Analysis I

Wintersemester 2025/2026

Ernst Kuwert

Mathematisches Institut
Universität Freiburg

Inhaltsverzeichnis

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 den 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

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

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: $M = \{\alpha, \beta, \ldots, \omega\}$.

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

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

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

(c) Eine Variante der aufzählenden Beschreibung haben wir bei

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

M wird durch die Menge $\mathbb{N} = \{1, 2, 3, \ldots\}$ 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. [1] 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.[2]

[1] soll das betont werden schreibe ich $M \overset{\neq}{\subset} N$.

[2] 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:

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

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 $A^C = X \backslash A$ 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:[3]

$\neg P$	*nicht P*	wahr genau wenn P falsch
		falsch genau wenn P wahr
$P \wedge Q$	*P und Q*	wahr genau wenn P, Q beide wahr
		falsch genau wenn mindestens eine der Aussagen P, Q falsch
$P \vee Q$	*P oder Q*	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:

$P \Rightarrow Q$	*aus P folgt Q*	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:

[3] 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

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

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

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

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 $f : X \to Y$. 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 $\mathrm{id}_X : X \to X$, $\mathrm{id}_X(x) = x$. Die Verkettung (Hintereinanderschaltung) von Abbildungen $f : X \to Y$ und $g : Y \to Z$ 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[4]

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

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

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

[4]René Descartes, 1596–1650

Als nächstes drei Begriffe zum Abbildungsverhalten: $f : X \to Y$ 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 $f : X \to Y$ 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 $f : X \to Y$ bijektiv. Dann können wir $y \in Y$ die eindeutig bestimmte Lösung von $f(x) = y$ zuordnen. Wir kriegen eine Abbildung $g : Y \to X$, $y \mapsto g(y)$, wobei $g(y)$ eben diese Lösung ist, das heißt

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

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

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

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

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

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 [5] 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 Integralrechung. 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*[6].

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:

$$\begin{array}{ll}
\mathbb{N} = \{1, 2, 3, \ldots\} & \text{natürliche Zahlen} \\[2ex]
\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\} & \text{ganze Zahlen} \\[2ex]
\mathbb{Q} = \left\{ \dfrac{p}{q} \,\middle|\, q \in \mathbb{N},\, p \in \mathbb{Z} \right\} & \text{rationale Zahlen}
\end{array}$$

Die rellen 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:

---

[5] Archimedes von Syrakus, 287-212 v.C.
[6] H.-D. Ebbinghaus (Hrsg.): Zahlen, Springer 1992

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

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

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

$\square$

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:*

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

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

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

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

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. $\square$

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) $\dfrac{a}{c} + \dfrac{b}{d} = \dfrac{ad + bc}{cd}$,

(2) $\dfrac{a}{c} \cdot \dfrac{b}{d} = \dfrac{ab}{cd}$,

(3) $\dfrac{a/c}{b/d} = \dfrac{ad}{bc}$, *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). □

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

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

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

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

□

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
$\lvert I\rvert = b - a$ 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 $\{k_0, \ldots, k_1\}$ 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 $\{k_0, \ldots, k_1\}$ und es folgt $k' > k_1 > b$, Widerspruch zu b obere Schranke von M. □

3 Vollständige Induktion

Am Amazonas gibt es ein sehr kleines indigenes Volk, die Pirahã, die nur die Zählbegriffe *eins*, *wenige* und *viele* kennen. Im Gegensatz dazu haben wir die natürlichen Zahlen $1, 2 = 1 + 1, 3 = 2 + 1, \ldots$ zur Verfügung. Laut Kronecker[7] 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) $n \in M \quad \Rightarrow \quad n+1 \in M$.

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 $1, 2, 3, \ldots$ 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[8] 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 $\quad \Rightarrow \quad A(n+1)$ 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. $\square$

[7] Leopold Kronecker, 1823-1891

[8] 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

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

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 $A(n)$:

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

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:

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

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 durchlaufend nummerierte Zahlen $a_m, \ldots, a_n$ setzen wir

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

Der Index i duchläuft 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

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

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 $m, \ldots, n$, so muss j die Werte $m-1, \ldots, n-1$ 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 $n > m$

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

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 $n \in \mathbb{N}_0$

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

Dabei ist $x^k = x \cdot \ldots \cdot x$ 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

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

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) [9] *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

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

$\square$

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 [10].

Definition 3.1 (Zahl der Elemente) *Eine Menge M hat $n \in \mathbb{N}$ Elemente (Notation: $\#M = n$) wenn eine bijektive Abbildung $\{1,\ldots,n\} \xrightarrow{\sim} M$ 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 $\{1,\ldots,m\}$ als auch zu $\{1,\ldots,n\}$. Wir hätten dann eine Bijektion

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

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) [11] *Für alle $n \in \mathbb{N}$ gilt: ist $f : \{1,\ldots,m\} \to \{1,\ldots,n\}$ 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

$$f : \{1,\ldots,m\} \to \{1,\ldots,n+1\} \quad \text{injektiv.}$$

[9] Jakob Bernoulli, Basel 1689

[10] 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).

[11] 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

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

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 $k = 1, \ldots, m-1$; das geht gut falls $f(k) \in \{1, \ldots, n\}$ für alle $k = 1, \ldots, m-1$. Andernfalls ist $f(i) = n+1$ für genau ein $i \in \{1, \ldots, m-1\}$. Aber dann ist $f(m) \neq n+1$ da f injektiv, als Bildpunkt von i können wir also $f(m) \in \{1, \ldots, n\}$ nehmen:

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

Die $f(1), \ldots, f(m)$ 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 $f : \{1, \ldots, m\} \to \{1, \ldots, n\}$ surjektiv für $m, n \in \mathbb{N}$, so folgt $m \geq n$.*

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

$$g : \{1, \ldots, n\} \to \{1, \ldots, m\},\ g(k) = i_k,$$

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 $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$. Die Ziffern $\sigma_1 = \sigma(1), \ldots, \sigma_n = \sigma(n)$ sind eine Umordnung (oder Permutation) von $1, \ldots, n$, wir identifizieren jedes σ mit dem Tupel $(\sigma_1, \ldots, \sigma_n)$.

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

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:

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

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

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

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

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

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

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

$\square$

Für das nächste Zählproblem brauchen wir die

Definition 3.2 (Binomialkoeffizienten) *Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}_0$ setzen wir*

$$\binom{\alpha}{k} = \frac{\alpha \cdot (\alpha-1) \cdot \ldots \cdot (\alpha-k+1)}{1 \cdot 2 \cdot \ldots \cdot k} = \prod_{i=1}^{k} \frac{\alpha - i + 1}{i} \qquad \textit{sowie} \qquad \binom{\alpha}{0} = 1.$$

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

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

Beweis: Für $k = 1$ sieht man das direkt. Für $k \geq 2$ berechnen wir

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

$\square$

Blaise Pascal (1623-1662) hat die Zahlen $\binom{n}{k}$ in folgendem Schema angegeben. Dabei sind die Zeilen durch $n = 0, 1, 2, \ldots$ nummeriert, die Werte $k = 0, 1, \ldots, n$ entsprechen den Einträgen der n-ten Zeile. Nach Lemma 3.1 ist jeder Eintrag die Summe der beiden darüber stehenden Einträge, damit können die Zeilen sukzessive berechnet werden.

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}$$

Ebenfalls für $\alpha = n \in \mathbb{N}_0$ folgt durch Erweitern der Binomialkoeffizienten mit $(n-k)!$ die alternative Darstellung

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \text{ für } n \in \mathbb{N}_0,\ k \in \{0, 1, \ldots, n\}, \tag{3.1}$$

und daraus weiter die am Diagramm ersichtliche Symmetrieeigenschaft

$$\binom{n}{k} = \binom{n}{n-k} \text{ für } n \in \mathbb{N}_0,\ k \in \{0, 1, \ldots, n\}. \tag{3.2}$$

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

Beweis: Wir erledigen zunächst den Fall $k = 0$. Es gibt genau eine null-elementige Teilmenge von $\{1, \ldots, n\}$, 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 $k = 0, 1, \ldots, n$ ist die Zahl der k-elementigen Teilmengen von $\{1, \ldots, n\}$ gleich $\binom{n}{k}$.

Wir oben festgestellt gilt $A(0)$. Für den Induktionsschluss müssen wir die Anzahl der k-elementigen Teilmengen von $\{1, \ldots, n+1\}$ bestimmen, wobei $n \in \mathbb{N}_0$ und $k \in \{1, \ldots, n+1\}$. 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 $\{1, \ldots, n\}$. Klasse 2 ergibt sich, indem wir zu jeder $(k-1)$-elementigen Teilmenge von $\{1, \ldots, n\}$ das Element $n + 1$ dazunehmen. Induktiv ist die Zahl der k-elementigen Teilmengen von $\{1, \ldots, n+1\}$ also gleich

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

Zum Schlus haben wir Lemma 3.1 benutzt. Das Argument gilt auch im Extremfall $k = n+1$: es gibt keine $(n+1)$-elementigen Teilmengen von $\{1, \ldots, n\}$, 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*

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

Das Produkt $(a+b)^n$ besteht aus n Faktoren. Beim Ausmultiplizieren ergeben sich Terme der Form $a^k b^{n-k}$ mit $k = 0, 1, \ldots, n$. 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

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

mit geigneten 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

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

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

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

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,

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

□

4 Grenzwerte von Folgen

Wir definieren hier den Begriff des Grenzwerts im Fall von Folgen reeller Zahlen. Die zugrundliegende 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 verewigen wollte; leider ist das Grab aber verschollen. Sein Ansatz wurde von Weierstraß 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 $a_n = 1, 4, 9, \ldots$ die Folge der Quadratzahlen. Die Folge $a_n = 1, 1, \ldots$ 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 $n \to \infty$, falls gilt:*

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

Die Zahl a heißt Grenzwert der Folge und wir schreiben

$$\lim_{n\to\infty} a_n = a \qquad \textit{oder} \qquad a_n \to a \textit{ für } n \to \infty.$$

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 Modikikation von Definition 4.1 lautet wie folgt:

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

Diese Version folgt aus Definition 4.1: angenommen wir haben dort zu $\varepsilon > 0$ ein K gefundeen. Dann wählen wir $N > K$ als die nächst größere natürliche Zahl. Für $n \geq N$ ist $n > K$ und es 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 $\lim_{n\to\infty} a_n = a$. 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 $\lim_{n\to\infty} q^n = 0$. 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

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

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 ± 1, insbesondere ist $a_n - a_{n+1} = (-1)^n(1-(-1)) = \pm 2$. Angenommen es gilt $\lim_{n\to\infty} a_n = a$ 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 $n > K$

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

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

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

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 3ε auch mit 10ε oder mit $C\varepsilon$ für eine Konstante C, auch die Abschätzung $|a_n - a| \leq \varepsilon$ wäre ausreichend.

Durch den Begriff der ε-Umgebung gewinnt der Grenzwertbegriff an Anschaulichkeit.

Definition 4.2 (ε-Umgebung) *Die ε-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 ε 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 $\lvert a_n\rvert \leq C$ 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 beschänkt.*

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

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

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

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

$\square$

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 $a_n \to a, b_n \to b$ mit $n \to \infty$.*

a) *Für alle $\lambda, \mu \in \mathbb{R}$ ist $(\lambda a_n + \mu b_n)_{n \in \mathbb{N}}$ konvergent mit Grenzwert $\lim_{n \to \infty}(\lambda a_n + \mu b_n) = \lambda a + \mu b$.*

b) *Die Folge $(a_n \cdot b_n)_{n \in \mathbb{N}}$ ist konvergent mit Grenzwert $\lim_{n \to \infty}(a_n \cdot b_n) = a \cdot b$.*

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 $\lim_{n \to \infty} a_n/b_n = a/b$.*

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

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

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

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

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

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

□

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 $\lim_{n\to\infty} a_n = 0$.

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

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) $a_n = 1 + q + \ldots + q^n = \sum_{k=0}^n q^k$. Dann ergibt sich aus Beispiel 3.2, Beispiel 4.3 und Satz 4.3

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

Wir schreiben hierfür auch

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

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* $\lim_{n\to\infty} a_n = a$ *und* $\lim_{n\to\infty} b_n = b$*. 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 $a-\varepsilon < b+\varepsilon$ beziehungsweis $(a-b)/2 < \varepsilon$ 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 $\lim_{n\to\infty} c_n = a$ 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 $\lim_{n\to\infty} 1/n = 0$.

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

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

Wir schreiben $\lim_{n\to\infty} a_n = +\infty$ *oder* $a_n \to +\infty$ *mit* $n \to \infty$*. 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 $a_n \to +\infty$ wie folgt ausgedrückt werden:

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

Bei Konvergenz gegen $a \in \mathbb{R}$ liegen für $n > K$ alle Folgenglieder in $B_\varepsilon(a)$. Bei Konvergenz gegen $+\infty$ tritt anstelle der ε-Umgebung das rechtsseitig unendliche Intervall (C,∞). 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 $\lim_{n\to\infty} q^n = +\infty$. 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 $n \to \infty$ folgende Tabelle:

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

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

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

Also folgt induktiv $a_n \geq n(c - \frac{1}{4}) \to \infty$. 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 $|a_n| \to +\infty$ folgt $1/a_n \to 0$.*

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

Beweis: Um (1) zu zeigen sei $\varepsilon > 0$ gegeben. Wegen $|a_n| \to +\infty$ 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 $a_n \to 0$, $a_n > 0$, 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 $n > K$. □

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 rellen 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 [12]. Wir brauchen 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, inbesondere 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 Widerspuch. □

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.

[12] 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:

$$x_{n+1} = \frac{1}{2}\Big(x_n + \frac{a}{x_n}\Big), \qquad \text{Startwert } x_0 > 0. \tag{5.1}$$

Wir können x_n und a/x_n als Seiten eines Rechtecks mit Flächeninhalt a auffassen. Im Fall eines Quadrats wäre $x^2 = a$, die Lösung wäre also gefunden. Im allgemeinen sind die Seiten aber relativ zum Quadrat zu lang bzw. zu kurz, der Übergang zum arithmetische Mittel 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

$$x_n - x_{n+1} = \frac{1}{2}\Big(x_n - \frac{a}{x_n}\Big) = \frac{x_n^2 - a}{2x_n} \geq 0 \text{ für } n \geq 1, \qquad \text{also } x_1 \geq x_2 \geq \ldots \geq 0.$$

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

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

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,

$$f_n(x) = \Big(1 + \frac{x}{n}\Big)^n \qquad \text{für } x \in \mathbb{R},\, n \in \mathbb{N}. \tag{5.2}$$

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

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

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

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

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

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

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

Eine gegebene Zahl a hat somit eine Darstellung als Dezimalbruch. Umgekehrt stellt sich aber die Frage: seien $a_0 \in \mathbb{Z}$ und $k_1, k_2, \ldots$ beliebig mit $k_j \in \{0, 1, \ldots, 9\}$. Konvergiert dann die Dezimalbruchfolge $a_n = a_0, k_1 k_2 \ldots k_n$ 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[13] besteht darin, die Glieder der Folge nicht mit dem unbekannten 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 $\lim_{n\to\infty} a_n = a$. 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.

[13] Augustin Louis Cauchy, 1789–1857

Beispiel 5.4 (Konvergenz von Dezimalbrüchen) Seien $a_0 \in \mathbb{Z}$, $k_j \in \{0, 1, \ldots, 9\}$ 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 $0,99\ldots = 1,00\ldots$. 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

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

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

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

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 $x_1 \geq x_2 \geq \ldots \geq 0$. 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

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

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

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

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:

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

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*

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{1}{k!}. \tag{5.3}$$

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 $e_3 = 2{,}666\ldots$ und $e_5 = 2{,}716\ldots$. Wir zeigen nun, dass die Folge f_n ebenfalls den Grenzwert e hat. Aus $f_n \leq e_n$ ergibt sich $\lim_{n\to\infty} f_n \leq e$. Wir zeigen andererseits für beliebiges $m \in \mathbb{N}$

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

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

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

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

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

Die Eulersche Zahl hat somit die alternative Darstellung

$$e = \lim_{n\to\infty} \Big(1 + \frac{1}{n}\Big)^n. \tag{5.4}$$

Wird ein Euro für ein Jahr mit 100% angelegt, so hat man bei linearer Verzinsung 2 Euro, bei kontinuierlichem Zinseszins dagegen $e = 2,71828\ldots$ 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

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

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 $|I_n| = b_n - a_n \to 0$ mit $n \to \infty$.*

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 $x = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$.

Beweis: Nach Voraussetzung ist $a_1 \leq a_2 \leq \ldots$ und $b_1 \geq b_2 \geq \ldots$. 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 $a = \lim_{n\to\infty} a_n$ bzw. $b = \lim_{n\to\infty} b_n$. Dann folgt nach Satz 4.3 und Satz 4.4

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

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 $f : I \to \mathbb{R}$ auf einem Intervall I heißt *streng monoton wachsend*, wenn folgende Implikation gilt:

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

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 $f : I \to f(I)$, also jetzt mit Zielbereich $f(I)$ statt $\mathbb{R}$, ist dann injektiv und surjektiv, das heißt es gibt die Umkehrfunktioa $g : f(I) \to I$. 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

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

In unserem Fall ist $f : I = [0, \infty) \to \mathbb{R}$, $f(x) = x^n$, 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 $g : f(I) \to I$ 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:

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

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 $k+1$:

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

Weiter ist $I_{k+1} \subset I_k$ und $|I_k| = 2^{1-k}|I_1| \to 0$, 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 $a_k, b_k \to x$ mit $k \to \infty$. Durch Grenzübergang in (5.7) folgt

$$x^n = \lim_{n\to\infty} a_k^n \le y \le \lim_{n\to\infty} b_k^n = x^n.$$

Die Eindeutigkeit der n-ten Wurzel und die Monotonie der Funktion wurden oben gezeigt. □

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

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

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

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

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 $s = p/q$

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

Dies bedeutet

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

Die anderen beiden Potenzgesetze werden ähnlich verifiziert.

Beispiel 5.7 Ein Grenzwert, der häufiger auftritt, ist

$$\lim_{n\to\infty} a^{1/n} = 1 \qquad \text{für } a > 0. \tag{5.8}$$

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

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

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

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

Es folgt $\lim_{n\to\infty} a^{1/n} = 1/\lim_{n\to\infty}(a^{-1})^{1/n} = 1$ 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 $n_1 < n_2 < n_3 \ldots$ 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 $n \longmapsto a_n$ und der Folge $k \longmapsto n_k$ die die Nummern aussucht. Am Beispiel $a_n = (-1)^n/n^3$ und $n_k = 2k-1$ sieht das so aus:

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

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 $k \to \infty$ gegen a konvergiert.*

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

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 $a_{n_k} \to a$ mit $k \to \infty$. 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 $n_1 < n_2 < \ldots$ ist die Menge $\{n_k : k > K\}$ nicht endlich, die eine Richtung ist gezeigt. Für die umgekehrte Aussage wählen wir induktiv n_k mit $n_1 < n_2 < \ldots$, 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 $|a_{n_k} - a| < 1/k \to 0$ mit $k \to \infty$. $\square$

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 $k > K$

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

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 $\limsup_{n\to\infty} x_n = x^*$, falls folgende zwei Bedingungen erfüllt sind:*

(i) *es gibt eine Teilfolge x_{n_k} mit $x_{n_k} \to x^*$ für $k \to \infty$,*

(ii) *für alle $x > x^*$ ist die Menge $\{n \in \mathbb{N} : x_n > x\}$ endlich.*

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

(i) *es gibt eine Teilfolge x_{n_k} mit $x_{n_k} \to x_*$ für $k \to \infty$,*

(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 $x_{2k} = 1 + \frac{1}{4k^2} \to 1$, 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 $\limsup_{n\to\infty} x_n = 1$.

Ist $x^* = \limsup_{n\to\infty} x_n \in \mathbb{R}$, 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 $\limsup_{n\to\infty} x_n = -\infty$ gilt schon $\lim_{n\to\infty} x_n = -\infty$, wieder nach (ii). Dagegen muss im Fall $\limsup_{n\to\infty} x_n = +\infty$ 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 $\limsup_{n\to\infty} x_n = x^*$.*

Beweis: Wir behandlen erst die Fälle wenn $\limsup_{n\to\infty} x_n = \pm\infty$ 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 $n_1 < n_2 < \ldots$ mit $x_{n_k} \geq k$. Es folgt $\limsup_{n\to\infty} x_n = +\infty$.

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 $x_n \to -\infty$ und es folgt $\limsup_{n\to\infty} x_n = -\infty$.

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

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

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:

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

Nun gilt $b_k \to x^*$ für $k \to \infty$. 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 $\limsup_{n\to\infty} x_n = x^*$, 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:

Vollständigkeitsaxiom:
Cauchyfolgen sind konvergent
$\Downarrow$
Konvergenzkriterium der
Monotonie und Beschränktheit
$\Downarrow$
Intervallschachtelungsprinzip
$\Downarrow$
Bolzano-Weierstraß: Auswahlsatz
$\Downarrow$
Cauchyfolgen sind konvergent

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 $a_{n_k} \to a$ für $k \to \infty$. 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 $k \to \infty$, also $n_k \to \infty$, folgt

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

Damit ist die Implikation bewiesen. $\square$

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

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

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

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

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

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

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

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

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 $a_n \to S$ sowie $b_n \to S$. Für $x \in M$ sieht man durch Induktion $x \leq b_n$ für alle n, also $x \leq \lim_{n\to\infty} b_n = S$. 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. □

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

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 $x_n \to +\infty = \sup M$. □

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) $\varphi : A \to B$ 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) $A \sim B \quad \Rightarrow \quad B \sim A$

(iii) $A \sim B$ und $B \sim C \quad \Rightarrow \quad A \sim C$.

Beweis:

(i) Wähle als Bijektion die Identität $id_A : A \to A$, $a \longmapsto a$.

(ii) Sei $\varphi : A \to B$ Bijektion. Wähle dann $\varphi^{-1} : B \to A$.

(iii) Seien $\varphi : A \to B$ und $\psi : B \to C$ bijektiv. Wähle dann $\psi \circ \varphi : A \to C$.

□

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 $\{1, \ldots, n\}$. Eine weitere Äquivalenzklasse liefert die Menge $\mathbb{N}$, denn nach dem Schubfachprinzip, Satz 3.3, gibt es für $n \in \mathbb{N}$ keine Bijektion $\mathbb{N} \leftrightarrow \{1, \ldots, n\}$.

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 $\varphi : \mathbb{N} \to A$ existiert, so ist A abzählbar.*

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

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

Falls die Rekursion nicht abbricht erhalten wir $f : \mathbb{N} \to A$, $f(j) = a_{n_j}$. 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 $f : \{1, \ldots, k\} \to A$, $f(j) = a_{n_j}$. 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

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

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

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

Durch Verkettung mit $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}^+$, $f(p,q) = p/q$, 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:

$$\begin{array}{llcccccc}
 & p = & 1 & 2 & 3 & 4 & 5 & 6 \\
q = & & & & & & & \\
1 & & (1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1) \\
 & & & \swarrow & \swarrow & \swarrow & \swarrow & \swarrow \\
2 & & (1,2) & (2,2) & (3,2) & (4,2) & \cdot & \\
 & & & \swarrow & \swarrow & \swarrow & \cdot & \\
3 & & (1,3) & (2,3) & (3,3) & \cdot & & \\
 & & & \swarrow & \swarrow & \cdot & & \\
4 & & (1,4) & (2,4) & \cdot & & & \\
 & & & \swarrow & \cdot & & & \\
5 & & (1,5) & & & & &
\end{array}$$

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 $1, \ldots, k-1$ enthalten insgesamt $1 + \ldots + (k-1) =: N_k$ Einträge ($N_k = k(k-1)/2$). Damit bekommt (p, q) die Nummer

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

Um $\varphi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ 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 $N_1 < N_2 < \ldots$.
- 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 φ 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 $\varphi : \mathbb{N} \to \mathbb{R}$.*

Beweis: Sei $\varphi : \mathbb{N} \to \mathbb{R}$ 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 φ 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 bezeichet 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 Mengenaxiomatik (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 $90°$ 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 $\overline{z} = x - iy \in \mathbb{C}$ die zu z konjugiert komplexe Zahl. Anschaulich ergibt sich $\overline{z}$ aus z durch Spiegelung an der x-Achse, insbesondere gilt $\overline{\overline{z}} = \overline{x - iy} = x + iy = z$.

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

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

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

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

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

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

□

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) $|z|^2 = z\,\overline{z}$.

(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 $z = x + iy$

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

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

$$|z_1 z_2|^2 = (z_1 z_2)(\overline{z_1 z_2}) = z_1\,\overline{z_1}\,z_2\overline{z_2} = |z_1|^2\,|z_2|^2.$$

Die Ungleichungen (3) sind offensichtlich. $\square$

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

$$\frac{1}{z} = \frac{\overline{z}}{z\,\overline{z}} = \frac{x-iy}{x^2+y^2} \quad \text{für } z = x+iy \neq 0. \tag{6.3}$$

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

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

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 $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ ist*

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

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 $x_1^2+\ldots+x_n^2 = 0$, also $x_i = 0$ für alle $i = 1,\ldots,n$ und damit $x = 0$. Die Skalarmultiplikation im $\mathbb{R}^n$ ist komponentenweise definiert durch $\lambda(x_1,\ldots,x_n) = (\lambda x_1,\ldots,\lambda x_n)$, 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 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*

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

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*

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

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 $\|\xi\|, \|\eta\| = 1$,

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

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. □

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 Dreeick 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 Beschränkheit, ε-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 $x_1, \ldots, x_n$ 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 $i = 1, \ldots, n$ die Koordinatenfolgen $\big((x_k)_i\big)_{k\in\mathbb{N}}$ konvergent (beschränkt, Cauchyfolgen) sind.*

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

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

Sei nun $x_k \to x$, also $\|x_k - x\| \to 0$. Dann folgt $|(x_k)_i - x_i| \leq \|x_k - x\|_{\max} \to 0$ für $i = 1, \ldots n$. Umgekehrt: angenommen $(x_k)_i \to x_i$ für $i = 1, \ldots, n$. 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 $x = (x_1, \ldots, x_n)$ 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 Beschränkheit 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* $x_k \to a$ *mit* $k \to \infty$*.*

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

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 $f : D \to \mathbb{R}$, 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:*

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

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 $f : D \to \mathbb{R}^m$ ist in jedem $x_0 \in D$ stetig, denn

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

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

Beispiel 7.2 Für $f : \mathbb{R} \to \mathbb{R}$, $f(x) = ax + b$ 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 δ 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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, 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

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

Jetzt wählen wir $\delta = \min\big(1, \varepsilon/(1 + 2|x_0|)\big)$. 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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$, 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

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

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

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

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

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

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

- 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 $(x_0 - \delta, x_0 + \delta) \subset \mathbb{R}^n \backslash E$.

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 $\mathbb{R} \backslash \mathbb{Q}$ ist ebenfalls dicht: zum Beispiel ist $\sqrt{2} + \mathbb{Q}$ dicht in $\mathbb{R}$ und Teilmenge von $\mathbb{R} \backslash \mathbb{Q}$, 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 $\mathbb{R} \backslash \mathbb{Q}$. 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 $x = (x_1, \ldots, x_n)$. Eventuell kann $f(x)$ auch Werte in $\mathbb{R}^m$ haben, das heißt es ist $f(x) = (f_1(x), \ldots, f_m(x))$. Es ist naheliegend, in der Definition der Stetigkeit den Betrag durch die Euklidische Norm zu ersetzen.

Definition 7.2 (Stetigkeit) *Die Funktion $f : D \to \mathbb{R}^m$, 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:*

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

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:

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

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

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

Wir können $\delta = \varepsilon$ nehmen.

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

Beispiel 7.7 $f : D \to \mathbb{R}^m$ heißt Lipschitzstetig mit Konstante $L \in [0,\infty)$, falls gilt:[14]

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

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

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

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

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

Durch Vertauschen von x_1 mit x_2 ergibt sich dist_E Lipschitzstetig, genauer

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

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

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

Eine lineare Abbildung vertauscht mit Linearkombinationen, das heißt es gilt

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

Wir schätzen mit der Dreiecksungleichung und Cauchy-Schwarz ab:

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

Also erhalten wir die Abschätzung

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

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

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

[14] 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 $e_1, \ldots, e_m$ des $\mathbb{R}^m$:

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

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

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

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 $f : D \to \mathbb{R}^m$ 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 $\lim_{k\to\infty} x_k = x_0$ gilt: $\lim_{k\to\infty} f(x_k) = f(x_0)$.*

Beweis: Für die Implikation (1) $\Rightarrow$ (2) sei $x_k \in D$ mit $x_k \to x_0$ für $k \to \infty$. 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 $k = 1, 2, \ldots$, 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 $f : D \to \mathbb{R}^m$, $f(D) \subset E \subset \mathbb{R}^m$, und $g : E \to \mathbb{R}^k$. Ist f stetig in x_0 und g stetig in $y_0 = f(x_0)$, so ist $g \circ f : D \to \mathbb{R}^k$ stetig in x_0.*

Beweis: Wir verwenden Satz 7.1. Ist $x_n \in D$ eine Folge mit $\lim_{n\to\infty} x_n = x_0$, so folgt $f(x_n) \to f(x_0)$ aus der Stetigkeit von f in x_0, und weiter $g(f(x_n)) \to g(f(x_0))$ 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 $f, g : D \to \mathbb{R}$ stetig. Dann ist auch die Funktion $|f| : D \to \mathbb{R}$ 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 $f : D \to \mathbb{R}^m$ ist genau dann stetig in $x_0 \in D$, wenn jede Koordinatenfunktion $f_i : D \to \mathbb{R}$, $1 \leq i \leq m$, 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 $c : I \to \mathbb{R}^m$, $c = c(t)$, 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:

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

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

Satz 7.4 (Stetigkeitsregeln) *Seien $f, g : D \to \mathbb{R}$ 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 $f/g : D \cap B_\delta(x_0) \to \mathbb{R}$ 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 $\lim_{k\to\infty} x_k = x_0$. Nach Satz 7.1 gilt dann $f(x_k) \to f(x_0)$ und $g(x_k) \to g(x_0)$ mit $k \to \infty$. Es folgt mit Satz 4.3 für $k \to \infty$

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

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

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

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

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

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

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 $f : D \to \mathbb{R}$ 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 Unterverktorraum des Raums aller Funktionen $f : D \to \mathbb{R}$, 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 $f : D \to \mathbb{R}^m$ konvergiert für $x \to x_0$ 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: $\lim_{x\to x_0} f(x) = a$ oder $f(x) \to a$ für $x \to x_0$.

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 $\lim_{x\to x_0} f(x)$ 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 $f : D \to \mathbb{R}^m$ sind folgende Aussagen äquivalent:*

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

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

(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. $\square$

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

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

Zum Beispiel ist $\lim_{x\searrow 0} \operatorname{sign}(x) = +1$ und $\lim_{x\nearrow 0} \operatorname{sign}(x) = -1$, während der Grenzwert $\lim_{x\to 0} \operatorname{sign}(x)$ 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 $x \to x_0$.

Der Grenzwertbegriff für Funktionen lässt sich wieder auf Folgen zurückführen.

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

(1) *$f(x) \to a$ für $x \to x_0$.*

(2) *$\lim_{k\to\infty} f(x_k) = a$ für jede Folge $x_k \in D\backslash\{x_0\}$ mit $x_k \to x_0$.*

Beweis: Der Beweis ist analog zu Satz 7.1. In Definition 7.3 kommen nur 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 $f, g : D \to \mathbb{R}$ mit $f(x) \to a$, $g(x) \to b$ für $x \to x_0$. Dann folgt*

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

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

(3) *Sei $f : D \to \mathbb{R}$ mit $f(x) \to a$ für $x \to x_0$. Ist $f \geq 0$ auf $B_\delta(x_0) \backslash \{x_0\}$, 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 $f : D \to \mathbb{R}$. Dann gilt $\lim_{x \to x_0} f(x) = +\infty$, wenn es zu jedem $K > 0$ ein $\delta > 0$ gibt mit*

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

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

Beispiel 7.12 Die Funktion $f(x) = 1/|x|$, $x \neq 0$, hat den Grenzwert $\lim_{x \to 0} f(x) = +\infty$. 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 $f : D \to (0, \infty)$, so ist $\lim_{x \to x_0} f(x) = +\infty$ äquivalent zu $\lim_{x \to x_0} 1/f(x) = 0$.

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

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

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

Analog wird der Grenzwert für $x \to -\infty$ 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 $p : \mathbb{K} \to \mathbb{K}$ heißt (reelles oder komplexes) Polynom vom Grad $n \in \mathbb{N}_0$, wenn es $a_0, a_1, \ldots, a_n \in \mathbb{K}$ mit $a_n \neq 0$ gibt, so dass gilt:*

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

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 $p : \mathbb{K} \to \mathbb{K}$ Polynom vom Grad n mit Koeffizienten $a_0, \ldots, a_n$. Ist $p(\lambda) = 0$ für ein $\lambda \in \mathbb{K}$, so gibt es ein Polynom $q : \mathbb{K} \to \mathbb{K}$ vom Grad $n-1$ mit Koeffizienten $b_0, \ldots, b_{n-1}$ wobei $b_{n-1} = a_n$, so dass gilt:*

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

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,

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

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 $a_0 = a_1 = \ldots = a_n = 0$ und $p(z)$ ist die Nullfunktion. Manche Autoren sprechen auch vom *Nullpolynom*, aber das ist Geschmackssache.

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

$$p(z) = \sum_{i=0}^{m} a_i z^i \ \textit{und}\ q(z) = \sum_{i=0}^{n} b_i z^i \quad \textit{für alle } z \in \mathbb{K}.$$

Ist $p(z) = q(z)$ an mehr als $\max(m, n)$ Stellen, so folgt $n = m$ und $b_i = a_i$ für $i = 0, 1, \ldots, m$.

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 $p : \mathbb{K} \to \mathbb{K}$ vom Grad n hat eine eindeutige Zerlegung*

$$p(z) = (z - \lambda_1)^{\nu_1} \cdot \ldots \cdot (z - \lambda_r)^{\nu_r}\, q(z) \quad \textit{für alle } z \in \mathbb{K}. \tag{7.5}$$

Dabei sind $\{\lambda_1, \ldots, \lambda_r\} \subset \mathbb{K}$ die Nullstellen von p in $\mathbb{K}$ (eventuell $r = 0$), und $q : \mathbb{K} \to \mathbb{K}$ ist ein Polynom vom Grad $n - (\nu_1 + \ldots + \nu_r) \in \{0, \ldots, n\}$ mit $q(z) \neq 0$ für alle $z \in \mathbb{K}$. Die Zahlen $\nu_i \geq 1$ heißen Vielfachheiten der Nullstellen λ_i.

Beweis: Die Existenz der Zerlegung folgt induktiv aus Lemma 7.3. Für die Eindeutigkeit betrachte eine zweite solche Zerlegung $p(z) = (z-\lambda_1)^{\mu_1} \cdot \ldots \cdot (x-\lambda_r)^{\mu_r} h(z)$, ohne Einschränkung mit $\nu_1 \geq \mu_1$. Indem wir durch $(z-\lambda_1)^{\mu_1}$ teilen, folgt für alle $z \neq \lambda_1$

$$(z-\lambda_1)^{\nu_1-\mu_1} \cdot (z-\lambda_2)^{\nu_2} \cdot \ldots \cdot (z-\lambda_r)^{\nu_r} \cdot q(z) = (z-\lambda_2)^{\mu_2} \cdot \ldots \cdot (z-\lambda_r)^{\mu_r} \cdot h(z).$$

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 $i \in \{1, \ldots, r\}$. Nach Division haben wir schließlich $q(z) = h(z)$ für alle $z \in \mathbb{K} \backslash \{\lambda_1, \ldots, \lambda_r\}$, und Folgerung 7.2 liefert $q = h$. □

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

$$f : \mathbb{R} \backslash N \to \mathbb{R},\ f(x) = \frac{p(x)}{q(x)}.$$

In welchen $x_0 \in N$ hat f eine stetige Fortsetzung? Sei x_0 eine μ-fache Nullstelle von $p(x)$ und eine ν-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

$$f(x) = (x-x_0)^{\mu-\nu} \frac{\tilde{p}(x)}{\tilde{q}(x)} \qquad \text{für alle } x \in \mathbb{R} \backslash N,$$

Aus den Rechenregeln für Grenzwerte folgt

$$\lim_{x \to x_0} f(x) = \begin{cases} 0 & \text{falls } \mu > \nu, \\ \dfrac{\tilde{p}(x_0)}{\tilde{q}(x_0)} \neq 0 & \text{falls } \mu = \nu. \end{cases}$$

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 $x \to \pm\infty$ ist ähnlich. Wir schreiben für $|x|$ groß

$$f(x) = x^{m-n} \frac{a_m + a_{m-1} x^{-1} + \ldots + a_0 x^{-m}}{b_n + b_{n-1} x^{-1} + \ldots + b_0 x^{-n}}.$$

Es folgt $f(x) \to 0$ wenn $m < n$, und $f(x) \to a_m/b_n$ wenn $m = n$. Für $m > n$ sei zum Beispiel $a_m/b_n > 0$, Dann gilt $f(x) \to +\infty$ für $x \to +\infty$ und $f(x) \to \pm\infty$ für $x \to -\infty$, je nachdem ob $m - n$ gerade oder ungerade ist.

8 Zwischenwertsatz und monotone Funktionen

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:

$$a = \inf I,\, b = \sup I \quad \text{ und } (a,b) \subset I.$$

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 $f : [a,b] \to \mathbb{R}$ 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 $x_n \to x_0$ und wegen Stetigkeit $f(x_0) = \lim_{n\to\infty} f(x_n) \leq y_0$. 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 $f : I \to \mathbb{R}$ 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 $f : I \to \mathbb{R}$ monoton wachsend, so existieren in jedem $x_0 \in I$ die einseitigen Grenzwerte, genauer gilt*

$$\begin{aligned} \lim_{x\nearrow x_0} f(x) &= \sup\{f(x) : x \in I,\, x < x_0\} \quad \text{falls } x_0 > a, \\ \lim_{x\searrow x_0} f(x) &= \inf\{f(x) : x \in I,\, x > x_0\} \quad \text{falls } x_0 < b. \end{aligned}$$

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 $f : I \to \mathbb{R}$ streng monoton wachsend. Aus $f(x_1) = f(x_2)$ folgt dann $x_1 = x_2$, das heißt f ist injektiv und die Umkehrfunktion $g : f(I) \to I$ 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

$$y_2 = f\big(g(y_2)\big) \leq f\big(g(y_1)\big) = y_1,$$

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 $f : I \to \mathbb{R}$ 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 $g : f(I) \to \mathbb{R}$ 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

$$a \in I \Leftrightarrow \alpha \in f(I) \quad \text{und} \quad b \in I \Leftrightarrow \beta \in f(I). \tag{8.1}$$

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 $f : [0, \infty) \to \mathbb{R}$, $f(x) = x^n$, 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 $g : [0, \infty) \to \mathbb{R}$, $g(y) = y^{1/n}$, 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

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 $f : I \to \mathbb{R}^n$ hat in $x_0 \in I$ die Ableitung $a \in \mathbb{R}^n$ (Notation: $f'(x_0) = a$), falls gilt:*

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = a. \tag{9.1}$$

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 $x = x_0 + h$:

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

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 $f : I \to \mathbb{R}$ 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 $x \to x_0$ 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 $f : I \to \mathbb{R}^n$ ist genau dann in $x_0 \in I$ differenzierbar, wenn alle Koordinatenfunktionen $f_i : I \to \mathbb{R}$, $i = 1, \ldots, n$, in x_0 differenzierbar sind. Die Ableitung kann dann koordinatenweise berechnet werden, das heißt es gilt:*

$$f'(x_0) = ((f_1)'(x_0), \ldots, (f_n)'(x_0)) \in \mathbb{R}^n.$$

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

$$f : I \to \mathbb{R}^3, \quad f(t) = \big(x(t), y(t), z(t)\big),$$

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 $f : I \to \mathbb{R}^n$ heißt differenzierbar auf I (oder einfach differenzierbar), falls f in jedem Punkt $x_0 \in I$ differenzierbar ist. Die hierdurch gegebene Funktion*

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

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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$, 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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$, 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 $f : I \to \mathbb{R}^n$ differenzierbar in x_0, so ist f auch stetig in x_0.*

Beweis: Mit den Grenzwertregeln, Satz 7.6(1), folgt für $x \to x_0$

$$f(x) = f(x_0) + \underbrace{\frac{f(x) - f(x_0)}{x - x_0}}_{\to f'(x_0)} \underbrace{(x - x_0)}_{\to 0} \longrightarrow f(x_0).$$

Dies zeigt $\lim_{x \to x_0} f(x) = f(x_0)$, also f stetig in x_0 nach Lemma 7.2. $\square$

Satz 9.2 (Differentiationsregeln) *Seien $f, g : I \to \mathbb{R}$ 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 $x \to x_0$ gegen das gewünschte konvergieren. Für (1) haben wir

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

Natürlich gilt die Aussage mit demselben Argument auch für vektorwertige Funktionen. Die Produktregel folgt durch „Mischen der Terme“:

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

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

$$\begin{aligned}
\frac{1}{x - x_0}\left(\frac{1}{g(x)} - \frac{1}{g(x_0)}\right) &= -\frac{1}{g(x)g(x_0)} \frac{g(x) - g(x_0)}{x - x_0} \\
&\to -\frac{1}{g(x_0)^2}\, g'(x_0) \quad \text{mit } x \to x_0.
\end{aligned}$$

Für beliebiges f schreiben wir $f/g = f \cdot \dfrac{1}{g}$ und verwenden die Produktregel. □

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 $f : I \to \mathbb{R}$, $g : J \to \mathbb{R}$ 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 $g \circ f : I \to \mathbb{R}$ in x_0 differenzierbar und hat die Ableitung*

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

Beweis: Wir verwenden die Charakterisierung des Grenzwerts mittels Folgen. Sei $x_n \neq x_0$ mit $x_n \to x_0$, also $f(x_n) \to f(x_0)$ wegen f stetig in x_0. Ist $f(x_n) \neq f(x_0)$ für alle n, so folgt

$$\frac{g(f(x_n)) - g(f(x_0))}{x_n - x_0} = \frac{g(f(x_n)) - g(f(x_0))}{f(x_n) - f(x_0)}\,\frac{f(x_n) - f(x_0)}{x_n - x_0} \to g'(f(x_0))f'(x_0).$$

Ist dagegen $f(x_n) = f(x_0)$ für alle n, so gilt $f'(x_0) = \lim\limits_{n\to\infty} \dfrac{f(x_n) - f(x_0)}{x_n - x_0} = 0$ und es folgt

$$\lim_{n\to\infty} \frac{g(f(x_n)) - g(f(x_0))}{x_n - x_0} = 0 = g'(f(x_0))f'(x_0).$$

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. □

Satz 9.4 (Differenzierbarkeit der Umkehrfunktion) *Sei $f : I \to \mathbb{R}$ 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 $g : I^* \to I$ ist differenzierbar in $y_0 = f(x_0)$ mit Ableitung*

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

Beweis: Nach Satz 8.2 ist I^* ein offenes Intervall und $g : I^* \to \mathbb{R}$ streng monoton und stetig, insbesondere $g(y) \to g(y_0) = x_0$ mit $y \to y_0$. Wir verwenden wieder die Charakterisierung des Grenzwerts mittels Folgen. Sei $y_n \neq y_0$ mit $y_n \to y_0$, dann ist $g(y_n) \neq g(y_0)$ und

$$\frac{g(y_n) - g(y_0)}{y_n - y_0} = \frac{g(y_n) - g(y_0)}{f(g(y_n)) - f(g(y_0))} = \frac{1}{\dfrac{f(g(y_n)) - f(g(y_0))}{g(y_n) - g(y_0)}} \to \frac{1}{f'(x_0)}.$$

□

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 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, die Ableitung $f'(0) = 0$, die Umkehrfunktion g kann daher im Nullpunkt nicht differenzierbar sein. Das lässt sich auch direkt überprüfen, es gilt

$$g(y) = \begin{cases} y^{1/3} & \text{für } y \geq 0 \\ -|y|^{1/3} & \text{für } y \leq 0 \end{cases} \qquad \text{also} \qquad \frac{g(y) - g(0)}{y - 0} = |y|^{-2/3} \to +\infty \text{ mit } y \to 0.$$

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 $f : (0, \infty) \to \mathbb{R}$, $f(x) = x^n$, $n \in \mathbb{N}$, ist streng monoton wachsend und stetig mit $f'(x) = nx^{n-1} > 0$, siehe Beispiel 9.4. Die Umkehrfunktion ist $g : (0, \infty) \to \mathbb{R}$, $g(y) = y^{1/n}$. 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 vorangegangengen 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:

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

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 affinlineare 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 $f : I \to \mathbb{R}^n$ in $x_0 \in I$ die Ableitung $a \in \mathbb{R}^n$, wenn gilt:*

$$\lim_{x \to x_0} \frac{f(x) - \big(f(x_0) + a(x - x_0)\big)}{x - x_0} = 0. \tag{9.2}$$

Beweis: Folgt sofort aus der Umformung

$$\frac{f(x) - \big(f(x_0) + a(x - x_0)\big)}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} - a.$$

□

Um für zwei Funktionen f, g das asymptotische Verhalten für $x \to x_0$ zu vergleichen, werden manchmal die Landauschen Symbole benutzt:

$$\begin{aligned} f = \mathcal{O}(g) \text{ für } x \to x_0 \quad &\Leftrightarrow \quad \limsup_{x \to x_0} \frac{|f(x)|}{|g(x)|} < \infty \quad (f \text{ ist abgeschätzt durch } g), \\ f = o(g) \text{ für } x \to x_0 \quad &\Leftrightarrow \quad \lim_{x \to x_0} \frac{|f(x)|}{|g(x)|} = 0 \qquad (f \text{ wird klein relativ zu } g). \end{aligned}$$

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:

$$f(x) = f(x_0) + a(x - x_0) + o(x - x_0) \qquad \text{für } x \to x_0.$$

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

Es ist eine zentrales Problem der Analysis, aus Eigenschaften der Ableitung auf Eigenschaften der Funktion selbst zu schließen. Der Mittelwertsatz ist dafür ein einfaches und effektives Hilfsmittel. Für seinen Beweis müssen wir allerdings erst über Extremwerte – Maxima und Minima – von stetigen Funktionen sprechen.

Wir benötigen einen Satz, der die Existenz von Extremalstellen für eine stetige Funktion $f : I \to \mathbb{R}$ (natürlich mit $I \neq \emptyset$) allgemein garantiert. Oft wird das so formuliert, dass die Funktion ihr Minimum bzw. Maximum annimmt. Das ist allerdings misverständlich, 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:

$$f : I = (0, \infty) \to \mathbb{R},\ f(x) = \frac{1}{1+x}.$$

Weder das Supremum von f noch das Infimum von f wird hier angenommen.

Satz 10.1 (Existenz von Extremalstellen) *Sei $I = [a, b]$ ein abgeschlossenes und beschränktes Intervall, und $f : I \to \mathbb{R}$ sei stetig. Dann ist f beschränkt und nimmt Infimum und Supremum an, das heißt es gibt $x_0, x_1 \in I$ mit*

$$f(x_0) = \inf_{x \in I} f(x) \quad \text{und} \quad f(x_1) = \sup_{x \in I} f(x).$$

BEWEIS: Wir zeigen die Existenz eines $x_0 \in I$ mit $f(x_0) = \inf_I f$. Insbesondere ist f dann nach unten beschränkt, denn $f(x) \geq f(x_0)$ für alle $x \in I$. Nach Definition des Infimums, siehe Folgerung 6.1, gibt es nun eine Folge $x_k \in I$ mit $f(x_k) \to \inf_I f$, eine sogenannte *Minimalfolge*. Bolzano-Weierstraß, siehe Satz 5.7, liefert eine Teilfolge x_{k_j} und ein $x_0 \in \mathbb{R}$ mit $x_{k_j} \to x_0$ für $j \to \infty$. Es folgt $x_0 \in [a, b]$, und wegen f stetig gilt

$$f(x_0) = \lim_{j \to \infty} f(x_{k_j}) = \inf_I f.$$

Für das Supremeum argumentieren wir entsprechend. □

Definition 10.1 (Lokale Extrema) *Die Funktion $f : (a, b) \to \mathbb{R}$ 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 10.1 konstruiert. Bei einer Wanderung freut man sich zum Beispiel, wenn man auf dem Weg einen schönen Aussichtspunkt erreicht. Muss man dann aber noch 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 10.2 (notwendige Bedingung für Extrema) *Die Funktion $f : (a,b) \to \mathbb{R}$ 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

$$\begin{aligned} f(x) \geq f(x_0) \text{ auf } [x_0, x_0 + \delta) \quad &\Rightarrow \quad f'_+(x_0) := \lim_{x \searrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \geq 0, \\ f(x) \geq f(x_0) \text{ auf } (x_0 - \delta, x_0] \quad &\Rightarrow \quad f'_-(x_0) := \lim_{x \nearrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \leq 0. \end{aligned}$$

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 10.3 (Mittelwertsatz der Differentialrechnung) *Sei $f : [a,b] \to \mathbb{R}$ 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 10.1 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 10.2 und wir können $\xi = \xi_1$ wählen. Analog, wenn $\xi_2 \in (a,b)$. Es bleibt der Fall wenn ξ_1, ξ_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 $h : [a,b] \to \mathbb{R}$ 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:*

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

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:

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

□