Taschenrechner liefert zum Beispiel die Näherungswerte  $e_3 = 2,666\dots$  und  $e_5 = 2,716\dots$ . Wir zeigen nun, dass die Folge  $f_n$  ebenfalls den Grenzwert  $e$  hat. Aus  $f_n \leq e_n$  ergibt sich  $\lim_{n \rightarrow \infty} f_n \leq e$ . Wir zeigen andererseits für beliebiges  $m \in \mathbb{N}$

$$\lim_{n \rightarrow \infty} f_n \geq e_m.$$

Mit  $m \rightarrow \infty$  folgt dann  $\lim_{n \rightarrow \infty} f_n \geq e$  wie behauptet. Für  $n \geq m$  schätze wie folgt ab:

$$f_n = \sum_{k=0}^n \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \geq \sum_{k=0}^m \frac{1}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n}.$$

Der  $k$ -te Summand auf der rechten Seite ist das Produkt von  $1/k!$  mit den Faktoren  $(n-j)/n$  für  $j = 0, \dots, k-1$ . Jeder dieser Faktoren konvergiert mit  $n \rightarrow \infty$  gegen Eins, also folgt

$$\lim_{n \rightarrow \infty} f_n \geq \sum_{k=0}^m \frac{1}{k!} = e_m.$$

Die Eulersche Zahl hat somit die alternative Darstellung

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n. \quad (5.4)$$

Wird ein Euro für ein Jahr mit 100% angelegt, so hat man bei linearer Verzinsung 2 Euro, bei kontinuierlichem Zinseszins dagegen  $e = 2,71828 \dots$  Euro. Der Zinseszins ist ein spezielles Wachstumsgesetz, die Folge  $f_n$  ist dadurch gut motiviert. Die Folge  $e_n$  kommt hier nur technisch ins Spiel, sie tritt in der Abschätzung auf. Bei der Diskussion der Exponentialfunktion werden wir aber dieser Reihe wieder begegnen.

**Satz 5.4 (Irrationalität der Eulerschen Zahl)**  $e = \sum_{k=0}^{\infty} 1/k!$  ist nicht rational.

BEWEIS: Wir zeigen, dass die  $e_n$  die Zahl  $e$  gut approximieren. Es gilt

$$\begin{aligned} e - e_n &= \sum_{k=n+1}^{\infty} \frac{1}{k!} \\ &= \frac{1}{(n+1)!} \left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \dots\right) \\ &\leq \frac{1}{(n+1)!} (1 + 2^{-1} + 2^{-2} + \dots) \\ &= \frac{2}{(n+1)!}. \end{aligned}$$

Nach Multiplikation mit  $n!$  ergibt sich hieraus

$$0 < n!(e - e_n) \leq \frac{2}{n+1} < 1 \quad \text{für } n \geq 2.$$

Nun ist  $n!e_n = \sum_{k=0}^n n!/k! \in \mathbb{N}$ . Wäre  $e$  rational, so wäre auch  $n!e \in \mathbb{N}$  für  $n$  hinreichend groß. Dann wäre  $n!(e - e_n)$  eine ganze Zahl, Widerspruch.  $\square$

Als nächste Anwendung des Vollständigkeitsaxioms nun das Intervallschachtelungsprinzip.

**Definition 5.4 (Intervallschachtelung)** Eine Intervallschachtelung ist eine Folge von Intervallen  $I_n = [a_n, b_n] \subset \mathbb{R}$  mit  $I_{n+1} \subset I_n$  für alle  $n$  und  $|I_n| = b_n - a_n \rightarrow 0$  mit  $n \rightarrow \infty$ .

Es ist nicht verlangt dass die Intervalle den gleichen Mittelpunkt haben.

**Satz 5.5 (Intervallschachtelungsprinzip)** Zu jeder Intervallschachtelung  $(I_n)_{n \in \mathbb{N}}$  mit  $I_n = [a_n, b_n]$  gibt es genau  $x \in \mathbb{R}$  so dass

$$x \in \bigcap_{n \in \mathbb{N}} I_n, \text{ also } x \in I_n \text{ für alle } n \in \mathbb{N}.$$

Es gilt  $x = \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$ .

BEWEIS: Nach Voraussetzung ist  $a_1 \leq a_2 \leq \dots$  und  $b_1 \geq b_2 \geq \dots$ . Ferner gilt

$$a_n \leq b_n \leq b_1 \quad \text{und} \quad b_n \geq a_n \geq a_1 \quad \text{für alle } n \in \mathbb{N}.$$

Aus Satz 5.3 folgt die Existenz der Grenzwerte  $a = \lim_{n \rightarrow \infty} a_n$  bzw.  $b = \lim_{n \rightarrow \infty} b_n$ . Dann folgt nach Satz 4.3 und Satz 4.4

$$0 \leq b - a = \lim_{n \rightarrow \infty} b_n - \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} (b_n - a_n) = 0.$$

Setze  $x := a = b$ . Dann ist  $a_n \leq a = x = b \leq b_n$ , also  $x \in I_n$  für alle  $n$ . Sei  $x' \in \mathbb{R}$  mit  $x' \in I_n$  für alle  $n$ , das heißt  $a_n \leq x' \leq b_n$ . Durch Grenzübergang ergibt sich nach Satz 4.4  $a \leq x' \leq b$ , also  $x' = x$ .  $\square$

Wir wollen eine Intervallschachtelung verwenden, um für  $y \geq 0$  und  $n \in \mathbb{N}$  die Zahl  $y^{1/n}$  zu konstruieren, also die Lösung  $x \geq 0$  der Gleichung  $x^n = y$ . Dazu machen wir ein paar Vorüberlegungen zu monotonen Funktionen. Eine Funktion  $f : I \rightarrow \mathbb{R}$  auf einem Intervall  $I$  heißt *streng monoton wachsend*, wenn folgende Implikation gilt:

$$x_2 > x_1 \quad \Rightarrow \quad f(x_2) > f(x_1). \quad (5.5)$$

$f$  ist dann injektiv: ist  $f(x_1) = f(x_2)$  so kann weder  $x_2 > x_1$  noch  $x_1 > x_2$  sein, es folgt also  $x_1 = x_2$ . Die Abbildung  $f : I \rightarrow f(I)$ , also jetzt mit Zielbereich  $f(I)$  statt  $\mathbb{R}$ , ist dann injektiv und surjektiv, das heißt es gibt die Umkehrfunktion  $g : f(I) \rightarrow I$ . Diese Umkehrfunktion ist ebenfalls streng monoton wachsend: wäre  $y_2 > y_1$  mit  $g(y_2) \leq g(y_1)$ , so folgt aus der Monotonie von  $f$

$$y_1 = f(g(y_1)) \geq f(g(y_2)) = y_2, \quad \text{Widerspruch.} \quad (5.6)$$

In unserem Fall ist  $f : I = [0, \infty) \rightarrow \mathbb{R}$ ,  $f(x) = x^n$ , streng monoton wachsend nach Anordnungsaxiom (A2). Somit ist  $f$  injektiv, es gibt also höchstens eine Lösung  $x \geq 0$  der Gleichung  $f(x) = y$  bzw.  $x^n = y$ . Die Umkehrfunktion  $g : f(I) \rightarrow I$  erfüllt  $y = f(g(y)) = g(y)^n$ , also ist  $x = g(y)$  die gesuchte Lösung, es gilt  $g(y) = y^{1/n}$ . Allerdings ist  $g(y)$  nur für  $y \in f(I)$  definiert, wir müssen noch  $f(I) = [0, \infty)$  zeigen. Das ist genau das Problem der Existenz der Wurzel, das wir nun angehen.

**Satz 5.6 (*n*-te Wurzel)** *Zu jedem  $y \geq 0$  und  $n \in \mathbb{N}$  gibt es genau ein  $x \geq 0$  mit  $x^n = y$ , Bezeichnung  $\sqrt[n]{y}$  oder  $y^{1/n}$ . Die Funktion  $y \mapsto y^{1/n}$  ist streng monoton wachsend auf  $[0, \infty)$ .*

BEWEIS: Mit der Methode der fortgesetzten Halbierung konstruieren wir eine Intervallschachtelung  $I_k = [a_k, b_k]$ , so dass gilt:

$$a_k^n \leq y \leq b_k^n \quad \text{für alle } k \in \mathbb{N}. \quad (5.7)$$

Als Start wählen wir  $I_1 = [0, b_1]$  mit  $b_1 \in \mathbb{N}$  hinreichend groß. Ist  $I_k$  schon gefunden, so setzen wir  $m_k = (a_k + b_k)/2$  und wählen

$$I_{k+1} = [a_{k+1}, b_{k+1}] = \begin{cases} [a_k, m_k] & \text{falls } m_k^n \geq y \\ [m_k, b_k] & \text{falls } m_k^n < y. \end{cases}$$

Damit gilt (5.7) auch für  $k + 1$ :

$$\begin{aligned} \text{Fall 1:} & \quad b_{k+1}^n = m_k^n \geq y, \quad \text{und} \quad a_{k+1}^n = a_k^n \leq y \quad \text{nach Induktion,} \\ \text{Fall 2:} & \quad a_{k+1}^n = m_k^n < y, \quad \text{und} \quad b_{k+1}^n = b_k^n \geq y \quad \text{nach Induktion.} \end{aligned}$$

Weiter ist  $I_{k+1} \subset I_k$  und  $|I_k| = 2^{1-k}|I_1| \rightarrow 0$ , also ist  $(I_k)$  eine Intervallschachtelung. Nach Satz 5.5 gibt es ein  $x \in \mathbb{R}$  mit  $x \in I_k$  für alle  $k$ , und es gilt  $a_k, b_k \rightarrow x$  mit  $k \rightarrow \infty$ . Durch Grenzübergang in (5.7) folgt

$$x^n = \lim_{n \rightarrow \infty} a_k^n \leq y \leq \lim_{n \rightarrow \infty} b_k^n = x^n.$$

Die Eindeutigkeit der  $n$ -ten Wurzel und die Monotonie der Funktion wurden oben gezeigt.  $\square$

Für  $a \geq 0$  können nun Potenzen  $a^r$  mit Exponenten  $r \in \mathbb{Q}$  erklärt werden. Potenzen mit ganzzahligen Exponenten seien vorausgesetzt, inklusive Rechenregeln. Wir definieren

$$a^r = (a^p)^{1/q} \quad \text{für } r = p/q \text{ mit } p \in \mathbb{Z}, q \in \mathbb{N}.$$

Dies hängt nicht von der Darstellung von  $r$  ab. Denn sei  $r = p'/q'$ , also  $pq' = p'q$ , dann gilt

$$\left( (a^{p'})^{1/q'} \right)^{qq'} = \left( ((a^{p'})^{1/q'})^{q'} \right)^q = (a^{p'})^q = a^{p'q} = a^{pq'} = \left( (a^p)^{1/q} \right)^{qq'},$$

also  $(a^{p'})^{1/q'} = (a^p)^{1/q}$  wegen Eindeutigkeit. Weiter zeigt man die Rechenregeln

$$(i) a^r a^s = a^{r+s} \quad (ii) (a^r)^s = a^{rs} \quad (iii) a^r b^r = (ab)^r.$$

Zum Beispiel gilt mit  $r = k/m$  und  $s = p/q$

$$(a^r a^s)^{mq} = (a^{k/m})^{mq} (a^{p/q})^{mq} = \left( (a^k)^{1/m} \right)^m \cdot \left( (a^p)^{1/q} \right)^q = (a^k)^q \cdot (a^p)^m = a^{kq+pm}.$$

Dies bedeutet

$$a^r a^s = (a^{kq+pm})^{1/mq} = a^{(kq+pm)/mq} = a^{k/m+p/q} = a^{r+s}.$$

Die anderen beiden Potenzgesetze werden ähnlich verifiziert.

**Beispiel 5.7** Ein Grenzwert, der häufiger auftritt, ist

$$\lim_{n \rightarrow \infty} a^{1/n} = 1 \quad \text{für } a > 0. \quad (5.8)$$

Betrachte erst den Fall  $a \geq 1$ . Dann ist  $a^{1/n} \geq 1^{1/n} = 1$  wegen Monotonie der Wurzelfunktion, also  $a^{1/n} = 1 + \xi_n$  mit  $\xi_n \geq 0$ . Aus der Bernoulli-Ungleichung, Satz 3.2, folgt

$$a = (1 + \xi_n)^n \geq 1 + n\xi_n \quad \Rightarrow \quad 0 \leq \xi_n \leq (a - 1)/n.$$

Also  $\lim_{n \rightarrow \infty} \xi_n = 0$  bzw.  $\lim_{n \rightarrow \infty} a^{1/n} = \lim_{n \rightarrow \infty} (1 + \xi_n) = 1$ . Im Fall  $0 < a < 1$  verwende

$$\left( a^{1/n} \cdot (a^{-1})^{1/n} \right)^n = a \cdot a^{-1} = 1 \quad \Rightarrow \quad a^{1/n} = \frac{1}{(a^{-1})^{1/n}}.$$

Es folgt  $\lim_{n \rightarrow \infty} a^{1/n} = 1 / \lim_{n \rightarrow \infty} (a^{-1})^{1/n} = 1$  wegen  $a^{-1} > 1$ .

**Definition 5.5 (Teilfolge)** Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge und  $(n_k)_{k \in \mathbb{N}}$  eine Folge natürlicher Zahlen mit  $n_1 < n_2 < n_3 \dots$ . Dann heißt die Folge  $(a_{n_k})_{k \in \mathbb{N}}$  Teilfolge von  $(a_n)_{n \in \mathbb{N}}$ .

Die Teilfolge entsteht aus der ursprünglichen Folge durch *Auswahl der Nummern*  $n_k$ , zum Beispiel ergibt sich mit  $n_k = 2k$  die Teilfolge  $a_{2k}$  mit geraden Nummern. Es gilt allgemein  $n_k \geq k$ , denn  $n_1 \geq 1$  und per Induktion  $n_{k+1} \geq n_k + 1 \geq k + 1$ . Streng genommen sind Folgen Abbildungen von  $\mathbb{N}$  nach  $\mathbb{R}$ . Formal ergibt sich eine Teilfolge damit als Verkettung der Ausgangsfolge  $n \mapsto a_n$  und der Folge  $k \mapsto n_k$  die die Nummern aussucht. Am Beispiel  $a_n = (-1)^n/n^3$  und  $n_k = 2k - 1$  sieht das so aus:

$$\mathbb{N} \xrightarrow{n_k} \mathbb{N} \xrightarrow{a_n} \mathbb{R}, \quad k \mapsto 2k - 1 \mapsto \frac{(-1)^{2k-1}}{(2k-1)^3} = -\frac{1}{(2k-1)^3}.$$

Wir interessieren uns für die Frage ob eine gegebene Folge eine Teilfolge hat die konvergiert.

**Definition 5.6 (Häufungspunkt von Folgen)**  $a \in \mathbb{R}$  heißt *Häufungspunkt der Folge*  $(a_n)_{n \in \mathbb{N}}$ , wenn es eine Teilfolge  $(a_{n_k})_{k \in \mathbb{N}}$  gibt, die mit  $k \rightarrow \infty$  gegen  $a$  konvergiert.

Beispielsweise hat die Folge  $a_n = (-1)^n + 1/n^2$  den Häufungspunkt  $+1$ , denn mit  $n_k = 2k$  gilt  $a_{n_k} = a_{2k} = 1 + 1/(2k)^2 \rightarrow 1$  mit  $k \rightarrow \infty$ . Auch  $-1$  ist ein Häufungspunkt der Folge, denn für  $n_k = 2k - 1$  ist  $a_{n_k} = a_{2k-1} = -1 + 1/(2k-1)^2 \rightarrow -1$  mit  $k \rightarrow \infty$ .

*Bemerkung:* konvergiert die Folge  $(a_n)$  gegen  $a$ , so konvergiert auch jede Teilfolge  $(a_{n_k})$  gegen  $a$ . Denn wähle zu  $\varepsilon > 0$  ein  $N \in \mathbb{N}$  mit  $|a_n - a| < \varepsilon$  für alle  $n \geq N$ . Es folgt  $|a_{n_k} - a| < \varepsilon$  für  $k \geq N$ , denn  $n_k \geq k \geq N$ .

**Lemma 5.1** Die Zahl  $a \in \mathbb{R}$  ist genau dann Häufungspunkt der Folge  $(a_n)_{n \in \mathbb{N}}$ , wenn für jedes  $\varepsilon > 0$  die Menge  $\{n \in \mathbb{N} : a_n \in B_\varepsilon(a)\}$  unendlich viele Elemente hat.

BEWEIS: Sei  $a \in \mathbb{R}$  Häufungspunkt von  $(a_n)_{n \in \mathbb{N}}$ . Es gibt also eine Teilfolge  $a_{n_k} \rightarrow a$  mit  $k \rightarrow \infty$ . Zu jedem  $\varepsilon > 0$  gibt es dann ein  $K \in \mathbb{R}$  mit  $a_{n_k} \in B_\varepsilon(a)$  für alle  $k > K$ . Wegen  $n_1 < n_2 < \dots$  ist die Menge  $\{n_k : k > K\}$  nicht endlich, die eine Richtung ist gezeigt. Für die umgekehrte Aussage wählen wir induktiv  $n_k$  mit  $n_1 < n_2 < \dots$ , so dass  $a_{n_k} \in B_{1/k}(a)$ . Die Induktion bricht nicht ab, da  $a_n \in B_{1/k}(a)$  für unendlich viele  $n$ . Es folgt  $|a_{n_k} - a| < 1/k \rightarrow 0$  mit  $k \rightarrow \infty$ .  $\square$

**Satz 5.7 (Bolzano-Weierstraß)** Jede beschränkte Folge reeller Zahlen hat eine konvergente Teilfolge, also mindestens einen Häufungspunkt.

BEWEIS: Sei  $(x_n)_{n \in \mathbb{N}}$  eine beschränkte Folge, also  $x_n \in [a_1, b_1] = I_1$  für alle  $n \in \mathbb{N}$ . Wir suchen den Häufungspunkt mittels fortgesetzter Intervallhalbierung: sei  $I_k = [a_k, b_k]$  schon gefunden mit der Eigenschaft

$$(*) \quad x_n \in I_k \quad \text{für unendlich viele } n.$$

Setze  $m_k = \frac{1}{2}(a_k + b_k)$  und definiere

$$I_{k+1} = [a_{k+1}, b_{k+1}] = \begin{cases} [m_k, b_k], & \text{falls } x_n \in [m_k, b_k] \text{ für unendlich viele } n, \\ [a_k, m_k] & \text{sonst.} \end{cases}$$

Durch Inspektion der Fälle sieht man, dass Aussage (\*) auch für  $I_{k+1}$  stimmt. Die  $I_k$  bilden

eine Intervallschachtelung, nach Satz 5.5 gibt es also ein  $x \in \mathbb{R}$  mit  $x \in I_k$  für alle  $k$ . Sei  $\varepsilon > 0$  gegeben. Bestimme  $K \in \mathbb{R}$  so dass  $|I_k| < \varepsilon$  für alle  $k > K$ . Dann gilt für  $k > K$

$$I_k \subset B_\varepsilon(x), \quad \text{denn für } x' \in I_k \text{ gilt } |x' - x| \leq |I_k| < \varepsilon.$$

Es folgt  $x_n \in B_\varepsilon(x)$  für unendlich viele  $n$ , und  $x$  ist Häufungspunkt nach Lemma 5.1.  $\square$

Wir kommen jetzt zu einem technischen Werkzeug das etwas schwieriger zu bedienen ist, aber sehr nützlich sein kann.

**Definition 5.7 (Limes superior/inferior)** Für eine reelle Folge  $(x_n)_{n \in \mathbb{N}}$  und  $x^* \in \mathbb{R} \cup \{\pm\infty\}$  gilt  $\limsup_{n \rightarrow \infty} x_n = x^*$ , falls folgende zwei Bedingungen erfüllt sind:

- (i) es gibt eine Teilfolge  $x_{n_k}$  mit  $x_{n_k} \rightarrow x^*$  für  $k \rightarrow \infty$ ,
- (ii) für alle  $x > x^*$  ist die Menge  $\{n \in \mathbb{N} : x_n > x\}$  endlich.

Entsprechend bedeutet  $\liminf_{n \rightarrow \infty} x_n = x_*$  mit  $x_* \in \mathbb{R} \cup \{\pm\infty\}$ :

- (i) es gibt eine Teilfolge  $x_{n_k}$  mit  $x_{n_k} \rightarrow x_*$  für  $k \rightarrow \infty$ ,
- (ii) für alle  $x < x_*$  ist die Menge  $\{n \in \mathbb{N} : x_n < x\}$  endlich.

Um zu sehen, wie die Definition funktioniert, betrachten wir ein einfaches

**Beispiel 5.8** Sei  $x_n = (-1)^n + \frac{1}{n^2}$ . Dann gilt  $x_{2k} = 1 + \frac{1}{4k^2} \rightarrow 1$ , also Bedingung (i). Für  $x > 1$  folgern wir

$$x_n > x \quad \Rightarrow \quad \frac{1}{n^2} > x - (-1)^n \geq x - 1 \quad \Rightarrow \quad n^2 < \frac{1}{x-1}.$$

Also ist Bedingung (ii) ebenfalls erfüllt, es gilt also  $\limsup_{n \rightarrow \infty} x_n = 1$ .

Ist  $x^* = \limsup_{n \rightarrow \infty} x_n \in \mathbb{R}$ , so ist  $x^*$  der größte Häufungspunkt. Denn zu  $x > x^*$  gibt es ein  $\varepsilon > 0$  mit  $x - \varepsilon > x^*$ , also nach (ii)

$$\{n \in \mathbb{N} : x_n \in B_\varepsilon(x)\} \subset \{n \in \mathbb{N} : x_n > x - \varepsilon\} = \text{endlich.}$$

Im Fall  $\limsup_{n \rightarrow \infty} x_n = -\infty$  gilt schon  $\lim_{n \rightarrow \infty} x_n = -\infty$ , wieder nach (ii). Dagegen muss im Fall  $\limsup_{n \rightarrow \infty} x_n = +\infty$  nicht die ganze Folge konvergieren, etwa  $x_n = (-1)^n n$ . Während wir den Grenzwert nur für konvergente Folgen haben, sind  $\limsup$  bzw.  $\liminf$  für jede beliebige Folge  $x_n$  definiert. Das wollen wir nun zeigen, wobei wir uns auf den Limes superior beschränken.

**Satz 5.8 (Existenz des Limes superior)** Für jede reelle Folge  $(x_n)_{n \in \mathbb{N}}$  gibt es genau ein  $x^* \in \mathbb{R} \cup \{\pm\infty\}$  mit  $\limsup_{n \rightarrow \infty} x_n = x^*$ .

BEWEIS: Wir behandeln erst die Fälle wenn  $\limsup_{n \rightarrow \infty} x_n = \pm\infty$  ist.

**Fall 1:**  $(x_n)$  ist nicht nach oben beschränkt.

Dann ist  $\{n : x_n \geq b\}$  unendlich für alle  $b \in \mathbb{R}$ . Bestimme induktiv  $n_1 < n_2 < \dots$  mit  $x_{n_k} \geq k$ . Es folgt  $\limsup_{n \rightarrow \infty} x_n = +\infty$ .

**Fall 2:** Es gibt ein  $b_1 \in \mathbb{R}$  mit  $x_n \leq b_1$  für alle  $n \in \mathbb{N}$ .

**Fall 2.1:**  $\{n : x_n \geq a\}$  ist endlich für alle  $a \in \mathbb{R}$ .

Dann gilt  $x_n \rightarrow -\infty$  und es folgt  $\limsup_{n \rightarrow \infty} x_n = -\infty$ .

**Fall 2.2:** Es gibt ein  $a_1 \in \mathbb{R}$ , so dass  $\{n : x_n \in [a_1, b_1]\}$  unendlich ist.

Sei  $[a_k, b_k]$  die in Satz 5.7 konstruierte Intervallschachtelung und  $x^*$  der zugehörige Häufungspunkt. Bedingung (i) aus Definition 5.7 ist also erfüllt, wir behaupten

$$\{n \in \mathbb{N} : x_n > b_k\} \text{ ist endlich für jedes } k = 1, 2, \dots$$

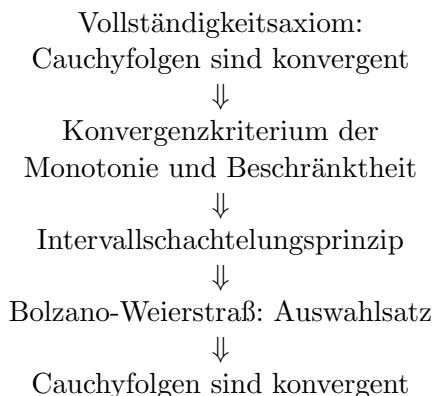
Für  $k = 1$  ist das richtig, da  $b_1$  obere Schranke ist (Fall 2). Induktiv ergibt sich, je nachdem ob bei der Halbierung das rechte oder das linke Teilintervall gewählt wird:

$$\begin{aligned} b_{k+1} = b_k &\Rightarrow \{n : x_n > b_{k+1}\} = \{n : x_n > b_k\} = \text{endlich (Induktion),} \\ b_{k+1} = m_k &\Rightarrow \{n : x_n > b_{k+1}\} = \{n : x_n \in (m_k, b_k]\} \cup \{n : x_n > b_k\} = \text{endlich,} \\ &\text{(Fallunterscheidung und Induktion).} \end{aligned}$$

Nun gilt  $b_k \rightarrow x^*$  für  $k \rightarrow \infty$ . Ist  $x > x^*$ , so ist  $b_k < x$  für  $k$  hinreichend groß. Es folgt  $\{n : x_n > x\} \subset \{n : x_n > b_k\}$ , also ist  $\{n : x_n > x\}$  endlich. Dies zeigt  $\limsup_{n \rightarrow \infty} x_n = x^*$ , womit die Existenz des Limes superior bewiesen ist.

Angenommen es gibt  $x_1^* < x_2^*$  mit beiden Eigenschaften (i), (ii) aus Definition 5.7. Wähle  $x \in (x_1^*, x_2^*)$ . Wegen (ii) für  $x_1^*$  ist dann  $\{n : x_n > x\}$  endlich. Dann kann aber (i) für  $x_2^*$  nicht gelten, ein Widerspruch.  $\square$

Die Begriffe Häufungspunkt, Limes superior und Limes inferior sind gewöhnungsbedürftig, und wir werden bei Gelegenheit mehr Beispiele betrachten. Die logische Abfolge der *zentralen theoretischen Aussagen* in diesem Abschnitt war folgende:



Die Implikationen sind so gemeint, dass in jedem Beweis nur die direkt vorangehende Eigenschaft von  $\mathbb{R}$  benutzt wurde. Die letzte Implikation werden wir gleich noch zeigen. Es folgt, dass jede der vier Eigenschaften als Vollständigkeitsaxiom benutzt werden könnte - die anderen würden dann als Sätze folgen. Im nächsten Abschnitt werden wir eine weitere äquivalente Eigenschaft kennenlernen, nämlich den Satz vom Supremum (Satz 6.1). Diese Eigenschaft wird in vielen Texten als Axiom der Vollständigkeit eingeführt, wir haben uns aber für die Konvergenz der Cauchyfolgen als grundlegendes Axiom entschieden.

BEWEIS: *Auswahlsatz von Bolzano-Weierstraß*  $\Rightarrow$  *Cauchyfolgen sind konvergent*

Sei  $(a_n)_{n \in \mathbb{N}}$  Cauchyfolge. Dann ist  $(a_n)$  beschränkt, denn es gibt ein  $N \in \mathbb{N}$  mit  $|a_n - a_m| \leq 1$  für  $n, m \geq N$ , also

$$|a_n| \leq |a_n - a_N| + |a_N| \leq 1 + |a_N| \quad \text{für alle } n \geq N.$$

Nach Bolzano-Weierstraß gibt es eine Teilfolge mit  $a_{n_k} \rightarrow a$  für  $k \rightarrow \infty$ . Wir zeigen, dass die ganze Folge  $a_n$  gegen  $a$  konvergiert. Dazu schätze ab

$$|a_n - a| \leq |a_n - a_{n_k}| + |a_{n_k} - a| < \varepsilon + |a_{n_k} - a| \quad \text{für } n, n_k > K.$$

Hier wurde benutzt, dass  $(a_n)$  eine Cauchyfolge ist. Mit  $k \rightarrow \infty$ , also  $n_k \rightarrow \infty$ , folgt

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

Damit ist die Implikation bewiesen. □

## 6 Teilmengen von $\mathbb{R}$ und von $\mathbb{R}^n$

Dieses Kapitel ist ein Sammelsurium zum Thema Mengen im  $\mathbb{R}^n$ . Im ersten Teil geht es um die Existenz von Supremum und Infimum von Teilmengen von  $\mathbb{R}$ . Dies liefert eine weitere äquivalente Charakterisierung der Vollständigkeit. Zweites Thema ist die Mächtigkeit der reellen Zahlen:  $\mathbb{N}$ ,  $\mathbb{Z}$  und  $\mathbb{Q}$  sind abzählbar,  $\mathbb{R}$  dagegen überabzählbar. Im letzten Teil des Kapitels verallgemeinern wir das Konzept der Konvergenz auf Folgen im  $\mathbb{R}^n$ , dazu definieren wir die Euklidische Norm und zeigen die Ungleichung von Cauchy-Schwarz. Diese Begriffe finden auch Anwendung in den komplexen Zahlen  $\mathbb{C} \cong \mathbb{R}^2$ .

Die für Folgen eingeführten Begriffe gelten ähnlich auch für Teilmengen von  $\mathbb{R}$ , mit kleinen Abwandlungen.

**Definition 6.1** Die Menge  $M \subset \mathbb{R}$  heißt

$$\begin{aligned} \text{nach oben beschränkt} &\Leftrightarrow \text{es gibt ein } b \in \mathbb{R} \text{ mit } x \leq b \text{ für alle } x \in M, \\ \text{nach unten beschränkt} &\Leftrightarrow \text{es gibt ein } a \in \mathbb{R} \text{ mit } x \geq a \text{ für alle } x \in M. \end{aligned}$$

$b$  bzw.  $a$  heißt dann obere bzw. untere Schranke. Weiter heißt

$$M \text{ beschränkt} \Leftrightarrow \text{es gibt ein } K \geq 0 \text{ mit } |x| \leq K \text{ für alle } x \in M.$$

Eine Menge ist genau dann beschränkt, wenn sie nach oben und unten beschränkt ist. Denn aus  $|x| \leq K$  folgt  $-K \leq x \leq K$ , und aus  $a \leq x \leq b$  folgt umgekehrt  $|x| \leq \max(|a|, |b|)$ .

**Beispiel 6.1** Die Menge  $[0, 1)$  ist nach oben beschränkt, eine obere Schranke ist zum Beispiel  $K = 2025$ . Es gibt aber in  $[0, 1)$  kein größtes Element, denn es gilt der Schluss

$$x \in [0, 1) \Rightarrow x < \frac{x+1}{2} \in [0, 1).$$

Unter den oberen Schranken von  $[0, 1)$  gibt es jedoch eine kleinste, nämlich die Zahl 1. Wir bemerken dass diese kleinste obere Schranke nicht Element der Menge  $[0, 1)$  ist.

**Definition 6.2 (Supremum/Infimum)** Die Zahl  $a \in \mathbb{R} \cup \{\pm\infty\}$  heißt Supremum (bzw. Infimum) der Menge  $M \subset \mathbb{R}$ , wenn folgendes gilt:

- (1)  $x \leq a$  (bzw.  $x \geq a$ ) für alle  $x \in M$ ,
- (2) Für alle  $a' < a$  (bzw.  $a' > a$ ) gibt es ein  $x \in M$  mit  $x > a'$  ( $x < a'$ ).

*Notation:*  $a = \sup M$  (bzw.  $a = \inf M$ ). Bedingung (1) besagt, dass  $a$  eine obere Schranke ist, nach Bedingung (2) gibt es keine kleinere obere Schranke. Deshalb wird  $\sup M$  auch als kleinste obere Schranke bezeichnet, und analog  $\inf M$  als größte untere Schranke.

**Satz 6.1** Jede Menge  $M \subset \mathbb{R}$  hat genau ein Supremum  $S \in \mathbb{R} \cup \{\pm\infty\}$ .

**BEWEIS:** Wir zeigen als erstes die Eindeutigkeit. Angenommen es gibt  $S_{1,2} \in \mathbb{R} \cup \{\pm\infty\}$ , die beide die Definition 6.2 erfüllen; wir können  $S_1 < S_2$  annehmen. Nach Eigenschaft (2) bzgl.  $S_2$  gibt es ein  $x \in M$  mit  $x > S_1$ , im Widerspruch zur Eigenschaft (1) bezüglich  $S_1$ .

Man sieht leicht  $\sup \emptyset = -\infty$ , und  $\sup M = +\infty$  falls  $M$  nicht nach oben beschränkt

ist. Im verbleibenden Fall wählen wir ein Element  $a_1$  von  $M$  und eine obere Schranke  $b_1$  von  $M$ , und konstruieren wie folgt eine Intervallschachtelung  $I_n = [a_n, b_n]$  für  $n \geq 1$ , wobei  $m_n = (a_n + b_n)/2$ :

$$I_{n+1} = [a_{n+1}, b_{n+1}] = \begin{cases} [m_n, b_n], & \text{falls } [m_n, b_n] \cap M \neq \emptyset \\ [a_n, m_n], & \text{sonst.} \end{cases}$$

Nach dem Intervallschachtelungsprinzip, Satz 5.5, gibt es genau ein  $S \in \mathbb{R}$  mit  $S \in I_n$  für alle  $n$ , und genauer gilt  $a_n \rightarrow S$  sowie  $b_n \rightarrow S$ . Für  $x \in M$  sieht man durch Induktion  $x \leq b_n$  für alle  $n$ , also  $x \leq \lim_{n \rightarrow \infty} b_n = S$ . Andererseits gilt, ebenfalls induktiv,  $M \cap I_n \neq \emptyset$  für alle  $n$ , das heißt es gibt  $x_n \in M$  mit  $a_n \leq x_n \leq b_n$ . Ist  $S' < S$ , so gilt also  $x_n > S'$  für hinreichend große  $n$ . Damit ist  $\sup M = S$  gezeigt.  $\square$

**Folgerung 6.1** Sei  $M \subset \mathbb{R}$  nichtleer. Dann gibt es eine Folge  $x_n \in M$  (bzw.  $x'_n \in M$ ) mit  $x_n \rightarrow \sup M$  (bzw.  $x'_n \rightarrow \inf M$ ).

BEWEIS: Wir zeigen die Aussage für das Supremum. Ist  $M$  nach oben beschränkt, so wurde eine solche Folge im Beweis des vorangehenden Satzes konstruiert. Ist  $M$  nicht nach oben beschränkt, so gibt es zu  $n \in \mathbb{N}$  ein  $x_n \in M$  mit  $x_n \geq n$ , also  $x_n \rightarrow +\infty = \sup M$ .  $\square$

Viele Autoren verwenden die Existenz des Supremums als Vollständigkeitsaxiom. Das Archimedische Axiom (A3) ist dann eine Folgerung: es gibt ein  $n \in \mathbb{N}$  mit  $n > 1/\varepsilon$ . Andernfalls ist  $S = \sup \mathbb{N} \leq 1/\varepsilon$ , und es gibt ein  $n \in \mathbb{N}$  mit  $n > S - 1$ , also  $n + 1 > S$ , Widerspruch. Außerdem impliziert die Existenz des Supremums das Konvergenzkriterium der Monotonie und Beschränktheit, wie man leicht sieht.

Jetzt zu einem ganz anderen Thema: wie kann man die unendlichen Mengen  $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$  im Bezug auf ihre Größe vergleichen? Gibt es wirklich mehr rationale Zahlen als natürliche? Mehr reelle als rationale? Die Antwort lautet witzigerweise, dass es gleich viele natürliche, ganze und rationale Zahlen gibt, aber mehr reelle Zahlen. Endliche Mengen haben genau dann gleichviele Elemente, wenn es zwischen ihnen eine Bijektion gibt. Darauf stützt sich auch der Vergleich bei unendlichen Mengen, der von Georg Cantor 1872 eingeführt wurde.

**Definition 6.3** Eine Menge  $A$  heißt gleichmächtig zur Menge  $B$  (Notation:  $A \sim B$ ), wenn es eine bijektive Abbildung (Bijektion)  $\varphi : A \rightarrow B$  gibt.

**Lemma 6.1** Die Relation  $A \sim B$  ( $A$  ist gleichmächtig zu  $B$ ) ist eine Äquivalenzrelation, das heißt für alle Mengen  $A, B, C$  gilt:

- (i)  $A \sim A$
- (ii)  $A \sim B \Rightarrow B \sim A$
- (iii)  $A \sim B$  und  $B \sim C \Rightarrow A \sim C$ .

BEWEIS:

- (i) Wähle als Bijektion die Identität  $id_A : A \rightarrow A$ ,  $a \mapsto a$ .
- (ii) Sei  $\varphi : A \rightarrow B$  Bijektion. Wähle dann  $\varphi^{-1} : B \rightarrow A$ .
- (iii) Seien  $\varphi : A \rightarrow B$  und  $\psi : B \rightarrow C$  bijektiv. Wähle dann  $\psi \circ \varphi : A \rightarrow C$ .

$\square$

Durch die Relation *gleichmächtig* werden die Mengen in Äquivalenzklassen eingeteilt. Für jedes  $n \in \mathbb{N}_0$  hat man die Klasse der Mengen mit  $n$  Elementen, ein Repräsentant ist die Menge  $\{1, \dots, n\}$ . Eine weitere Äquivalenzklasse liefert die Menge  $\mathbb{N}$ , denn nach dem Schubfachprinzip, Satz 3.3, gibt es für  $n \in \mathbb{N}$  keine Bijektion  $\mathbb{N} \leftrightarrow \{1, \dots, n\}$ .

**Definition 6.4** Eine Menge  $A$  heißt

- abzählbar unendlich, wenn sie gleichmächtig zu  $\mathbb{N}$  ist;
- abzählbar, wenn sie endlich oder abzählbar unendlich ist;
- überabzählbar, wenn sie nicht abzählbar ist.

**Lemma 6.2** Falls eine surjektive Abbildung  $\varphi : \mathbb{N} \rightarrow A$  existiert, so ist  $A$  abzählbar.

BEWEIS: Sei  $a_n = \varphi(n)$ . Die naheliegende Idee ist, bei der Abzählung induktiv diejenigen  $a_n$  wegzuworfen, die bereits vorher auftraten, und die übrigen entsprechend zu nummerieren. Dazu setzen wir  $n_1 = 1$  und konstruieren  $n_1 < n_2 < \dots$  induktiv durch

$$n_j = \min\{n > n_{j-1} : a_n \neq a_{n_i} \text{ für alle } i < j\} \quad \text{für } j \geq 2. \quad (6.1)$$

Falls die Rekursion nicht abbricht erhalten wir  $f : \mathbb{N} \rightarrow A$ ,  $f(j) = a_{n_j}$ . Die Abbildung ist injektiv: nach (6.1) gilt  $f(j) = a_{n_j} \neq a_{n_i} = f(i)$  für alle  $i < j$ . Weiter gibt es zu  $n \in \mathbb{N}$  ein  $i \in \mathbb{N}$  mit  $a_n = f(i)$ . Im Fall  $n = n_j$  gilt das mit  $i = j$ . Ist  $n_{j-1} < n < n_j$  mit  $j \geq 2$  so gibt es nach (6.1) ein  $i < j$  mit  $a_n = a_{n_i} = f(i)$ . Damit ist  $f$  bijektiv.

Falls die Rekursion nach  $n_k$  abbricht, erhalten wir  $f : \{1, \dots, k\} \rightarrow A$ ,  $f(j) = a_{n_j}$ . Man sieht wie oben dass  $f$  injektiv ist. Der Abbruch besagt wegen (6.1): zu  $n > n_k$  gibt es ein  $i \leq k$  mit  $a_n = a_{n_i} = f(i)$ . Für  $n \leq n_k$  folgt das wie oben, also ist  $f$  bijektiv.  $\square$

**Satz 6.2** Die Mengen  $\mathbb{Z}$  und  $\mathbb{Q}$  sind abzählbar.

BEWEIS: Für  $\mathbb{Z}$  wähle die surjektive (sogar bijektive) Abbildung

$$\varphi : \mathbb{N} \rightarrow \mathbb{Z}, \varphi(n) = \begin{cases} (n-1)/2 & n \text{ ungerade} \\ -n/2 & n \text{ gerade.} \end{cases}$$

Das Argument zeigt, dass wir uns für  $\mathbb{Q}$  auf die Menge  $\mathbb{Q}^+ = \{p/q : p, q \in \mathbb{N}\}$  beschränken können. Wir konstruieren nun eine surjektive Abbildung

$$\varphi : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N} = \{(p, q) : p, q \in \mathbb{N}\}.$$

Durch Verkettung mit  $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q}^+$ ,  $f(p, q) = p/q$ , erhalten wir dann eine Surjektion von  $\mathbb{N}$  nach  $\mathbb{Q}^+$ , und der Satz ist bewiesen. Die Idee ist, in folgendem Schema die Diagonalen

nacheinander von rechts oben nach links unten abzuzählen:

$p =$	1	2	3	4	5	6
$q =$						
1	(1, 1)	(2, 1)	(3, 1)	(4, 1)	(5, 1)	(6, 1)
2	(1, 2)	(2, 2)	(3, 2)	(4, 2)	.	
3	(1, 3)	(2, 3)	(3, 3)	.	.	
4	(1, 4)	(2, 4)	.	.	.	
5	(1, 5)	.	.	.	.	

Der Eintrag  $(p, q)$  steht in der Diagonale mit Nummer  $k = p + q - 1$ , und zwar an der  $q$ -ten Stelle von rechts oben. Die Diagonalen mit Nummern  $1, \dots, k - 1$  enthalten insgesamt  $1 + \dots + (k - 1) =: N_k$  Einträge ( $N_k = k(k - 1)/2$ ). Damit bekommt  $(p, q)$  die Nummer

$$n(p, q) = N_k + q \quad \text{wobei } k = p + q - 1 \in \mathbb{N}, 1 \leq q \leq k. \quad (6.2)$$

Um  $\varphi : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$  zu definieren, müssen wir umgekehrt zu  $n \in \mathbb{N}$  den Eintrag  $(p, q)$  bestimmen. Das machen wir in drei Schritten:

- es gibt genau ein  $k \in \mathbb{N}$  mit  $N_k < n \leq N_{k+1}$ . Beachte dazu  $N_1 = 0$  und  $N_1 < N_2 < \dots$
- setze  $q = n - N_k$ . Wegen  $N_{k+1} - N_k = k$  gilt also  $1 \leq q \leq k$ .
- setze  $p = k - q + 1$ , also  $1 \leq p \leq k$  und  $k = p + q - 1$ .

Wir definieren nun  $\varphi(n) = (p, q)$ . Man sieht leicht dass  $\varphi(n(p, q)) = (p, q)$ , also ist  $\varphi$  surjektiv und damit  $\mathbb{N} \times \mathbb{N}$  abzählbar. Das Argument zeigt allgemein, dass eine abzählbare Vereinigung von jeweils abzählbaren Mengen wiederum abzählbar ist.  $\square$

**Satz 6.3 ( $\mathbb{R}$  ist nicht abzählbar)** *Es gibt keine surjektive Abbildung  $\varphi : \mathbb{N} \rightarrow \mathbb{R}$ .*

BEWEIS: Sei  $\varphi : \mathbb{N} \rightarrow \mathbb{R}$  eine beliebige Abbildung und  $\varphi(n) = x_n$ . Wir konstruieren eine Intervallschachtelung  $(I_n)_{n \in \mathbb{N}}$  mit  $x_n \notin I_n$  für jedes  $n$ . Nach Satz 5.5 gibt es dann ein  $x \in \mathbb{R}$  mit  $x \in I_n$  für alle  $n$ , insbesondere folgt  $x \neq x_n$  für alle  $n$ , das heißt  $\varphi$  ist nicht surjektiv. Hier die Konstruktion: ist  $I_n$  schon bestimmt, so zerlege  $I_n$  in drei abgeschlossene Teilintervalle gleicher Länge und wähle für  $I_{n+1}$  ein Teilintervall, das  $x_{n+1}$  nicht enthält (im Zweifelsfall das rechte). Um  $I_1$  zu definieren, wende das Argument an auf  $I_0 = [0, 1]$ .  $\square$

Die Menge  $\mathbb{R}$  repräsentiert also eine weitere Äquivalenzklasse der Relation gleichmächtig, die als Mächtigkeit des Kontinuums bezeichnet wird. Cantor hat 1878 vermutet, dass jede Teilmenge von  $\mathbb{R}$  entweder abzählbar ist oder schon gleichmächtig zu  $\mathbb{R}$ . Es hat sich aber mit Arbeiten von Gödel (1938) und Cohen (1963) ergeben, dass diese sogenannte Kontinuumshypothese im Rahmen unserer Mengensexiomatik (die wir nicht behandelt haben) nicht entschieden werden kann, ein merkwürdiges Ergebnis.

Bisher haben wir nur Folgen und Teilmengen in  $\mathbb{R}$  betrachtet. Um später die Exponentialfunktion und die trigonometrischen Funktionen zu verstehen brauchen wir aber auch Folgen

und Teilmengen in  $\mathbb{C}$ , den komplexen Zahlen. Hier führen wir diese kurz ein.

In der Euklidischen Ebene gibt es neben der Vektoraddition eine Multiplikation, und mit dieser wird  $\mathbb{R}^2$  zum Körper  $\mathbb{C} = (\mathbb{R}^2, +, \cdot)$  der komplexen Zahlen. Dazu wird  $\mathbb{R}$  als Teilmenge des  $\mathbb{R}^2$  aufgefasst, indem  $x \in \mathbb{R}$  mit dem Punkt  $(x, 0) \in \mathbb{R}^2$  identifiziert wird. Schreibt man weiter  $(0, 1) = i$ , so sieht die Darstellung von  $(x, y) \in \mathbb{R}^2$  bezüglich der Standardbasis wie folgt aus:

$$(x, y) = (x, 0) + (0, 1)y = x + iy.$$

Für  $z = x + iy$  heißt  $x = \operatorname{Re} z$  der Realteil und  $y = \operatorname{Im} z$  der Imaginärteil von  $z$ . Die Addition der Vektoren  $x_1 + iy_1$  und  $x_2 + iy_2$  ist komponentenweise definiert, also

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

Die Multiplikation ist wie folgt gegeben: auf  $\mathbb{R}$  wählen wir die gegebene Multiplikation. Als wesentliche Regel fordern wir dann  $i^2 = -1$ . Damit ist die Multiplikation allgemein festgelegt, denn durch Ausmultiplizieren folgt

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 + i^2y_1y_2 + i(x_1y_2 + x_2y_1) = x_1x_2 - y_1y_2 + i(x_1y_2 + x_2y_1).$$

Multiplikation mit  $\lambda \in \mathbb{R}$  ist die Skalarmultiplikation im Vektorraum  $\mathbb{R}^2$ , denn

$$\lambda(x + iy) = \lambda x + i\lambda y = (\lambda x, \lambda y) = \lambda(x, y).$$

Multiplikation mit  $i$  ergibt dagegen

$$i(x + iy) = -y + ix = (-y, x).$$

Der Vektor  $(-y, x)$  entsteht aus  $(x, y)$  durch Drehung um  $90^\circ$  im mathematisch positiven Sinn, das heißt gegen den Uhrzeigersinn. Die Körpergesetze in  $\mathbb{C}$  lassen sich leicht verifizieren, nur das Assoziativgesetz der Multiplikation erfordert etwas Rechenarbeit. Für das inverse Element der Multiplikation ist ein weiterer Begriff nützlich: für  $z = x + iy \in \mathbb{C}$  heißt  $\bar{z} = x - iy \in \mathbb{C}$  die zu  $z$  konjugiert komplexe Zahl. Anschaulich ergibt sich  $\bar{z}$  aus  $z$  durch Spiegelung an der  $x$ -Achse, insbesondere gilt  $\bar{\bar{z}} = \overline{x - iy} = x + iy = z$ .

**Lemma 6.3** *Für die komplexe Konjugation gelten folgende Regeln:*

$$(1) \quad \overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2,$$

$$(2) \quad \bar{z} = z \quad \Leftrightarrow \quad z \in \mathbb{R},$$

$$(3) \quad \text{Für } z = x + iy \text{ ist } \operatorname{Re} z = \frac{1}{2}(z + \bar{z}) \text{ und } \operatorname{Im} z = \frac{1}{2i}(z - \bar{z}).$$

BEWEIS: Die Beweise erfolgen alle durch Nachrechnen, zum Beispiel gilt

$$\overline{z_1 z_2} = (x_1 - iy_1)(x_2 - iy_2) = x_1x_2 - y_1y_2 - i(x_1y_2 + y_1x_2) = \overline{z_1 z_2}.$$

□

Der Betrag von  $z = x + iy$  ist definiert als Euklidische Länge des Vektors  $(x, y)$ , das heißt

$$|z| = \sqrt{x^2 + y^2} \quad \text{für } z = x + iy.$$

**Lemma 6.4** Für den Betrag gilt:

- (1)  $|z|^2 = z \bar{z}$ .
- (2)  $|z_1 z_2| = |z_1| |z_2|$ .
- (3)  $|\operatorname{Re} z| \leq |z|$ ,  $|\operatorname{Im}(z)| \leq |z|$ .

BEWEIS: Für Aussage (1) berechnen wir mit  $z = x + iy$

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

Die Formel (2) folgt aus (1), Lemma 6.3(1) und den Körpergesetzen:

$$|z_1 z_2|^2 = (z_1 z_2)(\overline{z_1 z_2}) = z_1 \bar{z}_1 z_2 \bar{z}_2 = |z_1|^2 |z_2|^2.$$

Die Ungleichungen (3) sind offensichtlich. □

Damit können wir die Zahl  $1/z$  leicht hinschreiben, und zwar folgt aus Lemma 6.4(1)

$$\frac{1}{z} = \frac{\bar{z}}{z \bar{z}} = \frac{x - iy}{x^2 + y^2} \quad \text{für } z = x + iy \neq 0. \quad (6.3)$$

Wir kommen nun zur Konvergenz in  $\mathbb{C}$ . Für diesen Begriff spielt die multiplikative Struktur keine Rolle, sondern nur die Darstellung  $z = (x, y)$  als Punkte im  $\mathbb{R}^2$ . Wir behandeln die Konvergenz gleich für Punkte im  $\mathbb{R}^n$ , das wird später sicher gebraucht. Nach Definition ist

$$\mathbb{R}^n = \{x = (x_1, \dots, x_n) : x_i \in \mathbb{R}\} = \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ mal}}.$$

Die Definition der Konvergenz in  $\mathbb{R}$  verwendet den Betrag bzw. den daraus abgeleiteten Abstand  $|x_k - x|$  von  $x_k, x \in \mathbb{R}$ , Konvergenz bedeutet dass dieser Abstand gegen Null geht. Im  $\mathbb{R}^n$  ersetzen wir den Betrag durch die Euklidische Norm.

**Definition 6.5** Die Euklidische Norm eines Vektors  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  ist

$$\|x\| = \sqrt{x_1^2 + \dots + x_n^2}.$$

Der Euklidische Abstand von  $x, y \in \mathbb{R}^n$  ist  $\|x - y\|$ .

Die Doppelstriche  $\|\cdot\|$  sollen die Norm vom Betrag  $|\cdot|$  in  $\mathbb{R}$  abgrenzen, allerdings sind auch einfache Striche üblich, das gilt vor allem für den Fall  $\mathbb{R}^2 \cong \mathbb{C}$ . Das Argument der Wurzel ist als Summe von Quadraten nichtnegativ, also ist  $\|x\|$  definiert. Der folgende Satz fasst wesentliche Eigenschaften der Euklidischen Norm zusammen.

**Satz 6.4** Für die Euklidische Norm  $\|x\|$  gilt:

- (1) Positivität:  $\|x\| \geq 0$  mit Gleichheit genau wenn  $x = 0$ .
- (2) Halblinearität:  $\|\lambda x\| = |\lambda| \|x\|$  für alle  $\lambda \in \mathbb{R}$ ,  $x \in \mathbb{R}^n$ .
- (3) Dreiecksungleichung:  $\|x + y\| \leq \|x\| + \|y\|$  für alle  $x, y \in \mathbb{R}^n$ .

Gleichheit in (3) gilt genau wenn  $x$  und  $y$  gleichsinnig parallel sind.

BEWEIS: Nach Definition der Wurzel ist  $\|x\| \geq 0$ . Im Gleichheitsfall ist  $x_1^2 + \dots + x_n^2 = 0$ , also  $x_i = 0$  für alle  $i = 1, \dots, n$  und damit  $x = 0$ . Die Skalarmultiplikation im  $\mathbb{R}^n$  ist komponentenweise definiert durch  $\lambda(x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$ , also folgt

$$\|\lambda x\| = \left( \sum_{i=1}^n (\lambda x_i)^2 \right)^{1/2} = \left( \lambda^2 \sum_{i=1}^n x_i^2 \right)^{1/2} = |\lambda| \left( \sum_{i=1}^n x_i^2 \right)^{1/2} = |\lambda| \|x\|.$$

Die Ungleichung (3) erfordert etwas Vorarbeit. Wir zeigen erst die Ungleichung von Cauchy-Schwarz, dazu brauchen wir das Euklidische Skalarprodukt.

**Definition 6.6** Das Standardskalarprodukt von  $x, y \in \mathbb{R}^n$  ist

$$\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n.$$

Offenbar gilt  $\langle x, x \rangle = \|x\|^2$  für  $x \in \mathbb{R}^n$ .

**Lemma 6.5** Für das Standardskalarprodukt  $\langle x, y \rangle$  gilt:

- (1) Positivität:  $\langle x, x \rangle \geq 0$  mit Gleichheit genau wenn  $x = 0$ .
- (2) Symmetrie:  $\langle x, y \rangle = \langle y, x \rangle$  für alle  $x, y \in \mathbb{R}^n$ .
- (3) Bilinearität: Für alle  $x, y, z \in \mathbb{R}^n$  und  $\lambda, \mu \in \mathbb{R}$  gilt

$$\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle \quad \text{und} \quad \langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle.$$

BEWEIS: Die Positivität folgt aus der Positivität der Norm, und die Symmetrie ist offensichtlich. Auch die Bilinearität ergibt sich leicht aus der Definition:

$$\langle \lambda x + \mu y, z \rangle = \sum_{i=1}^n (\lambda x_i + \mu y_i) z_i = \lambda \sum_{i=1}^n x_i z_i + \mu \sum_{i=1}^n y_i z_i = \lambda \langle x, z \rangle + \mu \langle y, z \rangle.$$

□

Im folgenden Beweis verwenden wir die Normierung eines Vektors: ist  $x \in \mathbb{R}^n$ ,  $x \neq 0$ , so hat  $\xi = x/\|x\|$  dieselbe Richtung aber Norm Eins. Denn mit Satz 6.4(2) (Halblinearität) gilt

$$\|\xi\| = \left\| \frac{1}{\|x\|} x \right\| = \frac{1}{\|x\|} \|x\| = 1.$$

**Satz 6.5 (Ungleichung von Cauchy-Schwarz)** Für  $x, y \in \mathbb{R}^n$  gilt

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \tag{6.4}$$

Gleichheit gilt genau wenn  $x$  und  $y$  linear abhängig sind.

BEWEIS: Wir berechnen erst für  $\xi, \eta \in \mathbb{R}^n$  normiert, das heißt  $\|\xi\|, \|\eta\| = 1$ ,

$$0 \leq \frac{1}{2} \|\xi - \eta\|^2 = \frac{1}{2} (\|\xi\|^2 + \|\eta\|^2) - \langle \xi, \eta \rangle = 1 - \langle \xi, \eta \rangle, \quad \text{also } \langle \xi, \eta \rangle \leq 1.$$

Gleichheit gilt genau wenn  $\eta = \xi$ . Für beliebige  $x, y \in \mathbb{R}^n$  folgt die Ungleichung von Cauchy-Schwarz durch Skalierung: ist  $x = 0$  oder  $y = 0$  so sind beide Seiten gleich Null. Andernfalls setze  $\xi = x/\|x\|$  und  $\eta = y/\|y\|$ , dann folgt

$$\langle x, y \rangle = \langle \|x\| \xi, \|y\| \eta \rangle = \|x\| \|y\| \langle \xi, \eta \rangle \leq \|x\| \|y\|.$$

Dies zeigt (6.4) wenn  $\langle x, y \rangle \geq 0$ , und durch Anwendung auf  $-y$  auch wenn  $\langle x, y \rangle \leq 0$ . Bei Gleichheit ist entweder  $x = 0$  bzw.  $y = 0$ , oder es gilt  $\eta = \pm\xi$ , also  $x, y$  parallel.  $\square$

Wir können nun leicht den Beweis von Satz 6.4 abschließen:

*Beweis der Dreiecksungleichung:* Mit Satz 6.5 gilt

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

Mit Monotonie der Wurzel folgt  $\|x + y\| \leq \|x\| + \|y\|$  wie behauptet. Im Gleichheitsfall müssen  $x, y$  linear abhängig sein und zusätzlich  $\langle x, y \rangle \geq 0$ , also sind  $x, y$  gleichsinnig parallel.  $\square$

Zur Bezeichnung *Dreiecksungleichung*: für drei Punkte  $x, y, z$  folgt aus (3)

$$\|x - z\| = \|x - y + y - z\| \leq \|x - y\| + \|y - z\|,$$

das heißt im Dreieck mit den Eckpunkten  $x, y, z$  ist jede Seite kürzer als die Summe der beiden anderen Seiten. Im Gleichheitsfall sind  $x - y$  und  $y - z$  parallel, das heißt die Punkte  $x, y, z$  liegen auf einer Geraden, genauer liegt  $y$  zwischen  $x$  und  $z$ .

Im  $\mathbb{R}^n$  mit  $n \geq 2$  haben wir keine Anordnung zur Verfügung. Deshalb verwenden wir die Ungleichungszeichen  $<, >, \leq, \geq$ , Formulierungen wie "nach oben bzw. unten beschränkt", Supremum/Infimum, und Monotonie *ausschließlich* für reelle Zahlen und *niemals* für Punkte im  $\mathbb{R}^n$ . Eine Ungleichung  $a < b$  ist nur für reelle Zahlen sinnvoll (später auch mal für gewisse Matrizen, aber das tut jetzt nichts zur Sache). Der Begriff der Konvergenz beruht aber nicht auf der Anordnung. Wie angekündigt verallgemeinern wir ihn, indem wir den Betrag in  $\mathbb{R}$  durch die Euklidische Norm in  $\mathbb{R}^n$  ersetzen.

**Definition 6.7 (Konvergenz im  $\mathbb{R}^n$ )** Die Folge  $x_k \in \mathbb{R}^n$  konvergiert gegen  $a \in \mathbb{R}^n$ , falls gilt: für alle  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{R}$ , so dass  $\|x_k - a\| < \varepsilon$  für alle  $k > K$ .

Entsprechend ergeben sich auch die Konzepte Beschränktheit,  $\varepsilon$ -Umgebung, Cauchyfolge, und damit Vollständigkeit:

- $M \subset \mathbb{R}^n$  beschränkt:  $\|x\| \leq K$  für alle  $x \in M$ .
- $B_\varepsilon(x) = \{y \in \mathbb{R}^n : \|y - x\| < \varepsilon\}$ ,
- $x_k \in \mathbb{R}^n$  Cauchyfolge: für alle  $\varepsilon > 0$  gibt es ein  $K \in \mathbb{R}$  mit  $\|x_k - x_l\| < \varepsilon$  für  $k, l > K$ .

Jeder Punkt  $x \in \mathbb{R}^n$  ist durch seine  $n$  Koordinaten  $x_1, \dots, x_n$  gegeben, und Fragen der Konvergenz lassen sich auf die Konvergenz der Koordinaten reduzieren.

**Satz 6.6 (Euklidische Norm versus Koordinaten)** *Eine Folge  $x_k \in \mathbb{R}^n$  ist genau dann konvergent (beschränkt, Cauchyfolge), wenn für alle  $i = 1, \dots, n$  die Koordinatenfolgen  $((x_k)_i)_{k \in \mathbb{N}}$  konvergent (beschränkt, Cauchyfolgen) sind.*

BEWEIS: Für  $x \in \mathbb{R}^n$  gilt  $(\max_{1 \leq i \leq n} |x_i|)^2 \leq |x_1|^2 + \dots + |x_n|^2 \leq n (\max_{1 \leq i \leq n} |x_i|)^2$ . Wir setzen  $\max_{1 \leq i \leq n} |x_i| = \|x\|_{\max}$ , es gilt also

$$\|x\|_{\max} \leq \|x\| \leq \sqrt{n} \|x\|_{\max} \quad \text{für alle } x \in \mathbb{R}^n. \quad (6.5)$$

Sei nun  $x_k \rightarrow x$ , also  $\|x_k - x\| \rightarrow 0$ . Dann folgt  $|(x_k)_i - x_i| \leq \|x_k - x\|_{\max} \rightarrow 0$  für  $i = 1, \dots, n$ . Umgekehrt: angenommen  $(x_k)_i \rightarrow x_i$  für  $i = 1, \dots, n$ . Dann gibt es zu  $\varepsilon > 0$  ein  $K_i \in \mathbb{R}$  mit  $|(x_k)_i - x_i| < \varepsilon/\sqrt{n}$  für  $k > K_i$ . Mit  $x = (x_1, \dots, x_n)$  folgt

$$\|x_k - x\| \leq \sqrt{n} \|x_k - x\|_{\max} < \sqrt{n} \frac{\varepsilon}{\sqrt{n}} = \varepsilon \quad \text{für } k > \max_{1 \leq i \leq n} K_i =: K.$$

Die Argumente für die Beschränktheit und die Cauchyfolgen Eigenschaft sind ähnlich, wir behandeln sie in den Übungsaufgaben.  $\square$

**Folgerung 6.2** *Mit der Euklidischen Norm ist  $\mathbb{R}^n$  vollständig: zu jeder Cauchyfolge  $(x_k)_{k \in \mathbb{N}}$  gibt es ein  $a \in \mathbb{R}^n$ , so dass  $x_k \rightarrow a$  mit  $k \rightarrow \infty$ .*

BEWEIS: Nach Satz 6.6 sind die  $((x_k)_i)_{k \in \mathbb{N}}$  Cauchyfolgen in  $\mathbb{R}$ . Laut Vollständigkeitsaxiom gibt es  $a_i \in \mathbb{R}$  mit  $\lim_{k \rightarrow \infty} (x_k)_i = a_i$  für  $i = 1, \dots, n$ . Setze  $a = (a_1, \dots, a_n)$ . Nach Satz 6.6 folgt  $\lim_{k \rightarrow \infty} x_k = a$ .  $\square$



## 7 Stetigkeit

Der Begriff der stetigen Funktion ist ein fundamentales Konzept der Analysis. In der Schule wird oft die Charakterisierung gegeben, dass sich der Graph der Funktion zeichnen lässt ohne mit dem Stift abzusetzen. Dann kann jedenfalls die Funktion keinen Sprung haben, und das ist schon ganz zutreffend. Aber in mathematischen Argumenten können wir diese anschauliche Beschreibung nicht einsetzen, wie schon beim Konzept des Grenzwerts brauchen wir eine quantitative Fassung. Wir betrachten erst reelle Funktionen  $y = f(x)$  einer Variablen  $x \in D$  wobei  $D \subset \mathbb{R}$ . Später behandeln wir aber auch Funktionen mehrerer Variabler sowie vektorwertige Funktionen.

**Definition 7.1 (Stetigkeit)** Die Funktion  $f : D \rightarrow \mathbb{R}$ , wobei  $D \subset \mathbb{R}$ , heißt stetig in  $x_0 \in D$ , falls es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt, so dass gilt:

$$|f(x) - f(x_0)| < \varepsilon \quad \text{für alle } x \in D \text{ mit } |x - x_0| < \delta. \quad (7.1)$$

$f$  heißt stetig, falls  $f$  in allen  $x_0 \in D$  stetig ist.

Wir wollen einige einfache Beispiele betrachten.

**Beispiel 7.1** Eine konstante Funktion  $f : D \rightarrow \mathbb{R}^m$  ist in jedem  $x_0 \in D$  stetig, denn

$$|f(x) - f(x_0)| = 0 \quad \text{für alle } x \in D.$$

Zu gegebenem  $\varepsilon > 0$  können wir jedes  $\delta > 0$  wählen.

**Beispiel 7.2** Für  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = ax + b$  mit  $a \neq 0$ , berechnen wir

$$|f(x) - f(x_0)| = |(ax + b) - (ax_0 + b)| = |a| |x - x_0|.$$

Sei  $\varepsilon > 0$  gegeben. Dann wählen wir  $\delta = \varepsilon/|a| > 0$ , es folgt für  $|x - x_0| < \delta$

$$|f(x) - f(x_0)| = |a| |x - x_0| < |a| \delta = \varepsilon.$$

Also ist  $f$  stetig auf ganz  $\mathbb{R}$ . Die Wahl von  $\delta$  ist tatsächlich maximal, denn es gilt

$$|f(x_0 \pm \delta) - f(x_0)| = |a| |(x_0 \pm \delta) - x_0| = |a| \delta = \varepsilon.$$

Die Schranke  $\delta = \varepsilon/|a|$  hängt von  $a$  ab, sie wird klein für  $|a|$  groß. Dies entspricht der Tatsache dass die Gerade  $y = ax + b$  steiler wird.

**Beispiel 7.3** Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^2$ , ist auf ganz  $\mathbb{R}$  stetig. Es gilt

$$|f(x) - f(x_0)| = |x^2 - x_0^2| = |x + x_0| |x - x_0| \leq (|x| + |x_0|) |x - x_0|.$$

Um den ersten Faktor unabhängig von  $x$  abzuschätzen nehmen wir  $|x - x_0| \leq 1$  an. Dann gilt

$$|x| \leq |x - x_0| + |x_0| \leq 1 + |x_0|, \quad \text{also } |x| + |x_0| \leq 1 + 2|x_0|.$$

Jetzt wählen wir  $\delta = \min(1, \varepsilon/(1 + 2|x_0|))$ . Für  $|x - x_0| < \delta$ , insbesondere  $|x - x_0| \leq 1$ , folgt

$$|f(x) - f(x_0)| < (1 + 2|x_0|) \delta \leq \varepsilon.$$

Die Abhängigkeit  $\delta = \delta(x_0)$  spiegelt wieder, dass der Graph von  $f(x)$  für  $x_0$  groß steil wird.

**Beispiel 7.4** Die Betragsfunktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = |x|$ , ist in jedem Punkt  $x_0 \in \mathbb{R}$  stetig. Das folgt aus der Dreiecksungleichung: es gilt  $|x| \leq |x - x_0| + |x_0|$ , und durch Vertauschen

$$|f(x) - f(x_0)| = ||x| - |x_0|| \leq |x - x_0|.$$

Zu  $\varepsilon > 0$  können wir also  $\delta = \varepsilon$  nehmen.

**Beispiel 7.5** Die charakteristische Funktion (Indikatorfunktion) einer Menge  $E \subset \mathbb{R}$  ist

$$\chi_E : \mathbb{R} \rightarrow \mathbb{R}, \chi_E(x) = \begin{cases} 1 & \text{für } x \in E, \\ 0 & \text{für } x \notin E. \end{cases}$$

Sei  $\chi_E$  stetig im Punkt  $x_0 \in \mathbb{R}$ . Dann gibt es zu  $\varepsilon = 1$  ein  $\delta > 0$  mit

$$|\chi_E(x) - \chi_E(x_0)| < 1 \quad \text{für alle } x \in (x_0 - \delta, x_0 + \delta).$$

- ist  $\chi_E(x_0) = 1$ , so folgt  $\chi_E(x) = 1$  für  $x \in (x_0 - \delta, x_0 + \delta)$ , also  $(x_0 - \delta, x_0 + \delta) \subset E$ ,
- ist  $\chi_E(x_0) = 0$ , so folgt  $\chi_E(x) = 0$  für  $x \in (x_0 - \delta, x_0 + \delta)$ , also  $(x_0 - \delta, x_0 + \delta) \subset \mathbb{R}^n \setminus E$ .

Ein interessanter Spezialfall ist  $E = \mathbb{Q}$ . Die Funktion  $\chi_{\mathbb{Q}}$  nennt man auch Dirichletfunktion, weil Dirichlet sie in seinen Vorlesungen als Beispiel eingeführt hat.  $\mathbb{Q}$  ist dicht in  $\mathbb{R}$  nach Satz 2.6. Aber  $\mathbb{R} \setminus \mathbb{Q}$  ist ebenfalls dicht: zum Beispiel ist  $\sqrt{2} + \mathbb{Q}$  dicht in  $\mathbb{R}$  und Teilmenge von  $\mathbb{R} \setminus \mathbb{Q}$ , sonst wäre  $\sqrt{2} \in \mathbb{Q}$ . Also gibt es in jedem Intervall  $(x_0 - \delta, x_0 + \delta)$  Punkte aus  $\mathbb{Q}$  und Punkte aus  $\mathbb{R} \setminus \mathbb{Q}$ . Die Funktion  $\chi_{\mathbb{Q}}$  ist daher in keinem Punkt  $x_0 \in \mathbb{R}$  stetig.

Jetzt betrachten wir Funktionen bzw. Abbildungen  $f(x)$  auf  $D \subset \mathbb{R}^n$ , das heißt  $f$  ist Funktion von  $x = (x_1, \dots, x_n)$ . Eventuell kann  $f(x)$  auch Werte in  $\mathbb{R}^m$  haben, das heißt es ist  $f(x) = (f_1(x), \dots, f_m(x))$ . Es ist naheliegend, in der Definition der Stetigkeit den Betrag durch die Euklidische Norm zu ersetzen.

**Definition 7.2 (Stetigkeit)** Die Funktion  $f : D \rightarrow \mathbb{R}^m$ , wobei  $D \subset \mathbb{R}^n$ , heißt stetig in  $x_0 \in D$ , falls es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt, so dass gilt:

$$\|f(x) - f(x_0)\| < \varepsilon \quad \text{für alle } x \in D \text{ mit } \|x - x_0\| < \delta. \quad (7.2)$$

$f$  heißt stetig, falls  $f$  in allen  $x_0 \in D$  stetig ist.

Mithilfe von Umgebungen lässt sich die Stetigkeit in  $x_0$  auch so fassen:

$$\text{Für alle } \varepsilon > 0 \text{ gibt es ein } \delta > 0 \text{ mit } f(D \cap B_\delta(x_0)) \subset B_\varepsilon(f(x_0)).$$

**Beispiel 7.6** Die Euklidische Norm  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = \|x\| = (x_1^2 + \dots + x_n^2)^{1/2}$ , ist stetig. Denn es gilt  $\|x\| \leq \|x - x_0\| + \|x_0\|$ , also mittels Vertauschen von  $x$  und  $x_0$

$$|f(x) - f(x_0)| = ||\|x\| - \|x_0\|| \leq \|x - x_0\|.$$

Wir können  $\delta = \varepsilon$  nehmen.

Ein sehr nützliches hinreichendes Kriterium für die Stetigkeit ist wie folgt.

**Beispiel 7.7**  $f : D \rightarrow \mathbb{R}^m$  heißt Lipschitzstetig mit Konstante  $L \in [0, \infty)$ , falls gilt:<sup>15</sup>

$$\|f(x) - f(y)\| \leq L\|x - y\| \quad \text{für alle } x, y \in D.$$

Zum Beispiel ist die Euklidische Norm Lipschitzstetig mit Konstante  $L = 1$ , vgl. Beispiel 7.6. Es gilt allgemein: *jede Lipschitzstetige Funktion ist stetig*. Sei dazu  $L > 0$  die Lipschitzkonstante. Zu gegebenem  $\varepsilon > 0$  wählen wir  $\delta = \varepsilon/L > 0$ , und erhalten

$$\|x - x_0\| < \delta \quad \Rightarrow \quad \|f(x) - f(x_0)\| \leq L\|x - x_0\| < L\delta = \varepsilon.$$

**Beispiel 7.8** Die Abstandsfunktion von einer Menge  $E \subset \mathbb{R}^n$  lautet

$$\text{dist}_E : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \text{dist}_E(x) = \inf_{y \in E} \|y - x\|.$$

Diese Funktion ist Lipschitzstetig mit Konstante Eins. Denn es gilt nach Dreiecksungleichung

$$\|y - x_2\| \leq \|y - x_1\| + \|x_1 - x_2\| \quad \text{für alle } x_{1,2} \in \mathbb{R}^n, y \in E.$$

Bilden wir auf beiden Seiten das Infimum über  $y \in E$ , so folgt

$$\text{dist}_E(x_2) \leq \text{dist}_E(x_1) + \|x_1 - x_2\|.$$

Durch Vertauschen von  $x_1$  mit  $x_2$  ergibt sich  $\text{dist}_E$  Lipschitzstetig, genauer

$$|\text{dist}_E(x_1) - \text{dist}_E(x_2)| \leq \|x_1 - x_2\|.$$

**Beispiel 7.9** Lineare Abbildungen  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  sind Lipschitzstetig. Dazu verwenden wir für  $x \in \mathbb{R}^n$  die Darstellung bezüglich der Standardbasis:

$$x = x_1 e_1 + \dots + x_n e_n \quad \text{wobei } e_j = (0, \dots, 0, \overset{j}{1}, 0, \dots, 0).$$

Eine lineare Abbildung vertauscht mit Linearkombinationen, das heißt es gilt

$$A(x) = A(x_1 e_1 + \dots + x_n e_n) = x_1 A(e_1) + \dots + x_n A(e_n) \quad \text{mit } A(e_j) \in \mathbb{R}^m.$$

Wir schätzen mit der Dreiecksungleichung und Cauchy-Schwarz ab:

$$\|A(x)\|^2 = \left\| \sum_{j=1}^n x_j A(e_j) \right\|^2 \leq \left( \sum_{j=1}^n \|A(e_j)\| |x_j| \right)^2 \leq \left( \sum_{j=1}^n \|A(e_j)\|^2 \right) \left( \sum_{j=1}^n |x_j|^2 \right).$$

Also erhalten wir die Abschätzung

$$\|A(x)\| \leq \|A\| \|x\| \quad \text{wobei } \|A\| = \left( \sum_{j=1}^n \|A(e_j)\|^2 \right)^{1/2}.$$

Die Zahl  $\|A\| \in [0, \infty)$  nennt man die Euklidische Norm von  $A$ . Es folgt, wieder mit  $A$  linear,

$$\|A(x) - A(y)\| = \|A(x - y)\| \leq \|A\| \|x - y\| \quad \text{für alle } x, y \in \mathbb{R}^n.$$

---

<sup>15</sup>Rudolph Lipschitz, 1832-1903

Somit ist  $A$  Lipschitzstetig mit Konstante  $\|A\|$ , und insbesondere stetig. Wir können die Euklidische Norm auch mit der Matrixdarstellung von  $A$  ausdrücken. Dazu entwickeln wir die  $A(e_j)$  in die Standardbasis  $e_1, \dots, e_m$  des  $\mathbb{R}^m$ :

$$A(e_j) = \sum_{i=1}^m a_{ij} e_i \quad \text{für } j = 1, \dots, n.$$

Die  $a_{ij}$  sind die Elemente der Darstellungsmatrix. Es folgt  $\|A(e_j)\|^2 = \sum_{i=1}^m a_{ij}^2$  und damit

$$\|A\| = \left( \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \right)^{1/2}.$$

Der Beweis verwendet, dass  $\mathbb{R}^n$  eine endliche Basis hat. Das ist auch wesentlich, lineare Abbildungen auf einem Vektorraum  $V$  mit  $\dim V = \infty$  sind nicht notwendig stetig.

Bisher haben wir direkt mit der Definition argumentiert. Jetzt wollen zeigen, dass die Stetigkeit unter gewissen Operationen erhalten bleibt. Dazu wollen wir die Konvergenzregeln für Folgen anwenden, siehe Satz 4.3; der Zusammenhang wird durch folgende Aussage geleistet.

**Satz 7.1 (Folgenkriterium der Stetigkeit)** *Für  $f : D \rightarrow \mathbb{R}^m$  und  $x_0 \in D$  sind folgende Aussagen äquivalent:*

- (1)  $f$  ist stetig in  $x_0$ .
- (2) Für jede Folge  $(x_k)_{k \in \mathbb{N}}$  in  $D$  mit  $\lim_{k \rightarrow \infty} x_k = x_0$  gilt:  $\lim_{k \rightarrow \infty} f(x_k) = f(x_0)$ .

BEWEIS: Für die Implikation (1)  $\Rightarrow$  (2) sei  $x_k \in D$  mit  $x_k \rightarrow x_0$  für  $k \rightarrow \infty$ . Wähle zu  $\varepsilon > 0$  ein  $\delta > 0$  mit  $\|f(x) - f(x_0)\| < \varepsilon$  für alle  $x \in D$  mit  $\|x - x_0\| < \delta$ . Es gibt ein  $K \in \mathbb{R}$  mit  $\|x_k - x_0\| < \delta$  für  $k > K$ , also folgt  $\|f(x_k) - f(x_0)\| < \varepsilon$  für  $k > K$ .

Jetzt gelte (2). Wir nehmen indirekt an,  $f$  sei nicht stetig in  $x_0$ . Es gibt dann ein  $\varepsilon > 0$ , so dass (7.2) für kein  $\delta > 0$  erfüllt ist. Wählen wir  $\delta_k = 1/k$  mit  $k = 1, 2, \dots$ , so gibt es jeweils ein  $x_k \in D$  mit  $\|x_k - x_0\| < 1/k$ , aber  $\|f(x_k) - f(x_0)\| \geq \varepsilon$ . Die Folge  $x_k$  konvergiert gegen  $x_0$ , aber  $f(x_k)$  konvergiert nicht gegen  $f(x_0)$ , Widerspruch zu (2). Damit ist der Satz bewiesen.  $\square$

**Satz 7.2 (Verkettung stetiger Funktionen)** *Seien  $f : D \rightarrow \mathbb{R}^m$ ,  $f(D) \subset E \subset \mathbb{R}^m$ , und  $g : E \rightarrow \mathbb{R}^k$ . Ist  $f$  stetig in  $x_0$  und  $g$  stetig in  $y_0 = f(x_0)$ , so ist  $g \circ f : D \rightarrow \mathbb{R}^k$  stetig in  $x_0$ .*

BEWEIS: Wir verwenden Satz 7.1. Ist  $x_n \in D$  eine Folge mit  $\lim_{n \rightarrow \infty} x_n = x_0$ , so folgt  $f(x_n) \rightarrow f(x_0)$  aus der Stetigkeit von  $f$  in  $x_0$ , und weiter  $g(f(x_n)) \rightarrow g(f(x_0))$  wegen der Stetigkeit von  $g$  in  $y_0 = f(x_0)$ . Nach Satz 7.1 ist die Stetigkeit von  $g \circ f$  in  $x_0$  gezeigt.  $\square$

**Beispiel 7.10** Seien  $f, g : D \rightarrow \mathbb{R}$  stetig. Dann ist auch die Funktion  $|f| : D \rightarrow \mathbb{R}$  stetig, und ebenso die Funktionen  $\max(f, g) = (f + g + |f - g|)/2$  und  $\min(f, g) = (f + g - |f - g|)/2$ .

**Satz 7.3** *Eine Funktion  $f : D \rightarrow \mathbb{R}^m$  ist genau dann stetig in  $x_0 \in D$ , wenn jede Koordinatenfunktion  $f_i : D \rightarrow \mathbb{R}$ ,  $1 \leq i \leq m$ , in  $x_0$  stetig ist.*

BEWEIS: Nach Satz 6.6 ist eine Folge von Punkten im  $\mathbb{R}^m$  genau dann konvergent, wenn die einzelnen Koordinatenfolgen konvergieren. Mit Satz 7.1 ergibt sich die Behauptung.  $\square$

**Beispiel 7.11** Sei  $I \subset \mathbb{R}$  ein Intervall. Funktionen  $c : I \rightarrow \mathbb{R}^m$ ,  $c = c(t)$ , nennt man auch Wege oder Kurven. Die Bezeichnung der Variablen geht auf Newton zurück, er betrachtete die Bahnkurve ( $c = curva$ ) eines Massenpunkts als Funktion der Zeit ( $t = tempus$ ), und schrieb in Koordinaten  $c(t) = (x(t), y(t), z(t))$ . Ein konkretes Beispiel ist der horizontale Wurf aus Höhe  $h$  mit Geschwindigkeit  $v$ :

$$c : I = [0, \sqrt{2h/g}] \rightarrow \mathbb{R}^3, c(t) = \left( vt, 0, h - \frac{1}{2}gt^2 \right) \quad (g = \text{Erdbeschleunigung}).$$

Der Einfachheit halber beschränken wir uns im folgenden Satz auf reellwertige Funktionen.

**Satz 7.4 (Stetigkeitsregeln)** Seien  $f, g : D \rightarrow \mathbb{R}$  stetig in  $x_0 \in D$ . Dann gilt:

- (1) Für beliebige  $\lambda, \mu \in \mathbb{R}$  ist die Funktion  $\lambda f + \mu g$  stetig in  $x_0$ .
- (2) Die Funktion  $fg$  ist stetig in  $x_0$ .
- (3) Ist  $g(x_0) \neq 0$ , so ist die Funktion  $f/g : D \cap B_\delta(x_0) \rightarrow \mathbb{R}$  für  $\delta > 0$  hinreichend klein definiert und stetig in  $x_0$ .

BEWEIS: Wir führen die Aussagen auf die entsprechenden Regeln für Folgen zurück. Sei  $(x_k)_{k \in \mathbb{N}}$  eine beliebige Folge mit  $x_k \in D$  und  $\lim_{k \rightarrow \infty} x_k = x_0$ . Nach Satz 7.1 gilt dann  $f(x_k) \rightarrow f(x_0)$  und  $g(x_k) \rightarrow g(x_0)$  mit  $k \rightarrow \infty$ . Es folgt mit Satz 4.3 für  $k \rightarrow \infty$

$$\begin{aligned} (\lambda f + \mu g)(x_k) &= \lambda f(x_k) + \mu g(x_k) \rightarrow \lambda f(x_0) + \mu g(x_0) = (\lambda f + \mu g)(x_0), \\ (fg)(x_k) &= f(x_k)g(x_k) \rightarrow f(x_0)g(x_0) = (fg)(x_0). \end{aligned}$$

Nach Satz 7.1 sind  $\lambda f + \mu g$  sowie  $fg$  stetig in  $x_0$ . Wir nehmen nun an dass  $g \neq 0$  auf  $D \cap B_\delta(x_0)$ , und  $x_k \in D \cap B_\delta(x_0)$ . Dann folgt auch

$$\frac{f}{g}(x_k) = \frac{f(x_k)}{g(x_k)} \rightarrow \frac{f(x_0)}{g(x_0)} = \frac{f}{g}(x_0),$$

also gilt Aussage (3). Wir zeigen unten  $g \neq 0$  auf einer geeigneten Umgebung  $D \cap B_\delta(x_0)$ .  $\square$

**Lemma 7.1** Sei  $g : D \rightarrow \mathbb{R}$  stetig in  $x_0$  mit  $g(x_0) \neq 0$ . Dann gibt es ein  $\delta > 0$  mit

$$|g(x)| \geq \frac{1}{2}|g(x_0)| > 0 \quad \text{für alle } x \in D \cap B_\delta(x_0).$$

BEWEIS: Da  $g$  stetig in  $x_0$ , gibt es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  mit  $|g(x) - g(x_0)| < \varepsilon$  für alle  $x \in D \cap B_\delta(x_0)$ , also folgt mit der Dreiecksungleichung für diese  $x$

$$|g(x)| = |g(x_0) - (g(x_0) - g(x))| \geq |g(x_0)| - |g(x) - g(x_0)| > |g(x_0)| - \varepsilon.$$

Mit der Wahl  $\varepsilon = \frac{1}{2}|g(x_0)| > 0$  folgt die Behauptung.  $\square$

Die Menge der stetigen Funktionen  $f : D \rightarrow \mathbb{R}$  wird mit  $C^0(D)$  bezeichnet. Eine offensichtliche Konsequenz von Satz 7.4(1) ist

**Folgerung 7.1 ( $C^0$ -Räume)**  $C^0(D)$  ist ein Untervektorraum des Raums aller Funktionen  $f : D \rightarrow \mathbb{R}$ , mit der punktweisen Addition und Skalarmultiplikation.

Wir behandeln jetzt noch das Konzept des Grenzwerts einer Funktion. Die Unterschiede zu Folgen sind gering, deshalb fassen wir uns kurz.

**Definition 7.3 (Grenzwert für Funktionen)** Sei  $x_0$  ein Häufungspunkt von  $D \subset \mathbb{R}^n$ . Die Funktion  $f : D \rightarrow \mathbb{R}^m$  konvergiert für  $x \rightarrow x_0$  gegen  $a \in \mathbb{R}^m$ , falls es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt so dass gilt:

$$\|f(x) - a\| < \varepsilon \quad \text{für alle } x \in D \text{ mit } 0 < \|x - x_0\| < \delta.$$

Notation:  $\lim_{x \rightarrow x_0} f(x) = a$  oder  $f(x) \rightarrow a$  für  $x \rightarrow x_0$ .

Beachten Sie dass hier nicht  $x_0 \in D$  sein muss, die Ungleichung  $\|f(x) - a\| < \varepsilon$  wird auch nicht im Punkt  $x = x_0$  verlangt. Es ist sowohl für die Existenz des Grenzwerts als auch für den Wert von  $\lim_{x \rightarrow x_0} f(x)$  egal, ob die Funktion  $f(x)$  in  $x_0$  definiert ist bzw. welchen Funktionswert sie dort hat. Die Beziehung zwischen Grenzwert und Stetigkeit ist wie folgt:

**Lemma 7.2 (Stetigkeit und Grenzwert)** Sei  $D \subset \mathbb{R}^n$ , und  $x_0$  ein Häufungspunkt mit  $x_0 \in D$ . Für  $f : D \rightarrow \mathbb{R}^m$  sind folgende Aussagen äquivalent:

- (1)  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ .
- (2)  $f$  ist stetig in  $x_0$ .

BEWEIS: Die beiden Aussagen lauten, wobei wir ausnahmsweise die Abkürzungen  $\forall$  (für alle) und  $\exists$  (es gibt) verwenden:

- (1)  $\forall \varepsilon > 0 \exists \delta > 0 : \|f(x) - f(x_0)\| < \varepsilon \quad \forall x \in D \cap B_\delta(x_0) \setminus \{x_0\}$ ,
- (2)  $\forall \varepsilon > 0 \exists \delta > 0 : \|f(x) - f(x_0)\| < \varepsilon \quad \forall x \in D \cap B_\delta(x_0)$ .

Da sowieso  $\|f(x) - f(x_0)\| = 0$  für  $x = x_0$ , sind die Aussagen identisch. □

Sowohl bei der Stetigkeit als auch beim Grenzwert spielt der zugrundeliegende Definitionsbereich eine Rolle. Wir schreiben  $\lim_{x \rightarrow x_0, x \in D} f(x)$ , wenn wir den gewählten Definitionsbereich hervorheben möchten. In  $\mathbb{R}$  werden oft einseitige Grenzwerte gebraucht:

$$\lim_{x \searrow x_0} f(x) = \lim_{x \rightarrow x_0, x > x_0} f(x) \quad \text{und} \quad \lim_{x \nearrow x_0} f(x) = \lim_{x \rightarrow x_0, x < x_0} f(x).$$

Zum Beispiel ist  $\lim_{x \searrow 0} \text{sign}(x) = +1$  und  $\lim_{x \nearrow 0} \text{sign}(x) = -1$ , während der Grenzwert  $\lim_{x \rightarrow 0} \text{sign}(x)$  nicht existiert. Sind aber der links- und rechtsseitige Grenzwert einer Funktion  $f(x)$  im Punkt  $x_0$  gleich, so ist dies der Grenzwert von  $f(x)$  für  $x \rightarrow x_0$ .

Der Grenzwertbegriff für Funktionen lässt sich wieder auf Folgen zurückführen.

**Satz 7.5 (Definition von  $\lim_{x \rightarrow x_0} f(x)$  mit Folgen)** Sei  $x_0$  Häufungspunkt von  $D \subset \mathbb{R}^n$ , und  $f : D \rightarrow \mathbb{R}^m$ . Für  $a \in \mathbb{R}^m$  sind äquivalent:

- (1)  $f(x) \rightarrow a$  für  $x \rightarrow x_0$ .
- (2)  $\lim_{k \rightarrow \infty} f(x_k) = a$  für jede Folge  $x_k \in D \setminus \{x_0\}$  mit  $x_k \rightarrow x_0$ .

BEWEIS: Der Beweis ist analog zu Satz 7.1. In Definition 7.3 kommen nur  $x \in D$  vor mit  $\|x - x_0\| > 0$ . Entsprechend werden hier in (2) nur Folgen mit  $x_k \neq x_0$  betrachtet.  $\square$

Für Grenzwerte von Funktionen gelten Rechenregeln analog zu Satz 7.4. Der Beweis wird den Lesern/innen überlassen.

**Satz 7.6 (Rechenregeln für Grenzwerte)** Sei  $x_0$  Häufungspunkt von  $D \subset \mathbb{R}^n$ . Es gilt:

(1) Seien  $f, g : D \rightarrow \mathbb{R}$  mit  $f(x) \rightarrow a$ ,  $g(x) \rightarrow b$  für  $x \rightarrow x_0$ . Dann folgt

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

(2) Seien  $f : D \rightarrow \mathbb{R}^m$  mit  $f(D) \subset E \subset \mathbb{R}^m$ , und  $g : E \rightarrow \mathbb{R}$ . Gilt  $f(x) \rightarrow y_0$  mit  $x \rightarrow x_0$  und ist  $g$  stetig in  $y_0$ , so folgt  $(g \circ f)(x) \rightarrow g(y_0)$  mit  $x \rightarrow x_0$ .

(3) Sei  $f : D \rightarrow \mathbb{R}$  mit  $f(x) \rightarrow a$  für  $x \rightarrow x_0$ . Ist  $f \geq 0$  auf  $B_\delta(x_0) \setminus \{x_0\}$ , so folgt  $a \geq 0$ .

Für reellwertige Funktionen können wir als (uneigentliche) Grenzwerte auch  $\pm\infty$  zulassen.

**Definition 7.4 (Uneigentlicher Grenzwert)** Sei  $x_0 \in \mathbb{R}^n$  ein Häufungspunkt von  $D$ , und  $f : D \rightarrow \mathbb{R}$ . Dann gilt  $\lim_{x \rightarrow x_0} f(x) = +\infty$ , wenn es zu jedem  $K > 0$  ein  $\delta > 0$  gibt mit

$$f(x) > K \quad \text{für alle } x \in D \text{ mit } 0 < \|x - x_0\| < \delta.$$

Entsprechend wird  $\lim_{x \rightarrow x_0} f(x) = -\infty$  erklärt.

**Beispiel 7.12** Die Funktion  $f(x) = 1/|x|$ ,  $x \neq 0$ , hat den Grenzwert  $\lim_{x \rightarrow 0} f(x) = +\infty$ . Dagegen hat  $g(x) = 1/x$ ,  $x \neq 0$ , keinen Grenzwert bei  $x = 0$ . Es existieren aber die einseitigen Grenzwerte  $\lim_{x \searrow 0} g(x) = +\infty$  und  $\lim_{x \nearrow 0} g(x) = -\infty$ . Ist  $f : D \rightarrow (0, \infty)$ , so ist  $\lim_{x \rightarrow x_0} f(x) = +\infty$  äquivalent zu  $\lim_{x \rightarrow x_0} 1/f(x) = 0$ .

Im Fall  $D \subset \mathbb{R}$  besteht weiter die Möglichkeit, Grenzwerte für  $x \rightarrow \pm\infty$  zu betrachten.

**Definition 7.5 (Grenzwert bei  $\pm\infty$ )** Sei  $f : D \rightarrow \mathbb{R}^m$  mit  $D \subset \mathbb{R}$  nicht nach oben beschränkt. Dann gilt  $\lim_{x \rightarrow +\infty} f(x) = a \in \mathbb{R}^m$ , wenn es zu  $\varepsilon > 0$  ein  $K \in \mathbb{R}$  gibt mit

$$\|f(x) - a\| < \varepsilon \quad \text{für alle } x \in D \text{ mit } x > K.$$

Analog wird der Grenzwert für  $x \rightarrow -\infty$  erklärt.

Wir wollen als Beispiel für diese Regeln das Verhalten von rationalen Funktionen untersuchen, dazu brauchen wir etwas Vorwissen über Polynome. Im Folgenden sei  $\mathbb{K} = \mathbb{R}$  oder  $\mathbb{C}$ .

**Definition 7.6** Eine Funktion  $p : \mathbb{K} \rightarrow \mathbb{K}$  heißt (reelles oder komplexes) Polynom vom Grad  $n \in \mathbb{N}_0$ , wenn es  $a_0, a_1, \dots, a_n \in \mathbb{K}$  mit  $a_n \neq 0$  gibt, so dass gilt:

$$p(z) = \sum_{k=0}^n a_k z^k \quad \text{für alle } z \in \mathbb{K}. \quad (7.3)$$

Aus der Definition ist nicht sofort klar, daß der Grad  $n$  und die Koeffizienten  $a_i$  eindeutig bestimmt sind. Dies wird in Folgerung 7.2 gezeigt.

**Lemma 7.3 (Abspaltung von Linearfaktoren)** Sei  $p : \mathbb{K} \rightarrow \mathbb{K}$  Polynom vom Grad  $n$  mit Koeffizienten  $a_0, \dots, a_n$ . Ist  $p(\lambda) = 0$  für ein  $\lambda \in \mathbb{K}$ , so gibt es ein Polynom  $q : \mathbb{K} \rightarrow \mathbb{K}$  vom Grad  $n - 1$  mit Koeffizienten  $b_0, \dots, b_{n-1}$  wobei  $b_{n-1} = a_n$ , so dass gilt:

$$p(z) = (z - \lambda)q(z) \quad \text{für alle } z \in \mathbb{K}. \quad (7.4)$$

BEWEIS: Wegen  $p(\lambda) = 0$  gilt für alle  $z \in \mathbb{K}$ , da sich der Summand  $a_0$  weghebt,

$$p(z) = p(z) - p(\lambda) = \sum_{k=1}^n a_k(z^k - \lambda^k).$$

Wir wollen aus  $z^k - \lambda^k$  einen Faktor  $z - \lambda$  ausklammern. Im Fall  $\lambda = 0$  ist das klar, sonst berechnen wir mit der geometrischen Summe, siehe Beispiel 3.2,

$$z^k - \lambda^k = \lambda^k \left( \left( \frac{z}{\lambda} \right)^k - 1 \right) = \lambda^k \left( \frac{z}{\lambda} - 1 \right) \sum_{j=0}^{k-1} \left( \frac{z}{\lambda} \right)^j = (z - \lambda) \sum_{j=0}^{k-1} \lambda^{k-j-1} z^j.$$

Einsetzen ergibt eine Darstellung  $p(z) = (z - \lambda)q(z)$  wie verlangt. Die Potenz  $z^{n-1}$  kommt nur vor für  $k = n$  und  $j = n - 1$ . Es ist dann  $\lambda^{k-j-1} = \lambda^0 = 1$ , damit folgt  $b_{n-1} = a_n$ .  $\square$

**Lemma 7.4** Ein Polynom  $p$  vom Grad  $n$  hat höchstens  $n$  verschiedene Nullstellen.

BEWEIS: Durch Induktion über  $n \in \mathbb{N}_0$ . Für  $n = 0$  gilt  $p(z) = a_0 \neq 0$  für alle  $z$ , also hat  $p$  keine Nullstelle. Ist  $p(z)$  Polynom vom Grad  $n \in \mathbb{N}$ , so hat entweder  $p$  keine Nullstelle oder nach Lemma 7.3 gilt  $p(z) = (z - \lambda)q(z)$  für alle  $z \in \mathbb{K}$ , mit einem Polynom  $q$  vom Grad  $n - 1$ . Nach Induktion hat  $q$  höchstens  $n - 1$  Nullstellen, also  $p$  höchstens  $n$  Nullstellen.  $\square$

Die Aussage kann auch so formuliert werden: gilt  $p(z) = \sum_{k=0}^n a_k z^k$  für alle  $z \in \mathbb{K}$  und hat  $p(z)$  mehr als  $n$  Nullstellen, so folgt  $a_0 = a_1 = \dots = a_n = 0$  und  $p(z)$  ist die Nullfunktion. Manche Autoren sprechen auch vom *Nullpolynom*, aber das ist Geschmackssache.

**Folgerung 7.2 (Koeffizientenvergleich)** Seien  $p, q : \mathbb{K} \rightarrow \mathbb{K}$  Polynome vom Grad  $m$  bzw.  $n$ , das heißt es gilt mit  $a_m, b_n \neq 0$

$$p(z) = \sum_{i=0}^m a_i z^i \quad \text{und} \quad q(z) = \sum_{i=0}^n b_i z^i \quad \text{für alle } z \in \mathbb{K}.$$

Ist  $p(z) = q(z)$  an mehr als  $\max(m, n)$  Stellen, so folgt  $n = m$  und  $b_i = a_i$  für  $i = 0, 1, \dots, m$ .

BEWEIS: Wäre  $m \neq n$  oder  $a_i \neq b_i$  für ein  $i$ , so wäre  $p - q$  Polynom vom Grad höchstens  $\max(m, n)$  mit mehr als  $\max(m, n)$  Nullstellen, im Widerspruch zu Lemma 7.4.  $\square$

**Satz 7.7** Jedes Polynom  $p : \mathbb{K} \rightarrow \mathbb{K}$  vom Grad  $n$  hat eine eindeutige Zerlegung

$$p(z) = (z - \lambda_1)^{\nu_1} \cdot \dots \cdot (z - \lambda_r)^{\nu_r} q(z) \quad \text{für alle } z \in \mathbb{K}. \quad (7.5)$$

Dabei sind  $\{\lambda_1, \dots, \lambda_r\} \subset \mathbb{K}$  die Nullstellen von  $p$  in  $\mathbb{K}$  (eventuell  $r = 0$ ), und  $q : \mathbb{K} \rightarrow \mathbb{K}$  ist ein Polynom vom Grad  $n - (\nu_1 + \dots + \nu_r) \in \{0, \dots, n\}$  mit  $q(z) \neq 0$  für alle  $z \in \mathbb{K}$ . Die Zahlen  $\nu_i \geq 1$  heißen Vielfachheiten der Nullstellen  $\lambda_i$ .

BEWEIS: Die Existenz der Zerlegung folgt induktiv aus Lemma 7.3. Für die Eindeutigkeit betrachte eine zweite solche Zerlegung  $p(z) = (z - \lambda_1)^{\mu_1} \cdot \dots \cdot (z - \lambda_r)^{\mu_r} h(z)$ , ohne Einschränkung mit  $\nu_1 \geq \mu_1$ . Indem wir durch  $(z - \lambda_1)^{\mu_1}$  teilen, folgt für alle  $z \neq \lambda_1$

$$(z - \lambda_1)^{\nu_1 - \mu_1} \cdot (z - \lambda_2)^{\nu_2} \cdot \dots \cdot (z - \lambda_r)^{\nu_r} \cdot q(z) = (z - \lambda_2)^{\mu_2} \cdot \dots \cdot (z - \lambda_r)^{\mu_r} \cdot h(z).$$

Beide Seiten sind Polynome, haben nach Folgerung 7.2 also dieselben Koeffizienten. Deshalb stimmen sie auch in  $z = \lambda_1$  überein; es folgt  $\nu_1 = \mu_1$  und analog  $\nu_i = \mu_i$  für alle  $i \in \{1, \dots, r\}$ . Nach Division haben wir schließlich  $q(z) = h(z)$  für alle  $z \in \mathbb{K} \setminus \{\lambda_1, \dots, \lambda_r\}$ , und Folgerung 7.2 liefert  $q = h$ .  $\square$

**Beispiel 7.13 (Rationale Funktionen)** Seien  $p(x) = \sum_{i=0}^m a_i x^i$  und  $q(x) = \sum_{j=0}^n b_j x^j$  reelle Polynome mit  $a_m, b_n \neq 0$ . Setze  $N = \{x \in \mathbb{R} : q(x) = 0\}$  und definiere

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

In welchen  $x_0 \in N$  hat  $f$  eine stetige Fortsetzung? Sei  $x_0$  eine  $\mu$ -fache Nullstelle von  $p(x)$  und eine  $\nu$ -fache Nullstelle von  $q(x)$ , eventuell mit  $\mu = 0$  falls  $p(x_0) \neq 0$ . Also gilt  $p(x) = (x - x_0)^\mu \tilde{p}(x)$  und  $q(x) = (x - x_0)^\nu \tilde{q}(x)$  für Polynome  $\tilde{p}, \tilde{q}$  mit  $\tilde{p}(x_0), \tilde{q}(x_0) \neq 0$ . Es folgt

$$f(x) = (x - x_0)^{\mu - \nu} \frac{\tilde{p}(x)}{\tilde{q}(x)} \quad \text{für alle } x \in \mathbb{R} \setminus N,$$

Aus den Rechenregeln für Grenzwerte folgt

$$\lim_{x \rightarrow x_0} f(x) = \begin{cases} 0 & \text{falls } \mu > \nu, \\ \frac{\tilde{p}(x_0)}{\tilde{q}(x_0)} \neq 0 & \text{falls } \mu = \nu. \end{cases}$$

In diesen beiden Fällen ist  $f(x)$  stetig fortsetzbar in  $x_0$  nach Lemma 7.2. Im Fall  $\mu < \nu$  ist  $x_0$  dagegen eine Polstelle. Sei zum Beispiel  $\tilde{p}(x_0)/\tilde{q}(x_0) > 0$ , dann ergibt sich für die einseitigen Grenzwerte bei  $x_0$

	$x \searrow x_0$	$x \nearrow x_0$
$\nu - \mu$ gerade	$+\infty$	$+\infty$
$\nu - \mu$ ungerade	$+\infty$	$-\infty$

Die Diskussion der Grenzwerte  $x \rightarrow \pm\infty$  ist ähnlich. Wir schreiben für  $|x|$  groß

$$f(x) = x^{m-n} \frac{a_m + a_{m-1}x^{-1} + \dots + a_0x^{-m}}{b_n + b_{n-1}x^{-1} + \dots + b_0x^{-n}}.$$

Es folgt  $f(x) \rightarrow 0$  wenn  $m < n$ , und  $f(x) \rightarrow a_m/b_n$  wenn  $m = n$ . Für  $m > n$  sei zum Beispiel  $a_m/b_n > 0$ , Dann gilt  $f(x) \rightarrow +\infty$  für  $x \rightarrow +\infty$  und  $f(x) \rightarrow \pm\infty$  für  $x \rightarrow -\infty$ , je nachdem ob  $m - n$  gerade oder ungerade ist.



## 8 Zwischenwertsatz und Monotonie

Das Lösen von Gleichungen ist ohne Zweifel ein zentrales Thema der Mathematik. Lösungen von quadratischen oder kubischen Gleichungen können mithilfe von Wurzeln berechnet werden. Aber Arbeiten von Abel und Galois um 1815 haben gezeigt, dass das für Polynome vom Grad 5 oder mehr im allgemeinen nicht funktioniert, und ohnehin sind viele Gleichungen gar nicht algebraisch. Es hat sich die Erkenntnis durchgesetzt, dass meistens die Lösung nur durch einen abstrakten Existenzbeweis gefunden werden kann. Von diesem Typ ist der Zwischenwertsatz, der zuerst von Bolzano 1817 formuliert wurde, obwohl er eigentlich ziemlich offensichtlich ist.

In diesem Abschnitt haben wir es mit reellen Funktionen zu tun, die auf einem Intervall definiert sind. Wir verwenden dabei folgende Tatsache: eine Menge  $I \subset \mathbb{R}$  ist genau dann ein Intervall mit Grenzen  $a \leq b$ , wenn gilt:

$$a = \inf I, b = \sup I \quad \text{und} \quad (a, b) \subset I.$$

Hier sind  $\pm\infty$  als Grenzen erlaubt. Der Durchschnitt von zwei Intervallen  $I_{1,2}$  mit Grenzen  $a_{1,2}$  und  $b_{1,2}$  ist wieder ein Intervall, und zwar mit Grenzen  $a = \max(a_1, a_2)$  und  $b = \min(b_1, b_2)$ .

**Satz 8.1 (Zwischenwertsatz)** Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig. Dann gibt es zu jedem  $y_0$  zwischen  $f(a)$  und  $f(b)$  ein  $x_0 \in [a, b]$  mit  $f(x_0) = y_0$ .

*Bemerkung.* Die Gleichung  $f(x) = y_0$  kann mehrere Lösungen in  $[a, b]$  besitzen, das heißt  $x_0$  ist im allgemeinen nicht eindeutig bestimmt.

BEWEIS: Durch Übergang zu  $-f$  und  $-y_0$  können wir annehmen dass  $f(a) \leq y_0 \leq f(b)$ . Die Menge  $M = \{x \in [a, b] : f(x) \leq y_0\}$  ist nichtleer, da  $a \in M$ . Wir wählen  $x_0 = \sup M \in [a, b]$  und zeigen zuerst  $f(x_0) \geq y_0$ . Im Fall  $x_0 = b$  gilt das nach Annahme. Andernfalls ist  $f(x) > y_0$  für  $x_0 < x \leq b$ , und es folgt mit Stetigkeit  $f(x_0) = \lim_{x \searrow x_0} f(x) \geq y_0$ . Jetzt zeigen wir  $f(x_0) \leq y_0$ , das gilt wieder nach Annahme wenn  $x_0 = a$ . Andernfalls gibt es  $x_n > x_0 - 1/n$  mit  $x_n \in M$ , also  $f(x_n) \leq y_0$ . Es folgt  $x_n \leq x_0$ , also  $x_n \rightarrow x_0$  und wegen Stetigkeit  $f(x_0) = \lim_{n \rightarrow \infty} f(x_n) \leq y_0$ . Damit ist der Satz bewiesen.  $\square$

*Bemerkung.* Der Beweis liefert die größte Lösung  $x_0$  der Gleichung  $f(x) = y_0$ .

**Folgerung 8.1** Sei  $I \subset \mathbb{R}$  ein Intervall und  $f : I \rightarrow \mathbb{R}$  stetig. Dann ist  $f(I)$  ein Intervall mit Endpunkten  $\alpha = \inf_{x \in I} f(x)$  und  $\beta = \sup_{x \in I} f(x)$ .

BEWEIS: Wir verwenden das Kriterium von oben, zu zeigen ist also  $(\alpha, \beta) \subset f(I)$ . Zu  $y \in (\alpha, \beta)$  gibt es  $x_1, x_2 \in I$  mit  $f(x_1) < y < f(x_2)$ . Dann gibt es nach Satz 8.1 ein  $x \in [x_1, x_2]$  (bzw.  $x \in [x_2, x_1]$ ) mit  $f(x) = y$ , also  $y \in f(I)$ .  $\square$

Jetzt zu monotonen Funktionen. Nach Satz 5.3 haben monotone, beschränkte Folgen einen Grenzwert; hier eine analoge Aussage für monotone Funktionen.

**Lemma 8.1 (einseitige Grenzwerte)** Sei  $I$  Intervall mit Grenzen  $a < b$ . Ist  $f : I \rightarrow \mathbb{R}$  monoton wachsend, so existieren in jedem  $x_0 \in I$  die einseitigen Grenzwerte, genauer gilt

$$\begin{aligned} \lim_{x \nearrow x_0} f(x) &= \sup\{f(x) : x \in I, x < x_0\} && \text{falls } x_0 > a, \\ \lim_{x \searrow x_0} f(x) &= \inf\{f(x) : x \in I, x > x_0\} && \text{falls } x_0 < b. \end{aligned}$$

Für  $f$  monoton fallend gilt die entsprechende Aussage.

BEWEIS: Setze  $S = \sup\{f(x) : x \in I, x < x_0\}$ . Nach Definition des Supremums gibt es zu jedem  $\alpha < S$  ein  $x' \in I$ ,  $x' < x_0$ , mit  $f(x') > \alpha$ . Es folgt für alle  $x \in I$  mit  $x' \leq x < x_0$

$$\alpha < f(x') \leq f(x) \leq S.$$

Dies zeigt  $\lim_{x \nearrow x_0} f(x) = S$ . Die zweite Aussage folgt analog.  $\square$

Sei nun  $f : I \rightarrow \mathbb{R}$  streng monoton wachsend. Aus  $f(x_1) = f(x_2)$  folgt dann  $x_1 = x_2$ , das heißt  $f$  ist injektiv und die Umkehrfunktion  $g : f(I) \rightarrow I$  ist definiert. Wie in Gleichung (5.6) gezeigt, ist  $g$  ebenfalls streng monoton wachsend, hier nochmal das Argument: wäre  $y_2 > y_1$  mit  $g(y_2) \leq g(y_1)$ , so folgt wegen  $f$  wachsend

$$y_2 = f(g(y_2)) \leq f(g(y_1)) = y_1,$$

ein Widerspruch. Nach Lemma 8.1 hat  $g$  also links- und rechtsseitige Grenzwerte. Wir interessieren uns nun für die Stetigkeit der Umkehrfunktion  $g$ .

**Satz 8.2 (Monotonie und Umkehrfunktion)** Sei  $I$  ein Intervall mit Endpunkten  $a < b$ , und  $f : I \rightarrow \mathbb{R}$  sei streng monoton wachsend und stetig. Dann gilt:

- (1)  $f(I)$  ist ein Intervall mit Endpunkten  $\alpha = \lim_{x \searrow a} f(x) < \lim_{x \nearrow b} f(x) = \beta$ .
- (2) Die Umkehrfunktion  $g : f(I) \rightarrow \mathbb{R}$  ist stetig.
- (3)  $\lim_{y \searrow \alpha} g(y) = a$  und  $\lim_{y \nearrow \beta} g(y) = b$ .

BEWEIS: (1): Nach Folgerung 8.1 ist  $f(I)$  ein Intervall mit Grenzen  $\inf_I f$  und  $\sup_I f$ . Offenbar ist  $\inf_I f \leq \lim_{x \searrow a} f(x)$ . Andererseits gilt nach Lemma 8.1

$$f(x) \geq \lim_{x \searrow a} f(x) \quad \text{für alle } x \in I, x > a.$$

Im Fall  $a \in I$  ist  $f(a) = \lim_{x \searrow a} f(x)$ , da  $f$  stetig in  $a$ . Also ist  $\lim_{x \searrow a} f(x)$  untere Schranke von  $f(I)$ , es folgt  $\inf_I f \geq \lim_{x \searrow a} f(x)$  und damit Gleichheit. Für die rechte Intervallgrenze argumentieren wir entsprechend.

(2): Sei  $y_0 \in f(I)$  mit  $y_0 > \alpha$ . Setze  $x_0 = g(y_0)$  und  $x_- := \lim_{y \nearrow y_0} g(y)$ . Für  $y < y_0$  gilt  $g(y) < g(y_0) = x_0$ , mit  $y \nearrow y_0$  folgt  $x_- \leq x_0$ . Angenommen es ist  $x_- < x_0$ , dann ergibt sich für  $x \in (x_-, x_0)$ :

- für  $y < y_0$  gilt  $g(y) \leq x_- < x$ , also  $y = f(g(y)) < f(x)$ . Mit  $y \nearrow y_0$  folgt  $y_0 \leq f(x)$ .
- wegen  $x < x_0 = g(y_0)$  gilt  $f(x) < f(g(y_0)) = y_0$ .

Der Widerspruch zeigt  $x_- = x_0$  und damit  $g$  linksseitig stetig in  $y_0$ . Analog folgt  $g$  rechtsseitig stetig in jedem Punkt  $y_0 \in f(I)$  mit  $y_0 < \beta$ , also ist  $g$  insgesamt stetig.

(3): Die Funktion  $g$  erfüllt die Voraussetzungen des Satzes, nach (1) hat das Intervall  $I = g(f(I))$  die Endpunkte  $a = \lim_{y \searrow \alpha} g(y)$  und  $b = \lim_{y \nearrow \beta} g(y)$ .  $\square$

Der Beweis der Stetigkeit präzisiert folgende Vorstellung: der Graph von  $f$  ergibt sich aus dem Graph von  $g$  durch Spiegelung an der Winkelhalbierenden. Ein Sprung von  $g$  würde dabei

einem Intervall entsprechen, auf dem  $f$  konstant ist, im Widerspruch zur strengen Monotonie. Als Ergänzung zu Satz 8.2 bemerken wir noch

$$a \in I \Leftrightarrow \alpha \in f(I) \quad \text{und} \quad b \in I \Leftrightarrow \beta \in f(I). \quad (8.1)$$

Für  $a \in I$  ist  $\alpha = \inf\{f(x) : x \in I\} = f(a)$ . Sei umgekehrt  $\alpha = f(x)$  für ein  $x \in I$ . Wäre  $x > a$ , so ist  $f(x') < f(x) = \alpha$  für  $x' \in (a, x)$ , Widerspruch. Also folgt  $a = x \in I$ .

**Beispiel 8.1**  $f : [0, \infty) \rightarrow \mathbb{R}$ ,  $f(x) = x^n$ , ist stetig und streng monoton wachsend mit

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

Der Satz liefert  $f([0, \infty)) = [0, \infty)$ , also (erneut) die Existenz der  $n$ -ten Wurzel (vgl. Satz 5.6). Weiter: die Umkehrfunktion  $g : [0, \infty) \rightarrow \mathbb{R}$ ,  $g(y) = y^{1/n}$ , ist stetig und es gilt

$$\lim_{y \searrow 0} y^{1/n} = 0 \quad \text{und} \quad \lim_{y \nearrow \infty} y^{1/n} = \infty.$$

Wir werden in Kürze weitere Anwendungen des Satzes sehen. Insbesondere werden wir die Logarithmusfunktion als Umkehrfunktion der Exponentialfunktion definieren.



## 9 Die Ableitung

Die Differentialrechnung wird Newton (1665) und Leibniz (1675) zugeschrieben. Ihre Arbeiten wurden erst später veröffentlicht, und zwar die von Newton nach denen von Leibniz, was zu einem heftigen Streit über die Urheberschaft geführt hat, Newton hat Leibniz des Plagiats bezichtigt. Das hat die historische Forschung widerlegt, ihre Arbeiten sind unabhängig entstanden. Newton nannte veränderliche Größen *fluents* und ihre momentane Änderung *fluxions*, dafür verwendete er einen Punkt wie noch heute in der Physik üblich. Leibniz schrieb die Ableitung als Differentialquotient, diese Notation hat sich weithin durchgesetzt. Beide hatten kein rigoroses Konzept des Grenzwerts, aber sie konnten mit der Idee von infinitesimalen Größen Rechenregeln ableiten, z.B. Produktregel, Quotientenregel und Kettenregel. Es war entscheidend dass diese nicht nur einem eingeweihten kleinen Kreis zugänglich waren, sondern allgemein Anwendung fanden. In England und USA spricht man daher vom *Calculus* (deutsch: *Kalkül*).

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

**Definition 9.1 (Ableitung)** Die Funktion  $f : I \rightarrow \mathbb{R}^n$  hat in  $x_0 \in I$  die Ableitung  $a \in \mathbb{R}^n$  (Notation:  $f'(x_0) = a$ ), falls gilt:

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = a. \quad (9.1)$$

Wir nennen  $f$  differenzierbar in  $x_0$ , falls es ein  $a \in \mathbb{R}^n$  mit (9.1) gibt, falls also der in (9.1) betrachtete Grenzwert existiert.

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

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

Leibniz interessierte sich für die Definition der Ableitung im Zusammenhang mit dem Problem, die Tangente an eine ebene Kurve in einem gegebenen Punkt zu definieren. Nehmen wir dazu an, dass die Kurve als Graph einer Funktion  $f : I \rightarrow \mathbb{R}$  gegeben ist, und dass die Tangente im Punkt  $(x_0, f(x_0))$  gesucht ist. Der Differenzenquotient

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

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

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

Im Fall von vektorwertigen Funktionen, also  $n \geq 2$ , ist der Grenzwert in (9.1) wie üblich bezüglich der Euklidischen Norm aufzufassen. Nach Satz 6.6 ist das aber äquivalent zur Konvergenz der einzelnen Koordinaten. Das besagt hier

**Lemma 9.1** Die Funktion  $f : I \rightarrow \mathbb{R}^n$  ist genau dann in  $x_0 \in I$  differenzierbar, wenn alle Koordinatenfunktionen  $f_i : I \rightarrow \mathbb{R}$ ,  $i = 1, \dots, n$ , in  $x_0$  differenzierbar sind. Die Ableitung kann dann koordinatenweise berechnet werden, das heißt es gilt:

$$f'(x_0) = ((f_1)'(x_0), \dots, (f_n)'(x_0)) \in \mathbb{R}^n.$$

Newton entwickelte den Differentialkalkül (Englisch: *Calculus*) unter anderem um die Keplerschen Gesetze für die Planetenbewegung zu begründen, er konnte sie damit aus dem Gravitationsgesetz ableiten. Dazu wird die Bewegung eines Planeten durch eine Abbildung

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

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

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

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

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

**Definition 9.2 (Ableitungsfunktion)** Die Funktion  $f : I \rightarrow \mathbb{R}^n$  heißt differenzierbar auf  $I$  (oder einfach differenzierbar), falls  $f$  in jedem Punkt  $x_0 \in I$  differenzierbar ist. Die hierdurch gegebene Funktion

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

heißt Ableitungsfunktion oder schlicht Ableitung von  $f$ .

**Beispiel 9.1** Für eine konstante Funktion, also  $f(x) = c \in \mathbb{R}^n$  für alle  $x \in I$ , gilt für  $x \neq x_0$ :

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

**Beispiel 9.2** Für  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x$ , gilt

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

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

**Beispiel 9.3** Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = |x|$ , ist nicht differenzierbar in  $x_0 = 0$ :

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

Die rechts- und linksseitige Ableitung existieren in  $x_0 = 0$ , sie sind aber verschieden.

**Satz 9.1 (differenzierbar  $\Rightarrow$  stetig)** Sei  $I$  ein offenes Intervall. Ist  $f : I \rightarrow \mathbb{R}^n$  differenzierbar in  $x_0$ , so ist  $f$  auch stetig in  $x_0$ .

BEWEIS: Mit den Grenzwertregeln, Satz 7.6(1), folgt für  $x \rightarrow x_0$

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

Dies zeigt  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ , also  $f$  stetig in  $x_0$  nach Lemma 7.2.  $\square$

**Satz 9.2 (Differentiationsregeln)** Seien  $f, g : I \rightarrow \mathbb{R}$  differenzierbar in  $x_0 \in I$ . Dann sind auch die Funktionen  $\alpha f + \beta g$  ( $\alpha, \beta \in \mathbb{R}$ ),  $fg$  und  $f/g$  (im Fall  $g(x_0) \neq 0$ ) in  $x_0$  differenzierbar mit folgenden Ableitungen:

(1) *Linearität:*

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

(2) *Produktregel:*

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

(3) *Quotientenregel:*

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

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

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

Natürlich gilt die Aussage mit demselben Argument auch für vektorwertige Funktionen. Die Produktregel folgt durch „Mischen der Terme“:

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

wobei die Stetigkeit von  $g$  in  $x_0$  benutzt wurde (Satz 9.1). Wir zeigen die Quotientenregel zunächst für die Funktion  $1/g$ , also  $f \equiv 1$ . Es gibt ein  $\delta > 0$  mit  $g(x) \neq 0$  für  $|x - x_0| < \delta$  nach Lemma 7.1. Für diese  $x \in I$  gilt

$$\begin{aligned} \frac{1}{x - x_0} \left( \frac{1}{g(x)} - \frac{1}{g(x_0)} \right) &= - \frac{1}{g(x)g(x_0)} \frac{g(x) - g(x_0)}{x - x_0} \\ &\rightarrow - \frac{1}{g(x_0)^2} g'(x_0) \quad \text{mit } x \rightarrow x_0. \end{aligned}$$

Für beliebiges  $f$  schreiben wir  $f/g = f \cdot \frac{1}{g}$  und verwenden die Produktregel.  $\square$

**Beispiel 9.4** Wir zeigen für  $f_n(x) = x^n$  per Induktion  $f'(x) = nx^{n-1}$ . Für  $n = 1$  gilt das, denn nach Beispiel 9.2 gilt  $f'_1(x) = 1$ . Jetzt berechne mit der Produktregel

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

Per Induktion gilt  $f'_{n-1}(x) = (n-1)x^{n-2}$  und damit  $f'_n(x) = nx^{n-1}$ . Allgemeiner ergibt sich mit Satz 9.2(1) für Polynome  $p(x) = \sum_{k=0}^n a_k x^k$  die Formel

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

**Beispiel 9.5** Für  $f(x) = x^{-n}$ ,  $n \in \mathbb{N}$ , gilt  $f'(x) = -nx^{-n-1}$  nach der Quotientenregel:

$$f'(x) = -\frac{nx^{n-1}}{(x^n)^2} = -nx^{-n-1}.$$

**Satz 9.3 (Kettenregel)** Seien  $f : I \rightarrow \mathbb{R}$ ,  $g : J \rightarrow \mathbb{R}$  mit  $f(I) \subset J$ . Ist  $f$  in  $x_0 \in I$  differenzierbar und  $g$  in  $y_0 = f(x_0) \in J$  differenzierbar, so ist auch  $g \circ f : I \rightarrow \mathbb{R}$  in  $x_0$  differenzierbar und hat die Ableitung

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

BEWEIS: Wir verwenden die Charakterisierung des Grenzwerts mittels Folgen. Sei  $x_n \neq x_0$  mit  $x_n \rightarrow x_0$ , also  $f(x_n) \rightarrow f(x_0)$  wegen  $f$  stetig in  $x_0$ . Ist  $f(x_n) \neq f(x_0)$  für alle  $n$ , so folgt

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

Ist dagegen  $f(x_n) = f(x_0)$  für alle  $n$ , so gilt  $f'(x_0) = \lim_{n \rightarrow \infty} \frac{f(x_n) - f(x_0)}{x_n - x_0} = 0$  und es folgt

$$\lim_{n \rightarrow \infty} \frac{g(f(x_n)) - g(f(x_0))}{x_n - x_0} = 0 = g'(f(x_0)) f'(x_0).$$

Für  $x_n$  beliebig betrachten wir die Teilfolgen mit  $f(x_n) \neq f(x_0)$  bzw.  $f(x_n) = f(x_0)$ . Falls beide Folgen nicht endlich sind, hat der Differenzenquotient aber denselben Limes, nämlich  $g'(f(x_0)) f'(x_0)$ . Also konvergiert die ganze Folge gegen  $g'(f(x_0)) f'(x_0)$ , der Satz ist bewiesen.  $\square$

**Satz 9.4 (Differenzierbarkeit der Umkehrfunktion)** Sei  $f : I \rightarrow \mathbb{R}$  streng monoton und stetig auf dem offenen Intervall  $I$ , und differenzierbar in  $x_0$  mit  $f'(x_0) \neq 0$ . Dann ist  $I^* = f(I)$  ein offenes Intervall, und die Umkehrfunktion  $g : I^* \rightarrow I$  ist differenzierbar in  $y_0 = f(x_0)$  mit Ableitung

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

BEWEIS: Nach Satz 8.2 ist  $I^*$  ein offenes Intervall und  $g : I^* \rightarrow \mathbb{R}$  streng monoton und stetig, insbesondere  $g(y) \rightarrow g(y_0) = x_0$  mit  $y \rightarrow y_0$ . Wir verwenden wieder die Charakterisierung des Grenzwerts mittels Folgen. Sei  $y_n \neq y_0$  mit  $y_n \rightarrow y_0$ , dann ist  $g(y_n) \neq g(y_0)$  und

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

□

*Bemerkung.* Die Bedingung  $f'(x_0) \neq 0$  ist nicht nur hinreichend sondern auch notwendig für die Existenz von  $g'(f(x_0))$ . Denn sind  $f, g$  Umkehrfunktionen, die an den Stellen  $x_0$  bzw.  $y_0 = f(x_0)$  differenzierbar sind, so folgt aus der Kettenregel

$$g(f(x)) = x \quad \Rightarrow \quad g'(f(x_0))f'(x_0) = 1,$$

insbesondere  $f'(x_0) \neq 0$ . Zum Beispiel hat die Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^3$ , die Ableitung  $f'(0) = 0$ , die Umkehrfunktion  $g$  kann daher im Nullpunkt nicht differenzierbar sein. Das lässt sich auch direkt überprüfen, es gilt

$$g(y) = \begin{cases} y^{1/3} & \text{für } y \geq 0 \\ -|y|^{1/3} & \text{für } y \leq 0 \end{cases} \quad \text{also} \quad \frac{g(y) - g(0)}{y - 0} = |y|^{-2/3} \rightarrow +\infty \text{ mit } y \rightarrow 0.$$

Wenn man die Formel für die Ableitung  $g'(y)$  vergessen hat, hat man sie mit der Kettenregel herleiten, es gilt (siehe oben)

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

**Beispiel 9.6** Die Funktion  $f: (0, \infty) \rightarrow \mathbb{R}$ ,  $f(x) = x^n$ ,  $n \in \mathbb{N}$ , ist streng monoton wachsend und stetig mit  $f'(x) = nx^{n-1} > 0$ , siehe Beispiel 9.4. Die Umkehrfunktion ist  $g: (0, \infty) \rightarrow \mathbb{R}$ ,  $g(y) = y^{1/n}$ . Nach Satz 9.4 ist  $g$  differenzierbar mit

$$g'(y) = \frac{1}{f'(g(y))} = \frac{1}{n(y^{1/n})^{n-1}} = \frac{1}{n}y^{\frac{1}{n}-1}.$$

Für  $h(y) = y^\alpha$  mit  $\alpha = m/n \in \mathbb{Q}$  gilt  $h(y) = g(y)^m$ , also folgt weiter aus der Kettenregel, siehe auch Beispiel 9.5,

$$h'(y) = mg(y)^{m-1}g'(y) = m y^{\frac{m}{n} - \frac{1}{n}} \frac{1}{n} y^{\frac{1}{n}-1} = \alpha y^{\alpha-1}.$$

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

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

Bei dieser saloppen Notation ist jedoch darauf zu achten, wo die jeweiligen Funktionen definiert sind. In jedem Fall ist die Bezeichnung  $\frac{d}{dx}$  für den Ableitungsoperator üblich und praktisch.

Differenzierbarkeit kann als Approximierbarkeit durch eine affin-lineare Funktion interpretiert werden, wobei der Fehler schneller als linear verschwindet. Diese Deutung wird uns bei Funktionen mehrerer Variabler erneut begegnen. Genauer ist Folgendes gemeint.

**Lemma 9.2** Genau dann hat  $f : I \rightarrow \mathbb{R}^n$  in  $x_0 \in I$  die Ableitung  $a \in \mathbb{R}^n$ , wenn gilt:

$$\lim_{x \rightarrow x_0} \frac{f(x) - (f(x_0) + a(x - x_0))}{x - x_0} = 0. \quad (9.2)$$

BEWEIS: Folgt sofort aus der Umformung

$$\frac{f(x) - (f(x_0) + a(x - x_0))}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - a.$$

□

Um für zwei Funktionen  $f, g$  das asymptotische Verhalten für  $x \rightarrow x_0$  zu vergleichen, werden manchmal die Landauschen Symbole benutzt:

$$f = \mathcal{O}(g) \text{ für } x \rightarrow x_0 \Leftrightarrow \limsup_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} < \infty \quad (f \text{ ist abgeschätzt durch } g),$$

$$f = o(g) \text{ für } x \rightarrow x_0 \Leftrightarrow \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = 0 \quad (f \text{ wird klein relativ zu } g).$$

Hier stehen  $o(g)$  bzw.  $\mathcal{O}(g)$  nicht für eine konkrete Funktion, sondern symbolisch für alle Funktionen mit dem Konvergenzverhalten, das rechts angegeben ist. Die Gleichung  $f'(x_0) = a$  kann damit äquivalent wie folgt geschrieben werden:

$$f(x) = f(x_0) + a(x - x_0) + o(x - x_0) \quad \text{für } x \rightarrow x_0.$$

Das ist unter anderem in der Physik höchst beliebt. Beim Rechnen mit dem Ausdruck  $o(x - x_0)$  ist Vorsicht angesagt, wie gesagt muss man sich an die Bedeutung erinnern. Eine Abhängigkeit der Funktion  $f$  von weiteren Parametern ist in der Notation  $o(x - x_0)$  nicht direkt sichtbar.

## 10 Mittelwertsatz

In vielen Anwendungen der Analysis geht es darum, Eigenschaften einer Funktion aus Eigenschaften der Ableitung herzuleiten. Führt man zum Beispiel eine Stunde lang auf der Autobahn mit Geschwindigkeit höchstens 100 km/h, so ist man vom Ausgangsort höchstens 100 km entfernt. So banal das ist, wir müssen zeigen dass unser analytisches Konzept diesen Schluss auch liefert, dafür ist der Mittelwertsatz ein einfaches und effektives Instrument. Um ihn zu beweisen brauchen wir die Existenz von Extremwerten – Maxima und Minima – von stetigen Funktionen. Das ist ebenfalls ein fundamentales Resultat.

Wir benötigen einen Satz, der die Existenz von Extremalstellen für eine stetige Funktion  $f : I \rightarrow \mathbb{R}$  (natürlich mit  $I \neq \emptyset$ ) allgemein garantiert. Oft wird das so formuliert, dass die Funktion ihr Minimum bzw. Maximum annimmt. Das ist allerdings missverständlich, denn diese Bezeichnungen implizieren schon die Existenz der Extremalstellen; die richtige Bezeichnung ist Infimum bzw. Supremum. Jedenfalls brauchen wir für die Existenz der Extremalstellen auch Voraussetzungen an  $I$ , wie das folgende Beispiel zeigt:

$$f : I = (0, \infty) \rightarrow \mathbb{R}, f(x) = \frac{1}{1+x}.$$

Hier ist  $\lim_{x \searrow 0} f(x) = 1 = \sup_I f$  und  $\lim_{x \nearrow \infty} f(x) = 0 = \inf_I f$ . Wir zeigen dass das Problem für ein abgeschlossenes und beschränktes Intervall nicht besteht. Allgemeiner bietet es sich an Funktionen  $f : D \rightarrow \mathbb{R}$  mit  $D \subset \mathbb{R}^n$  zu betrachten und folgenden Begriff einzuführen:

$D \subset \mathbb{R}^n$  heißt (folgen-)kompakt, wenn jede Folge  $x_k \in D$  eine Teilfolge  $x_{k_j}$  hat, die gegen ein  $x \in D$  konvergiert.

**Satz 10.1**  $D \subset \mathbb{R}^n$  ist genau dann kompakt, wenn  $D$  abgeschlossen und beschränkt ist.

BEWEIS: Sei  $D$  abgeschlossen und beschränkt, und  $x_k \in D$ . Bolzano-Weierstraß, siehe Satz 5.7, liefert eine Teilfolge  $x_{k_j}$  und ein  $x_0 \in \mathbb{R}^n$  mit  $x_{k_j} \rightarrow x_0$  für  $j \rightarrow \infty$ . Wegen  $D$  abgeschlossen gilt  $x_0 \in D$ . Sei umgekehrt  $D$  kompakt, und  $x_k \in D$  sei eine Folge mit  $x_k \rightarrow x_0 \in \mathbb{R}^n$ . Es gibt dann eine Teilfolge  $x_{k_j}$ , die gegen ein  $x'_0 \in D$  konvergiert. Aus der Eindeutigkeit des Grenzwerts folgt  $x_0 = x'_0 \in D$ , also ist  $D$  abgeschlossen. Wäre  $D$  nicht beschränkt, so gibt es zu  $k \in \mathbb{N}$  ein  $x_k \in D$  mit  $|x_k| \geq k$ . Da  $D$  kompakt existiert eine konvergente Teilfolge  $x_{k_j}$ , und diese ist nach Satz 4.2 beschränkt, ein Widerspruch.  $\square$

**Satz 10.2 (Existenz von Extremalstellen)** Sei  $D \subset \mathbb{R}^n$  kompakt, und  $f : D \rightarrow \mathbb{R}$  stetig. Dann ist  $f$  beschränkt und nimmt Infimum/Supremum an, das heißt es gibt  $x_0, x_1 \in D$  mit

$$f(x_0) = \inf_{x \in D} f(x) \quad \text{und} \quad f(x_1) = \sup_{x \in D} f(x).$$

BEWEIS: Wir zeigen die Existenz eines  $x_0 \in D$  mit  $f(x_0) = \inf_I D$ . Insbesondere ist  $f$  dann nach unten beschränkt, denn  $f(x) \geq f(x_0)$  für alle  $x \in D$ . Nach Definition des Infimums, siehe Folgerung 6.1, gibt es nun eine Folge  $x_k \in D$  mit  $f(x_k) \rightarrow \inf_D f$ , eine sogenannte *Minimalfolge*. Es gibt eine Teilfolge  $x_{k_j} \rightarrow x_0 \in D$ . Wegen  $f$  stetig gilt

$$f(x_0) = \lim_{j \rightarrow \infty} f(x_{k_j}) = \inf_D f.$$

$\square$

Jetzt zurück zu Funktionen einer reellen Variablen.

**Definition 10.1 (Lokale Extrema)** Die Funktion  $f : (a, b) \rightarrow \mathbb{R}$  hat in  $x_0 \in (a, b)$  ein lokales Minimum, falls es ein  $\delta > 0$  gibt, so dass gilt:

$$f(x_0) \leq f(x) \quad \text{für alle } x \in B_\delta(x_0) = (x_0 - \delta, x_0 + \delta).$$

Ist sogar  $f(x_0) < f(x)$  für  $x \in B_\delta(x_0)$ ,  $x \neq x_0$ , so heißt das lokale Minimum isoliert. Ein (isoliertes) lokales Maximum ist entsprechend definiert.

Oft gilt das Interesse weniger den lokalen Extrema, sondern den globalen Extrema wie in Satz 10.2 konstruiert. Bei einer Wanderung freut man sich zum Beispiel, wenn man auf dem Weg einen lokalen Aussichtspunkt erreicht. Muss man aber danach einen riesigen Berg bezwingen, so ist dessen Höhe relevanter. Der Begriff des lokalen Extremums wird hier eingeführt, um in folgendem Satz eine globale Voraussetzung zu vermeiden.

**Satz 10.3 (notwendige Bedingung für Extrema)** Die Funktion  $f : (a, b) \rightarrow \mathbb{R}$  habe in  $x_0 \in (a, b)$  ein lokales Extremum. Ist  $f$  in  $x_0$  differenzierbar, so gilt  $f'(x_0) = 0$ .

BEWEIS: Sei  $x_0$  lokales Minimum von  $f$ , also  $f(x) \geq f(x_0)$  auf  $B_\delta(x_0)$  für ein  $\delta > 0$ . Es folgt

$$\frac{f(x) - f(x_0)}{x - x_0} \quad \begin{cases} \geq 0 & \text{für } x \in (x_0, x_0 + \delta), \\ \leq 0 & \text{für } x \in (x_0 - \delta, x_0). \end{cases}$$

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

Der Beweis zeigt tatsächlich eine Aussage über einseitige lokale Minima, und zwar gilt

$$\begin{aligned} f(x) \geq f(x_0) \text{ auf } [x_0, x_0 + \delta) &\Rightarrow f'_+(x_0) := \lim_{x \searrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \geq 0, \\ f(x) \geq f(x_0) \text{ auf } (x_0 - \delta, x_0] &\Rightarrow f'_-(x_0) := \lim_{x \nearrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \leq 0. \end{aligned}$$

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

**Satz 10.4 (Mittelwertsatz der Differentialrechnung)** Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig, differenzierbar auf  $(a, b) \neq \emptyset$ . Dann gibt es ein  $\xi \in (a, b)$  mit

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

BEWEIS: Wir zeigen die Behauptung zuerst im Fall  $f(a) = f(b) = 0$  (Satz von Rolle). Gesucht ist dann ein  $\xi \in (a, b)$  mit  $f'(\xi) = 0$ . Nach Satz 10.2 gibt es  $\xi_1, \xi_2 \in [a, b]$  mit

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

Ist  $\xi_1 \in (a, b)$ , so folgt  $f'(\xi_1) = 0$  nach Satz 10.3 und wir können  $\xi = \xi_1$  wählen. Analog, wenn  $\xi_2 \in (a, b)$ . Es bleibt der Fall wenn  $\xi_1, \xi_2$  beides Randpunkte sind. Aber dann folgt  $\inf f = \sup f = 0$  und  $f$  ist die Nullfunktion, also gilt  $f'(x) = 0$  für alle  $x \in (a, b)$ . Seien nun  $f(a), f(b) \in \mathbb{R}$  beliebig. Definiere  $h : [a, b] \rightarrow \mathbb{R}$  durch Abziehen der Sekante:

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

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

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

□

**Folgerung 10.1 (Monotoniekriterien)** Sei  $f$  differenzierbar auf  $(a, b)$ , stetig auf  $[a, b]$ . Dann gelten folgende Aussagen:

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

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

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

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

□

**Folgerung 10.2 (Schrankensatz)** Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig und differenzierbar auf  $(a, b)$ . Dann gilt für  $a \leq x_1 \leq x_2 \leq b$

$$\begin{aligned} f'(x) \geq m \text{ auf } (a, b) &\Rightarrow f(x_2) - f(x_1) \geq m(x_2 - x_1) \\ f'(x) \leq M \text{ auf } (a, b) &\Rightarrow f(x_2) - f(x_1) \leq M(x_2 - x_1) \end{aligned}$$

BEWEIS: Wir zeigen die erste Aussage. Mit  $g(x) = mx$  gilt  $(f - g)' = f' - g' \geq m - m = 0$ . Nach Folgerung 10.1 ist  $f - g$  wachsend, das heißt  $f(x_2) - mx_2 \geq f(x_1) - mx_1$ . □

Der Mittelwertsatz ist für vektorwertige Funktionen so nicht richtig. Das zeigt das ad hoc Beispiel  $f(x) = (x^2, x^3)$  auf  $[0, 1]$ . Hätten wir

$$\frac{f(1) - f(0)}{1 - 0} = f'(\xi) \quad \text{für ein } \xi \in (0, 1),$$

so wäre  $(1, 1) = (2\xi, 3\xi^2)$ , das ist unmöglich. Das Problem ist, dass es zwar für jede Komponente eine Zwischenstelle gibt, im allgemeinen kann aber nicht dieselbe Stelle  $\xi$  gewählt werden. Aber es gilt auch im vektorwertigen Fall folgende Version des Schrankensatzes.

**Folgerung 10.3 ( $f'$  beschränkt  $\Rightarrow f$  Lipschitz)** Sei  $f : [a, b] \rightarrow \mathbb{R}^n$  stetig und auf  $(a, b)$  differenzierbar. Falls  $\|f'(x)\| \leq L$  für alle  $x \in (a, b)$ , so folgt

$$\|f(x_2) - f(x_1)\| \leq L|x_2 - x_1| \quad \text{für alle } x_1, x_2 \in [a, b].$$

BEWEIS: Um die Aussage auf den reellwertigen Fall zu reduzieren betrachten wir

$$\varphi : [a, b] \rightarrow \mathbb{R}, \varphi(x) = \langle v, f(x) \rangle \quad \text{für } v \in \mathbb{R}^n, \|v\| = 1.$$

Aus der Ungleichung von Cauchy-Schwarz, siehe Satz 6.5, folgt

$$\varphi'(x) = \langle v, f'(x) \rangle \leq \|v\| \|f'(x)\| \leq L.$$

Für  $x_1 < x_2$  erhalten wir aus Folgerung 10.2, dem Schrankensatz,

$$\langle v, f(x_2) - f(x_1) \rangle = \varphi(x_2) - \varphi(x_1) \leq L(x_2 - x_1) = L|x_2 - x_1|.$$

Sei nun  $f(x_2) \neq f(x_1)$ , wähle dann  $v = (f(x_2) - f(x_1))/\|f(x_2) - f(x_1)\|$ . Es folgt

$$\|f(x_2) - f(x_1)\| = \left\langle \frac{f(x_2) - f(x_1)}{\|f(x_2) - f(x_1)\|}, f(x_2) - f(x_1) \right\rangle = \langle v, f(x_2) - f(x_1) \rangle \leq L|x_2 - x_1|.$$

□

Wir kommen jetzt zu höheren Ableitungen.

**Definition 10.2** Die zweite Ableitung von  $f : (a, b) \rightarrow \mathbb{R}$  in  $x_0$  ist  $f''(x_0) = (f')'(x_0)$ , und allgemeiner ist die  $k$ -te Ableitung

$$f^{(k)}(x_0) = (f^{(k-1)})'(x_0).$$

Damit  $f^{(k)}(x_0)$  definiert ist, müssen also die Ableitungen bis Ordnung  $k-1$  in einer Umgebung von  $x_0$  definiert sein, und es muss  $f^{(k-1)}$  in  $x_0$  differenzierbar sein.

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

**Satz 10.5** Sei  $f : (a, b) \rightarrow \mathbb{R}$  differenzierbar. In  $x_0 \in (a, b)$  gelte  $f'(x_0) = 0$ , und  $f''(x_0)$  sei definiert. Dann gilt:

- (1) Ist  $f''(x_0) > 0$ , so hat  $f$  in  $x_0$  ein isoliertes, lokales Minimum.
- (2) Hat  $f$  in  $x_0$  ein lokales Minimum, so folgt  $f''(x_0) \geq 0$ .

Analoge Aussagen gelten mit umgekehrten Ungleichungen für Maxima.

BEWEIS: Da  $f'(x_0) = 0$  nach Voraussetzung, gilt

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

Sei nun  $f''(x_0) > 0$ . Dann gibt es ein  $\delta > 0$  mit  $f'(x) > 0$  auf  $(x_0, x_0 + \delta)$  und  $f'(x) < 0$  auf  $(x_0 - \delta, x_0)$ . Nach Folgerung 10.1 ist  $f$  dann streng monoton wachsend auf  $[x_0, x_0 + \delta)$  und streng monoton fallend auf  $(x_0 - \delta, x_0]$ , hat also in  $x_0$  ein isoliertes, lokales Minimum. Wäre  $f''(x_0) < 0$  in (2), so hätte  $f$  in  $x_0$  ein isoliertes lokales Maximum im Widerspruch zur Voraussetzung. □

Die Funktion  $f(x) = x^4$  zeigt, dass in einem isolierten Minimum  $f''(x_0) = 0$  gelten kann. Häufig ist man nicht wirklich an den lokalen Minima, sondern am globalen Minimum interessiert. Dafür ist der Begriff der Konvexität relevant.

**Definition 10.3** Sei  $I \subset \mathbb{R}$  ein Intervall.  $f : I \rightarrow \mathbb{R}$  heißt *konvex*, falls gilt:

$$f((1-t)x_0 + tx_1) \leq (1-t)f(x_0) + tf(x_1) \quad \forall x_0, x_1 \in I, t \in [0, 1]. \quad (10.1)$$

Gilt dies mit  $\geq$  statt mit  $\leq$ , so heißt  $f$  *konkav*.

Die Sekante durch  $(x_0, f(x_0))$  und  $(x_1, f(x_1))$  hat die Geradengleichung

$$y = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) =: g(x).$$

An der Stelle  $x(t) = (1-t)x_0 + tx_1$  gilt

$$g(x(t)) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}t(x_1 - x_0) = (1-t)f(x_0) + tf(x_1).$$

Die Konvexität bedeutet also, dass der Graph von  $f$  stets unterhalb der Sekanten liegt,

**Satz 10.6 (Konvexitätskriterien)** Sei  $f : (a, b) \rightarrow \mathbb{R}$  differenzierbar. Dann sind folgende Aussagen äquivalent:

- (1)  $f'$  ist monoton wachsend auf  $(a, b)$ .
- (2)  $f$  ist konvex.
- (3)  $f(x_1) \geq f(x_0) + f'(x_0)(x_1 - x_0)$  für alle  $x_0, x_1 \in (a, b)$ .

Ist  $f$  zweimal differenzierbar auf  $(a, b)$ , so ist außerdem äquivalent:

- (4)  $f'' \geq 0$ .

BEWEIS: Wir zeigen  $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$ . Sei (1) erfüllt. Angenommen  $f$  ist nicht konvex, das heißt es gibt  $x_0, x_1 \in I$  und ein  $t \in [0, 1]$  mit

$$g(t) = f((1-t)x_0 + tx_1) - ((1-t)f(x_0) + tf(x_1)) > 0.$$

Wir können  $x_1 > x_0$  annehmen, sonst vertausche  $x_0, x_1$  und ersetze  $t$  durch  $1-t$ . Die Funktion  $g : [0, 1] \rightarrow \mathbb{R}$  ist stetig mit  $g(0) = g(1) = 0$ . Nach Satz 10.2 nimmt  $g$  ihr strikt positives Maximum in einem  $t_0 \in (a, b)$  an, und es gilt  $g'(t_0) = 0$  nach Satz 10.3. Aber wegen  $f'$  wachsend ist auch  $g'$  wachsend:

$$g'(t) = \underbrace{f'(x_0 + t(x_1 - x_0))}_{\text{wachsend}} \underbrace{(x_1 - x_0)}_{>0} - \underbrace{(f(x_1) - f(x_0))}_{\text{konstant}}.$$

Insbesondere folgt  $g'(t) \geq 0$  für alle  $t \geq t_0$ , und daraus  $g(1) \geq g(t_0) > 0$ , Widerspruch.

Es gelte jetzt (2), also gilt für alle  $t \in (0, 1)$  und  $x_0, x_1 \in (a, b)$  mit  $x_0 \neq x_1$

$$(x_1 - x_0) \frac{f(x_0 + t(x_1 - x_0)) - f(x_0)}{t(x_1 - x_0)} \leq f(x_1) - f(x_0).$$

Mit  $t \searrow 0$  folgt  $(x_1 - x_0)f'(x_0) \leq f(x_1) - f(x_0)$ , das ist Ungleichung (3). Schließlich folgt aus (3), indem wir dort  $x_0$  und  $x_1$  vertauschen und addieren,

$$(f'(x_1) - f'(x_0))(x_1 - x_0) \geq 0,$$

womit wiederum (1) gezeigt ist. Schließlich: ist  $f$  zweimal differenzierbar, so impliziert  $f'' \geq 0$  Bedingung (1) nach Folgerung 10.1, und umgekehrt folgt  $f'' \geq 0$  aus (1) einfach durch Betrachtung des Differenzenquotienten.  $\square$

**Beispiel 10.1 (Youngsche Ungleichung)** Für  $x, y \geq 0$  gilt die Ungleichung

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \quad \text{falls } p, q \in (1, \infty) \text{ mit } \frac{1}{p} + \frac{1}{q} = 1.$$

Dazu betrachten wir für festes  $y > 0$  die Funktion

$$f : (0, \infty) \rightarrow \mathbb{R}, f(x) = xy - \frac{x^p}{p}.$$

Es gilt  $f'(x) = y - x^{p-1}$  und  $f''(x) = -(p-1)x^{p-2} \leq 0$ , das heißt  $f$  ist konkav nach Satz 10.6. Aber  $f'(x_0) = 0$  für  $x_0 = y^{1/(p-1)}$ , also folgt mit Satz 10.6(3) wegen  $p/(p-1) = q$

$$xy - \frac{x^p}{p} = f(x) \leq f(x_0) + f'(x_0)(x - x_0) = y^{1/(p-1)}y - \frac{y^{p/(p-1)}}{p} = \frac{y^q}{q}.$$

**Definition 10.4** Sei  $I \subset \mathbb{R}$  ein offenes Intervall und  $k \in \mathbb{N}_0$ . Wir bezeichnen mit  $C^k(I)$  den  $\mathbb{R}$ -Vektorraum der  $k$  mal stetig differenzierbaren Funktionen  $f : I \rightarrow \mathbb{R}$ , das heißt

$$C^k(I) = \{f : I \rightarrow \mathbb{R} : f^{(i)} : I \rightarrow \mathbb{R} \text{ sind definiert und stetig für } i = 0, 1, \dots, k\}.$$

Weiter definieren wir  $C^\infty(I)$  als den  $\mathbb{R}$ -Vektorraum der unendlich oft differenzierbaren Funktionen, also ist

$$C^\infty(I) = \bigcap_{k \geq 0} C^k(I).$$

Der Umgang mit  $C^\infty$ -Funktionen ist besonders angenehm, weil die Klasse im Gegensatz zu den Räumen  $C^k(I)$  unter der Bildung von Ableitungen abgeschlossen ist. Es ist klar, dass Polynome unendlich oft differenzierbar sind; es gibt aber sehr viel mehr Funktionen in  $C^\infty(I)$ . Wir werden weitere im nächsten Abschnitt mithilfe der Exponentialfunktion konstruieren.

## 11 Die reelle Exponentialfunktion

Der berühmteste Mathematiker des 18. Jahrhunderts war Euler<sup>16</sup>, der aus Basel stammte und das erste Lehrbuch der Analysis verfasst hat (*Introductio in analysin infinitorum*, 1748). Er hat dort den Begriff der Funktion einführt und diese mit Reihendarstellungen definiert, vor allem natürlich die Exponentialfunktion. Im Laufe seines Lebens war er auch am Hof von Friedrich dem Großen in Berlin, aber er mochte nicht dessen gelehrte Diskussionsrunden u.a. mit Voltaire. So ist er dann nach Sankt Petersburg an den Hof Katharina der Großen gegangen. Insgesamt hat Euler etwa 800 mathematische Arbeiten geschrieben.

Wachstums- oder Zerfallsprozesse spielen in vielen Bereichen der Natur und der Anwendungen eine Rolle. Dabei wird die zukünftige (oder vergangene) Entwicklung einer Größe  $K(t)$ ,  $t \in I$ , durch ein Wachstumsgesetz und einen Startwert  $K(t_0) = K_0$  charakterisiert. Ein diskretes Modell ist der Zinseszins, bei dem das Intervall in Zeitschritte der Länge  $\Delta t$  unterteilt wird. In jedem Zeitschritt  $[t, t + \Delta t]$  wird linear mit der Rate  $\lambda \in \mathbb{R}$  verzinst. Das bedeutet

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

Um ein kontinuierliches Modell herzuleiten, lassen wir hier  $\Delta t$  gegen Null gehen. Formal erhalten wir das Wachstumsgesetz, und die Anfangsbedingung,

$$K'(t) = \lambda K(t) \quad \text{für } t \in I, \quad K(t_0) = K_0. \quad (11.1)$$

Ob dieses Modell sinnvoll ist, muss mathematisch untersucht werden. Erstens ist die Existenz einer Lösung  $K(t)$  zu klären, sonst wäre alles Quatsch. Zweitens sollte die Lösung eindeutig sein, damit das Modell Aussagen über die Zukunft oder die Vergangenheit erlaubt. Weitere Eigenschaften wie Monotonie oder asymptotisches Verhalten für große Zeiten sind natürlich auch von Interesse. Die Aufgabe, zu einem Wachstumsgesetz und Startwert die zeitliche Entwicklung zu bestimmen, bezeichnet man allgemein als *Anfangswertproblem*. Wir kennen schon ein triviales aber fundamentales Beispiel: das Nullwachstum

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

Die eindeutige Lösung ist die konstante Funktion  $K(t) = K_0$ , siehe Folgerung 10.1. Enthält ein Wachstumsgesetz die Ableitung der gesuchten Funktion, so spricht man von einer Differentialgleichung. Speziell ist  $K' = \lambda K$  die Differentialgleichung des natürlichen Wachstums mit Rate  $\lambda \in \mathbb{R}$ . Wir studieren nun das Anfangswertproblem

$$f' = f \quad \text{auf } \mathbb{R}, \quad f(0) = 1. \quad (11.2)$$

Das ist der Spezialfall von (11.1) mit  $I = \mathbb{R}$ ,  $\lambda = 1$ ,  $t_0 = 0$  und  $K_0 = 1$ . Wir werden die Lösung des allgemeinen Problems (11.1) später durch Skalierung bekommen.

**Lemma 11.1** *Eine Lösung  $f : \mathbb{R} \rightarrow \mathbb{R}$  von (11.2) hat folgende Eigenschaften:*

- (a)  $f \in C^0(\mathbb{R})$ ,

---

<sup>16</sup>Leonhard Euler, 1707-1783

(b)  $f(x)f(-x) = 1$  für alle  $x \in \mathbb{R}$ ,

(c)  $f$  ist strikt positiv.

BEWEIS: Nach Voraussetzung ist  $f$  differenzierbar, also stetig mit Satz 9.1. Weiter gilt

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

Also ist  $f(x)f(-x) = f(0)f(-0) = 1$ , insbesondere  $f(x) \neq 0$  für alle  $x \in \mathbb{R}$ . Da  $f(0) = 1 > 0$  ist  $f(x) > 0$  für alle  $x$ , denn sonst gibt es eine Nullstelle nach dem Zwischenwertsatz.  $\square$

**Satz 11.1 (Eindeutigkeit)** *Es gibt höchstens ein  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit  $f' = f$  auf  $\mathbb{R}$  und  $f(0) = 1$ .*

BEWEIS: Seien  $f, g$  zwei Lösungen von (11.2). Dann ist  $g > 0$  nach Lemma 11.1, und es folgt

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

Also ist  $f/g$  konstant mit  $f(0)/g(0) = 1$ , und somit  $f = g$ .  $\square$

Zur Existenz machen wir mit Cauchy den Reihenansatz  $f(x) = \sum_{k=0}^{\infty} a_k x^k$ . Zur Motivation können wir hier nur die Idee anführen, die Lösung durch Approximation mit Polynomen zu gewinnen, wir kommen in einem späteren Kapitel darauf zurück (Stichwort Taylorreihe). Wir berechnen nun

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

Koeffizientenvergleich in  $f' = f$  liefert  $(k+1)a_{k+1} = a_k$  bzw.  $a_{k+1} = a_k/(k+1)$ . Mit  $a_0 = f(0) = 1$  folgt per Induktion  $a_k = 1/k!$  für  $k \in \mathbb{N}_0$ . Natürlich war diese Herleitung nicht rigoros, die Vertauschung der Ableitung mit der unendlichen Reihe sowie der Koeffizientenvergleich wurden nicht begründet. Und überhaupt ist ja unklar ob es eine Lösung von (11.2) gibt die so als Reihe geschrieben werden kann. Aber diese Einwände wischen wir beiseite, unser Ziel ist nur eine potentielle Formel für die Lösung, wir haben jetzt den *educated guess*<sup>17</sup>

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

Wir zeigen dass diese Reihe für alle  $x \in \mathbb{R}$  konvergiert und die Lösung des Anfangswertproblems (11.2) liefert. Dabei ist die Reihe definiert als Grenzwert der endlichen Summen

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

Wir behaupten dass für jedes  $x \in \mathbb{R}$  die Folge  $f_n(x)$  eine Cauchyfolge ist. Sei  $k \in \mathbb{N}_0$  mit  $k+1 \geq 2|x|$ . Für alle  $m > k$  gilt dann die Abschätzung

$$\frac{|x|^m}{m!} = \frac{|x|^k}{k!} \frac{|x|}{k+1} \cdots \frac{|x|}{m} \leq \frac{|x|^k}{k!} \left(\frac{1}{2}\right)^{m-k} \rightarrow 0 \quad \text{mit } m \rightarrow \infty.$$

---

<sup>17</sup>die fundierte Vermutung

Daraus folgt weiter für alle  $n > k$

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

Somit ist  $f_n(x)$  eine Cauchyfolge, der Grenzwert  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  existiert.

**Satz 11.2 (Existenz)** Die Reihe  $f(x) = \sum_{k=0}^{\infty} x^k/k!$  konvergiert für alle  $x \in \mathbb{R}$  und ist Lösung des Anfangswertproblems  $f' = f$  auf  $\mathbb{R}$ ,  $f(0) = 1$ .

BEWEIS: Es gilt  $f(0) = \lim_{n \rightarrow \infty} f_n(0) = 1$ . Die Funktionen  $f_n(x)$ ,  $n \geq 1$ , erfüllen

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

Hieraus wollen wir  $f' = f$  durch Grenzübergang  $n \rightarrow \infty$  herleiten, wir brauchen dazu einige Abschätzungen. Für  $|x| \leq R$  und alle  $n \in \mathbb{N}_0$  haben wir

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

Es folgt  $|f'_n(x)| = |f_{n-1}(x)| \leq C$  für  $|x| \leq R$ , also mit dem Schrankensatz

$$|f_n(x_1) - f_n(x_2)| \leq C |x_1 - x_2| \quad \text{für alle } n \in \mathbb{N}_0, x_1, x_2 \in [-R, R].$$

Weiter folgt  $|f'_n(x_1) - f'_n(x_2)| = |f_{n-1}(x_1) - f_{n-1}(x_2)| \leq C |x_1 - x_2|$ . Sei nun  $x_0 \in (-R, R)$  gegeben. Zu  $x \in (-R, R)$  gibt es ein  $\xi$  zwischen  $x_0$  und  $x$ , so dass

$$\left| \frac{f_n(x) - f_n(x_0)}{x - x_0} - f'_n(x_0) \right| = |f'_n(\xi) - f'_n(x_0)| \leq C |\xi - x_0| \leq C |x - x_0|.$$

Hier lassen wir  $n \rightarrow \infty$  gehen. Wegen  $f'_n(x_0) = f_{n-1}(x_0) \rightarrow f(x_0)$  erhalten wir

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

Mit  $x \rightarrow x_0$  folgt  $f'(x_0) = f(x_0)$ . Da  $R > 0$  beliebig, gilt  $f' = f$  auf  $\mathbb{R}$ . □

**Definition 11.1 (Exponentialfunktion)** Die Exponentialfunktion ist

$$\exp : \mathbb{R} \rightarrow \mathbb{R}, \exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{11.3}$$

Sie ist eindeutige Lösung des Anfangswertproblems  $f' = f$  auf  $\mathbb{R}$ ,  $f(0) = 1$ .

Wie angekündigt erhalten wir die Lösung von (11.1) nun durch Skalierung.

**Folgerung 11.1** Die eindeutige Lösung des Anfangswertproblems (11.1) ist

$$K(t) = K_0 \exp(\lambda(t - t_0)) \quad \text{für } t \in I.$$

BEWEIS: Für  $K(t)$  wie behauptet ist  $K(t_0) = K_0 \exp(0) = K_0$ , und nach Kettenregel

$$K'(t) = \lambda K_0 \exp'(\lambda(t - t_0)) = \lambda K_0 \exp(\lambda(t - t_0)) = \lambda K(t).$$

Für eine beliebige Lösung  $K(t)$  berechnen wir, vgl. den Beweis von Lemma 11.1,

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

Einsetzen von  $t = t_0$  ergibt  $\exp(-\lambda(t - t_0))K(t) = \exp(0)K(t_0) = K_0$ . Jetzt verwenden wir Lemma 11.1 (b), und zwar folgt durch Multiplikation mit  $\exp(\lambda(t - t_0))$

$$K(t) = \exp(\lambda(t - t_0)) \exp(-\lambda(t - t_0))K(t) = K_0 \exp(\lambda(t - t_0)).$$

□

Jetzt interessieren wir uns für weitere Eigenschaften der Exponentialfunktion. Dazu könnten wir die Reihendarstellung benutzen, es ist aber allemal einfacher mit der Differentialgleichung zu argumentieren.

**Satz 11.3 (Eigenschaften der Exponentialfunktion)** *Es gelten folgende Aussagen:*

(a)  $\exp(1) = e = 2,71828\dots$

(b)  $\exp$  ist in  $C^\infty(\mathbb{R})$ , strikt positiv und streng monoton wachsend mit Grenzwerten

$$\lim_{x \nearrow \infty} \exp(x) = +\infty \quad \text{und} \quad \lim_{x \searrow -\infty} \exp(x) = 0.$$

(c) Es gilt die Funktionalgleichung

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

insbesondere  $\exp(x) \exp(-x) = 1$ .

BEWEIS: (a) Nach Definitionen 11.3 und 5.3 gilt  $\exp(1) = \sum_{k=0}^{\infty} 1/k! = e$ .

(b) Es gilt  $\exp' = \exp$ , also ist  $\exp$  differenzierbar und insbesondere stetig auf  $\mathbb{R}$ , vgl. Lemma 11.1(a). Per Induktion folgt  $\exp \in C^k(\mathbb{R})$  und  $\exp^{(k)} = \exp$  für alle  $k \in \mathbb{N}_0$ . Weiter ist  $\exp > 0$  nach Lemma 11.1 (c). Damit ist auch  $\exp' = \exp > 0$ , also  $\exp$  streng monoton wachsend. Für  $x \geq 0$  ist dann  $\exp'(x) = \exp(x) \geq \exp(0) = 1$ , also nach Schrankensatz

$$\exp(x) \geq 1 + x \rightarrow +\infty \quad \text{für } x \nearrow \infty. \quad (11.4)$$

Schließlich gilt  $\exp(x) \exp(-x) = 1$  nach Lemma 11.1(b), also  $\exp(x) \rightarrow 0$  mit  $x \rightarrow -\infty$ .

(c) Für  $y \in \mathbb{R}$  fest ist  $g(x) = \exp(x + y)$  Lösung von  $g' = g$  auf  $\mathbb{R}$  mit Anfangswert  $g(0) = \exp(y)$ . Folgerung 11.1 ergibt

$$\exp(x + y) = g(x) = g(0) \exp(x) = \exp(x) \exp(y).$$

□

In der Sprache der Algebra besagt die Funktionalgleichung, dass  $\exp : (\mathbb{R}, +) \rightarrow (\mathbb{R}^+, \cdot)$  ein Gruppenhomomorphismus ist, wobei  $\mathbb{R}^+ = \{r \in \mathbb{R} : r > 0\}$ . Genauer ist  $\exp$  ein Isomorphismus, wie wir gleich feststellen, der inverse Gruppenisomorphismus ist der Logarithmus.

**Satz 11.4 (Definition des Logarithmus)**  $\exp : (-\infty, \infty) \rightarrow (0, \infty)$  ist streng monoton wachsend und bijektiv. Die Umkehrfunktion  $\log : (0, \infty) \rightarrow (-\infty, \infty)$  heißt (natürlicher) Logarithmus. Sie ist streng monoton wachsend und bijektiv, und hat die Grenzwerte

$$\lim_{r \searrow 0} \log(r) = -\infty \quad \text{und} \quad \lim_{r \rightarrow \infty} \log(r) = \infty. \quad (11.5)$$

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

$$\log'(r) = \frac{1}{r} \quad \text{für alle } r \in (0, \infty). \quad (11.6)$$

Schließlich gilt die Funktionalgleichung

$$\log(rs) = \log(r) + \log(s) \quad \text{für alle } r, s > 0. \quad (11.7)$$

Inbesondere gilt  $\log\left(\frac{1}{r}\right) = -\log(r)$ .

BEWEIS: Nach Satz 8.2 zur Umkehrfunktion ist  $\exp$  bijektiv, die Umkehrfunktion  $\log$  ist ebenfalls streng monoton wachsend. Die Grenzwerte (11.5) folgen aus der dortigen Aussage (8.1). Die Funktionswerte ergeben sich direkt aus  $\exp(0) = 1$  bzw.  $\exp(1) = e$ . Da  $\exp'(x) = \exp(x) > 0$ , ist die Umkehrfunktion differenzierbar und hat die Ableitung, siehe Satz 9.4,

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

Die Funktionalgleichung (11.7) ergibt sich schließlich aus  $\exp(x)\exp(y) = \exp(x+y)$ , indem wir  $x = \log r$ ,  $y = \log(s)$  einsetzen und dann den Logarithmus nehmen.  $\square$

Für  $a > 0$  hatten wir die Potenz  $a^x$  bisher nur im Fall  $x \in \mathbb{Q}$  definiert. Diese Einschränkung können wir mit dem Logarithmus nun abschaffen.

**Definition 11.2 (Potenz mit reellen Exponenten)** Für  $a > 0$ ,  $x \in \mathbb{R}$  setzen wir

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

Mit den Funktionalgleichungen von  $\log$  und  $\exp$  ergeben sich die Regeln

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

Induktion liefert  $a^{nx} = (a^x)^n$ . Mit  $x = \frac{1}{n}$  folgt  $(a^{\frac{1}{n}})^n = a^{n \cdot \frac{1}{n}} = a$ , das heißt  $a^{\frac{1}{n}} = \sqrt[n]{a}$ . Damit stimmt für  $x \in \mathbb{Q}$  die neue mit der bisherigen Definition der Potenz überein. Weiter gelten für  $x \in \mathbb{R}$  die Ableitungsregeln

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

**Satz 11.5 (Wachstum von  $\exp$  und  $\log$ )** Es gelten folgende Aussagen:

$$\lim_{x \rightarrow \infty} x^{-s} e^x = \infty \quad \text{und} \quad \lim_{x \rightarrow \infty} x^s e^{-x} = 0 \quad \text{für jedes } s > 0, \quad (11.8)$$

$$\lim_{y \rightarrow \infty} y^{-s} \log y = 0 \quad \text{und} \quad \lim_{y \searrow 0} y^s (-\log y) = 0 \quad \text{für jedes } s > 0. \quad (11.9)$$

BEWEIS: Zu  $s > 0$  wähle  $n \in \mathbb{N}$  mit  $n > s$ . Für  $x > 0$  folgt dann

$$x^{-s}e^x \geq \frac{1}{n!}x^{n-s} = \frac{1}{n!}e^{(n-s)\log x} \rightarrow \infty \quad \text{mit } x \rightarrow \infty.$$

Die zweite Aussage in (11.8) folgt durch Übergang zum Kehrwert. Für die zweite Aussage in (11.9) substituieren wir  $y = e^{-x/s}$ . Mit  $y \searrow 0$  geht  $x = -s \log y \rightarrow \infty$ , also folgt aus (11.8)

$$y^s(-\log y) = e^{-x} \frac{x}{s} \Big|_{x=-s \log y} \rightarrow 0 \quad \text{mit } y \searrow 0.$$

Setzen wir  $y = 1/z$  mit  $z \rightarrow \infty$ , so folgt auch der linke Grenzwert in (11.9).  $\square$

**Folgerung 11.2** *Die Funktion*

$$\eta(x) = \begin{cases} e^{-1/x} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

*ist unendlich oft differenzierbar.*

BEWEIS: Wir zeigen durch Induktion, dass es Polynome  $p_n$  gibt mit

$$\eta^{(n)}(x) = \begin{cases} p_n\left(\frac{1}{x}\right) e^{-1/x} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0. \end{cases}$$

Für  $n = 0$  ist das richtig mit  $p_0(s) \equiv 1$ . Ist die Aussage für ein  $n \in \mathbb{N}_0$  gezeigt, so folgt:

$$\eta^{(n+1)}(x) = \begin{cases} \left(\frac{1}{x^2} p_n\left(\frac{1}{x}\right) - \frac{1}{x^2} p_n'\left(\frac{1}{x}\right)\right) e^{-1/x} & \text{für } x > 0, \\ 0 & \text{für } x < 0. \end{cases}$$

Für  $x \neq 0$  gilt der Induktionsschluss also mit  $p_{n+1}(s) = s^2(p_n(s) - p_n'(s))$ . Zu zeigen bleibt  $\eta^{(n+1)}(0) = 0$ . Für den linksseitigen Differenzenquotienten ist das klar, für  $x > 0$  berechnen wir mit (11.8)

$$\frac{\eta^{(n)}(x) - \eta^{(n)}(0)}{x} = \frac{1}{x} p_n\left(\frac{1}{x}\right) e^{-1/x} \rightarrow 0 \quad \text{mit } x \searrow 0.$$

$\square$

Wir haben die Differentialgleichung  $f' = f$  als Grenzwert des diskreten Zinsmodells motiviert, wenn der Zeitschritt gegen Null geht. Die diskreten Lösungen sollten dann die kontinuierliche Lösung approximieren. Dies sollte zu einer alternativen Formel für die Exponentialfunktion führen. Unterteilen wir das Zeitintervall  $[0, t]$  in  $n$  Abschnitte der Länge  $\Delta t = t/n$ , so ergibt sich bei Zinsrate Eins und Anfangswert Eins mittels Induktion der Kontostand

$$(1 + \Delta t)^n = \left(1 + \frac{t}{n}\right)^n.$$

**Satz 11.6 (Eulerapproximation)** *Für alle  $x \in \mathbb{R}$  gilt*

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

BEWEIS: Wir zeigen  $\phi_n(x) := x - n \log(1 + \frac{x}{n}) \rightarrow 0$  mit  $n \rightarrow \infty$ . Da  $\exp$  stetig folgt

$$\left(1 + \frac{x}{n}\right)^n = \exp\left(n \log\left(1 + \frac{x}{n}\right)\right) = \exp(x - \phi_n(x)) \rightarrow \exp(x) \quad \text{mit } n \rightarrow \infty.$$

Es gilt  $\phi_n(0) = 0$  und  $\phi'_n(x) = 1 - (1 + \frac{x}{n})^{-1} = (1 + \frac{x}{n})^{-1} \frac{x}{n}$ . Sei  $n \geq 2|x|$ . Für  $|y| \leq |x|$  folgt  $1 + \frac{y}{n} \geq 1 - \frac{|x|}{n} \geq 1/2$  und damit  $|\phi'_n(y)| \leq 2\frac{|x|}{n}$ . Mit dem Schrankensatz ergibt sich

$$|\phi_n(x)| \leq \frac{2|x|}{n} |x| \rightarrow 0 \quad \text{mit } n \rightarrow \infty.$$

□

Wir können die Funktionalgleichung anschaulich aus dem Zins-Modell verstehen: bei konstantem Zinssatz sollte es egal sein, ob das Geld erst für einen Zeitraum  $t_1$  angelegt wird, und dann die ersparte Summe wieder für einen Zeitraum  $t_2$ , oder eben gleich für die Gesamtdauer  $t_1 + t_2$ . Bei Zinssatz  $x = 1$  bedeutet das genau

$$\exp(t_1 + t_2) = \exp(t_1) \exp(t_2).$$



## 12 Die trigonometrischen Funktionen

In diesem Abschnitt definieren wir die trigonometrischen Funktionen. Unser Ausgangspunkt ist die harmonische Schwingung, über den Energie-Erhaltungssatz ergibt sich der Bezug zur gleichförmigen Kreisbewegung  $c' = ic$ . Formal ist das die Differentialgleichung des natürlichen Wachstums, allerdings mit der Rate  $i = \sqrt{-1}$ . Mit der Exponentialreihe bekommen wir  $c(t) = e^{it}$  und die trigonometrischen Funktionen durch die Eulersche Formel  $e^{it} = \cos(t) + i \sin(t)$ .

Bei einer harmonischen Schwingung wird die Auslenkung eines schwingungsfähigen Systems aus seiner Ruhelage durch eine Funktion  $u : I \rightarrow \mathbb{R}$ ,  $u = u(t)$ , beschrieben. Ein Beispiel ist die Schwingung einer Masse  $m > 0$  an einer Feder. Nach dem Hookeschen Gesetz (1676) ist die elastische Rückstellkraft der Feder proportional zur Auslenkung, also  $F = -ku$  mit  $k > 0$ . Die Kraft bewirkt eine Beschleunigung der Masse, nach dem Newtonschen Grundgesetz (1687) ist  $F = mu''$ . Wir nehmen  $k/m = 1$  an, und geben eine Auslenkung  $x_0$  und Geschwindigkeit  $y_0$  zu einer Zeit  $t_0 \in I$  vor. Damit ergibt sich das Anfangswertproblem

$$u'' + u = 0 \text{ auf } I, \quad u(t_0) = x_0, \quad u'(t_0) = y_0. \quad (12.1)$$

Bei einer Schwingung findet ein Austausch zwischen kinetischer und potentieller Energie statt. Bis auf Konstanten ist die kinetische Energie das Quadrat der Geschwindigkeit. Die potentielle Energie ist das Quadrat der Auslenkung, das folgt aus dem Hookeschen Gesetz. Die Gesamtenergie bleibt konstant. Wir wollen das mathematisch begründen.

**Lemma 12.1** *Für Funktionen  $u, v : I \rightarrow \mathbb{R}$  sind folgende (1) und (2) äquivalent:*

(1)  $u$  ist Lösung von  $u'' + u = 0$ , und  $v = -u'$ .

(2)  $c = u + iv : I \rightarrow \mathbb{C}$  ist Lösung von  $c' = ic$ , wobei  $i = \sqrt{-1}$ .

Weiter: Für jede Lösung von (2) ist  $|c(t)|^2 = u(t)^2 + v(t)^2$  konstant.

BEWEIS: Die Gleichung (2), also  $c' = ic$ , lautet in Koordinaten  $c = u + iv$

$$u' + iv' = i(u + iv) \quad \Leftrightarrow \quad (u', v') = (-v, u). \quad (12.2)$$

Seien  $u, v$  Lösungen von (1). Nach der zweiten Gleichung gilt  $u' = -v$ , und dann aus der ersten  $v' = -u'' = u$ , also gilt (2). Sei andererseits (2) erfüllt, also  $u' = -v$  und  $v' = u$ . Dann folgt  $u'' = -u$ , also gilt (1). Schließlich berechnen wir mit (2)

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

□

Es gilt also  $|c(t)|^2 = u(t)^2 + v(t)^2 = R^2$  für einen festen Radius  $R \geq 0$ . Physikalisch ist das der Energieerhaltungssatz, denn  $u$  ist die Auslenkung und  $v$  ist (minus) die Geschwindigkeit. Wir bezeichnen  $c' = ic$  als Differentialgleichung der Kreisbewegung. Der Geschwindigkeitsvektor  $c' = (-v, u)$  steht senkrecht auf  $c = (u, v)$  und hat die Länge  $|c'| = R$ . Dadurch ist  $c'$  bis auf Richtung eindeutig festgelegt, diese ergibt sich aus

$$\det(c, c') = \det \begin{pmatrix} u & u' \\ v & v' \end{pmatrix} = uv' - u'v = u^2 + v^2 > 0 \quad (\text{außer wenn } R = 0).$$

Anschaulich liegt die Kreisscheibe bei dieser Fahrtrichtung auf der linken Seite. Man sagt: der Kreis wird im mathematisch positiven Sinn durchlaufen, also gegen den Uhrzeigersinn. Wir zeigen jetzt, dass die Kreisbewegung durch die Anfangsposition eindeutig festgelegt ist, und folgern die Eindeutigkeit für das Anfangswertproblem der harmonischen Schwingung.

**Satz 12.1 (Eindeutigkeit der Kreisbewegung)** Sei  $t_0 \in I = (a, b)$ .

(1) Zu  $z_0 \in \mathbb{R}^2$  gibt es höchstens ein  $c : I \rightarrow \mathbb{C}$  mit

$$c' = ic \quad \text{auf } I, \quad c(t_0) = z_0. \quad (12.3)$$

(2) Zu  $x_0, y_0 \in \mathbb{R}$  gibt es höchstens ein  $u : I \rightarrow \mathbb{R}$  mit

$$u'' + u = 0 \quad \text{auf } I, \quad u(t_0) = x_0, u'(t_0) = y_0, \quad (12.4)$$

BEWEIS: Seien  $c_{1,2} : I \rightarrow \mathbb{C}$  zwei Lösungen von (12.3). Dann gilt für  $c = c_1 - c_2$

$$c' = (c_1 - c_2)' = (c_1)' - (c_2)' = i(c_1 - c_2) = ic, \quad c(t_0) = c_1(t_0) - c_2(t_0) = 0.$$

Aus Lemma 12.1 folgt  $|c(t)|^2 = |c(t_0)|^2 = 0$  für alle  $t$ , also  $c_1 = c_2$ . Für (2) seien  $u_k : I \rightarrow \mathbb{R}$ ,  $k = 1, 2$ , Lösungen von (12.4). Nach Lemma 12.1 sind dann  $c_k = u_k - iu_k'$  Lösungen von  $c_k' = ic_k$ . Da  $c_k(t_0) = u_k(t_0) - iu_k'(t_0) = x_0 - iy_0$ , folgt  $c_1 = c_2$  aus (1), also  $u_1 = u_2$ .  $\square$

Die Gleichung  $c' = ic$  ist formal identisch zur reellen Gleichung  $f' = \lambda f$  aus dem vorigen Kapitel, nur dass  $c$  Werte in  $\mathbb{C}$  hat und auf der rechten Seite statt  $\lambda$  der komplexe Faktor  $i = \sqrt{-1}$  steht. Es ist naheliegend, die Definition der Exponentialfunktion auf  $\mathbb{C}$  zu verallgemeinern und damit eine Lösung zu konstruieren.

**Definition 12.1 (Exponentialfunktion)** Die komplexe Exponentialfunktion ist

$$\exp : \mathbb{C} \rightarrow \mathbb{C}, \quad \exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}. \quad (12.5)$$

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

Die endlichen Summen  $f_n(z) = \sum_{k=0}^n z^k/k!$ ,  $n \in \mathbb{N}_0$ , bilden eine Cauchyfolge in  $\mathbb{C}$ . Das folgt durch Abschätzung mit der geometrischen Reihe exakt wie im reellen Fall, siehe Satz 11.2. Also ist  $\exp(z) = \lim_{n \rightarrow \infty} f_n(z)$  definiert.

**Satz 12.2 (Existenz der Kreisbewegung)** Die Funktion  $c : \mathbb{R} \rightarrow \mathbb{C}$ ,  $c(t) = \exp(it)$ , ist Lösung des Anfangswertproblems

$$c' = ic \quad \text{auf } \mathbb{R}, \quad c(0) = 1. \quad (12.6)$$

BEWEIS: Nach Definition ist  $c(0) = \exp(0) = 1$ . Betrachte die Approximationen

$$c_n : \mathbb{R} \rightarrow \mathbb{C}, \quad c_n(t) = \sum_{k=0}^n \frac{(it)^k}{k!} = \sum_{k=0}^n i^k \frac{t^k}{k!}.$$

Es gilt  $c_n(t) = f_n(it) \rightarrow \exp(it) = c(t)$  mit  $n \rightarrow \infty$ . Wir berechnen

$$c'_n(t) = \sum_{k=1}^n i^k \frac{t^{k-1}}{(k-1)!} = i \sum_{k=0}^{n-1} \frac{(it)^k}{k!} = ic_{n-1}(t).$$

Jetzt argumentieren wir wie in Satz 11.2. Für  $t \in [-R, R]$  gilt

$$|c_n(t)| \leq \sum_{k=0}^n \frac{|t|^k}{k!} \leq \sum_{k=0}^{\infty} \frac{R^k}{k!} =: C < \infty.$$

Es folgt  $|c'_n(t)| = |c_{n-1}(t)| \leq C$ , also nach dem (vektorwertigen) Schrankensatz

$$|c_n(t_1) - c_n(t_2)| \leq C |t_1 - t_2| \quad \text{für } t_{1,2} \in [-R, R].$$

Seien nun  $t_0, t \in (-R, R)$ . Für  $\phi(t) := c_n(t) - c_n(t_0) - c'_n(t_0)(t - t_0)$  gilt  $\phi(t_0) = 0$ , und

$$|\phi'(t)| = |c'_n(t) - c'_n(t_0)| = |c_{n-1}(t) - c_{n-1}(t_0)| \leq C |t - t_0|.$$

Der Schrankensatz auf  $[t_0, t]$  bzw.  $[t, t_0]$  liefert  $|\phi(t)| = |\phi(t) - \phi(t_0)| \leq C |t - t_0|^2$ . Es folgt

$$\left| \frac{c_n(t) - c_n(t_0)}{t - t_0} - c'_n(t_0) \right| = \frac{|c_n(t) - c_n(t_0) - c'_n(t_0)(t - t_0)|}{|t - t_0|} \leq C |t - t_0|.$$

Da  $c'_n(t_0) = ic_{n-1}(t_0) \rightarrow ic(t_0)$  mit  $n \rightarrow \infty$ , folgt

$$\left| \frac{c(t) - c(t_0)}{t - t_0} - ic(t_0) \right| \leq C |t - t_0|.$$

Also ist  $c'(t_0) = ic(t_0)$ , der Satz ist bewiesen. □

**Satz 12.3 (Eigenschaften der Kreisbewegung)** Die Kreisbewegung  $c(t) = e^{it}$  ist in  $C^\infty(\mathbb{R}, \mathbb{C})$  und hat folgende Eigenschaften:

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

BEWEIS: Sei  $c(t) = e^{it}$  für  $t \in \mathbb{R}$ . Nach Satz 12.2 gilt  $c' = ic$  sowie  $c(0) = 1$ . Daraus folgt mit Satz 9.1 die Stetigkeit von  $c$ . Durch Induktion sieht man  $c \in C^k(\mathbb{R}, \mathbb{C})$  und  $c^{(k)} = i^k c$  für alle  $k \in \mathbb{N}_0$ , analog zum reellen Fall. Aussage (a) folgt aus Lemma 12.1, es ist

$$|c(t)| = |c(0)| = 1 \quad \text{für alle } t \in \mathbb{R}.$$

Für die Funktionalgleichung (b) berechne

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

Aus der Eindeutigkeit, Satz 12.1, folgt  $\overline{c(s+t)} = \overline{c(s)c(t)}$  wie behauptet. Insbesondere ist  $c(t)c(-t) = c(0) = 1$ . Andererseits ist auch  $\overline{c(t)c(t)} = |c(t)|^2 = 1$  nach (a), also  $\overline{c(t)} = c(-t)$ .  $\square$

Eine weitere wichtige Eigenschaft der Funktion  $c(t) = e^{it}$  ist die Periodizität. Eine Zahl  $p \in \mathbb{R}$  heißt Periode einer Funktion  $c : \mathbb{R} \rightarrow \mathbb{C}$ , wenn gilt:

$$c(t+p) = c(t) \quad \text{für alle } t \in \mathbb{R}.$$

Jede Funktion  $c$  hat die triviale Periode  $p = 0$ . Generell bilden die Perioden von  $c$  eine Untergruppe von  $(\mathbb{R}, +)$ , denn für zwei Perioden  $p_1, p_2 \in \mathbb{R}$  gilt

$$c(t+p_1-p_2) = c(t+p_1-p_2+p_2) = c(t+p_1) = c(t).$$

Eine Periode  $p > 0$  heißt kleinste Periode, wenn im Intervall  $(0, p)$  keine Perioden von  $c$  liegen. Die Menge aller Perioden ist dann das Gitter  $\mathbb{Z} \cdot p$ . Denn zu  $q \notin \mathbb{Z} \cdot p$  gibt es ein  $k \in \mathbb{Z}$  mit  $kp < q < (k+1)p$ . Wäre  $q$  eine Periode, so auch  $q - kp \in (0, p)$ , ein Widerspruch.

**Lemma 12.2** Die Funktion  $c(t) = e^{it}$  hat eine Periode  $p > 0$ .

BEWEIS: Es reicht zu zeigen dass ein  $\tau > 0$  existiert mit  $c(\tau) = i$ . Denn dann gilt

$$\begin{aligned} \frac{d}{dt} c(t+\tau) &= c'(t+\tau) = ic(t+\tau) \quad \text{und} \quad c(t+\tau)|_{t=0} = i, \\ \frac{d}{dt} ic(t) &= ic'(t) = i(ic(t)) \quad \text{und} \quad ic(t)|_{t=0} = i. \end{aligned}$$

Mit der Eindeutigkeit, Satz 12.1, folgt  $c(t+\tau) = ic(t)$  für alle  $t \in \mathbb{R}$ , und dann

$$c(t+4\tau) = i^4 c(t) = c(t) \quad \text{für alle } t \in \mathbb{R}.$$

Also ist  $p = 4\tau > 0$  eine Periode von  $c$ . Schreibe nun  $c(t) = u(t) + iv(t)$ , und setze

$$\tau = \sup\{t > 0 : u, v > 0 \text{ auf } (0, t)\}.$$

Es gilt  $u(0) = 1, v(0) = 0$  und  $\lim_{t \searrow 0} v(t)/t = v'(0) = 1$ , also gibt es ein  $\delta > 0$  mit  $u, v > 0$  auf  $(0, \delta]$ , insbesondere ist  $\tau \geq \delta$ . Wir behaupten  $\tau < \infty$ . Auf  $(0, \tau)$  gilt  $u, v > 0$ , also  $u' = -v < 0$  und  $v' = u > 0$ , das heißt  $u$  ist monoton fallend und  $v$  monoton wachsend. Mit  $\varepsilon = v(\delta) > 0$  ist für  $t \in [\delta, \tau)$  dann  $u'(t) = -v(t) \leq -\varepsilon$ , also nach Schrankensatz

$$0 < u(t) \leq u(\delta) - \varepsilon(t - \delta), \quad \text{und damit } t \leq \delta + \frac{u(\delta)}{\varepsilon} < \infty.$$

Diese Abschätzung folgt auch für  $\tau$ , also ist  $\tau < \infty$ . Mit  $t \nearrow \tau$  ergibt sich nun  $u(\tau) \geq 0, v(\tau) > 0$  und  $u(\tau)^2 + v(\tau)^2 = 1$ . Wäre  $u(\tau) > 0$ , so gilt  $u, v > 0$  auf einem größeren Intervall, Widerspruch. Also ist  $u(\tau) = 0$  und  $v(\tau) = 1$  bzw.  $c(\tau) = i$ .  $\square$

Wir wollen nun den Bezug zur Geometrie herstellen. Dazu brauchen wir den Begriff der Länge einer Kurve. Für eine Kurve  $c : [a, b] \rightarrow \mathbb{R}^n$  betrachten wir die äquidistante Unterteilung  $t_k = a + k \frac{b-a}{n}, k = 0, \dots, n$ , und setzen

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

Das ist die Länge des Streckenzugs mit den Punkten  $c(t_k)$ .

**Definition 12.2 (Bogenlänge)** Die Länge von  $c \in C^1([a, b], \mathbb{R}^n)$  ist  $L(c) = \lim_{n \rightarrow \infty} L_n(c)$ .

Man kann zeigen, dass der Grenzwert für jede  $C^1$ -Kurve  $c : [a, b] \rightarrow \mathbb{R}$  existiert, und dass er einen sinnvollen Begriff der Länge liefert. Wir beschränken uns hier auf den Fall  $c(t) = e^{it}$ .

**Lemma 12.3** Die Kurve  $c : \mathbb{R} \rightarrow \mathbb{C}$ ,  $c(t) = e^{it}$ , bildet längentreu ab: für  $I = [a, b]$  gilt

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

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

$$L_n(c|_I) = \sum_{k=1}^n |e^{i(a+k\tau_n)} - e^{i(a+(k-1)\tau_n)}| = n |e^{i\tau_n} - 1| = \left| \frac{e^{i\tau_n} - 1}{\tau_n} \right| (b-a).$$

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

$$L(c|_I) = \lim_{n \rightarrow \infty} L_n(c|_I) = |I|. \quad (12.7)$$

□

Auf dem Intervall  $[0, \tau]$  durchläuft  $c(t) = e^{it}$  den Viertelkreis im ersten Quadranten streng monoton von  $c(0) = 1$  bis  $c(\tau) = i$ . Das motiviert die Definition

$$\frac{\pi}{2} := L(c|_{[0, \tau]}) = 1,57\dots \quad (12.8)$$

Archimedes hat obere und untere Abschätzungen für die Länge des Einheitskreises berechnet mit einer Abweichung von nur 0,002. Dazu hat er Approximationen von außen und innen durch regelmäßige  $n$ -Ecke benutzt mit  $n = 6, 12, 24, 48, 96$ , wir wollen uns damit hier nicht befassen. Nach Lemma 12.3 haben wir jedenfalls  $\tau = \pi/2$  bzw.  $4\tau = 2\pi$ .

**Satz 12.4 (Periodizität von  $e^{it}$ )** Die Kreisbewegung  $c(t) = e^{it}$  hat die kleinste Periode  $2\pi$ . Auf  $[0, 2\pi]$  durchläuft  $c(t)$  den Einheitskreis  $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$  genau einmal entgegen dem Uhrzeigersinn.

BEWEIS: Nach Lemma 12.2 ist  $4\tau$  eine Periode von  $c$ . Wegen  $c(t + \tau) = ic(t)$  gehen die Zeiten  $k\tau$  für  $k = 0, 1, 2, 3, 4$  auf die Punkte  $1, i, -1, -i, 1$ . Die Intervalle dazwischen werden der Reihe nach bijektiv auf die Viertelkreise in den vier offenen Quadranten abgebildet. Somit ist  $4\tau$  die kleinste Periode, und  $c : [0, 4\tau] \rightarrow \mathbb{S}^1$  ist bijektiv bis auf  $c(0) = c(4\tau)$ . □

**Definition 12.3 (Cosinus und Sinus)** Die Funktionen  $\cos, \sin : \mathbb{R} \rightarrow \mathbb{R}$  sind definiert als

$$\cos t = \operatorname{Re}(e^{it}) = \frac{e^{it} + e^{-it}}{2}, \quad \sin t = \operatorname{Im}(e^{it}) = \frac{e^{it} - e^{-it}}{2i}. \quad (12.9)$$

Nach Definition gilt also die *Eulersche Formel*

$$e^{it} = \cos t + i \sin t \quad \text{für alle } t \in \mathbb{R}. \quad (12.10)$$

**Satz 12.5 (Eigenschaften von  $\cos$  und  $\sin$ )** Es gelten folgende Eigenschaften:

(1)  $\cos$  und  $\sin$  sind in  $C^\infty(\mathbb{R})$ , und es gelten die Reihendarstellungen

$$\cos t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \quad \text{und} \quad \sin t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}. \quad (12.11)$$

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

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

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

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

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

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

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

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

BEWEIS: Nach Satz 12.3 ist  $c(t) = e^{it}$  in  $C^\infty(\mathbb{R}, \mathbb{C})$ , also gilt das auch für die Komponenten  $\cos t$  und  $\sin t$ . Aus der Definition der Exponentialfunktion folgt

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

Es gilt  $c' = ic$  nach Satz 12.2, die Gleichungen (2) folgen daraus durch Komponentenvergleich, vgl. (12.2). In (3) ergeben sich die Anfangswerte aus  $c(0) = 1$  und  $c'(0) = i$ . Die Differentialgleichungen folgen aus (2) durch Differentiation, vgl. Lemma 12.1. Die Identität (4) ist gleichbedeutend mit  $|e^{it}| = 1$ , das wurde in Aussage (a) von Satz 12.3 bewiesen. Mit Aussage (c) dieses Satz bekommen wir

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

Also gilt Behauptung (5). Berechne für (6) mit der Funktionalgleichung, siehe Satz 12.3(b),

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

Nach Definition ist klar, dass jede Periode von  $e^{it}$  auch Periode von  $\cos$  und  $\sin$  ist. Es gilt aber auch die Umkehrung: ist zum Beispiel  $\cos(t+p) = \cos t$  für alle  $t \in \mathbb{R}$ , so folgt durch Ableiten  $\sin(t+p) = \sin t$  für alle  $t \in \mathbb{R}$ , und damit  $\exp i(t+p) = \exp it$ . Nach Satz 12.4 ist  $2\pi$  die kleinste Periode von  $c(t) = e^{it}$ , also gilt das auch für  $\cos t$  und  $\sin t$ .  $\square$

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

	$t = 0$	$\pi/2$	$\pi$	$3\pi/2$	$2\pi$
$e^{it}$	1	$i$	$-1$	$-i$	1
$\cos$	1	$\searrow$ 0	$\searrow$ $-1$	$\nearrow$ 0	$\nearrow$ 1
$\sin$	0	$\nearrow$ 1	$\searrow$ 0	$\searrow$ $-1$	$\nearrow$ 0

Die Nullstellen von Cosinus und Sinus sind wie folgt.

$$\begin{aligned} e^{it} = 1 &\Leftrightarrow t = 2k\pi \quad \text{mit } k \in \mathbb{Z}, \\ \cos t = 0 &\Leftrightarrow t = \pi/2 + k\pi \quad \text{mit } k \in \mathbb{Z}, \\ \sin t = 0 &\Leftrightarrow t = k\pi \quad \text{mit } k \in \mathbb{Z}. \end{aligned} \tag{12.12}$$

Die Funktionen  $\cos$  und  $\sin$  sind periodisch und damit nicht injektiv. Wir können sie aber auf ein Intervall einschränken, auf dem sie streng monoton sind und damit eine Umkehrfunktion *arcus cosinus* bzw. *arcus sinus*<sup>18</sup> haben. Man spricht genauer von einem *Zweig* dieser Umkehrfunktionen, denn Einschränkung auf ein anderes Monotonieintervall führt zu einer anderen Umkehrfunktion. Unsere Wahl hier wird oft als *Hauptzweig* bezeichnet.

**Satz 12.6 (Arcusfunktionen)**  $\cos : [0, \pi] \rightarrow [-1, 1]$  und  $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow [-1, 1]$  haben stetige Umkehrfunktionen, die wie folgt bezeichnet werden:

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

$$\arccos'(x) = -\frac{1}{\sqrt{1-x^2}} \quad \text{für } x \in (-1, 1). \tag{12.13}$$

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

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}} \quad \text{für } x \in (-1, 1). \tag{12.14}$$

BEWEIS: Es gilt  $\cos' = -\sin < 0$  auf  $(0, \pi)$ , und  $\sin' = \cos > 0$  auf  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ . Also ist  $\cos$  auf  $[0, \pi]$  streng monoton fallend, und  $\sin$  auf  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  streng monoton wachsend. Die Existenz und Stetigkeit der Umkehrfunktionen gilt nach Satz 8.2. Die Arcusfunktionen sind differenzierbar auf  $(-1, 1)$  nach Satz 9.4. Wir berechnen, wobei wir  $\sin(\arccos x) > 0$  haben wegen  $\arccos x \in (0, \pi)$ :

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

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

**Definition 12.4 (Winkel)** Der Winkel zwischen Vektoren  $v, w \in \mathbb{R}^n$ ,  $v, w \neq 0$ , ist

$$\angle(v, w) = \arccos \left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\rangle \in [0, \pi].$$

Das Skalarprodukt liegt in  $[-1, 1]$  nach der Ungleichung von Cauchy-Schwarz, Satz 6.5, also ist der Winkel definiert. In den Grenzfällen  $v, w$  parallel bzw. antiparallel ist der Winkel  $\arccos(1) = 0$  bzw.  $\arccos(-1) = \pi$ . Wir wollen die Definition geometrisch begründen, wobei

---

<sup>18</sup>Arcus = Bogen (Latein)

wir  $\|v\| = \|w\| = 1$  und  $v, w$  linear unabhängig annehmen. Sei  $E$  die von  $v, w$  aufgespannte Ebene und  $e_1, e_2$  eine Orthonormalbasis mit  $e_1 = v$ . Mit  $w = xe_1 + ye_2$  gilt dann

$$\langle v, w \rangle = \langle e_1, xe_1 + ye_2 \rangle = x.$$

Wir können  $E$  mit  $\mathbb{R}^2$  isometrisch identifizieren so dass  $e_1 = (1, 0)$  und  $e_2 = (0, 1)$ , dann sind  $v, w$  die Vektoren  $(1, 0)$  und  $(x, y)$ . Wähle  $t \in [0, 2\pi)$  mit  $(x, y) = (\cos t, \sin t)$ , es folgt

$$\arccos\langle v, w \rangle = \arccos(\cos t) = \begin{cases} t & \text{falls } t \in [0, \pi] \text{ bzw. } y \geq 0, \\ 2\pi - t & \text{falls } t \in [\pi, 2\pi) \text{ bzw. } y < 0. \end{cases}$$

Die Vektoren  $v, w$  zerlegen den Einheitskreis der Ebene  $E$  in zwei Bögen, der Winkel  $\arccos\langle v, w \rangle$  ist die Länge des kürzeren Bogens.

In der Schule werden die trigonometrischen Funktionen elementargeometrisch erklärt. Um für ein gegebenes  $\alpha \in \mathbb{R}$  die Werte  $\cos \alpha$  und  $\sin \alpha$  zu definieren, wählt man ein rechtwinkliges Dreieck mit Winkel  $\alpha$  und setzt

$$\cos \alpha = \frac{\text{Ankathete}}{\text{Hypotenuse}} \quad \text{und} \quad \sin \alpha = \frac{\text{Gegenkathete}}{\text{Hypotenuse}}.$$

Gewisse Funktionswerte lassen sich mit Pythagoras ermitteln, durch Einzeichnen der Winkelhalbierenden im gleichseitigen Dreieck und im Quadrat ergibt sich

$\alpha$	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$
$\cos \alpha$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
$\sin \alpha$	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1

Allgemein beruft sich die Winkelmessung in der Schule aber auf die Anschauung, der Grenzwert der Bogenlänge am Kreis wird nicht behandelt. Für allgemeine  $\alpha \in \mathbb{R}$  ist daher die Konstruktion eines rechtwinkligen Dreiecks mit Winkel  $\alpha$  nicht exakt möglich, höchstens praktisch mit dem Geodreieck. Ein Bewusstsein für diese Problematik kann nützlich sein.

**Folgerung 12.1 (Polarkoordinaten)** *Zu jedem  $z = x + iy \in \mathbb{C} \setminus \{0\}$  gibt es eindeutig bestimmte  $r > 0$ ,  $\vartheta \in [0, 2\pi)$  mit  $z = re^{i\vartheta}$ .*

BEWEIS: Sei zunächst  $|z| = 1$ , also  $x \in [-1, 1]$ . Wir definieren

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

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

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

Für  $z \in \mathbb{C} \setminus \{0\}$  beliebig setzen wir  $r = |z| > 0$  und definieren  $\vartheta \in [0, 2\pi]$  mit  $z/r = e^{i\vartheta}$ . Die Eindeutigkeit von  $r$  ist klar, denn  $r = |z|$ , für den Polarwinkel  $\vartheta$  folgt sie aus Satz 12.4.  $\square$

Die Punkte  $(1, 0)$  und  $z = x + iy$  zerlegen den Einheitskreis wieder in zwei Bögen. Der Winkel  $\vartheta \in [0, 2\pi)$  ist die Länge des positiv orientierten Bogens von  $(1, 0)$  nach  $z$ . Für Vektoren  $v, w$  im  $\mathbb{R}^n$  kann nur der Winkel in  $[0, \pi]$  definiert werden, weil die von  $v, w$  aufgespannte Ebene a priori nicht orientiert ist.

Zum Schluss des Kapitels diskutieren wir die wesentlichen Eigenschaften der Exponentialfunktion für beliebige komplexe Argumente  $z \in \mathbb{C}$ .

**Satz 12.7 (komplexe Exponentialfunktion)** *Die Funktion*

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

hat folgende Eigenschaften:

(1) Für  $\lambda \in \mathbb{C}$  ist  $z(t) = e^{\lambda t}$  die eindeutige Lösung des Anfangswertproblems

$$z'(t) = \lambda z(t) \text{ für } t \in \mathbb{R}, \quad z(0) = 1.$$

(2)  $e^{z_1+z_2} = e^{z_1}e^{z_2}$  für alle  $z_1, z_2 \in \mathbb{C}$  (Funktionalgleichung).

(3) Es gilt  $\exp(\mathbb{C}) = \mathbb{C} \setminus \{0\}$ , und  $e^{z_1} = e^{z_2}$  genau wenn  $z_1 - z_2 \in 2\pi i\mathbb{Z}$ .

BEWEIS:

zu (1): Dass  $e^{\lambda t}$  eine Lösung ist, wurde im Fall  $\lambda = i$  in Satz 12.2 gezeigt. Das dortige Argument lässt sich wörtlich übertragen, man muss nur die Zahl  $i$  durch  $\lambda$  ersetzen. Ist  $z(t)$  irgendeine Lösung des Anfangswertproblems, so folgt

$$\frac{d}{dt}(e^{-\lambda t}z(t)) = -\lambda e^{-\lambda t}z(t) + e^{-\lambda t}\lambda z(t) = 0,$$

also  $e^{-\lambda t}z(t) = e^{-\lambda t}z(t)|_{t=0} = 1$ . Wir können in der Rechnung auch  $e^{\lambda t}$  statt  $z(t)$  nehmen, damit folgt  $z(t) = 1/e^{-\lambda t} = e^{\lambda t}$ .

zu (2): Wir berechnen mit (1) und der Produktregel

$$\frac{d}{dt}(e^{tz_1}e^{tz_2}) = z_1 e^{tz_1}e^{tz_2} + e^{tz_1}z_2 e^{tz_2} = (z_1 + z_2)e^{tz_1}e^{tz_2} \quad \text{und} \quad e^{tz_1}e^{tz_2}|_{t=0} = 1.$$

Aus (1) folgt  $e^{tz_1}e^{tz_2} = e^{t(z_1+z_2)}$ , und mit  $t = 1$  die Behauptung.

zu (3): Für  $z = x + iy$  gilt  $e^z = e^x e^{iy}$ , also  $|e^z| = e^x > 0$ . Sei  $w \in \mathbb{C} \setminus \{0\}$  gegeben. Nach Folgerung 12.1 gibt es ein  $y \in [0, 2\pi)$  mit  $e^{iy} = w/|w|$ . Setzen wir  $x = \log |w|$ , so folgt

$$w = |w| \frac{w}{|w|} = e^x e^{iy} = e^z \quad \text{mit } z = x + iy.$$

Weiter gilt, da  $2\pi$  kleinste Periode von  $e^{iy}$  ist,

$$e^z = e^x e^{iy} = 1 \quad \Leftrightarrow \quad e^x = e^{iy} = 1 \quad \Leftrightarrow \quad x = 0, y \in 2\pi\mathbb{Z}.$$

Daraus folgt mit der Funktionalgleichung

$$e^{z_1} = e^{z_2} \quad \Leftrightarrow \quad e^{z_1 - z_2} = 1 \quad \Leftrightarrow \quad z_1 - z_2 \in 2\pi i\mathbb{Z}.$$

Die Exponentialabbildung hat also die Periode  $2\pi i$ , der Streifen  $0 \leq \text{Im } z < 2\pi$  wird bijektiv auf  $\mathbb{C} \setminus \{0\}$  abgebildet. Damit ist (3) gezeigt.  $\square$



## 13 Das Riemannsche Integral

In seiner Habilitationsschrift (1854) hat B. Riemann eine Definition des Integrals skizziert. Ich vermute dass er das nicht als wesentliche Erkenntnis angesehen hat, es war für ihn wohl eher ein Hilfsmittel bei der Diskussion von Fourier-Reihen. Später hat Darboux eine Variante des Riemanschen Integrals mittels Ober- und Untersummen entwickelt, die gewisse Vorteile hat. Letztlich werden beide Konzepte überholt durch die Definition des Integrals von Henri Lebesgue, das in Analysis 3 behandelt wird.

In diesem Kapitel wird das Integral für Funktionen  $f : I = [a, b] \rightarrow \mathbb{R}$  einer Variablen eingeführt. Anschaulich ist das Integral, für  $f \geq 0$ , der Flächeninhalt der Menge

$$\{(x, y) \in \mathbb{R}^2 : x \in I, 0 < y < f(x)\}.$$

Allerdings haben wir den Begriff des Flächeninhalts noch gar nicht behandelt, so dass uns diese Vorstellung nicht als Definition dienen kann.

Für Treppenfunktionen, also stückweise konstante Funktionen, kann das Integral elementar als Rechtecksumme definiert werden. Unsere Strategie ist, dies auf allgemeinere Funktionen  $f$  durch Approximation mit Treppenfunktionen  $f_k$  zu erweitern. Wir brauchen dazu ein geeignetes Konzept für die Konvergenz  $f_k \rightarrow f$  von Funktionen. Um die Definition des Integrals übersichtlich zu halten, erledigen wir das vorab. Der folgende Begriff ist zunächst naheliegend:

**Definition 13.1 (punktweise Konvergenz)** Eine Folge von Funktionen  $f_k : D \rightarrow \mathbb{R}$ , wobei  $D$  beliebige Menge, konvergiert punktweise gegen die Funktion  $f : D \rightarrow \mathbb{R}$ , falls gilt:

$$f_k(x) \rightarrow f(x) \quad \text{für alle } x \in D. \quad (13.1)$$

**Beispiel 13.1** Für  $c_k > 0$  betrachte  $f_k : I = [0, 1] \rightarrow \mathbb{R}$  mit

$$f_k(x) = \begin{cases} c_k & \text{für } 0 < x \leq \frac{1}{k}, \\ 0 & \text{sonst.} \end{cases}$$

Die Folge konvergiert punktweise gegen  $f = 0$ , unabhängig von der Wahl der  $c_k$ . Denn es ist  $f_k(0) = 0$  für alle  $k$ , und für  $x > 0$  ist  $f_k(x) = 0$  für  $k > \frac{1}{x}$ . Der Inhalt des Rechtecks  $[0, 1/k] \times (0, c_k)$  ist  $c_k/k$ , diese Folge kann gegen jede Zahl in  $[0, \infty]$  gehen je nach Wahl der  $c_k$ . Dagegen ist das Integral der Grenzfunktion  $f = 0$  natürlich Null.

Die punktweise Konvergenz reicht also nicht aus, um die Konvergenz der Integrale zu garantieren; wir brauchen mehr Kontrolle.

**Definition 13.2 (Supremumsnorm)** Sei  $D$  eine Menge. Für  $f : D \rightarrow \mathbb{R}$  definieren wir

$$\|f\|_D = \sup\{|f(x)| : x \in D\} \in [0, \infty].$$

**Definition 13.3 (Gleichmäßige Konvergenz)** Eine Folge von Funktionen  $f_k : D \rightarrow \mathbb{R}$  konvergiert gleichmäßig gegen  $f : D \rightarrow \mathbb{R}$ , falls gilt:

$$\|f_k - f\|_D \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

**Beispiel 13.2** Für die  $f_k$  in Beispiel 13.1 und  $f = 0$  ist  $\|f_k - f\|_I = c_k$ , also gilt  $f_k \rightarrow f$  gleichmäßig genau wenn  $c_k \rightarrow 0$ . Der Rechteckinhalt  $c_k/k$  geht dann erst recht gegen Null.

Das sieht, jedenfalls was das Beispiel angeht, besser aus. Der Unterschied zwischen den beiden Konvergenzbegriffen wird auch in folgender Formulierung deutlich:

$$\begin{aligned} \text{punktweise: } & \forall \varepsilon > 0 \quad \forall x \in D \quad \exists K : \quad |f_k(x) - f(x)| < \varepsilon \text{ für alle } k > K, \\ \text{gleichmäßig: } & \forall \varepsilon > 0 \quad \exists K \quad \forall x \in D : \quad |f_k(x) - f(x)| < \varepsilon \text{ für alle } k > K. \end{aligned}$$

Punktweise Konvergenz erlaubt dass  $K$  von  $x$  abhängt, also  $K = K(\varepsilon, x)$ , dagegen verlangt gleichmäßige Konvergenz eine gemeinsame Schranke  $K = K(\varepsilon)$  für alle  $x \in D$ . Kontraposition der Eigenschaft von  $K = K(\varepsilon, x)$  ergibt die Implikation

$$|f_k(x) - f(x)| \geq \varepsilon \quad \Rightarrow \quad K(\varepsilon, x) \geq k.$$

**Beispiel 13.3** Wähle  $c_k = 1$  in Beispiel 13.1. Für  $0 < \varepsilon \leq 1$  gilt mit  $x_n = \frac{1}{n}$

$$|f_n(x_n) - f(x_n)| = |f_n(x_n)| = 1 \geq \varepsilon, \quad \text{also} \quad K(\varepsilon, x_n) \geq n.$$

Es gibt somit keine gemeinsame Schranke  $K = K(\varepsilon)$  für alle  $x \in [0, 1]$ .

Wegen  $|f_k(x) - f(x)| \leq \|f_k - f\|_D$  ist klar: aus gleichmäßig folgt punktweise. Um eine Folge  $f_k$  auf gleichmäßige Konvergenz zu testen, kann man also in zwei Schritten vorgehen:

- Konvergiert die Folge punktweise? Wenn nicht, so erst recht nicht gleichmäßig. Wenn ja, so ist die punktweise Grenzfunktion die einzig mögliche Kandidatin für den gleichmäßigen Grenzwert.
- Nun bestimme  $\|f_k - f\|_D$  bzw. schätze diese Norm ab. Gilt  $\|f_k - f\|_D \rightarrow 0$  mit  $k \rightarrow \infty$ , so ist  $f_k$  gleichmäßig konvergent gegen  $f$ . Wenn nicht, so ist  $f_k$  zwar punktweise, aber nicht gleichmäßig konvergent.

Wir bezeichnen mit  $\mathcal{B}(D)$  den Raum der beschränkten Funktionen auf  $D$ , also

$$\mathcal{B}(D) = \{f : D \rightarrow \mathbb{R} : \|f\|_D < \infty\}. \quad (13.2)$$

Dies ist ein Untervektorraum des Raums aller Funktionen auf  $D$ . Die Supremumsnorm ist eine Norm auf  $\mathcal{B}(D)$ , das heißt sie hat folgende Eigenschaften.

$$\text{Positivität:} \quad \|f\|_D \geq 0 \text{ mit Gleichheit genau wenn } f = 0,$$

$$\text{Halblinearität:} \quad \|\lambda f\|_D = |\lambda| \|f\|_D \text{ für } \lambda \in \mathbb{R}.$$

$$\text{Dreiecksungleichung:} \quad \|f + g\|_D \leq \|f\|_D + \|g\|_D.$$

Wir kommen nun zum eigentlichen Thema, der der Konstruktion des Integrals.

**Definition 13.4 (Zerlegung)** Eine Zerlegung  $Z$  des Intervalls  $I = [a, b]$  ist eine geordnete Menge von Punkten  $a = x_0 \leq x_1 \leq \dots \leq x_N = b$ . Wir setzen  $I_k = [x_{k-1}, x_k]$  sowie  $\Delta x_k = x_k - x_{k-1}$  für  $k = 1, \dots, N$ , und definieren die Feinheit von  $Z$  durch

$$\Delta(Z) = \max_{1 \leq k \leq N} \Delta x_k. \quad (13.3)$$

**Definition 13.5 (Riemannsche Summe)** Sei  $f : I = [a, b] \rightarrow \mathbb{R}$ . Die Riemannsche Summe von  $f$  zur Zerlegung  $Z$  mit Stützstellen  $\xi_k \in I_k$  ist

$$S_{Z,\xi}(f) = \sum_{k=1}^N f(\xi_k) \Delta x_k \in \mathbb{R}.$$

Die Riemannsche Summe ist ein Näherungswert für das noch zu definierende Integral. Eine konkrete Wahl der Zerlegung und der Stützstellen würde ein numerisches Verfahren zur Approximation des Integrals ergeben, zum Beispiel könnte man die äquidistante Zerlegung wählen und die Mittelpunkte der Teilintervalle als Stützstellen. Aber um das Integral stabil zu definieren, brauchen wir beliebige Approximationen, siehe Beispiel 13.5 unten. Auf jeden Fall sollte bei Verfeinerung einer Zerlegung der Approximationsfehler kleiner werden. Dies führt auf Riemanns Begriff der Integrierbarkeit<sup>19</sup>.

**Definition 13.6 (Riemann-Integral)** Eine beschränkte Funktion  $f : I = [a, b] \rightarrow \mathbb{R}$  heißt Riemann-integrierbar mit Integral  $S \in \mathbb{R}$  (Notation:  $S = \int_I f$ ) falls es zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  gibt so dass

$$|S_{Z,\xi}(f) - S| < \varepsilon \quad \text{für jede Wahl von } Z, \xi \text{ mit } \Delta(Z) < \delta.$$

Wir bezeichnen mit  $\mathcal{R}(I)$  die Klasse der Riemann-integrierbaren Funktionen auf  $I$ .

Die Zahl  $S$  in Definition 13.6 ist eindeutig bestimmt, und zwar gilt

$$S_{Z_n, \xi_n}(f) \rightarrow S = \int_I f \quad \text{für jede Folge mit } \Delta(Z_n) \rightarrow 0. \quad (13.4)$$

Denn zu  $\varepsilon > 0$  wähle  $\delta > 0$  wie oben, und weiter  $K \in \mathbb{R}$  mit  $\Delta(Z_n) < \delta$  für  $n > K$ . Es folgt  $|S_{Z_n, \xi_n} - S| < \varepsilon$  für  $n > K$ , das heißt  $S_{Z_n, \xi_n}(f) \rightarrow S$ . In Definition 13.6 kann man die Voraussetzung  $f$  beschränkt auch weglassen, denn das folgt schon aus der Integrierbarkeit.

**Beispiel 13.4** Die konstante Funktion  $f : I = [a, b] \rightarrow \mathbb{R}$ ,  $f(x) = c$ , ist integrierbar mit

$$\int_I f = c(b - a).$$

Denn für jede Wahl der  $x_k$  und  $\xi_k \in [x_{k-1}, x_k]$  gilt  $f(\xi_k) = c$ , also

$$S_{Z,\xi}(f) = \sum_{k=1}^N c(x_k - x_{k-1}) = c(x_N - x_0) = c(b - a).$$

**Beispiel 13.5** Die Dirichletfunktion

$$\chi_{\mathbb{Q}} : [0, 1] \rightarrow \mathbb{R}, \chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{für } x \in \mathbb{Q}, \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases}$$

ist nicht Riemann-integrierbar. Sei zum Beispiel  $Z_n$  die äquidistante Unterteilung, also  $I_k = [\frac{k-1}{n}, \frac{k}{n}]$ , mit  $1 \leq k \leq n$ , also  $\Delta(Z_n) = \frac{1}{n} \rightarrow 0$ . Wählt man rationale Stützstellen  $\xi_k$  so ist  $S_{Z,\xi}(\chi_{\mathbb{Q}}) = 1$ , für  $\xi_k$  irrational ist dagegen  $S_{Z,\xi}(\chi_{\mathbb{Q}}) = 0$ .

<sup>19</sup>Bernhard Riemann, 1826-1866

**Satz 13.1 (Linearität des Integrals)**  $\mathcal{R}(I)$  ist ein Untervektorraum von  $\mathcal{B}(I)$ , und das Integral  $\mathcal{R}(I) \rightarrow \mathbb{R}, f \mapsto \int_I f$ , ist linear. Es gilt also für  $\lambda, \mu \in \mathbb{R}$

$$\int_I (\lambda f + \mu g) = \lambda \int_I f + \mu \int_I g.$$

BEWEIS: Seien  $f, g \in \mathcal{R}(I)$ . Dann ist  $\lambda f + \mu g$  beschränkt, und es gilt

$$S_{Z,\xi}(\lambda f + \mu g) = \lambda S_{Z,\xi}(f) + \mu S_{Z,\xi}(g).$$

Also folgt für  $\Delta(Z)$  hinreichend klein

$$\left| S_{Z,\xi}(\lambda f + \mu g) - \left( \lambda \int_I f + \mu \int_I g \right) \right| \leq |\lambda| \left| S_{Z,\xi}(f) - \int_I f \right| + |\mu| \left| S_{Z,\xi}(g) - \int_I g \right| < \varepsilon.$$

□

Integrierbarkeit ist eine globale Eigenschaft, es kommt auf die Funktion als Ganzes an. Deshalb ist nicht direkt klar, wie groß der Raum  $\mathcal{R}(I)$  ist. Wir zeigen in diesem Kapitel dass stetige Funktionen integrierbar sind. Unsere Strategie dazu ist wie folgt:

- Treppenfunktionen sind in  $\mathcal{R}(I)$ .
- $\mathcal{R}(I)$  ist abgeschlossen unter gleichmäßiger Konvergenz.
- Stetige Funktionen sind durch Treppenfunktionen gleichmäßig approximierbar.

Unser Beweis liefert dass auch stückweise stetige Funktionen integrierbar sind.

**Lemma 13.1** Seien  $f, \tilde{f} : I \rightarrow \mathbb{R}$  mit  $\tilde{f}(x) = f(x)$  für alle  $x \in I \setminus \{p_1, \dots, p_r\}$ . Mit  $f \in \mathcal{R}(I)$  ist dann auch  $\tilde{f} \in \mathcal{R}(I)$  und es gilt  $\int_I \tilde{f} = \int_I f$ .

BEWEIS: Wir zeigen die Aussage für einen Ausnahmepunkt  $p \in I$ ; der allgemeine Fall folgt daraus per Induktion. Es gilt für jede Zerlegung  $Z$  mit Stützstellen  $\xi_k$

$$S_{Z,\xi}(\tilde{f}) - S_{Z,\xi}(f) = \sum_{k=1}^N (\tilde{f}(\xi_k) - f(\xi_k)) \Delta x_k = (\tilde{f}(p) - f(p)) \sum_{\{k:\xi_k=p\}} \Delta x_k.$$

Es gibt höchstens zwei  $k$  mit  $p \in I_k$  und  $\Delta x_k > 0$ . Damit schätzen wir ab

$$\begin{aligned} \left| S_{Z,\xi}(\tilde{f}) - \int_I f \right| &\leq \left| S_{Z,\xi}(\tilde{f}) - S_{Z,\xi}(f) \right| + \left| S_{Z,\xi}(f) - \int_I f \right| \\ &\leq 2 \left| \tilde{f}(p) - f(p) \right| \Delta(Z) + \left| S_{Z,\xi}(f) - \int_I f \right|, \end{aligned}$$

und die rechte Seite geht gegen Null mit  $\Delta(Z) \rightarrow 0$ .

□

Für eine abzählbare Ausnahmemenge ist Lemma 13.1 i.a. falsch, siehe Beispiel 13.5.

**Lemma 13.2** Sei  $I = I_1 \cup \dots \cup I_n$  eine Zerlegung von  $I = [a, b]$  in Intervalle  $I_k = [a_{k-1}, a_k]$ . Ist  $f : I \rightarrow \mathbb{R}$  auf jedem der Teilintervalle  $I_k$  integrierbar, so folgt  $f \in \mathcal{R}(I)$  und

$$\int_I f = \sum_{k=1}^n \int_{I_k} f.$$

BEWEIS: Es reicht den Fall  $n = 2$  zu betrachten. Sei  $I = I' \cup I''$  mit  $I' = [a, p]$  und  $I'' = [p, b]$ , und sei  $f : I \rightarrow \mathbb{R}$  integrierbar auf  $I'$  und  $I''$ . Sei  $Z$  eine Zerlegung von  $I$  mit Punkten  $a = x_0 \leq \dots \leq x_N = b$  sowie Stützstellen  $\xi_1, \dots, \xi_N$ . Wähle  $r \in \{1, \dots, N\}$  mit  $p \in [x_{r-1}, x_r]$ , und definiere Zerlegungen  $Z', Z''$  von  $I', I''$  mit Stützstellen  $\xi', \xi''$  wie folgt:

$$\begin{aligned} Z' &= \{a = x_0 \leq \dots \leq x_{r-1} \leq p\} & \xi' &= \{\xi_1, \dots, \xi_{r-1}, p\}, \\ Z'' &= \{p \leq x_r \leq \dots \leq x_N = b\} & \xi'' &= \{p, \xi_{r+1}, \dots, \xi_N\}. \end{aligned}$$

Offenbar gilt  $\Delta(Z'), \Delta(Z'') \leq \Delta(Z)$ . Die Bilanz der Riemannschen Summen lautet

$$|S_{Z,\xi}(f) - (S_{Z',\xi'}(f) + S_{Z'',\xi''}(f))| = |(f(\xi_r) - f(p)) \Delta x_r| \leq 2 \|f\|_I \Delta(Z).$$

Es folgt mit der Dreiecksungleichung

$$\left| S_{Z,\xi}(f) - \left( \int_{I'} f + \int_{I''} f \right) \right| \leq 2 \|f\|_I \Delta(Z) + \left| S_{Z',\xi'}(f) - \int_{I'} f \right| + \left| S_{Z'',\xi''}(f) - \int_{I''} f \right|.$$

Die rechte Seite geht mit  $\Delta(Z) \rightarrow 0$  gegen Null.  $\square$

**Folgerung 13.1** Sei  $f : I \rightarrow \mathbb{R}$  eine Treppenfunktion, d. h. es gibt  $a = x_0 < x_1 < \dots < x_n = b$  und  $c_k \in \mathbb{R}$  mit  $f(x) = c_k$  für alle  $x \in (x_{k-1}, x_k)$ . Dann ist  $f$  integrierbar und es gilt

$$\int_I f = \sum_{k=1}^n c_k \Delta x_k.$$

BEWEIS: Nach Lemma 13.1 ist  $f : [x_{k-1}, x_k] \rightarrow \mathbb{R}$  integrierbar für alle  $k = 1, \dots, n$ . Aus Lemma 13.2 folgt die Behauptung.  $\square$

Im zweiten Schritt geht es um das Verhalten des Integrals unter Approximationen. Hier ein weiteres Beispiel dafür, dass punktweise Konvergenz nicht ausreicht.

**Beispiel 13.6** Sei  $q_1, q_2, \dots$  eine Abzählung der rationalen Zahlen in  $[0, 1]$ . Betrachte

$$\chi_n : [0, 1] \rightarrow \mathbb{R}, \chi_n(x) = \begin{cases} 1 & \text{für } x \in \{q_1, \dots, q_n\} \\ 0 & \text{sonst.} \end{cases}$$

Es gilt  $\chi_n \rightarrow \chi_{\mathbb{Q}}$  punktweise, aber  $\chi_{\mathbb{Q}}$  ist nicht Riemann-integrierbar nach Beispiel 13.5.

Wir betrachten deshalb nun die Supremumsnorm.

**Satz 13.2 (Abschätzung des Integrals)** Für  $f \in \mathcal{R}(I)$  gilt

$$\left| \int_I f \right| \leq \|f\|_I |I| \quad \text{mit } \|f\|_I = \sup_{x \in I} |f(x)|.$$

BEWEIS: Für jede Zerlegung  $Z$  mit Stützstellen  $\xi_k$  gilt nach Dreiecksungleichung

$$|S_{Z,\xi}(f)| = \left| \sum_{k=1}^N f(\xi_k) \Delta x_k \right| \leq \sum_{k=1}^N |f(\xi_k)| \Delta x_k \leq \|f\|_I \sum_{k=1}^N \Delta x_k = \|f\|_I |I|. \quad (13.5)$$

Die Abschätzung für das Integral folgt.  $\square$

**Bemerkung.** Der Beweis zeigt, falls zusätzlich  $|f| \in \mathcal{R}(I)$ , die genauere Abschätzung

$$\left| \int_I f \right| \leq \int_I |f| \leq \|f\| |I|. \quad (13.6)$$

Tatsächlich ist für  $f \in \mathcal{R}(I)$  immer auch  $|f| \in \mathcal{R}(I)$ , das zeigen wir in Folgerung 13.4.

Für  $f, f_k \in \mathcal{R}(I)$  haben wir jedenfalls nun

$$\left| \int_I f_k - \int_I f \right| = \left| \int_I (f_k - f) \right| \leq \|f_k - f\|_I |I|.$$

Aus  $\|f_k - f\|_I \rightarrow 0$  folgt somit  $\int_I f_k \rightarrow \int_I f$ .

**Satz 13.3 (Integral und gleichmäßige Konvergenz)** *Konvergiert die Folge  $f_k \in \mathcal{R}(I)$  gleichmäßig gegen  $f : I \rightarrow \mathbb{R}$ , so ist  $f \in \mathcal{R}(I)$  und es gilt*

$$\int_I f = \lim_{k \rightarrow \infty} \int_I f_k.$$

BEWEIS: Wegen  $\|f_k - f\|_I \rightarrow 0$  ist  $f$  beschränkt. Um die Definition der Integrierbarkeit für  $f$  zu verifizieren, brauchen wir den Wert  $S$  des Integrals. Mit  $S_k = \int_I f_k$  gilt nach Satz 13.2

$$|S_k - S_l| \leq \|f_k - f_l\|_I |I| \leq (\|f_k - f\|_I + \|f - f_l\|_I) |I| < \varepsilon \quad \text{für } k, l \text{ groß.}$$

Wir setzen  $S = \lim_{k \rightarrow \infty} S_k$  und zeigen, dass  $f$  integrierbar ist mit  $\int_I f = S$ . Für jede Zerlegung  $Z$  mit Stützstellen  $\xi$  und jedes  $k \in \mathbb{N}$  haben wir mit (13.5)

$$\begin{aligned} |S_{Z,\xi}(f) - S| &\leq |S_{Z,\xi}(f) - S_{Z,\xi}(f_k)| + |S_{Z,\xi}(f_k) - S_k| + |S_k - S| \\ &\leq \|f - f_k\|_I |I| + |S_{Z,\xi}(f_k) - S_k| + |S_k - S|. \end{aligned}$$

Zu  $\varepsilon > 0$  wählen wir erst  $k \in \mathbb{N}$  mit  $\|f - f_k\|_I |I| < \varepsilon/3$  und  $|S_k - S| < \varepsilon/3$ . Für  $\Delta(Z) < \delta$  ist dann auch  $|S_{Z,\xi}(f_k) - S_k| < \varepsilon/3$ , da  $f_k$  integrierbar ist. Damit ist der Satz bewiesen.  $\square$

Im dritten Schritt geht es um die Approximation von stetigen Funktionen durch Treppenfunktionen. Um die Stetigkeit einer Funktion  $f : D \rightarrow \mathbb{R}$  mit  $D \subset \mathbb{R}$  zu quantifizieren, betrachten wir die Oszillation (oder Schwankung) auf Skala  $\delta > 0$ :

$$\omega_f(\delta) = \sup_{x, x' \in D, |x-x'| < \delta} |f(x) - f(x')| \in [0, \infty]. \quad (13.7)$$

Wird  $\delta > 0$  verkleinert, so sind am Supremum weniger Punkte  $x, x'$  beteiligt, die Oszillation wird also höchstens kleiner. Insbesondere existiert nach Lemma 8.1 der Grenzwert  $\lim_{\delta \searrow 0} \omega_f(\delta)$ . Für  $f$  stetig sollte dieser Grenzwert Null sein, da  $f$  auf kleinen Skalen wenig schwankt. Das ist aber so nicht richtig, wie das folgende Beispiel zeigt.

**Beispiel 13.7** Sei  $f : I = (0, 1] \rightarrow \mathbb{R}$ ,  $f(x) = \sin \frac{1}{x}$ . Dann ist  $f$  stetig, und  $|f(x) - f(x')| \leq 2$  für alle  $x, x' \in (0, 1]$ . Es gilt sogar

$$|f(x_n^+) - f(x_n^-)| = 2 \quad \text{für} \quad x_n^\pm = \frac{1}{2n\pi \pm \pi/2} \xrightarrow{n \rightarrow \infty} 0.$$

Es folgt  $\omega_f(\delta) = 2$  für alle  $\delta > 0$ . Die Funktion kann nicht auf  $[0, 1]$  stetig fortgesetzt werden, da sie in jeder Umgebung von  $x = 0$  zwischen den Werten  $\pm 1$  oszilliert.

Das Problem hängt mit dem Definitionsbereich zusammen. Zur Erinnerung: eine Menge  $D \subset \mathbb{R}^n$  heißt kompakt, wenn jede Folge  $x_n \in D$  eine Teilfolge hat, die gegen ein  $x_0 \in D$  konvergiert. In Satz 10.1 haben wir gezeigt:  $D \subset \mathbb{R}^n$  ist genau dann kompakt, wenn  $D$  abgeschlossen und beschränkt ist.

**Satz 13.4** *Ist  $f : D \rightarrow \mathbb{R}$  stetig mit  $D \subset \mathbb{R}^n$  kompakt, so gilt  $\lim_{\delta \searrow 0} \omega_f(\delta) = 0$ .*

*Beweis.* Andernfalls gibt es ein  $\varepsilon > 0$  und  $\delta_n \searrow 0$  mit  $\omega_f(\delta_n) \geq \varepsilon > 0$ . Dann existieren  $x_n, x'_n \in D$  mit  $|x_n - x'_n| < \delta_n$ , so dass  $|f(x_n) - f(x'_n)| \geq \varepsilon/2$ . Nach Übergang zu einer Teilfolge gilt  $x_n \rightarrow x_0 \in D$ . Aber dann ist auch  $x'_n \rightarrow x_0$ , und wegen  $f$  stetig in  $x_0$  folgt

$$\frac{\varepsilon}{2} \leq \lim_{n \rightarrow \infty} |f(x_n) - f(x'_n)| = |f(x_0) - f(x_0)| = 0, \text{ Widerspruch.}$$

□

$f : D \rightarrow \mathbb{R}$  heißt gleichmäßig stetig genau wenn  $\lim_{\delta \searrow 0} \omega_f(\delta) = 0$ . Daher besagt der Satz:

eine stetige Funktion auf einer kompakten Menge  $D$  ist (sogar) gleichmäßig stetig.

Wir bemerken dass die gleichmäßige Stetigkeit auch wie folgt charakterisiert werden kann:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad |f(x) - f(x')| < \varepsilon \text{ für alle } x, x' \in D \text{ mit } |x - x'| < \delta.$$

Zur Erinnerung: die Stetigkeit von  $f$  auf  $D$  bedeutet

$$\forall x_0 \in D \quad \forall \varepsilon > 0 \quad \exists \delta > 0 : \quad |f(x) - f(x_0)| < \varepsilon \text{ für alle } x \in D \text{ mit } |x - x_0| < \delta.$$

Im allgemeinen kann keine gemeinsame Schranke  $\delta > 0$  für alle  $x_0 \in D$  gewählt werden, siehe Beispiel 13.7. Genau das ist aber möglich wenn  $f$  gleichmäßig stetig ist.

**Lemma 13.3** *Jede stetige Funktion  $f : I = [a, b] \rightarrow \mathbb{R}$  kann gleichmäßig durch Treppenfunktionen approximiert werden.*

BEWEIS: Sei  $x_k = a + k \frac{b-a}{N}$  für  $k = 0, \dots, N$ , und setze

$$\varphi_N(x) = \begin{cases} f(x_0) & \text{für } x = x_0, \\ f(x_k) & \text{für } x \in (x_{k-1}, x_k] \text{ mit } k \in \{1, \dots, N\}, \end{cases}$$

Für  $x \in (x_{k-1}, x_k]$  ist  $|x - x_k| \leq \frac{b-a}{N}$ , also

$$|\varphi_N(x) - f(x)| = |f(x_k) - f(x)| \leq \omega_f\left(\frac{b-a}{N}\right).$$

Da  $\varphi_N(x_0) = f(x_0)$ , folgt mit Satz 13.4

$$\|\varphi_N - f\|_I \leq \omega_f\left(\frac{b-a}{N}\right) \rightarrow 0 \quad \text{für } N \rightarrow \infty.$$

□

**Definition 13.7 (stückweise stetig)** *Eine Funktion  $f : [a, b] \rightarrow \mathbb{R}$  heißt stückweise stetig, wenn es eine Zerlegung  $a = a_0 < \dots < a_n = b$  gibt, so dass  $f$  auf jedem Teilintervall  $[a_{k-1}, a_k]$  stetig ist, nach eventueller Abänderung in den Endpunkten  $a_{k-1}$  und  $a_k$ .*

**Satz 13.5 (stückweise stetig  $\Rightarrow$  integrierbar)** Ist  $f : I = [a, b] \rightarrow \mathbb{R}$  stückweise stetig, so ist  $f$  Riemann-integrierbar.

BEWEIS: Sei zunächst  $f$  stetig. Dann ist  $f$  gleichmäßiger Limes von Treppenfunktionen nach Lemma 13.3. Nach Folgerung 13.1 sind Treppenfunktionen integrierbar, also auch  $f$  wegen Satz 13.3. Ist  $f$  nur stückweise stetig, so ist  $f$  auf jedem der Teilintervalle stetig nach eventueller Abänderung in den zwei Endpunkten. Wie bewiesen ist die stetige Funktion auf dem Teilintervall integrierbar, also auch  $f$  nach Lemma 13.1. Schließlich folgt aus Lemma 13.2, dass  $f$  auf ganz  $I$  integrierbar ist.  $\square$

**Satz 13.6 (Monotonie des Integrals)** Sind  $f, g \in \mathcal{R}(I)$ , so gilt:

$$f \leq g \quad \Rightarrow \quad \int_I f \leq \int_I g.$$

BEWEIS: Für jede Zerlegung  $Z$  mit Stützstellen  $\xi_k$  gilt

$$S_{Z,\xi}(f) = \sum_{k=1}^N f(\xi_k) \Delta x_k \leq \sum_{k=1}^N g(\xi_k) \Delta x_k = S_{Z,\xi}(g).$$

Sei  $Z_n, \xi_n$  Folge mit  $\Delta(Z_n) \rightarrow 0$ , dann gilt  $S_{Z_n, \xi_n}(f) \rightarrow \int_I f$ ,  $S_{Z_n, \xi_n}(g) \rightarrow \int_I g$  nach (13.4).  $\square$

**Folgerung 13.2 (Mittelwertsatz der Integralrechnung)** Sei  $\varphi : I = [a, b] \rightarrow [0, \infty)$  stetig mit  $\int_I \varphi > 0$ . Dann gibt es für jede Funktion  $f \in C^0(I)$  ein  $\xi \in I$  mit

$$f(\xi) = \frac{1}{\int_I \varphi} \int_I f \varphi.$$

Im Spezialfall  $\varphi = 1$  gilt also  $f(\xi) = \frac{1}{|I|} \int_I f$ .

BEWEIS: Setze  $m = \min_{x \in I} f(x)$  und  $M = \max_{x \in I} f(x)$ . Wegen  $m\varphi \leq f\varphi \leq M\varphi$  gilt

$$m = \frac{1}{\int_I \varphi} \int_I m\varphi \leq \frac{1}{\int_I \varphi} \int_I f\varphi \leq \frac{1}{\int_I \varphi} \int_I M\varphi = M.$$

Die Existenz von  $\xi$  folgt aus dem Zwischenwertsatz.  $\square$

Um mit Definition 13.6 die Integrierbarkeit zu zeigen, müssen wir den Wert  $S$  des Integrals a priori kennen. Wir diskutieren jetzt ein Kriterium für Integrierbarkeit nach Darboux<sup>20</sup>, das diese Kenntnis nicht voraussetzt. Das ist vergleichbar mit dem Cauchy Kriterium für die Konvergenz von Folgen. Aus Zeitgründen wurde die Version in der Vorlesung nicht behandelt.

Sei  $f : I = [a, b] \rightarrow \mathbb{R}$  beschränkt und  $Z$  Zerlegung von  $I$  in Intervalle  $I_k$  der Länge  $\Delta x_k$ . Das Kriterium basiert auf dem Konzept der Ober- und Untersumme

$$\bar{S}_Z(f) = \sum_{k=1}^N (\sup_{I_k} f) \Delta x_k \quad \text{und} \quad \underline{S}_Z(f) = \sum_{k=1}^N (\inf_{I_k} f) \Delta x_k.$$

<sup>20</sup>Gaston Darboux, 1842-1917

Zunächst zur Abhängigkeit der Ober-/Untersumme von  $Z$ . Nach Definition gilt

$$\underline{S}_Z(f) \leq \overline{S}_Z(f). \quad (13.8)$$

Nehmen wir den Punkt  $\xi \in [x_{k-1}, x_k]$  hinzu, so folgt mit  $I'_k = [x_{k-1}, \xi]$  und  $I''_k = [\xi, x_k]$

$$\begin{aligned} \underline{S}_{Z \cup \{\xi\}}(f) - \underline{S}_Z(f) &= (\inf_{I'_k} f)(\xi - x_{k-1}) + (\inf_{I''_k} f)(x_k - \xi) - (\inf_{I_k} f)(x_k - x_{k-1}) \\ &= (\inf_{I'_k} f - \inf_{I_k} f)(\xi - x_{k-1}) + (\inf_{I''_k} f - \inf_{I_k} f)(x_k - \xi). \end{aligned}$$

Mit  $\inf_{I'_k} f, \inf_{I''_k} f, -\inf_{I_k} f \leq \|f\|_I$  sowie  $0 \leq x_k - x_{k-1} \leq \Delta(Z)$  erhalten wir

$$0 \leq \underline{S}_{Z \cup \{\xi\}}(f) - \underline{S}_Z(f) \leq 2\|f\|_I \Delta(Z),$$

und analog für die Obersummen

$$0 \geq \overline{S}_{Z \cup \{\xi\}}(f) - \overline{S}_Z(f) \geq -2\|f\|_I \Delta(Z).$$

Die jeweils linken Ungleichungen liefern für beliebige Zerlegungen  $Z, Z'$  per Induktion

$$\underline{S}_Z(f) \leq \underline{S}_{Z \cup Z'}(f) \leq \overline{S}_{Z \cup Z'}(f) \leq \overline{S}_{Z'}(f). \quad (13.9)$$

Ebenfalls induktiv folgt rechts, wobei  $N$  die Zahl der Teilintervalle von  $Z'$  ist,

$$\underline{S}_{Z \cup Z'}(f) - 2N\|f\|_I \Delta(Z) \leq \underline{S}_Z(f) \leq \overline{S}_Z(f) \leq \overline{S}_{Z \cup Z'}(f) + 2N\|f\|_I \Delta(Z). \quad (13.10)$$

Wir definieren nun das Ober- bzw. Unterintegral von  $f$  durch

$$\begin{aligned} \overline{S}(f) &= \inf\{\overline{S}_Z(f) : Z \text{ ist Zerlegung von } I\}, \\ \underline{S}(f) &= \sup\{\underline{S}_Z(f) : Z \text{ ist Zerlegung von } I\}. \end{aligned}$$

Aus (13.9) folgt  $\underline{S}(f) \leq \overline{S}(f)$ .

**Satz 13.7 (Darboux)** *Eine beschränkte Funktion  $f : I \rightarrow \mathbb{R}$  ist genau dann Riemann-integrierbar, wenn ihr Ober- und Unterintegral gleich sind, und es gilt dann*

$$\int_I f = \overline{S}(f) = \underline{S}(f).$$

BEWEIS: Sei  $f$  Riemann-integrierbar mit Integral  $S$ . Zu  $\varepsilon > 0$  wähle  $\delta > 0$  wie in Definition 13.6. Sei  $Z$  eine Zerlegung mit  $\Delta(Z) < \delta$ , dann folgt

$$S - \varepsilon \leq \inf_{\xi} S_{Z,\xi}(f) = \underline{S}_Z(f) \leq \underline{S}(f) \leq \overline{S}(f) \leq \overline{S}_Z(f) = \sup_{\xi} S_{Z,\xi}(f) \leq S + \varepsilon.$$

Umgekehrt sei  $Z'$  eine Zerlegung mit  $\overline{S}_{Z'}(f) > \underline{S}(f) - \varepsilon$ , und  $N$  sei die Zahl der Teilintervalle von  $Z'$ . Für jede Zerlegung  $Z$  folgt, mit (13.10) und (13.9)

$$\begin{aligned} S_{Z,\xi}(f) \geq \underline{S}_Z(f) &\geq \underline{S}_{Z \cup Z'}(f) - 2N\|f\|_I \Delta(Z) \\ &\geq \underline{S}_{Z'}(f) - 2N\|f\|_I \Delta(Z) \\ &\geq \underline{S}(f) - \varepsilon - 2N\|f\|_I \Delta(Z). \end{aligned}$$

Für  $\Delta(Z) < \delta$  folgt  $S_{Z,\xi}(f) > \underline{S}(f) - 2\varepsilon$ . Die Abschätzung nach oben ist analog.  $\square$

**Folgerung 13.3 (Darboux-Kriterium)** Eine beschränkte Funktion  $f : I = [a, b] \rightarrow \mathbb{R}$  ist genau dann Riemann-integrierbar, wenn es zu jedem  $\varepsilon > 0$  eine Zerlegung  $Z$  gibt mit

$$\overline{S}_Z(f) - \underline{S}_Z(f) < \varepsilon. \quad (13.11)$$

BEWEIS: Sei  $f$  Riemann-integrierbar, und  $Z, Z'$  seien Zerlegungen mit  $\overline{S}_Z(f) < \overline{S}(f) + \varepsilon$  und  $\underline{S}_{Z'}(f) > \underline{S}(f) - \varepsilon$ . Mit (13.9) folgt

$$\overline{S}_{Z \cup Z'}(f) - \underline{S}_{Z \cup Z'}(f) \leq \overline{S}_Z(f) - \underline{S}_{Z'}(f) < (\overline{S}(f) + \varepsilon) - (\underline{S}(f) - \varepsilon) = 2\varepsilon.$$

Aus (13.11) folgt andererseits  $\overline{S}(f) - \underline{S}(f) \leq \overline{S}_Z(f) - \underline{S}_Z(f) < \varepsilon$ , also  $\overline{S}(f) = \underline{S}(f)$ . Damit ist  $f$  Riemann-integrierbar nach Satz 13.7.  $\square$

Wir zeigen jetzt unter anderem dass mit  $f$  auch  $|f|$  Riemann-integrierbar ist.

**Folgerung 13.4** Sei  $f : [a, b] \rightarrow \mathbb{R}$  Riemann-integrierbar und  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  Lipschitz mit Konstante  $L$ . Dann ist  $\phi \circ f$  auch Riemann-integrierbar.

BEWEIS: Für eine Zerlegung  $Z$  und  $x, x' \in I_k$  gilt

$$\phi(f(x)) - \phi(f(x')) \leq L|f(x) - f(x')| \leq L \left( \sup_{I_k} f - \inf_{I_k} f \right).$$

Bilde das Supremum/Infimum bzgl.  $x$  und  $x'$ , multipliziere mit  $\Delta x_k$  und summiere. Es folgt

$$\overline{S}_Z(\phi \circ f) - \underline{S}_Z(\phi \circ f) \leq L(\overline{S}_Z(f) - \underline{S}_Z(f)) < \varepsilon,$$

für eine geeignete Zerlegung  $Z$  nach Folgerung 13.3.  $\square$

Wir können schließlich eine Ergänzung zu Lemma 13.2 beweisen.

**Folgerung 13.5** Ist  $f$  Riemann-integrierbar auf  $[a, b]$ , so ist  $f$  auch integrierbar auf jedem Teilintervall  $[c, d] \subset [a, b]$ . Insbesondere gilt

$$\int_{[a,b]} f = \int_{[a,c]} f + \int_{[c,b]} f \quad \text{für jedes } c \in [a, b].$$

BEWEIS: Sei  $Z$  eine Zerlegung von  $[c, d]$  in Intervalle  $I_k$  mit  $\Delta(Z) < \delta$ . Wir ergänzen  $Z$  zu einer Zerlegung  $Z'$  von  $[a, b]$  in Intervalle  $I_{k'}$ , auch mit  $\Delta(Z') < \delta$ . Es gilt

$$\overline{S}_Z(f) - \underline{S}_Z(f) = \sum_k (\sup_{I_k} f - \inf_{I_k} f) \Delta x_k \leq \sum_{k'} (\sup_{I_{k'}} f - \inf_{I_{k'}} f) \Delta x_{k'} = \overline{S}_{Z'}(f) - \underline{S}_{Z'}(f).$$

Nun sind  $\overline{S}_{Z'}(f)$  und  $\underline{S}_{Z'}(f)$  Grenzwerte von Riemannschen Summen zu  $Z'$ , und diese approximieren das Integral  $\int_{[a,b]} f$ . Für  $\delta > 0$  klein folgt  $\overline{S}_{Z'}(f) - \underline{S}_{Z'}(f) < \varepsilon$ . Nach Folgerung 13.3 ist  $f$  integrierbar auf  $[c, d]$ , und die Additivität folgt mit Lemma 13.2.  $\square$

Zum Schluss kurz zum vektorwertigen Riemann-Integral.  $f : I = [a, b] \rightarrow \mathbb{R}^n$  heißt Riemann-integrierbar wenn die Komponentenfunktionen Riemann-integrierbar sind. Natürlich könnten wir alternativ auf die Definition mittels Riemannscher Summen zurückgreifen. Wir definieren

$$\int_I f = \sum_{j=1}^n \left( \int_I f_j \right) e_j.$$

Dabei ist  $e_1, \dots, e_n$  die Standardbasis. Wir wollen allgemeiner das Integral auch für Funktionen  $f : I = [a, b] \rightarrow V$  definieren, wobei  $V$  beliebiger  $\mathbb{R}$ -Vektorraum mit  $\dim V < \infty$ . Sei dazu  $\mathcal{A} = \{v_1, \dots, v_n\}$  eine Basis von  $V$ . Dann hat  $f$  die Koordinatendarstellung  $f_{\mathcal{A}} : I \rightarrow \mathbb{R}^n$ , und wir definieren das Integral in  $V$  durch seinen Koordinatenvektor

$$\left( \int_I f \right)_{\mathcal{A}} = \int_I f_{\mathcal{A}} \in \mathbb{R}^n.$$

Dabei nehmen wir wieder an, dass die Komponentenfunktionen Riemann-integrierbar sind. Das Konzept ist unabhängig von der Wahl einer Basis: sei  $\mathcal{B} = \{w_1, \dots, w_n\}$  eine andere Basis und  $C \in \text{GL}_n(\mathbb{R})$  die Transformationsmatrix, also  $v_j = \sum_{i=1}^n C_{ij} w_i$ . Dann folgt  $f_{\mathcal{B}} = C f_{\mathcal{A}}$ , insbesondere sind die Komponenten von  $f_{\mathcal{B}}$  Riemann-integrierbar, und

$$\left( \int_I f_{\mathcal{B}} \right)_i = \int_I \sum_{j=1}^n C_{ij} (f_{\mathcal{A}})_j = \sum_{j=1}^n C_{ij} \int_I (f_{\mathcal{A}})_j = \sum_{j=1}^n C_{ij} \left( \int_I f_{\mathcal{A}} \right)_j = \left( C \int_I f_{\mathcal{A}} \right)_i.$$

Somit transformieren sich die Integrale von  $f_{\mathcal{A}}$  und  $f_{\mathcal{B}}$  mit der Transformationsmatrix  $C$ , definieren also denselben Vektor in  $V$ . Soweit zum vektorwertigen Fall.



## 14 Ableitung und Integral

Wir kommen nun zu dem zentralen, von Newton und Leibniz studierten Zusammenhang zwischen Differentiation und Integration. Einerseits ergibt sich für stetige Funktionen die Existenz einer Stammfunktion, andererseits können mit einer gegebenen Stammfunktion Inhalte oder Integrale berechnet werden. Wir hatten schon bei der Ableitung bemerkt, dass ein wichtiger Aspekt des *Calculus* seine allgemeine Anwendbarkeit war und ist. Das gilt hier erst recht.

**Definition 14.1** Sei  $f : (a, b) \rightarrow \mathbb{R}$ . Eine Funktion  $F : (a, b) \rightarrow \mathbb{R}$  heißt *Stammfunktion* von  $f$ , wenn gilt:

$$F' = f \quad \Leftrightarrow \quad F'(x) = f(x) \quad \text{für alle } x \in (a, b).$$

**Satz 14.1** Ist  $F$  eine Stammfunktion von  $f$  auf  $(a, b)$ , so ist jede Stammfunktion von  $f$  auf  $(a, b)$  von der Form  $F + c$ , für eine Konstante  $c \in \mathbb{R}$ .

BEWEIS: Sei  $G$  auch Stammfunktion von  $f$  auf  $(a, b)$ . Es folgt

$$(G - F)' = G' - F' = f - f = 0.$$

Nach Folgerung 10.1 gibt es eine Konstante  $c \in \mathbb{R}$  mit  $G - F = c$ , also  $G = F + c$ .  $\square$

Die Gleichung  $F' = f$  ist ein weiteres Beispiel für eine Differentialgleichung. Satz 14.1 sagt aus, dass eine Lösung  $F$  der Gleichung bis auf eine additive Konstante  $c \in \mathbb{R}$  eindeutig bestimmt ist. Es stellt sich nun das Problem der Existenz. Die Gleichung enthält auf der rechten Seite eine willkürliche Funktion, darum lautet hier die Frage:

Für welche  $f$  ist die Differentialgleichung  $F' = f$  auf dem Intervall  $(a, b)$  lösbar?

Ist die rechte Seite ein Polynom  $f(x) = \sum_{k=0}^n a_k x^k$  so können wir die Lösung direkt hinschreiben, es ist  $F(x) = \sum_{k=0}^n a_k x^{k+1} / (k+1)$ , das ist aber zu speziell. Mithilfe des Integrals werden wir eine Lösung für jede stetige rechte Seite angeben. Setze für  $f : [a, b] \rightarrow \mathbb{R}$  integrierbar

$$\int_a^b f(x) dx = \begin{cases} \int_I f & \text{falls } a \leq b, \\ -\int_I f & \text{falls } a \geq b. \end{cases}$$

Es folgt für beliebige  $a, b, c \in \mathbb{R}$ , sofern  $f$  auf allen Intervallen Riemann-integrierbar ist,

$$\int_a^b f(x) dx + \int_b^c f(x) dx + \int_c^a f(x) dx = 0. \quad (14.1)$$

Für  $a \leq b \leq c$  ist das Lemma 13.2. Die linke Seite bleibt nun gleich bei zyklischen Vertauschungen von  $a, b, c$ , und ändert nur das Vorzeichen bei antizyklischen Vertauschungen. Daher gilt (14.1) allgemein. Die Notation  $f(x) dx$  ist rein formal, ein  $dx$  wurde in der Vorlesung nicht definiert. Aber sie ist nützlich, zum Beispiel wenn  $f$  noch von weiteren Variablen  $y, z, a, \dots$  abhängt: es wird spezifiziert, bezüglich welcher Variablen integriert werden soll. Außerdem erinnert die Notation an die Riemannschen Summen, mit denen das Integral definiert wurde.

**Satz 14.2 (Hauptsatz der Differential- und Integralrechnung)** Sei  $f : I = [a, b] \rightarrow \mathbb{R}$  stetig. Dann ist für jedes  $x_0 \in I$  die Funktion

$$F : I \rightarrow \mathbb{R}, \quad F(x) = \int_{x_0}^x f(y) dy$$

eine Stammfunktion von  $f$ , das heißt es gilt  $F'(x) = f(x)$  für alle  $x \in I$ .

*Bemerkung.* In den Endpunkten des Intervalls ist dies im Sinne der einseitigen Ableitungen  $F'_+(a) = f(a)$  bzw.  $F'_-(b) = f(b)$  zu verstehen.

BEWEIS:  $F$  ist definiert nach Satz 13.5. Wir betrachten die rechtsseitige Ableitung, das heißt es seien  $x, x+h \in I$  mit  $h > 0$ . Dann gilt

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} - f(x) &= \frac{1}{h} \left( \int_{x_0}^{x+h} f(y) dy - \int_{x_0}^x f(y) dy - hf(x) \right) \\ &= \frac{1}{h} \int_x^{x+h} (f(y) - f(x)) dy. \end{aligned}$$

Mit Abschätzung 13.2 und der Stetigkeit von  $f$  im Punkt  $x$  folgt

$$\left| \frac{1}{h} \int_x^{x+h} (f(y) - f(x)) dy \right| \leq \frac{1}{|h|} |h| \sup_{[x, x+h]} |f(y) - f(x)| \rightarrow 0 \quad \text{mit } h \rightarrow 0.$$

Der Beweis im Fall  $h < 0$  ist identisch, nur im letzten Schritt muss  $\sup_{[x+h, x]}$  stehen.  $\square$

**Folgerung 14.1** Sei  $f \in C^0([a, b])$ . Ist  $F \in C^0([a, b])$  Stammfunktion von  $f$  auf  $(a, b)$ , so gilt für beliebiges  $x_0 \in [a, b]$

$$F(x) = F(x_0) + \int_{x_0}^x f(y) dy \quad \text{für alle } x \in [a, b]. \quad (14.2)$$

BEWEIS: Nach Satz 14.1 und Satz 14.2 gibt es ein  $c \in \mathbb{R}$  mit

$$F(x) = c + \int_{x_0}^x f(y) dy \quad \text{für alle } x \in (a, b).$$

Durch Grenzübergang folgt die Gleichung auch für  $x = a, b$ . Denn  $F$  ist stetig in den Randpunkten nach Voraussetzung, und das Integral ist ebenfalls stetig, zum Beispiel gilt

$$\left| \int_{x_0}^b f(y) dy - \int_{x_0}^x f(y) dy \right| = \left| \int_x^b f(y) dy \right| \leq |b-x| \|f\|_I \rightarrow 0 \quad \text{mit } x \nearrow b.$$

Durch Einsetzen von  $x = x_0$  sehen wir schließlich  $c = F(x_0)$ .  $\square$

**Folgerung 14.2 (Berechnung von Integralen mit Stammfunktionen)** Sei  $f \in C^0([a, b])$ . Ist  $F \in C^0([a, b])$  Stammfunktion von  $f$  auf  $(a, b)$ , so gilt

$$\int_a^b f(x) dx = F(b) - F(a) =: [F(x)]_{x=a}^{x=b}$$

BEWEIS: Folgt aus (14.2) mit  $x_0 = a, x = b$ .  $\square$

Mit dem Hauptsatz bzw. mit Folgerung 14.2 lassen sich die Differentiationsregeln aus Kapitel 9 in Integrationsregeln übersetzen. Die resultierenden Formeln sind sowohl theoretisch wichtig als auch zur Berechnung konkreter Integrale. In der folgenden Regel verwenden wir die Notation  $f \in C^1([a, b])$ . Damit ist gemeint:  $f \in C^1((a, b))$  und  $f, f'$  haben Grenzwerte in den Randpunkten  $a, b$ . Somit sind  $f, f'$  durch Fortsetzung stetig auf  $[a, b]$  definiert.<sup>21</sup>

<sup>21</sup>Es existieren dann die einseitigen Ableitungen in  $a, b$ , und sind gleich den Werten der Fortsetzung.

**Satz 14.3 (Partielle Integration)** Seien  $f, g \in C^1(I)$  mit  $I = [a, b]$ . Dann gilt

$$\int_a^b fg' = [f(x)g(x)]_{x=a}^{x=b} - \int_a^b f'g.$$

BEWEIS: Es gilt nach Produktregel  $(fg)' = f'g + fg'$  auf  $(a, b)$ . Dabei sind alle Funktionen  $f, g, f', g'$  in  $C^0([a, b])$ , also liefert Folgerung 14.2 die Behauptung.  $\square$

**Satz 14.4 (Substitutions- oder Transformationsregel)** Sei  $I = [a, b]$ ,  $I^* = [\alpha, \beta]$  und  $\varphi \in C^1(I)$  mit  $\varphi(I) \subset I^*$ . Dann gilt für  $f \in C^0(I^*)$

$$\int_{\varphi(a)}^{\varphi(b)} f(y) dy = \int_a^b f(\varphi(x))\varphi'(x) dx.$$

BEWEIS: Wähle nach Satz 14.2 eine Stammfunktion  $F \in C^1(I^*)$  von  $f$ . Dann gilt

$$\begin{aligned} \int_{\varphi(a)}^{\varphi(b)} f(y) dy &= [F(y)]_{y=\varphi(a)}^{y=\varphi(b)} \quad (\text{Folgerung 14.2}) \\ &= [F(\varphi(x))]_{x=a}^{x=b} \\ &= \int_a^b (F \circ \varphi)'(x) dx \quad (\text{Folgerung 14.2}) \\ &= \int_a^b F'(\varphi(x))\varphi'(x) dx \quad (\text{Kettenregel}) \\ &= \int_a^b f(\varphi(x))\varphi'(x) dx. \end{aligned}$$

$\square$

Für die Anwendung der Substitutionsregel ist folgendes *Kochrezept* nützlich: bei einem gegebenen Integral  $\int_{\alpha}^{\beta} f(y) dy$  möchten wir  $y = y(x)$  substituieren. Berechne dazu

$$y = y(x) \quad \Rightarrow \quad dy = y'(x) dx.$$

Bestimme die neuen Intervallgrenzen  $a, b$  durch Auflösen der Gleichungen  $\alpha = y(a)$ ,  $\beta = y(b)$ . Es ergibt sich die Substitutionsformel

$$\int_{\alpha}^{\beta} f(y) dy = \int_a^b f(y(x)) y'(x) dx.$$

Heutzutage können Stammfunktionen mit künstlicher Intelligenz gefunden werden, insofern sind per Hand Integrationen weniger bedeutsam. Hier trotzdem ein paar Beispiele.

**Beispiel 14.1** Direkte Kenntnis der Ableitungen ergibt

$$\begin{aligned} \int_a^b x^\alpha dx &= \left[ \frac{x^{\alpha+1}}{\alpha+1} \right]_{x=a}^{x=b} \quad (\alpha \in \mathbb{R} \setminus \{-1\}, a, b > 0), \\ \int_a^b \frac{dx}{x} &= [\log x]_{x=a}^{x=b} \quad (a, b > 0), \\ \int_a^b \frac{dx}{\sqrt{1-x^2}} &= [\arcsin x]_{x=a}^{x=b}, \quad (-1 < a, b < 1). \end{aligned}$$

Ein Beispiel für die Anwendung der partiellen Integration ist

**Beispiel 14.2** Ein Beispiel für partielle Integration ist

$$\int_1^x \log u \, du = \int_1^x 1 \cdot \log u \, du = [u \log u]_{u=1}^{u=x} - \int_1^x u \frac{1}{u} \, du = x \log x - (x - 1).$$

Eine weitere Anwendung der partiellen Integration ist das

**Beispiel 14.3 (Wallis-Produkt)** <sup>22</sup> Wir berechnen hier  $A_n = \int_0^{\pi/2} \sin^n x \, dx$ . Offenbar gilt

$$A_0 = \pi/2 \quad \text{und} \quad A_1 = [-\cos x]_{x=0}^{x=\pi/2} = 1.$$

Für  $n \geq 1$  leiten wir durch partielle Integration eine Rekursionsformel her:

$$\begin{aligned} A_{n+1} &= \int_0^{\pi/2} \sin x \sin^n x \, dx \\ &= \underbrace{[-\cos x \sin^n x]_{x=0}^{x=\pi/2}}_{=0} + n \int_0^{\pi/2} \cos^2 x \sin^{n-1} x \, dx \\ &= n \int_0^{\pi/2} \sin^{n-1} x \, dx - n \int_0^{\pi/2} \sin^{n+1} x \, dx, \end{aligned}$$

wobei wir  $\cos^2 x = 1 - \sin^2 x$  benutzt haben. Es folgt

$$A_{n+1} = \frac{n}{n+1} A_{n-1} \quad (n \geq 1).$$

Durch Induktion erhalten wir

$$\begin{aligned} A_{2n} &= \frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdot \dots \cdot \frac{1}{2} \cdot A_0 = \left( \prod_{k=1}^n \frac{2k-1}{2k} \right) \frac{\pi}{2}, \\ A_{2n+1} &= \frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdot \dots \cdot \frac{2}{3} \cdot A_1 = \left( \prod_{k=1}^n \frac{2k}{2k+1} \right) 1. \end{aligned}$$

Es folgt weiter

$$\frac{A_{2n}}{A_{2n+1}} = \frac{\pi}{2} \cdot \prod_{k=1}^n \frac{4k^2 - 1}{4k^2}.$$

Nun gilt  $A_{n+1} = \int_0^{\pi/2} \sin^{n+1} x \, dx \leq \int_0^{\pi/2} \sin^n x \, dx = A_n$ , und es folgt

$$1 \leq \frac{A_n}{A_{n+1}} \leq \frac{A_{n-1}}{A_{n+1}} = \frac{n+1}{n} \rightarrow 1 \quad \text{mit } n \rightarrow \infty.$$

Daraus ergibt sich die hübsche Produktdarstellung von Wallis

$$\frac{\pi}{2} = \prod_{k=1}^{\infty} \frac{4k^2}{4k^2 - 1} = \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdot \dots$$

---

<sup>22</sup>John Wallis, 1616-1703

Als nächstes behandeln wir Beispiele zur Substitutionsregel.

**Beispiel 14.4 (Lineare Parameterwechsel)** Mit  $y = (x - x_0)/m$  haben wir  $dy = 1/m dx$  und  $x = x_0 + my$ , also neue Grenzen  $a = x_0 + m\alpha$  und  $b = x_0 + m\beta$ . Es folgt

$$\int_{\alpha}^{\beta} f(y) dy = \frac{1}{m} \int_{x_0+m\alpha}^{x_0+m\beta} f\left(\frac{x-x_0}{m}\right) dx.$$

**Beispiel 14.5 (Integration von Ableitungen)**

$$\begin{aligned} \int_a^b \frac{f'(x)}{f(x)} dx &= \int_a^b (\log f)'(x) dx = [\log f(x)]_{x=a}^{x=b} \quad (f > 0), \\ \int_a^b u\sqrt{1+u^2} du &= \frac{1}{3} \int_a^b ((1+u^2)^{\frac{3}{2}})' du = \left[\frac{1}{3}(1+u^2)^{\frac{3}{2}}\right]_{u=a}^{u=b}. \\ \int_a^b F'(f(x))f'(x) dx &= [F(f(x))]_{x=a}^{x=b}. \end{aligned}$$

**Beispiel 14.6 (Flächeninhalt unter Hyperbel)** Zu berechnen ist das Integral

$$A(x) = \int_1^x \sqrt{u^2 - 1} du \quad \text{für } x \geq 1.$$

Hierfür brauchen wir die hyperbolischen Funktionen

$$\cosh(t) = \frac{e^t + e^{-t}}{2} \quad \text{und} \quad \sinh(t) = \frac{e^t - e^{-t}}{2}.$$

Aus den Definitionen folgt leicht

$$\cosh'(t) = \sinh(t) \quad \sinh'(t) = \cosh(t) \quad \cosh^2(t) - \sinh^2(t) = 1.$$

Man sieht leicht  $\sinh(t) > 0$  für  $t > 0$ , also ist  $\cosh(t)$  streng monoton wachsend auf  $[0, \infty)$  und hat die Umkehrfunktion  $\text{Arcosh} : [1, \infty) \rightarrow [0, \infty)$ . Wir substituieren nun  $u = \cosh t$ , also  $du = \sinh t dt$ , und erhalten

$$\begin{aligned} A(x) &= \int_0^{\text{Arcosh } x} \sinh^2 t dt \\ &= \frac{1}{4} \int_0^{\text{Arcosh } x} (e^{2t} + e^{-2t} - 2) dt \\ &= \left[ \frac{1}{8}(e^{2t} - e^{-2t}) \right]_{t=0}^{t=\text{Arcosh } x} - \frac{1}{2} \text{Arcosh } x \\ &= \frac{1}{2} \left( \left[ \frac{e^t + e^{-t}}{2} \frac{e^t - e^{-t}}{2} \right]_{t=0}^{t=\text{Arcosh } x} - \text{Arcosh } x \right) \\ &= \frac{1}{2} (x\sqrt{x^2 - 1} - \text{Arcosh } x). \end{aligned}$$

Für rationale Funktionen, also Quotienten von Polynomen, hat man ein spezielles Integrationsverfahren, die Partialbruchzerlegung, die wir hier nur an einem Beispiel vorführen:

**Beispiel 14.7 (Partialbruchzerlegung)** Um das Integral  $\int_{-1/2}^{1/2} \frac{dx}{1-x^2}$  zu berechnen, machen wir den Ansatz

$$\frac{1}{1-x^2} = \frac{1}{(1+x)(1-x)} \stackrel{!}{=} \frac{A}{1-x} + \frac{B}{1+x} = \frac{(A-B)x + (A+B)}{1-x^2}.$$

Der Koeffizientenvergleich ergibt  $A = B = 1/2$ , also folgt

$$\int_{-1/2}^{1/2} \frac{dx}{1-x^2} = \int_{-1/2}^{1/2} \left( \frac{1/2}{1-x} + \frac{1/2}{1+x} \right) dx = \frac{1}{2} \left[ \log \frac{1+x}{1-x} \right]_{x=-1/2}^{x=1/2} = \frac{1}{2} (\log 3 - \log 1/3) = \log 3.$$

Bei der Definition des Riemannsches Integrals ist das Definitionsintervall  $I = [a, b]$  nach Voraussetzung kompakt. Wir wollen ganz kurz erläutern, wie auch unendliche Integrationsintervalle  $-\infty \leq a < b \leq \infty$  und in den Intervallgrenzen unbeschränkte Funktionen im Rahmen des Riemann-Integrals behandelt werden können.

**Definition 14.2 (uneigentliches Riemann-Integral)** Sei  $I = [a, b)$  mit  $a < b \leq \infty$ . Die Funktion  $f : I \rightarrow \mathbb{R}$  sei Riemann-integrierbar auf  $[a, b']$  für alle  $b' < b$ . Im Fall der Existenz des Grenzwerts setzen wir

$$\int_a^b f(y) dy = \lim_{x \nearrow b} \int_a^x f(y) dy,$$

und nennen das Integral konvergent (oder existent).

Für die Konvergenz des uneigentlichen Integrals haben wir das folgende Cauchy Kriterium.

**Satz 14.5** In der Situation von Definition 14.2 ist  $\int_a^b f(x) dx$  genau dann konvergent, wenn es zu jedem  $\varepsilon > 0$  ein  $b' < b$  gibt mit

$$\left| \int_{x_1}^{x_2} f(x) dx \right| < \varepsilon \quad \text{für alle } x_1, x_2 \in (b', b). \quad (14.3)$$

BEWEIS: Setze  $F(x) = \int_a^x f(\xi) d\xi$  für  $x \in [a, b)$ . Existiert das uneigentliche Integral, das heißt  $F(x) \rightarrow S$  mit  $x \nearrow b$ , so gibt es ein  $b' < b$  mit  $|F(x) - S| < \varepsilon/2$  für alle  $x \in (b', b)$ . Es folgt

$$|F(x_1) - F(x_2)| \leq |F(x_1) - S| + |F(x_2) - S| < \varepsilon \quad \text{für } x_{1,2} \in (b', b).$$

Jetzt folgern wir umgekehrt aus (14.3) die Existenz des Grenzwerts  $\lim_{x \nearrow b} F(x)$ . Sei  $x_k \in [a, b)$ ,  $x_k \rightarrow b$ , eine Folge. Zu  $\varepsilon > 0$  sei  $b' < b$  wie in (14.3) gewählt. Dann folgt

$$|F(x_l) - F(x_k)| = \left| \int_{x_k}^{x_l} f(x) dx \right| < \varepsilon \quad \text{für } x_k, x_l \in (b', b).$$

Da die Folge beliebig war, existiert der Grenzwert  $\lim_{x \nearrow b} F(x)$ . □

Hier sind einige Beispiele von uneigentlichen Riemann-Integralen.

### Beispiel 14.8

$$\int_1^\infty x^\alpha dx = \begin{cases} \lim_{R \nearrow \infty} \left[ \frac{x^{\alpha+1}}{\alpha+1} \right]_{x=1}^{x=R} = -\frac{1}{\alpha+1} & \text{für } \alpha < -1, \\ \text{divergent} & \text{für } \alpha \geq -1. \end{cases}$$

$$\int_0^1 x^\alpha dx = \begin{cases} \text{divergent} & \text{für } \alpha \leq -1, \\ \lim_{\varepsilon \searrow 0} \left[ \frac{x^{\alpha+1}}{\alpha+1} \right]_{x=\varepsilon}^{x=1} = \frac{1}{\alpha+1} & \text{für } \alpha > -1. \end{cases}$$

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} = \lim_{x \nearrow 1} \int_0^x \frac{dy}{\sqrt{1-y^2}} = \lim_{x \nearrow 1} \arcsin x = \pi/2.$$

In den Beispielen oben sind die Integranden positiv. Das uneigentliche Integral  $\int_a^b f(x) dx$  heißt absolut konvergent, wenn  $\int_a^b |f(x)| dx$  konvergiert. Aus Satz 14.5 folgt sofort, dass ein absolut konvergentes Integral konvergiert. Umgekehrt impliziert die Konvergenz aber nicht die absolute Konvergenz. Ein simples Beispiel ist das uneigentliche Integral  $\int_1^\infty f(x) dx$ , wobei

$$f(x) = \frac{(-1)^{k-1}}{k} \text{ für } k \leq x < k+1.$$

Für die Existenz des uneigentlichen Integrals auf einem beidseitig offenen Intervall  $(a, b)$  wählt man einen Zwischenpunkt  $c \in (a, b)$  und verlangt die Existenz der Integrale auf  $(a, c]$  und  $[c, b)$ . Das Integral über  $(a, b)$  ergibt sich dann als Summe. Es ist leicht zu sehen, dass diese Definition nicht von der Wahl des Zwischenpunkts  $c$  abhängt: für  $a < c_1 < c_2 < b$  gilt

$$\int_{c_1}^x f(y) dy = \int_{c_1}^{c_2} f(y) dy + \int_{c_2}^x f(y) dy.$$

Mit  $x \nearrow b$  ergibt sich, wobei die Konvergenz auf einer Seite die auf der anderen impliziert,

$$\int_{c_1}^b f(y) dy = \int_{c_1}^{c_2} f(y) dy + \int_{c_2}^b f(y) dy.$$

Analog sieht man

$$\int_a^{c_2} f(y) dy = \int_a^{c_1} f(y) dy + \int_{c_1}^{c_2} f(y) dy,$$

und durch Subtraktion folgt

$$\int_a^{c_1} f(y) dy + \int_{c_1}^b f(y) dy = \int_a^{c_2} f(y) dy + \int_{c_2}^b f(y) dy.$$

Ein Beispiel ist, mit  $c = 0$ ,

$$\int_{-\infty}^\infty \frac{dx}{1+x^2} = 2 \lim_{b \nearrow \infty} \int_0^b \frac{dx}{1+x^2} = 2 \lim_{b \nearrow \infty} [\arctan x]_{x=0}^{x=b} = \pi.$$



## 15 Reihen

Unendliche Summen, sogenannte Reihen, spielen in der Analysis eine wichtige Rolle. Zum Beispiel ist die Eulersche Zahl als die Reihe  $e = \sum_{k=0}^{\infty} 1/k!$  definiert. Die Exponentialfunktion und die trigonometrischen Funktionen können als Reihen definiert beziehungsweise dargestellt werden, das haben wir in den Kapiteln 11 und 12 gesehen. Es ist wichtig, die Konvergenzfrage für Reihen allgemein zu untersuchen, je nach Fragestellung in  $\mathbb{C}$  oder  $\mathbb{R}$ .

**Definition 15.1** Gegeben sei eine Folge  $a_0, a_1, a_2, \dots$  mit  $a_k \in \mathbb{C}$ . Die Folge der Summen  $S_n = \sum_{k=0}^n a_k$ ,  $n \in \mathbb{N}_0$ , nennen wir die Reihe mit Gliedern  $a_k$ . Die Reihe heißt konvergent mit Wert  $S \in \mathbb{C}$ , wenn die Folge  $S_n$  gegen  $S$  konvergiert:

$$\sum_{k=0}^{\infty} a_k := \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k = \lim_{n \rightarrow \infty} S_n = S.$$

$S_n$  wird allgemein als  $n$ -te Partialsumme der Reihe bezeichnet. Oft wird die Reihe kurz mit  $a_0 + a_1 + a_2 + \dots$  angegeben, also durch die ersten Glieder. Leider ist es auch üblich von der Reihe  $\sum_{k=0}^{\infty} a_k$  zu sprechen wobei die ganze Folge  $(S_n)$  gemeint ist, egal ob der Grenzwert existiert oder nicht. Das ist natürlich etwas missverständlich, aber so ist es nun einmal.

**Beispiel 15.1** Die Reihe  $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots$  wird auch mit  $\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$  bezeichnet.

Gemeint ist jeweils die Folge  $(S_n)_{n \in \mathbb{N}}$  mit den Gliedern

$$S_1 = \frac{1}{1 \cdot 2} = \frac{1}{2}, \quad S_2 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} = \frac{2}{3}, \quad S_3 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} = \frac{3}{4}, \dots$$

Für die betrachtete Reihe gilt, für alle  $n \in \mathbb{N}$ ,

$$S_n = \sum_{k=1}^n \left( \frac{1}{k} - \frac{1}{k+1} \right) = \sum_{k=1}^n \frac{1}{k} - \sum_{k=2}^{n+1} \frac{1}{k} = 1 - \frac{1}{n+1}.$$

Also ist die Reihe konvergent mit Wert

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$

Wie hier im Beispiel kann eine Reihe statt bei  $k = 0$  auch bei  $k = 1$  oder einer anderen Zahl starten. Konvergente Reihen können addiert und mit komplexen Zahlen multipliziert werden. Genauer: sind  $\sum_{k=0}^{\infty} a_k$  und  $\sum_{k=0}^{\infty} b_k$  konvergente Reihen und  $\lambda, \mu \in \mathbb{C}$ , so folgt mit der Regel für Folgen, Satz 4.3(a),

$$\begin{aligned} \sum_{k=0}^n (\lambda a_k + \mu b_k) &= \lambda \sum_{k=0}^n a_k + \mu \sum_{k=0}^n b_k \rightarrow \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k, \quad \text{also} \\ \sum_{k=0}^{\infty} (\lambda a_k + \mu b_k) &= \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k. \end{aligned}$$

Werden endlich viele Glieder einer Reihe geändert, hinzugefügt oder weggelassen, so wird dadurch die Konvergenz oder Divergenz nicht beeinflusst. Allerdings wird sich der Wert der Reihe natürlich entsprechend ändern. Zum Beispiel haben wir für  $n \in \mathbb{N}_0$  und  $m \geq n$

$$\sum_{k=n}^m a_k = \sum_{k=0}^m a_k - \sum_{k=0}^{n-1} a_k,$$

und mit  $m \rightarrow \infty$  folgt, falls die Reihe  $\sum_{k=0}^{\infty} a_k$  konvergiert,

$$\sum_{k=n}^{\infty} a_k = \sum_{k=0}^{\infty} a_k - \sum_{k=0}^{n-1} a_k. \quad (15.1)$$

Jede Folge  $b_0, b_1, \dots$  komplexer Zahlen kann auch als Reihe  $\sum_{k=0}^n a_k$  aufgefasst werden, denn mit  $a_0 = b_0$  und  $a_k = b_k - b_{k-1}$  für  $k \geq 1$  gilt

$$\sum_{k=0}^n a_k = b_0 + \sum_{k=1}^n (b_k - b_{k-1}) = b_0 + \sum_{k=1}^n b_k - \sum_{k=0}^{n-1} b_k = b_n.$$

Der Unterschied zwischen Folgen und Reihen besteht in der Sichtweise, Reihen werden durch die Zuwächse  $a_n$  beschrieben. Es stellen sich zwei Fragen:

- Wie kann ich den Gliedern  $a_n$  ansehen, ob die Reihe konvergiert bzw. divergiert?
- Im Fall der Konvergenz: welchen Wert hat die Reihe?

Bei der zweiten Frage ist zum Beispiel gemeint, ob eine bereits definierte Zahl wie  $\sqrt{2}$ ,  $e$ ,  $\pi$ , ... als Grenzwert einer Reihe dargestellt werden kann. Im Folgenden steht aber die erste Frage im Zentrum des Interesses. Dabei ist das nächste Beispiel fundamental.

**Beispiel 15.2 (Geometrische Reihe)** Die geometrische Reihe  $\sum_{k=0}^{\infty} z^k$  mit  $z \in \mathbb{C}$  konvergiert genau für  $|z| < 1$ . Der Konvergenzbeweis ist wie in Beispiel 4.6, und zwar gilt

$$\sum_{k=0}^n z^k = \frac{1 - z^{n+1}}{1 - z} \rightarrow \frac{1}{1 - z} \quad \text{mit } n \rightarrow \infty.$$

Dagegen gilt für  $|z| \geq 1$  mit  $S_n = \sum_{k=0}^n z^k$

$$|S_{n+1} - S_n| = |z^{n+1}| = |z|^{n+1} \geq 1,$$

so dass  $(S_n)_{n \in \mathbb{N}}$  keine Cauchyfolge sein kann.

**Beispiel 15.3 (Unendliche Dezimalbrüche)** Ist  $(k_j)_{j \in \mathbb{N}_0}$  eine Folge von Ziffern  $k_j \in \{0, 1, \dots, 9\}$ , so ist die Reihe  $\sum_{j=0}^{\infty} k_j 10^{-j}$  konvergent, vgl. Beispiel 5.4.

**Beispiel 15.4 (Harmonische Reihe)** Die harmonische Reihe  $\sum_{k=1}^{\infty} 1/k$  ist bestimmt divergent gegen  $+\infty$ . Dies zeigen wir, indem wir wie folgt Klammern setzen:

$$\underbrace{\left(\frac{1}{1}\right)}_{\geq 1 \cdot \frac{1}{2}} + \underbrace{\left(\frac{1}{2} + \frac{1}{3}\right)}_{\geq 2 \cdot \frac{1}{4}} + \underbrace{\left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}_{\geq 4 \cdot \frac{1}{8}} + \underbrace{\left(\frac{1}{8} + \frac{1}{9} + \dots + \frac{1}{15}\right)}_{\geq 8 \cdot \frac{1}{16}} + \dots$$

Allgemein enthält der Abschnitt  $2^m \leq k < 2^{m+1}$ , mit  $m \in \mathbb{N}_0$ ,  $2^m$  Summanden  $1/k$ , die jeweils größer sind als  $2^{-(m+1)}$ . Also ist der Beitrag dieses Abschnitts mindestens  $2^m \cdot 2^{-(m+1)} = 1/2$ , und die Reihe konvergiert gegen  $+\infty$ .

Nach diesen ersten Beispielen nun zur allgemeinen Konvergenzfrage für Reihen. Wir beginnen mit einem notwendigen Kriterium.

**Satz 15.1 (Nullfolgentest)** *Ist die Reihe  $\sum_{k=0}^{\infty} a_k$  konvergent, so folgt  $\lim_{k \rightarrow \infty} a_k = 0$ .*

BEWEIS: Nach Voraussetzung ist die Folge der Partialsummen  $S_n = \sum_{k=0}^n a_k$  konvergent mit Grenzwert  $S \in \mathbb{C}$ , also folgt  $a_n = S_n - S_{n-1} \rightarrow S - S = 0$ .  $\square$

Als nächstes formulieren wir unsere Konvergenzkriterien für Folgen neu in der Situation von Reihen. Die Cauchyfolgeneigenschaft sieht wie folgt aus.

**Satz 15.2 (Konvergenzkriterium von Cauchy)** *Eine Reihe  $\sum_{k=0}^{\infty} a_k$  konvergiert dann und nur dann, wenn es zu jedem  $\varepsilon > 0$  ein  $K \in \mathbb{R}$  gibt, so dass gilt:*

$$\left| \sum_{k=n}^m a_k \right| < \varepsilon \quad \text{für alle } n, m \in \mathbb{N}_0 \text{ mit } m \geq n > K.$$

BEWEIS: Satz 5.2.  $\square$

**Satz 15.3 (Reihen mit Gliedern  $a_k \geq 0$ )** *Eine reelle Reihe  $\sum_{k=0}^{\infty} a_k$  mit  $a_k \geq 0$  für alle  $k$  konvergiert genau dann, wenn die Folge der Partialsummen  $S_n = \sum_{k=0}^n a_k$  nach oben beschränkt ist.*

BEWEIS: Da  $a_k \geq 0$ , ist die Folge  $(S_n)$  der Partialsummen monoton wachsend. Die Behauptung folgt aus Satz 5.3 und Satz 4.2.  $\square$

**Beispiel 15.5** Für  $s > 1$  ist die Reihe  $\sum_{k=1}^{\infty} k^{-s}$  konvergent. Wir schätzen dazu die Abschnitte  $2^m \leq k < 2^{m+1}$ ,  $m \in \mathbb{N}_0$ , nach oben ab. Die Funktion  $f(x) = x^{-s} = e^{-s \log x}$  ist monoton fallend für  $x \geq 1$ , also gilt

$$\sum_{2^m \leq k < 2^{m+1}} k^{-s} \leq 2^m \cdot (2^m)^{-s} = (2^{1-s})^m.$$

Für die Summe der Abschnitte mit Nummern  $m = 0, \dots, M$  folgt

$$\sum_{1 \leq k < 2^{M+1}} k^{-s} = \sum_{m=0}^M \left( \sum_{2^m \leq k < 2^{m+1}} k^{-s} \right) \leq \sum_{m=0}^M (2^{1-s})^m \leq \frac{1}{1 - 2^{1-s}}.$$

Beachte dabei  $2^{1-s} < 1$  für  $s > 1$ , so dass die geometrische Reihe rechts konvergiert, siehe Beispiel 4.6. Da die Folge der Partialsummen  $S_n$  monoton wachsend ist, gilt die Abschätzung für alle Partialsummen  $S_n$ , und die Reihe konvergiert nach Satz 15.3. Die Zahl  $s \in (0, \infty)$  ist ein Parameter, wir haben also die Funktion

$$\zeta : (1, \infty) \rightarrow \mathbb{R}, \zeta(s) = \sum_{k=1}^{\infty} k^{-s}.$$

Dies ist die Riemannsche Zetafunktion, sie spielt bei der Untersuchung der Verteilung der Primzahlen eine fundamentale Rolle.

Die Reihe  $\sum_{k=1}^{\infty} (-1)^{k-1}/k = 1 - 1/2 + 1/3 - + \dots$  ist ein Beispiel für eine sogenannte alternierende Reihe. Während die entsprechende Reihe mit nur positiven Vorzeichen nicht konvergiert – es ist die harmonische Reihe aus Beispiel 15.4 – ist die Reihe mit dem Vorzeichenwechsel konvergent. Dies ergibt sich aus dem nächsten Satz, bei dessen Beweis wieder Monotonieargumente eine wesentliche Rolle spielen.

**Satz 15.4 (Leibniz-Kriterium für alternierende Reihen)** Sei  $a_k, k \in \mathbb{N}_0$ , eine reelle, monoton fallende Nullfolge (also insbesondere  $a_k \geq 0$ ). Dann ist die Reihe  $\sum_{k=0}^{\infty} (-1)^k a_k$  konvergent. Außerdem gilt für alle  $n \in \mathbb{N}_0$  die Abschätzung

$$0 \leq (-1)^n \sum_{k=n}^{\infty} (-1)^k a_k \leq a_n. \quad (15.2)$$

BEWEIS: Wir betrachten die Partialsummen  $S_n = \sum_{k=0}^n (-1)^k a_k$ . Für alle  $n \in \mathbb{N}_0$  gilt:

$$\begin{aligned} S_{2n+2} - S_{2n} &= (-1)^{2n+2} a_{2n+2} + (-1)^{2n+1} a_{2n+1} = a_{2n+2} - a_{2n+1} \leq 0, \\ S_{2n+3} - S_{2n+1} &= (-1)^{2n+3} a_{2n+3} + (-1)^{2n+2} a_{2n+2} = a_{2n+2} - a_{2n+3} \geq 0, \\ S_{2n+1} - S_{2n} &= (-1)^{2n+1} a_{2n+1} = -a_{2n+1} \leq 0. \end{aligned}$$

Also ist  $(S_{2n})_{n \in \mathbb{N}_0}$  monoton fallend, und  $(S_{2n+1})_{n \in \mathbb{N}_0}$  monoton wachsend. Weiter haben wir für alle  $n \in \mathbb{N}_0$  die untere bzw. obere Schranke

$$S_{2n} \geq S_{2n+1} \geq S_1 \quad \text{und} \quad S_{2n+1} \leq S_{2n} \leq S_0.$$

Nach Satz 5.3 sind die Folgen  $S_{2n}$  und  $S_{2n+1}$  konvergent. Aber  $S_{2n+1} - S_{2n} = -a_{2n+1} \rightarrow 0$  nach Voraussetzung. Also konvergieren beide Folgen, und damit die gesamte Folge, gegen denselben Grenzwert  $S \in \mathbb{R}$ . Zusätzlich sehen wir

$$0 \leq a_1 - a_0 = S_1 \leq \lim_{n \rightarrow \infty} S_{2n} = S = \lim_{n \rightarrow \infty} S_{2n+1} \leq S_0 = a_0,$$

das heißt Abschätzung (15.2) gilt im Fall  $n = 0$ . Für beliebiges  $n$  wenden wir das an auf  $b_l = a_{n+l}$ . Das ist ebenfalls eine monoton fallende Nullfolge, also gilt

$$0 \leq \sum_{l=0}^{\infty} (-1)^l b_l \leq b_0 = a_n.$$

Abschätzung (15.2) folgt nun wegen

$$\sum_{l=0}^{\infty} (-1)^l b_l = (-1)^n \sum_{l=0}^{\infty} (-1)^{n+l} a_{n+l} = (-1)^n \sum_{k=n}^{\infty} (-1)^k a_k.$$

□

Bei der alternierenden Reihe  $1 - 1/2 + 1/3 - + \dots$  stößt man auf Merkwürdigkeiten, wenn man die Summationsreihenfolge ändert. Während die Ausgangsreihe konvergent ist, ist die durch Umordnung entstehende Reihe

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{5} - \frac{1}{4}\right) + \left(\frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13} - \frac{1}{6}\right) + \dots$$

bestimmt divergent gegen  $+\infty$ , denn die Summe der positiven Zahlen in der  $m$ -ten Klammer ist mindestens  $2^{m-1} \cdot 2^{-(m+1)} = 1/4$ . Dies ist ein interessantes Phänomen, jedoch sind wir in erster Linie an Reihen interessiert, deren Konvergenz stabiler ist. Dabei ist der folgende Begriff zentral.

**Definition 15.2** Die Reihe  $\sum_{k=0}^{\infty} a_k$  mit  $a_k \in \mathbb{C}$  heißt *absolut konvergent*, wenn die Reihe  $\sum_{k=0}^{\infty} |a_k|$  konvergiert, das heißt  $\sum_{k=0}^{\infty} |a_k| < \infty$ .

**Satz 15.5 (absolut konvergent  $\Rightarrow$  konvergent)** Wenn die Reihe  $\sum_{k=0}^{\infty} a_k$  absolut konvergiert, so ist sie konvergent und es gilt

$$\left| \sum_{k=0}^{\infty} a_k \right| \leq \sum_{k=0}^{\infty} |a_k|.$$

BEWEIS: Die Dreiecksungleichung besagt

$$\left| \sum_{k=n}^m a_k \right| \leq \sum_{k=n}^m |a_k| \quad \text{für } m \geq n \geq 0.$$

Aus dem Cauchy Kriterium für  $\sum_{k=0}^{\infty} |a_k|$  folgt deshalb das Cauchy Kriterium für  $\sum_{k=0}^{\infty} a_k$ , das heißt die Reihe konvergiert nach Satz 15.2. Die Abschätzung folgt, indem wir in der Dreiecksungleichung  $n = 0$  setzen und  $m \rightarrow \infty$  gehen lassen.  $\square$

Der folgende Satz fasst drei wichtige Kriterien für die absolute Konvergenz zusammen.

**Satz 15.6 (Tests für absolute Konvergenz)** Sei  $\sum_{k=0}^{\infty} a_k$  eine Reihe mit  $a_k \in \mathbb{C}$ . Ist eine der drei folgenden Bedingungen erfüllt, so ist  $\sum_{k=0}^{\infty} a_k$  absolut konvergent:

- (a) Majorantenkriterium ( $M$ -Test): Es gilt  $|a_k| \leq c_k \in [0, \infty)$  mit  $\sum_{k=0}^{\infty} c_k < \infty$ .
- (b) Quotientenkriterium: Es gibt ein  $\theta \in [0, 1)$  und ein  $n \in \mathbb{N}$  mit

$$\frac{|a_{k+1}|}{|a_k|} \leq \theta \quad \text{für alle } k \geq n \quad (\text{wobei } a_k \neq 0 \text{ für } k \geq n).$$

- (c) Wurzelkriterium: Es gibt ein  $\theta \in [0, 1)$  und ein  $n \in \mathbb{N}$  mit  $\sqrt[k]{|a_k|} \leq \theta$  für alle  $k \geq n$ .

Umgekehrt ist die Reihe divergent, wenn für ein  $n \in \mathbb{N}$  gilt:

$$\frac{|a_{k+1}|}{|a_k|} \geq 1 \quad \text{oder} \quad \sqrt[k]{|a_k|} \geq 1 \quad \text{für alle } k \geq n.$$

Für die harmonische Reihe  $\sum_{k=1}^{\infty} 1/k$  haben wir

$$\frac{|a_{k+1}|}{|a_k|} = \frac{k}{k+1} < 1 \quad \text{für alle } k \quad \text{und} \quad \lim_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} = 1.$$

Also reicht es *nicht*, in (b) nur die Ungleichung  $|a_{k+1}|/|a_k| < 1$  vorauszusetzen. Andererseits folgt aus  $\lim_{k \rightarrow \infty} |a_{k+1}|/|a_k| = 1$  auch nicht notwendig Divergenz: die Reihe  $\sum_{k=1}^{\infty} 1/k^2$  hat  $|a_{k+1}|/|a_k| = (k/k+1)^2 \rightarrow 1$ , aber sie konvergiert nach Beispiel 15.5. Eine ganz analoge Diskussion gilt für das Wurzelkriterium. Die Bedingungen (b) beziehungsweise (c) sind äquivalent zu den folgenden:

$$\limsup_{k \rightarrow \infty} \frac{|a_{k+1}|}{|a_k|} < 1 \quad \text{bzw.} \quad \limsup_{k \rightarrow \infty} \sqrt[k]{|a_k|} < 1.$$

BEWEIS DES SATZES: (a) folgt aus Satz 15.3, denn für alle  $n \in \mathbb{N}$  hat man die obere Schranke

$$\sum_{k=0}^n |a_k| \leq \sum_{k=0}^n c_k \leq \sum_{k=0}^{\infty} c_k < \infty.$$

Voraussetzung (b) liefert per Induktion

$$|a_k| \leq \theta^{k-n} |a_n| \quad \text{für } k \geq n.$$

Da die geometrische Reihe wegen  $0 \leq \theta < 1$  nach Beispiel 4.6 konvergiert, folgt die Behauptung aus (a) und wir erhalten außerdem die Abschätzung

$$\sum_{k=n}^{\infty} |a_k| \leq \frac{|a_n|}{1-\theta}. \quad (15.3)$$

Unter der Voraussetzung (c) gilt

$$|a_k| \leq \theta^k \quad \text{für } k \geq n.$$

Wieder folgt die Behauptung durch M-Test mit der geometrischen Reihe. Die Abschätzung lautet hier

$$\sum_{k=n}^{\infty} |a_k| \leq \frac{\theta^n}{1-\theta}. \quad (15.4)$$

Die Divergenzaussagen folgen unmittelbar aus dem Nullfolgentest, Satz 15.1.  $\square$

Die folgenden zwei Sätze gehören aber zum Standardrepertoire, auch wenn wir sie in unserer Vorlesung nicht gebraucht haben. Erstens geht es um die Multiplikation von zwei Reihen  $\sum_{k=0}^{\infty} a_k$  und  $\sum_{k=0}^{\infty} b_k$ . Wir schreiben die Produkte  $a_k b_l$  mit  $k, l \in \mathbb{N}_0$  in einem Schema auf, so dass  $a_k b_l$  in der  $k$ -ten Zeile und  $l$ -ten Spalte steht. Es ist dann naheliegend, erst die  $n+1$  Produkte  $a_k b_l$  in jeder Diagonale  $k+l=n$  zu addieren, und dann die Konvergenz der resultierenden Reihe zu studieren. Im folgenden schreiben wir kurz  $\sum_{k+l=n}$  für die Summe über alle Paare  $k, l \in \mathbb{N}_0$  mit  $k+l=n$ .

**Satz 15.7 (Cauchyprodukt)** Die Reihen  $\sum_{k=0}^{\infty} a_k$  und  $\sum_{k=0}^{\infty} b_k$  seien absolut konvergent. Dann ist auch die Reihe

$$\sum_{k=0}^{\infty} c_n \quad \text{mit} \quad c_n = \sum_{k+l=n} a_k b_l = \sum_{k=0}^n a_k b_{n-k}$$

absolut konvergent, und es gilt

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty} \left( \sum_{k+l=n} a_k b_l \right) = \left( \sum_{k=0}^{\infty} a_k \right) \cdot \left( \sum_{k=0}^{\infty} b_k \right).$$

BEWEIS: Wir setzen für  $N \in \mathbb{N}_0$

$$A_N := \sum_{k=0}^N |a_k| \rightarrow \sum_{k=0}^{\infty} |a_k| =: A \quad \text{und} \quad B_N := \sum_{k=0}^N |b_k| \rightarrow \sum_{k=0}^{\infty} |b_k| =: B,$$

wobei nach Voraussetzung  $A, B < \infty$ . Es folgt

$$\sum_{n=0}^N |c_n| \leq \sum_{k+l \leq N} |a_k| |b_l| \leq \sum_{k, l \leq N} |a_k| |b_l| = A_N B_N \leq AB < \infty.$$

Nach Satz 15.3 ist  $\sum_{n=0}^{\infty} c_n$  absolut konvergent und insbesondere konvergent (Satz 15.5). Um den Grenzwert zu identifizieren, reicht es also die geraden Partialsummen  $\sum_{n=0}^{2N} c_n$  zu betrachten. Wir berechnen

$$\left| \left( \sum_{k=0}^{2N} a_k \right) \cdot \left( \sum_{l=0}^{2N} b_l \right) - \sum_{n=0}^{2N} c_n \right| = \left| \sum_{k, l \leq 2N} a_k b_l - \sum_{k+l \leq 2N} a_k b_l \right| \leq \sum_{k, l \leq 2N, \max(k, l) > N} |a_k| |b_l|,$$

Die rechte Seite ist gleich  $A_{2N} B_{2N} - A_N B_N$ , konvergiert also gegen Null für  $N \rightarrow \infty$ . Für die letzte Abschätzung ist es hilfreich, die Indexbereiche zu skizzieren. Jedenfalls ist damit die gewünschte Formel für das Cauchyprodukt bewiesen.  $\square$

Oft wird mit dem Cauchyprodukts die Funktionalgleichung der Exponentialfunktion bewiesen. Wir haben die Charakterisierung durch das Anfangswertproblem benutzt, siehe Satz 11.3(c).

Als zweites kommen wir zur Frage der Umordnung zurück und zeigen, dass absolut konvergente Reihen beliebig umgeordnet werden können.

**Satz 15.8 (Umordnungssatz)** *Sei  $\tau : \mathbb{N} \rightarrow \mathbb{N}$  bijektiv. Falls  $\sum_{k=1}^{\infty} a_k$  absolut konvergent ist, so konvergiert auch die Reihe  $\sum_{i=1}^{\infty} a_{\tau(i)}$  absolut und hat denselben Grenzwert.*

BEWEIS: Für  $K \geq \max\{\tau(i) : 1 \leq i \leq m\}$  ist

$$\sum_{i=1}^m |a_{\tau(i)}| \leq \sum_{k=1}^K |a_k| \leq \sum_{k=1}^{\infty} |a_k| < \infty.$$

Nach dem Majorantenkriterium konvergiert die umgeordnete Reihe absolut. Sei nun  $S = \sum_{k=1}^{\infty} a_k$ . Für gegebenes  $K \in \mathbb{N}$  und  $m \geq \max\{\tau^{-1}(k) : 1 \leq k \leq K\}$ , bzw. äquivalent dazu  $\{\tau(1), \dots, \tau(m)\} \supset \{1, \dots, K\}$ , gilt

$$\left| S - \sum_{i=1}^m a_{\tau(i)} \right| \leq \left| S - \sum_{k=1}^K a_k \right| + \left| \sum_{k=1}^K a_k - \sum_{i=1}^m a_{\tau(i)} \right| \leq \left| S - \sum_{k=1}^K a_k \right| + \sum_{k=K+1}^{\infty} |a_k|.$$

Die Behauptung folgt, indem wir erst  $m \rightarrow \infty$  und dann  $K \rightarrow \infty$  gehen lassen.  $\square$



## 16 Potenzreihen

Wir haben bereits die Reihendarstellungen der Exponentialfunktion und der Funktionen Cosinus bzw. Sinus kennengelernt. Jetzt wollen wir Reihen dieses Typs allgemein studieren.

**Definition 16.1** Eine Potenzreihe mit Koeffizienten  $a_k \in \mathbb{C}$  ist eine Funktion der Form

$$P(z) = \sum_{k=0}^{\infty} a_k z^k \quad \text{für } z \in \mathbb{C}.$$

Die Definition lässt offen für welche  $z \in \mathbb{C}$  die Reihe tatsächlich konvergiert. Es ist klar dass  $P(0) = a_0$ , aber  $z = 0$  könnte der einzige Punkt sein wo  $P(z)$  konvergiert. Ein Beispiel ist  $\sum_{k=0}^{\infty} k^k z^k$ , die Reihe divergiert für alle  $z \neq 0$  nach dem Nullfolgentest. Damit die Potenzreihe von Interesse ist sollte sie aber für Punkte  $z$  in einem gewissen Gebiet konvergieren und damit dort eine Funktion definieren. Wir beginnen damit, dieses Konvergenzgebiet zu charakterisieren.

**Lemma 16.1 (von Abel)** Es gebe  $\varrho > 0$  und  $M \in [0, \infty)$  mit  $|a_k| \varrho^k \leq M$  für alle  $k \in \mathbb{N}_0$ . Dann ist  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  absolut konvergent für  $|z| < \varrho$ , und für alle  $k \in \mathbb{N}_0$  gilt

$$|P(z) - P_k(z)| \leq \sum_{\ell=k+1}^{\infty} |a_\ell| |z|^\ell \leq \frac{M}{1 - \frac{|z|}{\varrho}} \left(\frac{|z|}{\varrho}\right)^{k+1}. \quad (16.1)$$

Hier ist  $P_k(z) = \sum_{\ell=0}^k a_\ell z^\ell$  die  $k$ -te Partialsumme.

BEWEIS: Es gilt  $|a_k z^k| = |a_k| \varrho^k (|z|/\varrho)^k \leq M (|z|/\varrho)^k$ . Für  $|z| < \varrho$  ist die geometrische Reihe eine Majorante, genauer gilt

$$|P(z) - P_k(z)| \leq \sum_{\ell=k+1}^{\infty} |a_\ell| |z|^\ell \leq M \sum_{\ell=k+1}^{\infty} \left(\frac{|z|}{\varrho}\right)^\ell = \frac{M}{1 - \frac{|z|}{\varrho}} \left(\frac{|z|}{\varrho}\right)^{k+1}.$$

□

Das Konvergenzgebiet einer Potenzreihe sieht allgemein wie folgt aus.

**Satz 16.1 (vom Konvergenzradius)** Zu jeder Potenzreihe  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  gibt es genau ein  $R \in [0, \infty]$ , den Konvergenzradius, mit folgender Eigenschaft:

$$P(z) \text{ ist } \begin{cases} \text{absolut konvergent} & \text{für } |z| < R, \\ \text{divergent} & \text{für } |z| > R. \end{cases}$$

BEWEIS: Die Eindeutigkeit von  $R$  ist klar. Zur Existenz setze

$$R = \sup\{|z| : P(z) \text{ konvergiert}\} \in [0, \infty].$$

Ist  $|z| < R$ , so gibt es nach Definition ein  $z_0 \in \mathbb{C}$  mit  $|z| < |z_0| \leq R$ , so dass  $P(z_0)$  konvergiert. Der Nullfolgentest liefert  $a_k z_0^k \rightarrow 0$ , es gibt also ein  $M \in [0, \infty)$  mit  $|a_k| |z_0|^k \leq M$  für alle  $k$ . Nach dem Lemma von Abel, Lemma 16.1 mit  $\varrho = |z_0|$ , ist  $P(z)$  absolut konvergent. Andererseits ist die Reihe divergent für  $|z| > R$  nach Definition von  $R$ . □

**Beispiel 16.1** Für die Exponentialreihe  $\sum_{k=0}^{\infty} z^k/k!$  gilt, außer im trivialen Fall  $z = 0$ ,

$$\frac{|z^{k+1}/(k+1)!|}{|z^k/k!|} = \frac{|z|}{k+1} \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Nach dem Quotientenkriterium, siehe Satz 15.6, konvergiert die Reihe absolut für alle  $z \in \mathbb{C}$ . Der Konvergenzradius der Reihe ist  $R = \infty$ .

**Beispiel 16.2** Die Binomialreihe zum Parameter  $\alpha \in \mathbb{C}$  lautet

$$B_{\alpha}(z) = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k = 1 + \alpha z + \frac{\alpha(\alpha-1)}{2} z^2 + \dots$$

Für  $\alpha \in \mathbb{N}_0$  bricht die Reihe nach  $k = \alpha$  ab und es gilt  $B_{\alpha}(z) = (1+z)^{\alpha}$  nach Binomischer Formel. Sei nun  $\alpha \notin \mathbb{N}_0$ , für  $z \neq 0$  ist dann  $a_k := \binom{\alpha}{k} z^k \neq 0$  für alle  $k \in \mathbb{N}_0$  und es gilt

$$\left| \frac{a_{k+1}}{a_k} \right| = \frac{|\alpha - k|}{k+1} |z| \rightarrow |z| \quad \text{mit } k \rightarrow \infty.$$

Nach dem Quotientenkriterium konvergiert die Reihe für  $|z| < 1$  und divergiert für  $|z| > 1$ , das heißt der Konvergenzradius ist  $R = 1$ .

Wir interessieren uns nun dafür, wie die Funktionen  $P_k(z)$  gegen die Funktion  $P(z)$  konvergieren. Lässt sich zum Beispiel die Stetigkeit der  $P_k(z)$  auf die Grenzfunktion  $P(z)$  übertragen? Das ist ein generelles und grundlegendes Problem: sei  $f_k : D \rightarrow \mathbb{R}^m$  eine Folge von Funktionen auf  $D \subset \mathbb{R}^n$ , die punktweise gegen eine Grenzfunktion  $f : D \rightarrow \mathbb{R}^m$  konvergiert:

$$f(x) = \lim_{k \rightarrow \infty} f_k(x) \quad \text{für alle } x \in D.$$

Folgt die Stetigkeit von  $f$  aus der Stetigkeit der  $f_k$ , unter geeigneten Voraussetzungen? Das folgende Beispiel zeigt, dass die punktweise Konvergenz alleine im allgemeinen nicht die Stetigkeit der Grenzfunktion garantiert.

**Beispiel 16.3** Betrachte die stetigen Funktionen  $f_k : [0, 1] \rightarrow \mathbb{R}$ ,

$$f_k(x) = \begin{cases} 1 - kx & \text{für } 0 \leq x < \frac{1}{k}, \\ 0 & \text{sonst.} \end{cases}$$

Die Folge konvergiert punktweise gegen  $f : [0, 1] \rightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} 1 & \text{für } x = 0, \\ 0 & \text{sonst.} \end{cases}$$

Die Grenzfunktion ist im Punkt  $x = 0$  nicht stetig.

Na gut, das hätten wir eigentlich auch nicht erwartet. Solche Probleme kennen wir schon von der Vertauschbarkeit des Integrals mit der Konvergenz.

**Satz 16.2 (Gleichmäßige Konvergenz und Stetigkeit)** Seien  $f_k : D \rightarrow \mathbb{R}^m$ ,  $k \in \mathbb{N}$ , stetige Funktionen auf  $D \subset \mathbb{R}^n$ , die gleichmäßig gegen  $f : D \rightarrow \mathbb{R}^m$  konvergieren, also

$$\|f_k - f\|_D \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Dann ist  $f$  ebenfalls stetig auf  $D$ .

BEWEIS: Sei  $x_0 \in D$  gegeben. Für  $x \in D$  beliebig und  $k \in \mathbb{N}$  gilt

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_k(x)| + |f_k(x) - f_k(x_0)| + |f_k(x_0) - f(x_0)| \\ &\leq |f_k(x) - f_k(x_0)| + 2\|f - f_k\|_D. \end{aligned}$$

Für  $x \rightarrow x_0$  ergibt sich, da  $f_k$  stetig ist,

$$\limsup_{x \rightarrow x_0} |f(x) - f(x_0)| \leq 2\|f - f_k\|_D.$$

Mit  $k \rightarrow \infty$  folgt  $\limsup_{x \rightarrow x_0} |f(x) - f(x_0)| = 0$ , die Stetigkeit von  $f$  in  $x_0$ .  $\square$

**Satz 16.3 (Stetigkeit von Potenzreihen)** Sei  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  eine Potenzreihe mit Konvergenzradius  $R > 0$ . Dann ist  $P(z)$  stetig auf  $B_R(0) = \{z \in \mathbb{C} : |z| < R\}$ .

BEWEIS: Wir zeigen, dass die  $P_k(z)$  auf jeder Kreisscheibe  $B_r(0)$  mit  $r < R$  gleichmäßig konvergieren. Die Behauptung folgt dann aus Satz 16.2. Sei  $\varrho \in (r, R)$  beliebig gewählt. Da  $P(\varrho)$  konvergiert, gibt es ein  $M \in [0, \infty)$  mit  $|a_k| \varrho^k \leq M$  für alle  $k$ . Nach Lemma 16.1 folgt

$$\|P - P_k\|_{B_r(0)} \leq \frac{M}{1 - \frac{r}{\varrho}} \left(\frac{r}{\varrho}\right)^{k+1} \rightarrow 0 \quad \text{mit } k \rightarrow \infty.$$

Damit ist der Satz bewiesen.  $\square$

Nun zur Frage der Differenzierbarkeit der Grenzfunktion, wobei wir uns hier in Analysis 1 auf Funktionen  $f : I \rightarrow \mathbb{R}^m$  einer Variablen beschränken.

**Beispiel 16.4** Betrachte für  $\varepsilon > 0$  die Funktionen  $f_\varepsilon \in C^1(\mathbb{R})$ ,  $f_\varepsilon(x) = (\varepsilon^2 + x^2)^{1/2}$ . Es gilt

$$|x| \leq f_\varepsilon(x) \leq \varepsilon + |x| \quad \text{für alle } x \in \mathbb{R},$$

also konvergieren die  $f_\varepsilon$  gleichmäßig gegen  $f(x) = |x|$ . In  $x = 0$  ist  $f$  nicht differenzierbar.

**Satz 16.4 (Vertauschung von Konvergenz und Ableitung)** Seien  $f_k \in C^1(I, \mathbb{R}^m)$ ,  $I = (a, b)$ , punktweise konvergent gegen  $f : I \rightarrow \mathbb{R}^m$ . Konvergiert die Folge  $f'_k$  gleichmäßig gegen ein  $g : I \rightarrow \mathbb{R}^m$ , also

$$\|f'_k - g\|_I \rightarrow 0 \quad \text{mit } k \rightarrow \infty,$$

so folgt  $f' = g$ , und  $f \in C^1(I, \mathbb{R}^m)$ .

BEWEIS: Wähle  $x_0 \in I$ . Dann besagt der Hauptsatz der Differential- und Integralrechnung

$$f_k(x) = f_k(x_0) + \int_{x_0}^x f'_k(y) dy \quad \text{für alle } x \in I.$$

Nach Analysis 1, Satz 13.3, vertauscht das Integral mit  $f'_k \rightarrow g$ , also folgt mit  $k \rightarrow \infty$

$$f(x) = f(x_0) + \int_{x_0}^x g(y) dy \quad \text{für alle } x \in I.$$

Nach Satz 16.2 ist  $g$  stetig. Es folgt  $f' = g$  aus dem Hauptsatz.  $\square$

Um den Satz auf Potenzreihen anzuwenden, brauchen wir folgende Hilfsaussage.

**Lemma 16.2** Sei  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  eine Potenzreihe mit Konvergenzradius  $R \in [0, \infty]$ . Dann hat die formal differenzierte Reihe

$$Q(z) = \sum_{k=1}^{\infty} k a_k z^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} z^k$$

denselben Konvergenzradius  $R$ .

BEWEIS:  $Q(z)$  hat höchstens den Konvergenzradius von  $P(z)$ , denn  $|a_k z^k| \leq |k a_k z^{k-1}| \cdot |z|$  für  $k \geq 1$ . Für  $\varrho \in (0, R)$  ist  $P(\varrho)$  konvergent, also gilt  $|a_k| \varrho^k \leq M$  für ein  $M \in [0, \infty)$ . Es folgt für alle  $z \in B_\varrho(0)$

$$|k a_k z^{k-1}| = k |a_k| \varrho^k \frac{|z|^{k-1}}{\varrho^k} \leq \frac{k M |z|^{k-1}}{\varrho^k}.$$

Die rechte Reihe konvergiert aber nach dem Quotientenkriterium, denn

$$\frac{(k+1)|z|^k}{\varrho^{k+1}} \left( \frac{k|z|^{k-1}}{\varrho^k} \right)^{-1} = \frac{(k+1)|z|}{k\varrho} \rightarrow \frac{|z|}{\varrho} < 1 \quad \text{mit } k \rightarrow \infty.$$

Also ist  $Q(z)$  für jedes  $z \in B_R(0)$  absolut konvergent.  $\square$

**Satz 16.5 (Differenzierbarkeit von Potenzreihen)** Die Potenzreihe  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  mit  $a_k \in \mathbb{C}$  habe den Konvergenzradius  $R > 0$ . Dann ist die Funktion

$$P : (-R, R) \rightarrow \mathbb{C}, P(x) = \sum_{k=0}^{\infty} a_k x^k$$

stetig differenzierbar, und ihre Ableitung ergibt sich durch gliedweise Differentiation:

$$P'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1} = \sum_{k=0}^{\infty} (k+1) a_{k+1} x^k \quad \text{für alle } x \in (-R, R). \quad (16.2)$$

*Hinweis.* Durch Induktion folgt  $P \in C^\infty((-R, R), \mathbb{C})$ .

BEWEIS: Sei  $\varrho \in (0, R)$ . Nach Beweis von Satz 16.3 konvergieren die  $P_k(x)$  auf  $(-\varrho, \varrho)$  gleichmäßig gegen  $P(x)$ . Nach Lemma 16.2 konvergieren die  $P'_k(x)$  auf  $(-\varrho, \varrho)$  ebenfalls gleichmäßig gegen  $Q(x)$ , die gliedweise differenzierte Reihe. Satz 16.4 impliziert  $P' = Q$  auf  $(-\varrho, \varrho)$ , also auf ganz  $(-R, R)$ .  $\square$

**Beispiel 16.5** Für die Potenzreihendarstellung des Logarithmus berechnen wir mit der geometrischen Reihe, für  $x \in (-1, 1)$ ,

$$\frac{d}{dx} \log(1+x) = \frac{1}{1+x} = \sum_{k=0}^{\infty} (-1)^k x^k =: Q(x).$$

$Q(x)$  ergibt sich durch gliedweise Differentiation der Reihe  $P(x) = \sum_{k=1}^{\infty} (-1)^{k-1} x^k / k$ . Nach Lemma 16.2 haben  $P(x)$  und  $Q(x)$  denselben Konvergenzradius, also  $R = 1$ , und Satz 16.5 ergibt  $P'(x) = Q(x)$  für  $x \in (-1, 1)$ . Nun ist  $\log(1+x) = P(x) = 0$  für  $x = 0$ , also folgt

$$\log(1+x) = P(x) = x - \frac{x^2}{2} + \frac{x^3}{3} - + \dots \quad \text{für } x \in (-1, 1).$$

**Beispiel 16.6** Die Ableitung der Funktion  $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$  ist nach Quotientenregel

$$\tan' = \left( \frac{\sin}{\cos} \right)' = \frac{\cos^2 + \sin^2}{\cos^2} = 1 + \tan^2.$$

Für die Umkehrfunktion  $\arctan$  folgt, für  $x \in (-1, 1)$ ,

$$\frac{d}{dx} \arctan x = \frac{1}{\tan'(\arctan x)} = \frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k} =: Q(x).$$

$Q(x)$  ist die gliedweise Ableitung von  $P(x) = \sum_{k=0}^{\infty} (-1)^k x^{2k+1} / (2k+1)$ . Der gemeinsame Konvergenzradius ist  $R = 1$ , und aus Satz 16.5 folgt  $P'(x) = Q(x)$  für alle  $x \in (-1, 1)$ . Da  $P(0) = \arctan 0 = 0$ , sehen wir

$$\arctan x = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - + \dots \quad \text{für } x \in (-1, 1).$$

Hat ein Polynom vom Grad höchstens  $k$  mehr als  $k$  Nullstellen, so ist es bekanntlich das Nullpolynom. Erstaunlicherweise hat man für Potenzreihen eine Aussage, die als verwandt betrachtet werden kann.

**Satz 16.6 (Identitätssatz für Potenzreihen)** Sei  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  eine Potenzreihe mit Konvergenzradius  $R > 0$ . Ist  $0 \in \mathbb{C}$  Häufungspunkt der Menge  $\{z \in \mathbb{C} : P(z) = 0\}$ , so folgt  $a_k = 0$  für alle  $k \in \mathbb{N}_0$ .

BEWEIS: Für  $\varrho \in (0, R)$  ist  $P(\varrho)$  konvergent, also  $|a_k| \varrho^k \leq M$  für alle  $k$ . Nach (16.1) gilt

$$|P(z) - P_k(z)| \leq \frac{M}{1 - \frac{|z|}{\varrho}} \left( \frac{|z|}{\varrho} \right)^{k+1} \quad \text{für } |z| < \varrho.$$

Für  $|z| \leq \varrho/2$  ergibt sich

$$|P(z) - P_k(z)| \leq \frac{2M}{\varrho^{k+1}} |z|^{k+1} =: C |z|^{k+1}.$$

Sei nun induktiv schon  $a_0 = \dots = a_{k-1} = 0$  gezeigt für ein  $k \in \mathbb{N}_0$ . Dann folgt

$$|P(z) - a_k z^k| = |P(z) - P_k(z)| \leq C |z|^{k+1} \quad \text{für } |z| \leq \varrho/2.$$

Sei nun  $P(z_i) = 0$  für eine Folge  $z_i \neq 0$  mit  $z_i \rightarrow 0$ . Einsetzen von  $z = z_i$  ergibt

$$|a_k| |z_i|^k = |P(z_i) - a_k z_i^k| \leq C |z_i|^{k+1},$$

also  $|a_k| \leq C |z_i| \rightarrow 0$  mit  $i \rightarrow \infty$ . Das zeigt  $a_k = 0$ . □

**Folgerung 16.1 (Koeffizientenvergleich)** Seien  $P(z) = \sum_{k=0}^{\infty} a_k z^k$  und  $Q(z) = \sum_{k=0}^{\infty} b_k z^k$  Potenzreihen mit positivem Konvergenzradius. Ist der Nullpunkt Häufungspunkt der Menge  $\{z \in \mathbb{C} : P(z) = Q(z)\}$ , so folgt  $a_k = b_k$  für alle  $k \in \mathbb{N}_0$ .

BEWEIS: Die Potenzreihe  $F(z) = \sum_{k=0}^{\infty} c_k z^k$  mit  $c_k = a_k - b_k$  hat positiven Konvergenzradius, und der Nullpunkt ist Häufungspunkt der Menge  $\{z \in \mathbb{C} : F(z) = 0\}$ . Die Behauptung folgt damit aus Satz 16.6. □

**Beispiel 16.7** Die Binomialreihe mit Parameter  $\alpha \in \mathbb{R}$  ist definiert durch

$$B_\alpha(z) = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k.$$

Sie hat Konvergenzradius  $R = 1$ , siehe Beispiel 16.2. Für  $x \in (-1, 1)$  gilt nach Satz 16.5

$$B'_\alpha(x) = \sum_{k=0}^{\infty} (k+1) \binom{\alpha}{k+1} x^k = \sum_{k=0}^{\infty} (k+1) \frac{\alpha-k}{k+1} \binom{\alpha}{k} x^k = \alpha B_\alpha(x) - x B'_\alpha(x),$$

das heißt  $B'_\alpha(x) = \frac{\alpha}{1+x} B_\alpha(x)$ . Es folgt mit  $f(x) = (1+x)^{-\alpha}$

$$(f B_\alpha)'(x) = f'(x) B_\alpha(x) + f(x) B'_\alpha(x) = f(x) B_\alpha(x) \left( -\frac{\alpha}{1+x} + \frac{\alpha}{1+x} \right) = 0.$$

Wegen  $B_\alpha(0) = 1 = f(0)$  ergibt sich die folgende Darstellung (Newton 1665)

$$(1+x)^\alpha = B_\alpha(x) = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{für alle } x \in (-1, 1).$$

In der Physik wird oft die Näherung  $(1+x)^\alpha = 1 + \alpha x$  benutzt, für  $|x| \ll 1$ .

Bisher haben wir uns immer in der offenen Kreisscheibe bewegt, auf der die Potenzreihe lokal gleichmäßig konvergiert. Zum Schluss des Kapitels betrachten wir die Situation am Rand.

**Satz 16.7 (Abelscher Grenzwertsatz)** *Ist die Potenzreihe  $P(x) = \sum_{k=0}^{\infty} a_k x^k$  für  $x = 1$  konvergent, so gilt*

$$\lim_{x \nearrow 1} P(x) = P(1).$$

BEWEIS: Nach Voraussetzung hat  $P$  Konvergenzradius  $R \geq 1$ . Berechne für  $0 \leq x < 1$

$$\begin{aligned} P_\ell(1) - P_\ell(x) &= \sum_{k=0}^{\ell} a_k (1 - x^k) \\ &= (1-x) \sum_{k=0}^{\ell} \sum_{j=0}^{k-1} a_k x^j \\ &= (1-x) \sum_{j=0}^{\ell-1} \sum_{k=j+1}^{\ell} a_k x^j \\ &= (1-x) \sum_{j=0}^{\ell-1} x^j (P_\ell(1) - P_j(1)). \end{aligned}$$

Zu  $\varepsilon > 0$  gibt es ein  $N \in \mathbb{N}$  mit  $|P_\ell(1) - P_j(1)| < \varepsilon$  für  $j, \ell \geq N$ , außerdem gilt  $|P_j(1)| \leq C$  für alle  $j$ . Damit schätzen wir ab

$$|P_\ell(1) - P_\ell(x)| \leq 2C(1-x) \sum_{j=0}^{N-1} x^j + \varepsilon(1-x) \sum_{j=N}^{\ell-1} x^j \leq 2C(1-x)N + \varepsilon,$$

wobei zuletzt die geometrische Reihe benutzt wurde. Mit  $\ell \rightarrow \infty$  und  $x \nearrow 1$  folgt

$$\limsup_{x \nearrow 1} |P(1) - P(x)| \leq \varepsilon.$$

□

**Beispiel 16.8** Nach dem Leibnizkriterium konvergiert die Reihe  $x - \frac{1}{2}x^2 + \frac{1}{3}x^3 \pm \dots$  auch für  $x = 1$ , also folgt aus Satz 16.7

$$\log 2 = \lim_{x \nearrow 1} \log(1 + x) = 1 - \frac{1}{2} + \frac{1}{3} - + \dots \quad (\text{Mercator 1668}).$$

Ebenso konvergiert die Reihe des arctan auch für  $x = 1$ , und es ergibt sich die Darstellung

$$\frac{\pi}{4} = \arctan 1 = \lim_{x \nearrow 1} \arctan x = 1 - \frac{1}{3} + \frac{1}{5} - + \dots \quad (\text{Gregory 1671, Leibniz 1674}).$$



## 17 Taylorentwicklung

Das letzte Thema in diesem Buch ist das Taylorpolynom bzw. die Taylorentwicklung<sup>23</sup>. Es geht dabei um den Vergleich einer gegebenen Funktion  $f(x)$  mit einem Polynom  $P(x)$ , das in einer Stelle  $x_0$  „von  $k$ -ter Ordnung“ mit  $f$  übereinstimmt, also  $f^{(i)}(x_0) = P^{(i)}(x_0)$  für  $i = 0, 1, \dots, k$ . Das Polynom  $P(x)$  sollte dann  $f(x)$  asymptotisch für  $x \rightarrow x_0$  gut approximieren, und das soll quantifiziert werden. Für  $k = 1$ , also affin-lineare Polynome, haben wir sowas schon gesehen bei der Interpretation der Ableitung.

Zur Erinnerung: eine Funktion  $P : \mathbb{R} \rightarrow \mathbb{R}$  heißt Polynom vom Grad  $k \in \mathbb{N}_0$ , wenn es  $a_0, \dots, a_k \in \mathbb{R}$  gibt mit  $a_k \neq 0$ , so dass gilt:

$$P(x) = \sum_{j=0}^k a_j x^j \quad \text{für alle } x \in \mathbb{R}.$$

Im Raum aller Funktionen  $f : \mathbb{R} \rightarrow \mathbb{R}$  ist die Menge  $\mathbb{P}_k$  der Polynome vom Grad  $\leq k$  der durch  $1, x, \dots, x^k$  erzeugte Unterraum. Es gilt: für jedes  $x_0 \in \mathbb{R}$  bilden die Funktionen  $1, x - x_0, \dots, (x - x_0)^k$  eine Basis von  $\mathbb{P}_k$ . Wegen  $\dim \mathbb{P}_k \leq k + 1$  müssen wir nur die lineare Unabhängigkeit zeigen. Dazu verwenden wir die Ableitungsregel

$$P(x) = \sum_{j=0}^k a_j (x - x_0)^j \quad \Rightarrow \quad \left(\frac{d}{dx}\right)^i P(x)|_{x=x_0} = i! a_i. \quad (17.1)$$

Ist  $P(x)$  die Nullfunktion, so folgt  $a_i = 0$  für  $i = 0, \dots, k$  wie behauptet.

**Lemma 17.1** Sei  $I = (a, b) \subset \mathbb{R}$ ,  $x_0 \in I$  und  $k \in \mathbb{N}_0$ . Zu  $f \in C^k(I)$  gibt es genau ein Polynom  $P \in \mathbb{P}_k$  mit  $P^{(i)}(x_0) = f^{(i)}(x_0)$  für  $i = 0, 1, \dots, k$ , und zwar

$$P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j. \quad (17.2)$$

$P_k$  heißt Taylorpolynom der Ordnung  $k$  von  $f$  mit Entwicklungspunkt  $x_0$ .

BEWEIS: Für das in (17.2) definierte Polynom gilt  $P^{(i)}(x_0) = f^{(i)}(x_0)$  für  $i = 0, \dots, k$ , wie man mit (17.1) sieht. Zur Eindeutigkeit sei  $P \in \mathbb{P}_k$  mit  $P^{(i)}(x_0) = 0$  für alle  $i = 0, \dots, k$ . Wie oben gezeigt gilt eine Darstellung  $P(x) = \sum_{j=0}^k a_j (x - x_0)^j$  für alle  $x \in \mathbb{R}$ . Mit (17.1) folgt  $a_i = 0$  für alle  $i = 0, \dots, k$ .  $\square$

**Folgerung 17.1** Das  $k$ -te Taylorpolynom mit Entwicklungspunkt  $x_0$  eines Polynoms  $f$  vom Grad höchstens  $k$  ist  $f$  selbst.

In der Situation von Lemma 17.1 heißt die Funktion

$$R_k : (a, b) \rightarrow \mathbb{R}, \quad R_k(x) = f(x) - P_k(x) \quad (17.3)$$

das Restglied  $k$ -ter Ordnung der Taylorentwicklung in  $x_0$ . Knackpunkt bei der Taylorentwicklung ist die Abschätzung dieses Restglieds und damit eine Aussage darüber, wie gut die Funktion durch das Taylorpolynom approximiert wird. Hierfür gibt es verschiedene mögliche Darstellungen von  $R_k$ .

---

<sup>23</sup>Brook Taylor, 1685-1731

**Satz 17.1 (Integraldarstellung des Restglieds)** Sei  $f \in C^{k+1}(I)$  für ein  $k \in \mathbb{N}_0$ , und  $P_k(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j$  das  $k$ -te Taylorpolynom im Punkt  $x_0 \in I$ . Dann gilt

$$f(x) = P_k(x) + R_k(x) \quad \text{mit} \quad R_k(x) = \frac{1}{k!} \int_{x_0}^x (x - y)^k f^{(k+1)}(y) dy.$$

BEWEIS: Durch Induktion über  $k \in \mathbb{N}_0$ . Für  $k = 0$  folgt aus dem Hauptsatz

$$f(x) = f(x_0) + \int_{x_0}^x f'(y) dy.$$

Für  $k \geq 1$  folgt induktiv mit partieller Integration, vgl. Lemma ?? für den Fall  $k = 1$ ,

$$\begin{aligned} f(x) &= P_{k-1}(x) + \frac{1}{(k-1)!} \int_{x_0}^x (x-y)^{k-1} f^{(k)}(y) dy \\ &= P_{k-1}(x) + \frac{1}{(k-1)!} \left( \left[ -\frac{(x-y)^k}{k} f^{(k)}(y) \right]_{y=x_0}^{y=x} + \int_{x_0}^x \frac{(x-y)^k}{k} f^{(k+1)}(y) dy \right) \\ &= P_k(x) + \frac{1}{k!} \int_{x_0}^x (x-y)^k f^{(k+1)}(y) dy. \end{aligned}$$

□

Die zweite Darstellung des Restglieds ist vielleicht etwas populärer.

**Satz 17.2 (Lagrange-Darstellung des Restglieds)** Sei  $f \in C^{k+1}(I)$  für ein  $k \in \mathbb{N}_0$ . Dann gibt es zu  $x_0, x \in I$  ein  $\xi$  zwischen  $x_0$  und  $x$ , so dass gilt:

$$f(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + R_k(x) \quad \text{mit} \quad R_k(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1}. \quad (17.4)$$

BEWEIS: Wir verwenden den Mittelwertsatz der Integralrechnung mit Gewicht, siehe Folgerung 13.2, Analysis 1: ist  $\varphi \in C^0(I)$  mit  $\varphi \geq 0$ , so gibt es zu  $f \in C^0(I)$  ein  $\xi \in I$  mit

$$\int_I f \varphi = f(\xi) \int_I \varphi.$$

Sei nun  $x > x_0$ . Dann können wir  $I = [x_0, x]$  und  $\varphi(y) = (x - y)^k$  wählen. Es folgt

$$\frac{1}{k!} \int_{x_0}^x (x - y)^k f^{(k+1)}(y) dy = \frac{1}{k!} f^{(k+1)}(\xi) \int_{x_0}^x (x - y)^k dy = \frac{f^{(k+1)}(\xi)}{(k+1)!} (x - x_0)^{k+1},$$

für ein  $\xi \in [x_0, x]$ . Der Fall  $x < x_0$  ist analog, der Satz ist bewiesen. □

**Beispiel 17.1** Betrachte für  $x \in (-1, 1)$  die Funktion  $f(x) = (1 - x)^{-1/2}$ , mit Ableitungen

$$f'(x) = \frac{1}{2}(1 - x)^{-3/2} \quad \text{und} \quad f''(x) = \frac{3}{4}(1 - x)^{-5/2}.$$

Es gilt  $f(0) = 1$  und  $f'(0) = 1/2$ , also lautet das Taylorpolynom der Ordnung Eins in  $x_0 = 0$

$$P_1(x) = f(0) + f'(0)x = 1 + \frac{1}{2}x,$$

mit der Lagrange-Restglieddarstellung

$$R_1(x) = \frac{f''(\xi)}{2}x^2 = \frac{3}{8}(1-\xi)^{-5/2}x^2 \quad \text{für ein } \xi \in [0, x].$$

Als Anwendung erhalten wir für die relativistische Energie eines Teilchens mit Ruhemasse  $m_0$  und Geschwindigkeit  $v$ , wenn wir  $\beta = v/c$  setzen,

$$E = \frac{m_0c^2}{\sqrt{1-\beta^2}} = m_0c^2 \left( 1 + \frac{1}{2}\beta^2 + \frac{f''(\xi)}{2}\beta^4 \right) = m_0c^2 + \frac{1}{2}m_0v^2 + \Delta E.$$

Dabei ist der erste Term die Ruheenergie und der zweite die klassische kinetische Energie. Für den relativistischen Korrekturterm ergibt sich aus der Restglieddarstellung die Abschätzung

$$\frac{\Delta E}{E_{kin}} = f''(\xi)\beta^2 \leq f''(\beta^2)\beta^2 < 0,008 \quad \text{für } \beta \leq 0,1.$$

Bei Geschwindigkeiten  $v \leq \frac{1}{10}c$  beträgt die relativistische Korrektur weniger als ein Prozent der klassischen kinetischen Energie.

Um das asymptotische Verhalten von Funktionen bei Grenzprozessen zu vergleichen, werden oft die Landauschen Symbole  $\mathcal{O}$  und  $o$  benutzt. Seien  $f, g$  zwei Funktionen, die für  $0 < |x - x_0| < \delta$  definiert sind, und es sei  $g(x) \neq 0$  für  $x$  nahe bei  $x_0$ . Dann schreibt man

$$\begin{aligned} f(x) = o(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \lim_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} = 0, \\ f(x) = \mathcal{O}(g(x)) \text{ für } x \rightarrow x_0 &\Leftrightarrow \limsup_{x \rightarrow x_0} \frac{|f(x)|}{|g(x)|} < \infty. \end{aligned}$$

In Worten: die Funktion  $f(x)$  ist klein- $o$  von  $g(x)$  beziehungsweise groß- $\mathcal{O}$  von  $g(x)$  für  $x \rightarrow x_0$ . Diese Begriffe sind analog für Grenzwerte  $|x| \rightarrow \infty$  usw. erklärt. Wir zeigen jetzt, dass sich eine Funktion  $f \in C^k(I)$  für  $x \rightarrow x_0$  asymptotisch wie das  $k$ -te Taylorpolynom  $P_k(x)$  verhält, bis auf einen Term der von der Ordnung  $k$  verschwindet:

$$f(x) = P_k(x) + o(|x - x_0|^k) \quad \text{für } x \rightarrow x_0.$$

Bei den Abschätzungen des Restlieds haben wir eine Ableitung mehr benötigt.

**Satz 17.3 (Approximation durch das Taylorpolynom)** Sei  $f \in C^k(I)$  für  $k \in \mathbb{N}_0$ , und  $P_k$  das  $k$ -te Taylorpolynom von  $f$  mit Entwicklungspunkt  $x_0 \in I$ . Dann ist  $P_k$  das eindeutig bestimmte Polynom vom Grad höchstens  $k$  mit

$$\lim_{x \rightarrow x_0} \frac{f(x) - P_k(x)}{(x - x_0)^k} = 0.$$

BEWEIS: Nach Satz 17.2 gibt es zu  $x \in I$  ein  $\xi$  zwischen  $x_0$  und  $x$  mit

$$\frac{f(x) - P_k(x)}{(x - x_0)^k} = \frac{f(x) - P_{k-1}(x)}{(x - x_0)^k} - \frac{1}{k!}f^{(k)}(x_0) = \frac{1}{k!}(f^{(k)}(\xi) - f^{(k)}(x_0)).$$

Da  $f^{(k)}$  stetig, ist  $|f^{(k)}(\xi) - f^{(k)}(x_0)| < \varepsilon$  für  $|x - x_0| < \delta$ , womit die Konvergenz gegen Null bewiesen ist. Für die Eindeutigkeit ist zu zeigen, dass für  $P(x) = \sum_{j=0}^k a_j(x - x_0)^j$  gilt:

$$\lim_{x \rightarrow x_0} \frac{P(x)}{(x - x_0)^k} = 0 \quad \Rightarrow \quad a_0 = \dots = a_k = 0.$$

Sei induktiv schon  $a_0 = \dots = a_{j-1} = 0$  gezeigt mit  $0 \leq j \leq k$ . Dann folgt

$$a_j = \lim_{x \rightarrow x_0} (x - x_0)^{-j} P(x) = \lim_{x \rightarrow x_0} (x - x_0)^{k-j} (x - x_0)^{-k} P(x) = 0.$$

□

Zum Schluss des Kapitels kommen wir zu einem anderen Aspekt der Taylorschen Formel. Wir betrachten eine Funktion  $f \in C^\infty(I)$ , wobei  $I$  offenes Intervall mit  $x_0 \in I$ . Wir haben dann für jedes  $k \in \mathbb{N}_0$  das Taylorpolynom  $P_k(x)$  mit Entwicklungspunkt  $x_0$ , nämlich

$$P_k(x) = \sum_{i=0}^k \frac{f^{(i)}(x_0)}{i!} (x - x_0)^i.$$

Wir fragen nun: konvergiert für  $x \in I$ , oder zumindest für  $x$  in einer Umgebung von  $x_0$ , die Folge  $P_k(x)$  für  $k \rightarrow \infty$ ? Und wenn ja, ist der Grenzwert der Funktionswert  $f(x)$ ? Die Reihe

$$P(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$$

heißt Taylorreihe von  $f(x)$  mit Entwicklungspunkt  $x_0$ . In Kapitel 16 wurden nur Potenzreihen mit Entwicklungspunkt  $x_0 = 0$  betrachtet, die Aussagen gelten aber sinngemäß.

**Definition 17.1 (analytische Funktion)** Sei  $I$  offenes Intervall und  $f : I \rightarrow \mathbb{R}$ . Dann heißt  $f$  analytisch, wenn jedes  $x_0 \in I$  eine Umgebung  $(x_0 - \delta, x_0 + \delta)$  hat, auf der  $f$  durch eine konvergente Potenzreihe mit Entwicklungspunkt  $x_0$  dargestellt wird:

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k \quad \text{für alle } x \in (x_0 - \delta, x_0 + \delta).$$

Man bezeichnet den entsprechenden Raum von Funktionen mit  $C^\omega(I)$ .

Die darstellende Potenzreihe kann nur die Taylorreihe sein. Denn  $f$  ist auf  $(x_0 - \delta, x_0 + \delta)$  von der Klasse  $C^\infty$  nach Satz 16.5, und gliedweise Differentiation ergibt

$$f^{(i)}(x_0) = \sum_{k=0}^{\infty} \delta_{ik} k! a_k = i! a_i.$$

Eine Funktion ist also genau dann analytisch, wenn sie  $C^\infty$  ist und für jedes  $x_0 \in I$  gilt: die Taylorreihe mit Entwicklungspunkt  $x_0$  konvergiert punktweise gegen  $f$  nahe bei  $x_0$ . Für die analytischen Funktionen gilt also eine Art Fernwirkung: hat man alle Ableitungen in einem Punkt  $x_0 \in I$ , so kennt man schon die ganze Funktion  $f : I \rightarrow \mathbb{R}$ . Hat die Menge der Nullstellen von  $f$  einen Häufungspunkt  $x_0 \in I$ , so ist  $f$  schon die Nullfunktion. Das folgt sofort aus Satz 16.6, angewandt im Punkt  $x_0$ . Eine analytische Funktion kann für jedes  $x_0 \in I$  auf die Kreisscheibe  $B_\delta(x_0) \subset \mathbb{C}$  fortgesetzt werden. Dazu muss man nur in die Potenzreihe Zahlen

$z \in B_\delta(x_0)$  einsetzen. In der komplexen Analysis betrachtet man komplex differenzierbare Funktionen. Es stellt sich heraus, dass diese immer analytisch sind, daraus ergibt sich die Bedeutung des Begriffs. Im allgemeinen sind aber Funktionen  $f \in C^\infty(I)$  nicht analytisch, wie das folgende Beispiel zeigt.

**Beispiel 17.2** Betrachte  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit

$$f(x) = \begin{cases} e^{-\frac{1}{x}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Es gilt  $f \in C^\infty(\mathbb{R})$  nach Übungsaufgabe, und zwar  $f^{(k)}(0) = 0$  für alle  $k \in \mathbb{N}_0$ . Also sind die Koeffizienten der Taylorreihe mit Entwicklungspunkt  $x_0 = 0$  alle Null und damit auch alle Partialsummen, die Reihe konvergiert somit gegen die Nullfunktion und nicht gegen  $f$ . In der Tat ist  $x_0 = 0$  Häufungspunkt der Nullstellenmenge von  $f$ .