Analysis I

Wintersemester 2019/2020

Ernst Kuwert

Mathematisches Institut
Universität Freiburg

Inhaltsverzeichnis

1 Arithmetik und Anordnung in $\mathbb{R}$

Startpunkt dieser Vorlesung ist eine axiomatische Beschreibung der rellen Zahlen. Danach sind diese eine Menge, bezeichnet mit $\mathbb{R}$, 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* $ab \in \mathbb{R}$ *zuordnet (*Multiplikation*),*
eine Relation $a > b$*, die für* $a, b \in \mathbb{R}$ *zutrifft oder nicht (*Anordnung*).*

Dabei sollen eine Reihe von Axiomen (Regeln) gelten, welche die reellen Zahlen charakterisieren. Diese sind in drei Gruppen unterteilt:

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

Explizite Konstruktionen einer Menge mit diesen Eigenschaften wurden von Cantor und Dedekind angegeben. Damit wissen wir dass sich die Axiome nicht widersprechen und dass es die reellen Zahlen wirklich gibt. Beim Umgang mit ihnen werden die speziellen Modelle von Cantor und Dedekind aber in aller Regel nicht benutzt, sondern die Analysis wird nur aus den Axiomen hergeleitet; das ist auch unser Zugang. Die Konstruktion von Cantor bzw. Dedekind gehört zum Fundament der Analysis. Es wäre abstrakt und würde uns ablenken, wenn wir uns jetzt damit beschäftigten. Lustigerweise ist die Analysis ein Haus, bei dem das Fundament erst zuletzt gebaut wurde. Zum Beispiel datiert der Hauptsatz der Differential- und Integralrechnung von Newton und Leibniz auf etwa 1680, während die Definition der reellen Zahlen erst 1895 nachgeholt wurde. Das hat damit zu tun, dass es den Schöpfern der Theorie vor allem um die vielen Anwendungen ging, die die Analysis hat. Die Arbeiten von Cantor und Dedekind sichern das Haus der Analysis sozusagen ab, spielen aber im Leben dieses Hauses keine Rolle. Bei Neugierde finden Sie Informationen zur Konstruktion von $\mathbb{R}$ in Kapitel 2 des Buchs *Zahlen* [1].

Wir beginnen nun mit den Körperaxiomen, die die arithmetischen Eigenschaften von $\mathbb{R}$ regeln, also Addition und Multiplikation. Sie lauten wie folgt:

	$+$	$\bullet$
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:	$a \cdot (b + c) = a \cdot b + a \cdot c.$	

[1] Hrsgb. H.-D. Ebbinghaus, 3. Auflage, Springer 1992

Eine Menge mit Verknüpfungen + und $\cdot$, so dass die Axiome (K) erfüllt sind, heißt Körper (englisch: *field*). $\mathbb{R}$ ist keineswegs der einzige Körper, zum Beispiel gelten die Axiome auch in den rationalen Zahlen $\mathbb{Q}$. 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 1.1 (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

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

Seien nun $a, b \in \mathbb{R}$ gegeben. Aus $x + a = b$ folgt

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

Die Zahl $b + (-a)$ löst tatsächlich die Gleichung, denn

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

$\square$

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

Satz 1.2 (Rechnen in $\mathbb{R}$) *Für reelle Zahlen a, b gelten folgende Aussagen:*

$$\left\{\begin{array}{lll} -(-a) = a & & -(a + b) = (-a) + (-b), \\ (a^{-1})^{-1} = a & (a \neq 0) & (ab)^{-1} = a^{-1}b^{-1} \quad (a, b \neq 0), \\ a \cdot 0 = 0 & & a(-b) = -(ab), \\ (-a)(-b) = ab & & a(b - c) = ab - ac. \end{array}\right. \tag{1.1}$$

$$ab = 0 \quad \Rightarrow \quad a = 0 \;\; \textit{oder} \;\; b = 0 \qquad \textit{(Nullteilerfreiheit)}. \tag{1.2}$$

Beweis: Die Zahl $-b$ ist das eindeutig bestimmte inverse Element zu b bezüglich der Addition. Die erste Zeile in (1.1) 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 1.1(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 (1.1) gilt mit

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

Ist in (1.2) $a \neq 0$, so folgt schließlich

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

Damit sind alle Aussagen des Satzes gezeigt. □

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

Satz 1.3 (Bruchrechnung) *Für $a, b, c, d \in \mathbb{R}$ mit $c, d \neq 0$ gilt:*

(1) $\frac{a}{c} + \frac{b}{d} = \frac{ad + bc}{cd}$,

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

(3) $\frac{a/c}{b/d} = \frac{ad}{bc}$, *falls zusätzlich $b \neq 0$.*

Wir kommen nun zu den Anordnungsaxiomen, und zwar formulieren wir

(**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 $ab > 0$.

Eine drittes Axiom, das nach Archimedes benannt ist, bringen wir später. Statt $-a > 0$ schreiben wir auch $a < 0$, und statt $a - b > 0$ auch $a > b$. Hier einige Folgerungen aus (A1) und (A2).

Satz 1.4 (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: (1) folgt direkt aus (A1) und der Definition von $a > b$. 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 (1.1) benutzt. Nach (A1) ist $a^2 > 0$ für $a \neq 0$ bewiesen. Aussage (6) ergibt sich nun mit

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

Zu guter Letzt haben wir für (7)

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

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

Definition 1.1 (Betrag einer reellen Zahl) *Der Betrag von $a \in \mathbb{R}$ ist*

$$|a| = \begin{cases} a & \text{falls } a \geq 0 \\ -a & \text{falls } a \leq 0. \end{cases} \tag{1.3}$$

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 1.5 (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 1.1 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 1.6 ($\mathbb{Q}$ 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 zwischen a und b ein Gitterpunkt liegt. Betrachte dazu

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

M ist nicht leer, denn nach Anmerkung zu A3 gibt es ein $k' \in \mathbb{N}$ mit $k' > -nb$ bzw. $(-k')/n < b$. Außerdem ist nb eine obere Schranke von M. Sei nun $m \in \mathbb{Z}$ die größte Zahl in M, also $m/n < b$ aber $(m+1)/n \geq b$. Es folgt

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

Es gilt also $m/n \in (a,b)$. □

Im Beweis wurde folgende Tatsache benutzt.

Satz 1.7 (maximale ganze Zahl) *Jede nach oben beschränkte, nichtleere Menge $M \subset \mathbb{Z}$ hat ein größtes Element.*

Beweis: Da $M \neq \emptyset$ gibt es ein $k_0 \in M$. Sei $C \in \mathbb{R}$ obere Schranke von M. Nach Axiom (A3) gibt es dann ein $N \in \mathbb{N}$ mit $N > C - k_0$ bzw. $k_0 + N > C$. Sei nun $i \in \{0, 1, \ldots, N-1\}$ maximal mit $k_0 + i \in M$. Dann ist $k_0 + i$ ein größtes Element von M, denn wäre $k \in M$ mit $k > k_0 + i$, so folgt $k \geq k_0 + N > C$, Widerspruch. □

2 Vollständige Induktion

Angeblich gibt es am Amazonas einen kleinen Stamm von Eingeborenen, den Piranhãs, die keine Zahlen kennen, sondern nur die beiden Begriffe *hói* und *hoí*, was wohl *wenige* und *viele* bedeutet. Der Stamm kennt auch keine Schrift, die beiden Wörter müssen sich irgendwie in der Aussprache unterscheiden. Im Gegensatz dazu hat unsere Kultur die natürlichen Zahlen $1, 2 = 1 + 1, 3 = 2 + 1, \ldots$; diese Menge wird üblicherweise mit $\mathbb{N}$ bezeichnet. Manche Autoren nehmen hier die Null dazu, ich finde das aber nicht so gut. Offenbar wird jedes $n \in \mathbb{N}$ erreicht, indem die 1 endlich oft addiert wird. Darauf beruht das

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) bedeutet 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. Das wäre aber kein schöner Weg, denn der Begriff des Zählens ist grundlegender als das Konzept der reellen Zahlen, und sollte unabhängig 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. Auch dies hat die Rolle als beruhigende Absicherung, die wir hier aber nicht brauchen. Stattdessen verwenden wir das Induktionsprinzip als Ausgangspunkt. Es hat folgende Umformulierung.

Satz 2.1 (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. $\square$

Man bezeichnet (1) als Induktionsanfang und (2) als Induktionsschluss. In den Naturwissenschaften bedeutet *Induktion* den Schluss vom Einzelfall auf ein allgemeines Gesetz. Wir sollten uns bewusst sein, dass es bei unserem Beweisverfahren natürlich um etwas ganz anderes geht.

Beispiel 2.1 (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*. In der ersten Klasse hatten die Schüler die Aufgabe bekommen, die Zahlen von Eins bis Hundert zu addieren. Der Lehrer hatte wohl gehofft eine Zeitlang in Ruhe auf seinem Handy zu daddeln. 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

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

Für $n > m$ kann die Summe rekursiv definiert und berechnet werden, das heißt

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

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 Produktzeichen ist ganz analog erklärt:

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

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

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

Wir zeigen das wieder durch vollständige Induktion, wobei wir hier bei $n = 0$ beginnen:

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

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 2.2 (Bernoullische Ungleichung) *Für $x \in \mathbb{R}$, $x \geq -1$, und $n \in \mathbb{N}$ gilt*

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

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

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

□

Als nächstes geht es um die Zahl der Elemente gewisser Mengen. Das Abzählen gehört zu den Dingen die jedem klar sind, nur die Mathematiker machen es kompliziert [2]. Eine Abbildung oder gleichbedeutend Funktion von einer Menge A in eine Menge B schreiben wir

$$f : A \to B,\ a \mapsto f(a).$$

Jedem $a \in A$ wird ein Bild $f(a) \in B$ zugeordnet. Es wird aber nicht verlangt, dass jedes $b \in B$ Bildpunkt ist. A heißt Definitionsbereich, B Zielbereich von f. Die Abbildung f ist

$$\begin{aligned}
\text{injektiv} &\Leftrightarrow \text{ aus } f(a) = f(a') \text{ folgt } a = a', \\
\text{surjektiv} &\Leftrightarrow \text{ zu jedem } b \in B \text{ gibt es ein } a \in A \text{ mit } f(a) = b, \\
\text{bijektiv} &\Leftrightarrow f \text{ ist injektiv und surjektiv.}
\end{aligned}$$

Man spricht auch von einer *Injektion, Surjektion* oder *Bijektion*. Ist f bijektiv, so ist die Umkehrabbildung f^{-1} definiert, die jedem $b \in B$ sein eindeutig bestimmtes Urbild zuordnet:

$$f^{-1} : B \to A,\ f^{-1}(b) = a \quad \text{wobei } f(a) = b.$$

Die Verkettung von zwei Abbildungen $f : A \to B$ und $g : B \to C$ ist gegeben durch

$$g \circ f : A \to C,\ (g \circ f)(a) = g(f(a)).$$

Ist zum Beispiel f bijektiv, so gilt $f^{-1} \circ f = \mathrm{id}_A$ und $f \circ f^{-1} = \mathrm{id}_B$. Dabei bezeichnet $\mathrm{id}_A : A \to A$ die identische Abbildung, sprich $\mathrm{id}_A(a) = a$ für alle $a \in A$.

Definition 2.1 (Zahl der Elemente) *Eine Menge M hat n Elemente ($n \in \mathbb{N}$), wenn eine Bijektion $\{1, \ldots, n\} \stackrel{\sim}{\to} M$ existiert, Notation $\#M = n$. Eine Menge M heißt endlich, wenn $\#M = n$ 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 gibt es Bijektionen zu $\{1, \ldots, m\}$ und zu $\{1, \ldots, n\}$. Durch Verkettung hätten wir eine Bijektion

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

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

[2] 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).

Satz 2.3 (Schubfachprinzip) *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

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

Zu zeigen ist $m \leq n + 1$, das ist klar für $m = 1$. Wir suchen eine injektive Abbildung

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

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, also können wir $f(m) \in \{1, \ldots, n\}$ als Bildpunkt 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. Es muss auch gelten: ist in jeder Schublade mindestens eine Socke, so haben wir mindestens soviel Socken wie Schubladen. Wir drücken das noch mathematisch aus.

Folgerung 2.1 *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 2.3 folgt $n \leq m$. □

Nach diesem Vorspann wollen wir für Mengen von Interesse die Anzahl der Elemente bestimmen. Als erstes betrachten wir die Menge S_n der bijektiven Abbildungen von $\{1, \ldots, n\}$ in sich. Jeder solchen Abbildung $\sigma \in S_n$ entspricht eindeutig eine Umordnung $(\sigma_1, \ldots, \sigma_n)$ der Ziffern von $(1, \ldots, n)$, nämlich $\sigma_i = \sigma(i)$. Man bezeichnet die Abbildungen auch als Permutationen von $1, \ldots, n$.

Satz 2.4 (Zahl der Permutationen) *Es gilt $\# S_n = n! = 1 \cdot 2 \cdot \ldots \cdot n$.*

Beweis: Die Behauptung ist klar für $n = 1$. Wir verwenden die Notation als Umordnungen, und zwar betrachten wir für jedes $k = 1, \ldots, n+1$ die Umordnungen $S_{n+1,k}$ in S_{n+1}, bei denen die Ziffer $n+1$ an der Stelle k steht:

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

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 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. Es folgt durch Induktion

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

□

Für das nächste Zählproblem brauchen wir die

Definition 2.2 (Binomialkoeffizienten) *Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}_0$ setzen wir*

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

Lemma 2.1 (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

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

□

Im Fall $\alpha = n \in \mathbb{N}_0$ erlaubt Lemma 2.1 die rekursive Berechnung der Binomialkoeffizienten nach dem Dreiecksschema von Blaise Pascal (1623-1662).

n=0							1							
n=1						1		1						
n=2					1		2		1					
n=3				1		3		3		1				
n=4			1		4		6		4		1			
n=5		1		5		10		10		5		1		
n=6	1		6		15		20		15		6		1	

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

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \text{ für } n \in \mathbb{N}_0,\ k \in \{0, 1, \ldots, n\}, \tag{2.1}$$

und daraus weiter die am Diagramm ersichtliche Symmetrieeigenschaft

$$\binom{n}{k} = \binom{n}{n-k} \text{ für } n \in \mathbb{N}_0,\ k \in \{0, 1, \ldots, n\}. \tag{2.2}$$

Satz 2.5 (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 2.1 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. □

Als Anwendung erhalten wir die allgemeine Binomische Formel.

Satz 2.6 (Binomische Formel) *Für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2.3}$$

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

$$(a+b)^n = \sum_{k=0}^{n} c(n,k)\, a^k b^{n-k} \tag{2.4}$$

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 (2.4) 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 2.1,

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

□

3 Grenzwerte von Folgen

Wir definieren hier den Begriff des Grenzwerts im Fall von Folgen reeller Zahlen. 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 3.1 (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 dann 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.

Mit den Quantoren $\forall$ (für alle), $\exists$ (existiert) und $\Rightarrow$ (daraus folgt) lässt sich die Definition der Konvergenz auch wie folgt fassen:

$$\forall \varepsilon > 0 \; \exists K \in \mathbb{R} : \quad \Big(n > K \quad \Rightarrow \quad |a_n - a| < \varepsilon\Big).$$

Wir schauen zuerst zwei grundlegende Beispiele an.

Beispiel 3.1 (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 $n > K$

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

Beispiel 3.2 (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.

In der Schule wird der Begriff des Grenzwerts verbal erklärt, etwa: *die a_n weichen von a beliebig wenig ab, wenn n hinreichend groß ist.* Hier wird das präzisiert: der Parameter $\varepsilon > 0$ gibt eine Fehlerschranke für die Abweichung vor. Die gesuchte Zahl K quantifiziert die Umschreibung *hinreichend groß*; die Fehlerschranke ist erfüllt wenn $n > K$ ist. Es soll für jedes $\varepsilon > 0$ ein solches K geben, dies entspricht der Umschreibung *beliebig wenig* in der verbalen Fassung. Der Vorteil der quantitativen Formulierung liegt darin, dass wir sie in mathematischen Argumenten benutzen können, ein rein verbales Konzept wäre dafür ungeeignet.

Je kleiner die gegebene Schranke $\varepsilon > 0$ ist, desto später wird sie durch die a_n erfüllt, desto größer muss also die Schranke K gewählt werden. Für die Folge $a_n = 1/n$ sei zum Beispiel $\varepsilon = 1/N$ für ein $N \in \mathbb{N}$ vorgegeben. Wie oben gezeigt kann dann $K = 1/\varepsilon$ gewählt werden. Dagegen reicht eine Zahl $K < 1/\varepsilon$ nicht aus, denn dann ist $N = 1/\varepsilon > K$, aber die Abschätzung ist für $n = N$ nicht erfüllt:

$$|a_N - 0| = \frac{1}{N} = \varepsilon.$$

Für $\varepsilon = 1/N$ ist daher $K = 1/\varepsilon$ die optimale, sprich kleinste mögliche Wahl, insbesondere wird K groß für $\varepsilon > 0$ klein. Manchmal wird das durch die Notation $K = K_\varepsilon$ hervorgehoben.

Im allgemeinen wäre es kompliziert, das optimale K zu berechnen. Um Konvergenz zu zeigen, ist das aber auch unnötig, wir brauchen nur irgendein K. Es reicht daher aus, die Differenz $|a_n - a|$ so abzuschätzen, dass eine geeignete Wahl von K erkennbar wird. Das vereinfacht die Sache erheblich, hier ein Beispiel.

Beispiel 3.3 (Geometrische Folge) Sei $q \in \mathbb{R}$ mit $|q| < 1$. Dann gilt $\lim_{n\to\infty} q^n = 0$. Um das zu zeigen, können wir $q \neq 0$ voraussetzen und haben dann $1/|q| > 1$, also gilt $1/|q| = 1+x$ für ein $x > 0$. Es folgt mit der Bernoulli-Ungleichung, Satz 2.2,

$$|q^n - 0| = |q|^n = \frac{1}{(1+x)^n} \leq \frac{1}{1+nx} \leq \frac{1}{nx} < \varepsilon$$

für alle $n > 1/(\varepsilon x)$. Wir können also $K = 1/(\varepsilon x)$ wählen.

Hier mal ein Beispiel einer Folge, die nicht konvergiert.

Beispiel 3.4 Die Folge $a_n = (-1)^n$, also $a_n = -1, 1, -1, \ldots$, ist nicht konvergent. Angenommen $\lim_{n\to\infty} a_n = a$ 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 $n > K$

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

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

Satz 3.1 (Eindeutigkeit des Grenzwerts) *Falls die Folge $(a_n)_{n\in\mathbb{N}}$ konvergent ist, so ist ihr Grenzwert eindeutig bestimmt.*

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

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

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

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

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

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.

An dieser Stelle bietet es sich an, den Begriff der ε-Umgebung einzuführen. Der Grenzwertbegriff gewinnt dadurch an Anschaulichkeit.

Definition 3.2 (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 einfach äquivalent, aber eventuell etwas anschaulicher. Der Name $B_\varepsilon(a)$ kommt übrigens aus dem Englischen: im $\mathbb{R}^3$ ist die Menge der Punkte, die zu einem festen Punkt a Abstand kleiner als ε 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 3.3 (Beschränktheit von Folgen) *Eine Folge $(a_n)_{n\in\mathbb{N}}$ heißt*

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

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

Beispiel 3.5 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 3.2 (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

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

Wähle zum Beispiel $\varepsilon = 1$. Dann folgt mit $C = \max(|a_1|, \ldots, |a_N|, |a| + 1)$

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

$\square$

Satz 3.3 (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 3.2 gibt es ein $C > 0$ mit $|a_n| \leq C$ für alle $n \in \mathbb{N}$, und außerdem mit $|b| \leq C$. Also

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

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

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

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

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

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

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

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

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

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

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

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

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

$\square$

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 Adition 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 Geraden und Ebenen durch den Nullpunkt, sowie der $\mathbb{R}^3$ selbst und der Nullvektor. Allgemein sind Unterräume 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:

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

Mit Regel a) aus Satz 3.3 und Satz 3.2 sieht man leicht, dass es sich auf jeder Stufe um Untervektorräume handelt. Dabei sind Nullfolgen definiert durch $\lim_{n\to\infty} a_n = 0$. Man kann sich überlegen, dass an keiner Stelle Gleichheit gilt, und dass keiner der Räume endlichdimensional ist. So weit dieser Ausflug.

Hier zwei Anwendungen der Rechenregeln für Grenzwerte.

Beispiel 3.6 (Grenzwerte rationaler Funktionen) Seien $p, q : \mathbb{R} \to \mathbb{R}$ Polynome vom Grad $m, n \in \mathbb{N}_0$, das heißt für alle $x \in \mathbb{R}$ gilt

$$p(x) = a_m x^m + a_{m-1}x^{m-1} + \ldots + a_0 \quad \text{und} \quad q(x) = b_n x^n + b_{n-1}x^{n-1} + \ldots + b_0,$$

wobei $a_i, b_j \in \mathbb{R}$ mit $a_m, b_n \neq 0$. Wir bestimmen im Fall $m \leq n$ das Verhalten von $p(k)/q(k)$ für $k \to \infty$, und zwar liefert Anwendung der Konvergenzregeln in Satz 3.3

$$\frac{p(k)}{q(k)} = k^{m-n}\frac{a_m + a_{m-1}k^{-1} + \ldots + a_0 k^{-m}}{b_n + b_{n-1}k^{-1} + \ldots + b_0 k^{-n}} \to \begin{cases} a_m/b_n & \text{falls } m = n,\\ 0 & \text{falls } m < n.\end{cases}$$

Beispiel 3.7 (geometrische Reihe) Für $-1 < q < 1$ betrachten wir die Folge

$$a_n = 1 + q + \ldots + q^n = \sum_{k=0}^{n} q^k.$$

Dann ergibt sich aus Beispiel 2.2, Beispiel 3.3 und Satz 3.3

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

Wir schreiben hierfür auch $\sum_{k=0}^{\infty} q^k = 1/(1-q)$.

Folgen, deren Folgenglieder Summen sind, heißen Reihen. Sie spielen eine große Rolle in der Analysis und werden noch ausführlicher untersucht.

Satz 3.4 (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 auch 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 3.4 (Uneigentliche Konvergenz) *Die Folge* $(a_n)_{n\in\mathbb{N}}$ *konvergiert uneigentlich (oder divergiert bestimmt) gegen* $+\infty$, *falls gilt:*

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

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:

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

Bei Konvergenz gegen $a \in \mathbb{R}$ liegen für $n > K$ alle Folgenglieder in $B_\varepsilon(a)$. Bei Konvergenz gegen $+\infty$ tritt anstelle der 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 3.8 Für $q > 1$ gilt $\lim_{n\to\infty} q^n = +\infty$. Denn zu gegebenem $C > 0$ gibt es nach Beispiel 3.3 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 3.3 behandelt, und der Fall $q \leq -1$ folgt mit etwas Überlegung aus Beispiel 3.4.

Beispiel 3.9 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 3.5 (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.

4 Vollständigkeit der reellen Zahlen

Die bisher eingeführten Axiome (K) sowie (A1) bis (A3) gelten selbstverständlich auch für die rationalen Zahlen. Dennoch sind die rationalen Zahlen für die Analysis ungeeignet. Wir beginnen mit folgender Beobachtung der Pythagoräer.

Satz 4.1 (Irrationalität von $\sqrt{2}$) *Die Gleichung $x^2 = 2$ ist in $\mathbb{Q}$ nicht lösbar.*

Beweis: (durch Widerspruch) Angenommen, die Gleichung $x^2 = 2$ hat eine rationale Lösung, also $x = p/q$ mit $p, q \in \mathbb{N}$. Durch fortgesetztes Kürzen können wir annehmen, dass höchstens eine der Zahlen p und q gerade ist. Nun gilt

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

Setze also $p = 2p_1$ mit $p_1 \in \mathbb{N}$, dann folgt weiter

$$2q^2 = 4p_1^2 \quad \Rightarrow \quad q^2 \text{ gerade} \quad \Rightarrow \quad q \text{ gerade .}$$

Also sind doch p, q beide gerade, ein Widerspuch. $\square$

In $\mathbb{R}$ ist die Gleichung $x^2 = 2$ lösbar, wie wir in Kürze sehen werden. Man könnte das als Grund für die Erweiterung $\mathbb{Q} \subset \mathbb{R}$ 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 falsch, geht aber am Kern der Sache vorbei: zum einen bleibt die Gleichung $x^2 = -1$ auch in $\mathbb{R}$ unlösbar. Andererseits bilden die Nullstellen aller Polynome mit rationalen Koeffizienten, die reellen algebraischen Zahlen, 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.

Allgemein ist es das Ziel der Analysis, neue Objekte – Zahlen, Funktionen, Operationen – durch Grenzprozesse zu konstruieren. Unsere Definition des Grenzwerts 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.

Beispiel 4.1 (Verfahren von Heron) Für $a > 0$ wollen wir eine Lösung der Gleichung $x^2 = a$ als Grenzwert einer Folge konstruieren, und zwar setzen wir rekursiv

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right), \qquad \text{Startwert } x_0 > 0. \tag{4.1}$$

Zur Motivation der Formel: wir suchen die positive Nullstelle der Funktion $f(x) = x^2 - a$. Sei $x_n > 0$ schon berechnet, dann hat der Graph von f im Punkt $(x_n, f(x_n))$ die Tangente

$$y = f'(x_n)(x - x_n) + f(x_n).$$

Wir wählen für x_{n+1} den Schnittpunkt der Tangente mit der x-Achse, also

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

Was können wir über die Iteration aussagen? Zunächst sehen wir induktiv $x_n > 0$ für alle n, die Iteration bricht nicht ab. Weiter gilt

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

Also ist $x_n^2 \geq a$ für $n \geq 1$. Weiter folgt

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

Hätten wir $x_n \to x$ mit $n \to \infty$, so folgt $x \geq 0$ und $x^2 \geq a$, insbesondere $x > 0$, und

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

Die Lösung von $x^2 = a$ wäre damit gefunden. Die Konvergenz der Folge x_n können wir aber nicht mit Definition 3.1 zeigen, denn dafür müssten wir den Grenzwert $\sqrt{a}$ schon kennen, und den wollen wir ja gerade konstruieren.

Beispiel 4.2 (Zinseszinsrechnung) Wird ein Euro für x Jahre mit dem Zinsfaktor 1 angelegt, so ist die Auszahlung $f_1(x) = 1 + x$. Die Idee des Zinseszinses ist es, den Zeitraum in kürzere Intervalle zu unterteilen und den Zins anteilig anzurechnen mit dem Effekt, dass der schon gesparte Zinsanteil des Guthabens seinerseits Zinsen liefert. Wird der Zeitraum x in n 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 im Zeitraum x,

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n \qquad \text{für } x \in \mathbb{R},\ n \in \mathbb{N}. \tag{4.2}$$

Was ergibt sich bei kontinuierlicher Verzinsung, existiert der Grenzwert $\lim_{n\to\infty} f_n(x)$?

Beispiel 4.3 (Dezimalbruchdarstellung) Jede Zahl $a \in \mathbb{R}$ hat eine Darstellung als Dezimalbruch, das heißt es gibt ein $a_0 \in \mathbb{Z}$ und Ziffern $k_j \in \{0, 1, \ldots, 9\}$, $j \in \mathbb{N}$, so das gilt

$$a = \lim_{n\to\infty} a_n \quad \text{mit} \quad a_n = a_0 + \sum_{j=1}^{n} k_j \cdot 10^{-j} = a_0, k_1 k_2 \ldots k_n.$$

Wähle dazu $a_0 \in \mathbb{Z}$ mit $a_0 \leq a < a_0 + 1$, und definiere induktiv für $n \in \mathbb{N}$

$$k_n = \max\{k \in \mathbb{Z} : a_{n-1} + k \cdot 10^{-n} \leq a\} \quad \text{und} \quad a_n = a_{n-1} + k_n \cdot 10^{-n}.$$

Wir zeigen $a_n \leq a < a_n + 10^{-n}$ für alle $n \in \mathbb{N}_0$. Der Fall $n = 0$ gilt nach Wahl von a_0. Sei $n \in \mathbb{N}$, und die Aussage sei schon bewiesen für $n - 1$. Dann ist

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

Es folgt $k_n \in \{0, 1 \ldots, 9\}$. Bei der Bildung des Maximums ist $k = k_n$ zulässig, aber $k = k_n + 1$ nicht. 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 ist die Behauptung bewiesen: a wird durch die Folge a_n approximiert, genauer

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

Umgekehrt ergibt sich aber die Frage: seien $a_0 \in \mathbb{Z}$ und $k_1, k_2, \ldots$ beliebig gegeben 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 Augustin Louis Cauchy (1789–1857) besteht darin, die Glieder der Folge nicht mit dem unbekannten Grenzwert, sondern *untereinander* zu vergleichen.

Definition 4.1 (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. Je nach Autor werden auch andere Aussagen als Vollständigkeitsaxiom zugrunde gelegt, die aber natürlich äquivalent sind und sich bei uns als Folgerungen ergeben werden.

Satz 4.2 *Eine Folge ist genau dann konvergent, wenn sie eine Cauchyfolge ist.*

Beweis: Eine Cauchyfolge ist konvergent nach dem Vollständigkeitsaxiom. Sei umgekehrt $\lim_{n \to \infty} a_n = a$. 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

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

□

Beispiel 4.4 (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 3.7, verwenden:

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

Nach Beispiel 3.3 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 4.3 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 4.2 (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 4.3 (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 $a_n \leq C$ für alle $n \in \mathbb{N}$. Für $\varepsilon > 0$ betrachten wir das Gitter $a_1 + \mathbb{Z} \cdot \varepsilon$, und setzen

$$M = \{j \in \mathbb{Z} : \text{es gibt ein } n \in \mathbb{N} \text{ mit } a_n \geq a_1 + j\varepsilon\}.$$

Wegen $a_1 \geq a_1 + 0 \cdot \varepsilon$ ist $0 \in M$. Weiter ist M ist nach oben beschränkt: zu $j \in M$ wähle $n \in \mathbb{N}$ mit $a_n \geq a_1 + j\varepsilon$, dann folgt

$$j \leq \frac{a_n - a_1}{\varepsilon} \leq \frac{C - a_1}{\varepsilon}.$$

Sei $k \in M$ maximales Eelemnt. Dann gibt es ein $N \in \mathbb{N}$ mit $a_N \geq a_1 + k\varepsilon$. Da k maximal, gilt aber $a_n < a_1 + (k+1)\varepsilon$ für alle n, und somit

$$a_1 + k\varepsilon \leq a_N \leq a_n < a_1 + (k+1)\varepsilon \quad \text{für alle } n \geq N.$$

Es folgt $|a_n - a_m| < \varepsilon$ für $n, m \geq N$, das heißt (a_n) ist eine Cauchyfolge. $\square$

Das Konvergenzkriterium der Monotonie und Beschränktheit ist nur hinreichend, denn es gibt auch konvergente Folgen die nicht monoton sind, etwa $a_n = (-1)^n/n$. Aber es ist ein enorm nützliches Kriterium, um Konvergenz zu zeigen.

Beispiel 4.5 (Konvergenz des Heronverfahrens) Die in Beispiel 4.1 konstruierte Folge x_n, $n \in \mathbb{N}_0$, erfüllte $x_n^2 \geq a$ und $x_1 \geq x_2 \geq \ldots$. Nach Satz 4.3 konvergiert die Folge gegen ein $x > 0$. Wir hatten bereits gesehen, dass dann die Gleichung $x^2 = a$ gelöst wird, das heißt wir haben eine Quadratwurzel von a gefunden.

Beispiel 4.6 (Die Zahl e und Zinsrechnung) Nach Beispiel 4.2 ergibt ein Euro mit Zinssatz Eins nach einem Jahr, wenn der Zins in n Intervallen berechnet wird,

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

Wir behaupten, dass die Folge f_n nach oben beschränkt ist. Berechne dazu mit der binomischen Formel, siehe Satz 2.6,

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

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:

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

Insgesamt sind beide Folgen nach oben beschränkt und monoton wachsend, also sind sie konvergent nach Satz 4.3.

Definition 4.3 (Eulersche Zahl) *Wir setzen*

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{1}{k!}. \tag{4.3}$$

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 $\lim_{n\to\infty} f_n \geq e$ 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

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n. \tag{4.4}$$

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. Dagegen kommt e_n hier nur technisch ins Spiel, die Folge tritt in der Abschätzung auf. Bei der Diskussion der Exponentialfunktion werden wir aber dieser Reihe wieder begegnen.

Satz 4.4 (Irrationalität der Eulerschen Zahl) $e = \sum_{k=0}^{\infty} 1/k!$ *ist nicht rational.*

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

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

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 der mittlere Term eine ganze Zahl für n hinreichend groß, ein Widerspruch. □

Als nächste Anwendung des Vollständigkeitsaxioms diskutieren wir nun das Intervallschachtelungsprinzip.

Definition 4.4 (Intervallschachtelung) *Eine Intervallschachtelung ist eine Folge von abgeschlossenen 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 $n \to \infty$.*

Satz 4.5 (Intervallschachtelungsprinzip) *Zu jeder Intervallschachtelung $(I_n)_{n\in\mathbb{N}}$ gibt es genau ein $x \in \mathbb{R}$ mit*

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

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 4.3 folgt die Existenz der Grenzwerte $a = \lim_{n\to\infty} a_n$ bzw. $b = \lim_{n\to\infty} b_n$. Dann folgt nach Satz 3.3 und Satz 3.4

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

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 3.4 $a \leq x' \leq b$, also $x' = x$. □

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 eine Vorüberlegung: eine Funktion $f : I \to \mathbb{R}$, $I \subset \mathbb{R}$, heißt *streng monoton wachsend*, wenn folgende Implikation gilt:

$$x_1 < x_2 \quad \Rightarrow \quad f(x_1) < f(x_2). \tag{4.5}$$

f ist dann injektiv, denn ist $f(x_1) = f(x_2)$, so kann weder $x_1 < x_2$ noch $x_1 > x_2$ sein, also $x_1 = x_2$. Die Funktion $f : I \to f(I)$ ist somit injektiv und surjektiv, das heißt es gibt die Umkehrfunktion $g : f(I) \to I$. Diese Umkehrfunktion ist dann ebenfalls streng monoton wachsend: wäre $a_1 < a_2$ mit $g(a_1) \geq g(a_2)$, so folgt aus der Monotonie von f

$$a_1 = f(g(a_1)) \geq f(g(a_2)) = a_2, \quad \text{Widerspruch.} \tag{4.6}$$

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

Satz 4.6 (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:

$$a_k^n \leq y \leq b_k^n \quad \text{für alle } k \in \mathbb{N}. \tag{4.7}$$

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 (4.7) auch für $k+1$:

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

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 4.5 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 (4.7) folgt

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

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

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

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

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

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

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 $s = p/q$

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

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 4.7 Ein Grenzwert, der häufiger auftritt, ist

$$\lim_{n\to\infty} a^{1/n} = 1 \qquad \text{für } a > 0. \tag{4.8}$$

Betrachte erst den Fall $a \geq 1$. Dann ist $a^{1/n} \geq 1^{1/n} = 1$ wegen Monotonie der Wurzelfunktion, also $a^{1/n} = 1 + \xi_n$ mit $\xi_n \geq 0$. Aus der Bernoulli-Ungleichung, Satz 2.2, folgt

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

Also $\lim_{n\to\infty} \xi_n = 0$ 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 4.5 (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 $(a_n)_{n\in\mathbb{N}}$.*

Durch Induktion ergibt sich sofort $n_k \geq k$: es ist $n_1 \geq 1$ und $n_{k+1} \geq n_k + 1 \geq k+1$. Die Teilfolge entsteht aus der ursprünglichen Folge durch *Auswahl der Nummern n_k*. Da Folgen Abbildungen von $\mathbb{N}$ nach $\mathbb{R}$ sind, ist eine Teilfolge formal als Verkettung von zwei Abbildungen definiert: der Ausgangsfolge $a : \mathbb{N} \to \mathbb{R}$, $n \longmapsto a_n$ und der Folge $\mathbb{N} \to \mathbb{N}$, $k \longmapsto n_k$, die die Indizes auswählt. Am Beispiel $a_n = (-1)^n/n^3$ und $n_k = 2k-1$ sieht das wie folgt aus:

$$\begin{array}{ccccc}
\mathbb{N} & \xrightarrow{n_k = 2k-1} & \mathbb{N} & \xrightarrow{a_n = (-1)^n/n^3} & \mathbb{R} \\
 & & & & \\
k & \longmapsto & n_k = 2k-1 & \longmapsto & a_{n_k} = -1/(2k-1)^3
\end{array}$$

Definition 4.6 (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 $k \to \infty$.

Lemma 4.1 *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: Wenn $a \in \mathbb{R}$ Häufungspunkt von $(a_n)_{n\in\mathbb{N}}$ ist, so gilt nach Definition $a_{n_k} \to a$ mit $k \to \infty$ für eine Teilfolge n_k. Zu jedem $\varepsilon > 0$ gibt es dann ein $K \in \mathbb{R}$ mit $a_{n_k} \in B_\varepsilon(a)$ für alle $k > K$. Die Abbildung $k \mapsto n_k$ ist injektiv wegen $n_1 < n_2 < \ldots$, also ist die Menge $\{n_k : k > K\}$ nicht endlich nach dem Schubfachprinzip. Dies beweist die eine Richtung der Äquivalenz. Umgekehrt 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 gilt. Es folgt dann $|a_{n_k} - a| < 1/k \to 0$ mit $k \to \infty$. □

Satz 4.7 (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 4.5 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 $k > K$

$$I_k \subset B_\varepsilon(x), \qquad \text{denn } x' \in I_k \;\Rightarrow\; x, x' \in I_k \;\Rightarrow\; |x' - x| \le |I_k| < \varepsilon.$$

Es folgt $x_n \in B_\varepsilon(x)$ für unendlich viele n, und x ist Häufungspunkt nach Lemma 4.1. □

Definition 4.7 (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 4.8 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

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

Also ist Bedingung (ii) ebenfalls erfüllt, und $\limsup_{n\to\infty} x_n = 1$.

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 lim sup bzw. lim inf für jede beliebige Folge x_n definiert. Das wollen wir nun zeigen, wobei wir uns auf den Limes superior beschränken.

Satz 4.8 (Existenz des Limes superior) *Für jede reelle Folge $(x_n)_{n\in\mathbb{R}}$ 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 4.7 konstruierte Intervallschachtelung und x^* der zugehörige Häufungspunkt. Bedingung (i) aus Definition 4.7 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 4.7. 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:

$$\begin{array}{c}\text{Vollständigkeitsaxiom:}\\ \text{Cauchyfolgen sind konvergent}\\ \Downarrow\\ \text{Konvergenzkriterium der}\\ \text{Monotonie und Beschränktheit}\\ \Downarrow\\ \text{Intervallschachtelungsprinzip}\\ \Downarrow\\ \text{Bolzano-Weierstraß: Auswahlsatz}\\ \Downarrow\\ \text{Cauchyfolgen sind konvergent}\end{array}$$

Die Implikationen sind so zu verstehen, dass im Beweis jedes Resultats nur die direkt vorangehende Eigenschaft von $\mathbb{R}$ benutzt wurde. Die letzte Implikation werden wir dabei gleich noch zeigen. Es folgt, dass jede der vier Eigenschaften als Axiom für $\mathbb{R}$ benutzt werden könnte - die anderen Eigenschaften würden als Sätze folgen. Im nächsten Abschnitt werden wir eine weitere, äquivalente Eigenschaft kennenlernen, nämlich den Satz vom Supremum (Satz 5.1). In dieser Vorlesung wird die Konvergenz der Cauchyfolgen als grundlegendes Axiom gewählt.

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. $\square$

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

Für Teilmengen von $\mathbb{R}$ haben wir analoge Begriffe wie bei reellen Folgen, mit kleinen Abwandlungen.

Definition 5.1 *Die Menge $M \subset \mathbb{R}$ heißt*

$$\begin{aligned} \textit{nach oben beschränkt} \quad &\Leftrightarrow \quad \exists\, K \in \mathbb{R} \textit{ mit } x \leq K \textit{ für alle } x \in M, \\ \textit{nach unten beschränkt} \quad &\Leftrightarrow \quad \exists\, K \in \mathbb{R} \textit{ mit } x \geq K \textit{ für alle } x \in M. \end{aligned}$$

Die Zahl K heißt dann obere bzw. untere Schranke. Weiter heißt

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

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 $K_1 \leq x \leq K_2$ folgt umgekehrt $|x| \leq \max(|K_1|, |K_2|)$.

Beispiel 5.1 Die Menge $[0, 1)$ ist nach oben beschränkt, eine obere Schranke ist zum Beispiel $K = 2019$. Es gibt 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 aber eine kleinste, nämlich die Zahl 1.

Definition 5.2 (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 $x > a'$ ($a' < a$).*

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 5.1 *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 5.2 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 4.5, 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 5.1 *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 $x_n \to +\infty = \sup M$. □

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

Lemma 5.1 *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 2.3, gibt es für $n \in \mathbb{N}$ keine Bijektion $\mathbb{N} \leftrightarrow \{1, \ldots, n\}$.

Definition 5.4 *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 5.2 *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 auszulassen, die bereits vorher auftraten, und die restlichen entsprechend neu zu nummerieren. Dazu setzen wir $n_1 = 1$ und konstruieren induktiv eine Teilfolge a_{n_k} durch die

Vorschrift

$$n_{k+1} = \min\{n > n_k : a_n \neq a_{n_j} \text{ für } j = 1, \ldots, k\}.$$

Falls die Rekursion nach einem n_k abbricht, ist die Abbildung $f : \{1, \ldots, k\} \to A$, $k \mapsto a_{n_k}$ bijektiv und damit ist A endlich. Andernfalls ist die Abbildung $f : \mathbb{N} \to A$, $k \mapsto a_{n_k}$ bijektiv und A ist abzählbar unendlich. $\square$

Satz 5.2 *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:

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

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. Damit bekommt (p, q) die Nummer

$$n(p, q) = N_k + q \quad \text{wobei } k = p + q - 1 \in \mathbb{N},\ 1 \leq q \leq k. \tag{5.1}$$

Um φ zu definieren, müssen wir umgekehrt zu $n \in \mathbb{N}$ den Eintrag (p, q) bestimmen. Wir zeigen dass $n \in \mathbb{N}$ eine eindeutige Darstellung wie in (5.1) besitzt. Die N_k sind streng monoton wachsend und es ist $N_1 = 0$, also gilt $N_k < n \leq N_{k+1}$ für ein $k \in \mathbb{N}$. Für $q = n - N_k$ folgt $1 \leq q \leq N_{k+1} - N_k = k$. Also ist $n = N_k + q$ wie verlangt. Umgekehrt folgt aus der Darstellung, dass $N_k < n \leq N_k + k = N_{k+1}$. Also ist k, und damit auch $q = n - N_k$, eindeutig. Weiter muss $k = p + q - 1$ gelten. Wir definieren $\varphi : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ durch

$$\varphi(n) = (k - q + 1, q) \quad \text{wobei } n = N_k + q \text{ mit } k \in \mathbb{N},\ q \in \{1, \ldots, k\}.$$

Die Abbildung φ ist wohldefiniert und surjektiv, denn es gilt

$$\varphi(n(p,q)) = (p,q) \quad \text{für alle } (p,q) \in \mathbb{N} \times \mathbb{N}.$$

Es ist eine gute Übung, das zu überprüfen. Das Argument zeigt allgemein, dass eine abzählbare Vereinigung von jeweils abzählbaren Mengen wiederum abzählbar ist. □

Satz 5.3 ($\mathbb{R}$ 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 4.5 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 $I_0 = [0,1]$. □

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 oder gleichmächtig zu $\mathbb{R}$ ist. 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.

In Zukunft brauchen wir den Begriff der Konvergenz auch für Punkte im $\mathbb{R}^n$. Für $n = 1$ haben wir als Modell die Zahlengerade benutzt, entsprechend betrachten wir für $n = 2$ die Ebene mit kartesischen Koordinaten $z = (x,y)$ und für $n = 3$ den dreidimensionalen Raum mit Koordinaten $p = (x,y,z)$; dabei finde ich den zweidimensionalen Fall besonders anschaulich. Der $\mathbb{R}^n$ ist eine mathematische Verallgemeinerung:

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

Die Vektoraddition und Skalarmultiplikation sind komponentenweise definiert:

$$\begin{aligned} x + y &= (x_1 + y_1, \ldots, x_n + y_n) \quad (x, y \in \mathbb{R}^n), \\ \lambda x &= (\lambda x_1, \ldots, \lambda x_n) \quad (x \in \mathbb{R}^n\ \lambda \in \mathbb{R}). \end{aligned}$$

Mit anderen Worten ist $(\lambda x + \mu y)_i = \lambda x_i + \mu y_i$ für $x, y \in \mathbb{R}^n$, $\lambda, \mu \in \mathbb{R}$. Die Verallgemeinerung des Betrags $|x|$ einer Zahl $x \in \mathbb{R}$ ist die Euklidische Länge oder Norm.

Definition 5.5 *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 zwei Punkten $x, y \in \mathbb{R}^n$ ist $\|x - y\|$.

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 5.4 *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 5.6 *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 5.3 *Für das Standardskalarprodukt $\langle x, y \rangle$ gilt:*

(1) Positivität: $\langle x, x \rangle \geq 0$ *mit Gleichheit genau wenn* $x = 0$.

(2) Symmetrie: $\langle x, y \rangle = \langle y, x \rangle$ *für alle* $x, y \in \mathbb{R}^n$.

(3) Bilinearität: *Für alle $x, y, z \in \mathbb{R}^n$ und $\lambda, \mu \in \mathbb{R}$ gilt*

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

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

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

□

Im folgenden Beweis wird die Normierung eines Vektors benutzt: ist $x \in \mathbb{R}^n$, $x \neq 0$, so hat $\xi = x/\|x\|$ dieselbe Richtung aber Länge Eins:

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

Satz 5.5 (Ungleichung von Cauchy-Schwarz) *Für $x, y \in \mathbb{R}^n$ gilt*

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|. \tag{5.2}$$

Gleichheit gilt genau dann, wenn x und y linear abhängig sind.

BEWEIS: Wird x oder y mit $\lambda \in \mathbb{R}$ multipliziert, so ergibt das auf beiden Seiten denselben Faktor $|\lambda|$. Dies legt nahe, erst Einheitsvektoren $\xi, \eta \in \mathbb{R}^n$ zu betrachten und die allgemeine Ungleichung im Anschluss durch Skalierung zu zeigen. Wir berechnen

$$0 \leq \|\xi \pm \eta\|^2 = \|\xi\|^2 \pm 2\langle \xi, \eta \rangle + \|\eta\|^2 = 2\big(1 \pm \langle \xi, \eta \rangle\big).$$

Also ist $-1 \leq \langle \xi, \eta \rangle \leq 1$ bzw. $|\langle \xi, \eta \rangle| \leq 1$. Seien nun $x, y \in \mathbb{R}^n$ beliebig. Ist einer der Vektoren Null, so sind beide Seiten gleich Null. Andernfalls folgt mit $\xi = x/\|x\|$ und $\eta = y/\|y\|$

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

Bei Gleichheit sind x, y linear abhängig: entweder ist einer der Vektoren Null, oder wir folgern $\eta = \mp\xi$, das heißt x, y sind parallel. □

Wir können nun leicht den Beweis von Satz 5.4 abschließen:

Beweis der Dreiecksungleichung: Mit Satz 5.5 gilt

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

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. □

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

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

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

Im Gegensatz zu $\mathbb{R}$ haben wir im $\mathbb{R}^n$ keine Anordnung zur Verfügung. Deshalb verwenden wir die Ungleichungszeichen $<, >, \leq, \geq$, Formulierungen wie "nach oben bzw. unten beschränkt" und den Begriff der 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). Ein Check zeigt jedoch, dass die Begriffe Konvergenz, ε-Umgebung, Beschränktheit, Cauchyfolge und Vollständigkeit in $\mathbb{R}$ alle nur die Betragsfunktion verwenden. Sie lassen sich dann auf den $\mathbb{R}^n$ verallgemeinern, indem wir statt der Betragsfunktion die Euklidische Norm einsetzen.

Definition 5.7 (Konvergenz im $\mathbb{R}^n$) *Eine Folge von Punkten $x_k \in \mathbb{R}^n$ konvergiert gegen $a \in \mathbb{R}^n$, falls gilt:*

Für alle $\varepsilon > 0$ gibt es ein $K \in \mathbb{R}$, so dass für alle $k > K$ gilt: $\|x_k - a\| < \varepsilon$.

Definition 5.8 (ε-Umgebung in $\mathbb{R}^n$) *Die ε-Umgebung von $a \in \mathbb{R}^n$ ist die Menge*

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

Sie wird auch als offene Kugel (für $n \geq 3$) bzw. offene Kreisscheibe (für $n = 2$) um a mit Radius $\varepsilon > 0$ bezeichnet. Die abgeschlossene Kugel ist

$$K_\varepsilon(a) = \{x \in \mathbb{R}^n : \|x - a\| \leq \varepsilon\}.$$

Definition 5.9 (Beschränkte Teilmengen von $\mathbb{R}^n$) *Eine Menge $M \subset \mathbb{R}^n$ heißt beschränkt, wenn gilt:*

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

Auch der Begriff der Cauchyfolge im $\mathbb{R}^n$ stützt sich auf die Norm:

$$\text{zu jedem } \varepsilon > 0 \text{ gibt es ein } K \in \mathbb{R} \text{ mit } \|x_k - x_l\| < \varepsilon \text{ für } k, l > K.$$

Die Vollständigkeit muss aber nicht wieder als Axiom gefordert werden, sondern folgt aus der Vollständigkeit von $\mathbb{R}$, Jeder Punkt $x \in \mathbb{R}^n$ ist durch seine n reellen Koordinaten $x_1, \ldots, x_n$ gegeben, und Fragen der Konvergenz lassen sich auf die Konvergenz der Koordinaten reduzieren.

Satz 5.6 (Euklidische Norm versus Koordinaten) *Eine Folge $x_k \in \mathbb{R}^n$ ist genau dann (norm-) 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: Es gilt $|x_i| \leq \|x\|$ für $i = 1, \ldots, n$, daraus folgen direkt die Implikationen $\Rightarrow$. Seien nun die Koordinatenfolgen jeweils beschränkt, also $|(x_k)_i| \leq K_i$ für alle k. Dann gilt

$$\|x_k\| = \left(\sum_{i=1}^n ((x_k)_i)^2 \right)^{1/2} \leq \left(\sum_{i=1}^n K_i^2 \right)^{1/2} =: K < \infty.$$

Das Argument für die Konvergenz ist ähnlich: sei $(x_k)_i \to a_i$ für $i = 1, \ldots, n$, das heißt zu $\varepsilon > 0$ gibt es $K_i \in \mathbb{R}$ mit $|(x_k)_i - a_i| < \varepsilon/\sqrt{n}$ für $k > K_i$. Es folgt mit $a := (a_1, \ldots, a_n)$

$$\|x_k - a\| = \left(\sum_{i=1}^n ((x_k)_i - a_i)^2 \right)^{1/2} < \left(n \, \frac{\varepsilon^2}{n} \right)^{1/2} = \varepsilon \qquad \text{für } k > K := \max_{i=1,\ldots,n} K_i.$$

Das Argument für die Cauchyfolgen ist analog: seien die $(x_k)_i$ Cauchyfolgen in $\mathbb{R}$ für jedes $i = 1, \ldots, n$. Zu $\varepsilon > 0$ gibt es dann $K_i \in \mathbb{R}$ mit $|(x_k)_i - (x_l)_i| < \varepsilon/\sqrt{n}$ für alle $k, l > K_i$. Es folgt für $k, l > \max_{i=1,\ldots,n} K_i$

$$\|x_k - x_l\| = \left(\sum_{i=1}^n ((x_k)_i - (x_l)_i)^2 \right)^{1/2} \leq \left(n \, \frac{\varepsilon^2}{n} \right)^{1/2} = \varepsilon.$$

□

Folgerung 5.2 *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 5.6 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 5.6 folgt $\lim_{k \to \infty} x_k = a$. □

Folgerung 5.3 (Bolzano-Weierstraß im $\mathbb{R}^n$) *Jede beschränkte Folge $(x_k)_{k \in \mathbb{N}}$ in $\mathbb{R}^n$ besitzt eine konvergente Teilfolge.*

Beweis: Die Koordinatenfolgen $((x_k)_i)_{k\in\mathbb{N}}$ sind beschränkt, siehe Satz 5.6, haben also jeweils eine konvergente Teilfolge nach Satz 4.7. Das Problem ist, dass diese Teilfolgen a priori verschieden sind. Wir konstruieren sukzessive Teilfolgen, damit schließlich alle n Koordinaten konvergieren. Sei für ein $j \in \{1,\ldots,n\}$ schon eine Teilfolge $k_1 < k_2 < \ldots$ gefunden, so dass

$$\lim_{l\to\infty} (x_{k_l})_i = a_i \qquad \text{für alle } i = 1,\ldots,j-1.$$

Der Fall $j = 1$ ist der Induktionsanfang. Es gibt dann eine weitere Teilfolge $l_1 < l_2 < \ldots$ und ein $a_j \in \mathbb{R}$, so dass noch zusätzlich

$$\lim_{m\to\infty} (x_{k_{l_m}})_j = a_j.$$

Nach n Schritten ist eine Teilfolge bestimmt, für die alle Koordinatenfolgen konvergieren. □

Definition 5.10 (Häufungspunkt von Mengen) *Ein Punkt $a \in \mathbb{R}^n$ heißt Häufungspunkt der Menge $M \subset \mathbb{R}^n$, falls für jedes $\varepsilon > 0$ die Menge $B_\varepsilon(a) \cap M$ unendlich viele Elemente hat.*

Ein Häufungspunkt von M ist nicht notwendig Element von M, zum Beispiel ist $0 \in \mathbb{R}$ Häufungspunkt von $M = \{1/n : n \in \mathbb{N}\}$.

Lemma 5.4 *Ein Punkt $a \in \mathbb{R}^n$ ist genau dann Häufungspunkt der Menge M, wenn es eine Folge von Punkten $x_k \in M\backslash\{a\}$ gibt mit $\lim_{k\to\infty} x_k = a$.*

Beweis: Das ist ähnlich zu Lemma 4.1. Sei a ein Häufungspunkt gemäß Definition 5.10. Zu $\varepsilon_k = 1/k$ gibt es dann ein $x_k \in B_{1/k}(a) \cap M\backslash\{a\}$. Die Folge $(x_k)_{k\in\mathbb{N}}$ ist wie verlangt. Sei umgekehrt eine Folge $x_k \in M\backslash\{a\}$ gegeben mit $\lim_{k\to\infty} x_k = a$. Angenommen, für ein $\varepsilon > 0$ ist die Menge $B_\varepsilon(a) \cap M$ endlich, also $B_\varepsilon(a) \cap M\backslash\{a\} = \{a_1,\ldots,a_m\}$ für ein $m \in \mathbb{N}_0$. Dann ist $\varrho := \min_{1\le i\le m} |a_i - a| > 0$. Wegen $\lim_{k\to\infty} \|x_k - a\| = 0$ folgt für k hinreichend groß

$$x_k \in B_\varrho(a) \cap M\backslash\{a\} = \emptyset,$$

ein Widerspruch. □

Beispiel 5.2 Die Menge $M = \{1/n : n \in \mathbb{N}\}$ hat $0 \in \mathbb{R}$ als einzigen Häufungspunkt. Dagegen ist die Menge der Häufungspunkte von $\mathbb{Q}$ gleich $\mathbb{R}$, denn in jeder Umgebung einer reellen Zahl gibt es eine rationale Zahl, siehe Satz 1.6. Es gibt eine kleine Differenz zwischen dem Begriff des Häufungspunkts für Folgen und für Mengen: die konstante Folge $x_k = a$, $k \in \mathbb{N}$, hat den Häufungspunkt a, dagegen hat die einelementige Menge $\{a\} \subset \mathbb{R}^n$ keinen Häufungspunkt. Überlegen Sie, dass für eine beliebige Folge $(x_k)_{k\in\mathbb{N}}$ im $\mathbb{R}^n$ die Häufungspunkte des Wertebereichs $M = \{x_k : k \in \mathbb{N}\}$ eine Teilmenge der Häufungspunkte der Folge bilden, die aber eine echte Teilmenge sein kann, wie das Beispiel der konstanten Folge zeigt.

Definition 5.11 *Eine Menge*

(i) *$U \subset \mathbb{R}^n$ heißt offen, falls es zu jedem $a \in U$ ein $\varepsilon > 0$ gibt mit $B_\varepsilon(a) \subset U$,*

(ii) *$A \subset \mathbb{R}^n$ heißt abgeschlossen, wenn folgende Implikation stets gilt:*

$$a = \lim_{n\to\infty} x_n \text{ mit } x_n \in A \quad \Rightarrow \quad a \in A.$$

Nach Lemma 5.4 ist eine Menge A genau dann abgeschlossen, wenn jeder Häufungspunkt von A schon Element von A ist.

Beispiel 5.3 Die Kugel $B_\varrho(a) = \{x \in \mathbb{R}^n : \|x-a\| < \varrho\}$ ist offen. Denn sei $x \in B_\varrho(a)$ und $y \in B_\varepsilon(x)$. Mit der Dreiecksungleichung folgt

$$\|y-a\| \leq \|y-x\| + \|x-a\| < \varepsilon + \|x-a\|.$$

Mit $\varepsilon = \varrho - \|x-a\| > 0$ folgt $\|y-a\| < \varrho$, also $B_\varepsilon(x) \subset B_\varrho(a)$, das heißt $B_\varrho(a)$ ist offen. Die Kugel $K_\varrho(a) = \{x \in \mathbb{R}^n : \|x-a\| \leq \varrho\}$ ist dagegen abgeschlossen: sind $x_k \in K_\varrho(a)$ und $x_k \to x \in \mathbb{R}^n$, so folgt

$$\|x-a\| \leq \|x-x_k\| + \|x_k-a\| \leq \|x-x_k\| + \varrho \to \varrho,$$

das heißt $x \in K_\varrho(a)$.

Satz 5.7 *Eine Menge $M \subset \mathbb{R}^n$ ist genau dann offen, wenn $\mathbb{R}^n \backslash M$ abgeschlossen ist.*

Beweis: Sei M offen und $x_k \in \mathbb{R}^n \backslash M$ mit $\lim_{k\to\infty} x_k = a \in \mathbb{R}^n$. Wäre $a \in M$, so folgt $B_\varepsilon(a) \subset M$ für ein $\varepsilon > 0$, und somit $x_k \in M$ für hinreichend große k, ein Widerspruch. Also ist $a \in \mathbb{R}^n \backslash M$, das heißt $\mathbb{R}^n \backslash M$ ist abgeschlossen.

Sei nun $\mathbb{R}^n \backslash M$ abgeschlossen. Wäre M nicht offen, so gibt es ein $a \in M$ mit $B_\varepsilon(a) \backslash M \neq \emptyset$ für alle $\varepsilon > 0$. Finde induktiv $x_k \in B_{1/k}(a)$ mit $x_k \in \mathbb{R}^n \backslash M$. Es folgt $a = \lim_{k\to\infty} x_k \in \mathbb{R}^n \backslash M$, ein Widerspruch. $\square$

Satz 5.8 *Die offenen Teilmengen des $\mathbb{R}^n$ bilden eine Topologie, das heißt es gilt:*

(1) *$\emptyset$ und $\mathbb{R}^n$ sind offen.*

(2) *Jede Vereinigung $\bigcup_{\lambda\in\Lambda} U_\lambda$ von offenen Mengen $U_\lambda \subset \mathbb{R}^n$ ist offen.*

(3) *Der Durchschnitt $\bigcap_{i=1}^k U_i$ von endlich vielen offenen Mengen $U_i \subset \mathbb{R}^n$ ist offen.*

Beweis: Aussage (1) ist evident. Ist $x \in \bigcup_{\lambda\in\Lambda} U_\lambda$, so gilt $x \in U_\mu$ für ein $\mu \in \Lambda$. Da U_μ offen, gibt es ein $\varepsilon > 0$ mit $B_\varepsilon(x) \subset U_\mu \subset \bigcup_{\lambda\in\Lambda} U_\lambda$, was zu zeigen war.

Sei nun $x \in \bigcap_{i=1}^k U_i$ wie in (3). Da U_i offen, gibt es ein $\varepsilon_i > 0$ mit $B_{\varepsilon_i}(x) \subset U_i$ für $i = 1,\ldots,k$. Es folgt $\varepsilon := \min_{i=1,\ldots,k} \varepsilon_i > 0$ und $B_\varepsilon(x) \subset B_{\varepsilon_i}(x) \subset U_i$ für alle $i = 1,\ldots,k$, also $B_\varepsilon(x) \subset \bigcap_{i=1}^k U_i$. $\square$

Wir betonen, dass im Gegensatz zur Vereinigung der Durchschnitt von unendlich vielen offenen Mengen in der Regel nicht offen ist, zum Beispiel gilt

$$\bigcap_{n=1}^\infty B_{1/n}(a) = \{a\}.$$

Eine Menge $M \subset \mathbb{R}^n$ muss weder offen noch abgeschlossen sein; dies trifft zum Beispiel auf ein halboffenes Intervall $[a,b)$ zu. Die Mengen $\mathbb{R}$ und $\emptyset$ sind die einzigen Teilmengen von $\mathbb{R}$, die sowohl offen als auch abgeschlossen sind. Dies soll in den Übungen gezeigt werden. Durch Übergang zu den Komplementen sieht man, dass beliebige Durchschnitte und endliche

Vereinigungen von abgeschlossenen Mengen abermals abgeschlossen sind.

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 5.5 *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

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

□

Man nennt die Euklidische Norm

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

auch den Betrag der komplexen Zahl z (Notation ohne Doppelstriche ist üblich).

Lemma 5.6 *Für den Betrag einer komplexen Zahl gelten folgende Regeln:*

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

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

(3) $|\mathrm{Re}\, z| \leq |z|$, $|\mathrm{Im}\,(z)| \leq |z|$.

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 5.5(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. □

Damit können wir nun das inverse Element der Multiplikation leicht hinschreiben, und zwar folgt aus Lemma 5.6(1) für $z = x + iy \neq 0$

$$\frac{1}{z} = \frac{\overline{z}}{z\,\overline{z}} = \frac{x - iy}{x^2 + y^2}. \tag{5.3}$$

Wir wollen jetzt einige Grundtatsachen über Nullstellen von Polynomen erarbeiten. Im folgenden sei stets $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$.

Definition 5.12 *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:*

$$p(z) = \sum_{k=0}^{n} a_k z^k \quad \textit{für alle } z \in \mathbb{K}. \tag{5.4}$$

Aus der Definition ist nicht unmittelbar ersichtlich, daß der Grad n und die Koeffizienten a_i eindeutig durch die Funktion $p : \mathbb{K} \to \mathbb{K}$ bestimmt sind. Dies wird in Satz 5.10 gezeigt.

Satz 5.9 (Abspaltung von Linearfaktoren) *Sei $p : \mathbb{K} \to \mathbb{K}$ Polynom vom Grad $n \in \mathbb{N}$ 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:*

$$p(z) = (z - \lambda)\,q(z) \quad \textit{für alle } z \in \mathbb{K}. \tag{5.5}$$

Beweis: Für $z, \lambda \in \mathbb{K}$ und $k \in \mathbb{N}$ gilt, vgl. Beispiel 2.2 zur geometrischen Summe,

$$\begin{aligned}
z^k - \lambda^k &= \sum_{j=1}^{k} \lambda^{k-j} z^j - \sum_{j=0}^{k-1} \lambda^{k-j} z^j \\
&= \sum_{j=0}^{k-1} (\lambda^{k-j-1} z^{j+1} - \lambda^{k-j} z^j) \\
&= (z - \lambda) \sum_{j=0}^{k-1} \lambda^{k-j-1} z^j.
\end{aligned}$$

Ist $p(\lambda) = 0$, so folgt weiter

$$\begin{aligned} p(z) = p(z) - p(\lambda) &= \sum_{k=1}^{n} a_k (z^k - \lambda^k) \\ &= (z-\lambda) \sum_{k=1}^{n} \sum_{j=0}^{k-1} a_k \lambda^{k-j-1} z^j \\ &= (z-\lambda) \sum_{j=0}^{n-1} \sum_{k=j+1}^{n} a_k \lambda^{k-j-1} z^j \\ &= (z-\lambda) \sum_{j=0}^{n-1} b_j z^j, \end{aligned}$$

wobei $b_j = \sum_{k=j+1}^{n} a_k \lambda^{k-j-1}$. Insbesondere $b_{n-1} = a_n$, was auch so klar ist. $\square$

Folgerung 5.4 *Sei $p : \mathbb{K} \to \mathbb{K}$ ein Polynom vom Grad $n \in \mathbb{N}_0$. Dann hat p höchstens n verschiedene Nullstellen.*

Beweis: Wir führen Induktion über $n \in \mathbb{N}_0$. Für $n = 0$ gilt $p(z) = a_0$ für alle $z \in \mathbb{K}$, wobei $a_0 \neq 0$, also hat p keine Nullstelle. Ist p Polynom vom Grad $n \in \mathbb{N}$, so hat entweder p keine Nullstelle oder es gilt nach Satz 5.9 $p(z) = (z - \lambda)q(z)$ 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. $\square$

Satz 5.10 (Koeffizientenvergleich) *Seien $p, q : \mathbb{K} \to \mathbb{K}$ Polynome vom Grad m bzw. n, das heißt es gilt mit $a_m, b_n \neq 0$*

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

Ist $p(\lambda_j) = q(\lambda_j)$ für paarweise verschiedene $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ mit $k > \max(n, m)$, so folgt $m = n$ und $a_i = b_i$ für $i = 0, \ldots, n$.

Beweis: Wäre $m \neq n$ oder $a_i \neq b_i$ für ein i, so wäre $p - q$ Polynom vom Grad höchstens $\max(n, m)$ mit $k > \max(n, m)$ Nullstellen, im Widerspruch zu Folgerung 5.4. $\square$

Lemma 5.7 *Jedes Polynom $p : \mathbb{K} \to \mathbb{K}$ vom Grad n hat eine eindeutige Zerlegung*

$$p(z) = (z - \lambda_1)^{\nu_1} \cdot \ldots \cdot (z - \lambda_r)^{\nu_r} \, q(z) \quad \text{für alle } z \in \mathbb{K}. \tag{5.6}$$

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.

Beweis: Die Existenz der Zerlegung folgt induktiv aus Satz 5.9. 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 Satz 5.10 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 Satz 5.10 liefert $q = h$. $\square$

Alles bisher geht von der Annahme aus, dass das Polynom eine Nullstelle hat, andernfalls ist es nutzlos. Es gibt aber reelle Polynome die tatsächlich keine Nullstellen haben, zum Beispiel $p : \mathbb{R} \to \mathbb{R}$, $p(x) = 1 + x^2$. Die Fortsetzung als komplexes Polynom, also $p : \mathbb{C} \to \mathbb{C}$ mit $p(z) = 1 + z^2$, hat allerdings die Nullstellen $z = \pm i$. Das lässt hoffen, dass wir im Komplexen allgemein Nullstellen finden.

Satz 5.11 (Fundamentalsatz der Algebra) *Jedes komplexe Polynom vom Grad $n \geq 1$ hat mindestens eine Nullstelle $\lambda \in \mathbb{C}$.*

Diesen Satz werden wir in Analysis II beweisen. Als Konsequenz ergibt sich jedenfalls

Folgerung 5.5 *Jedes Polynom $p : \mathbb{C} \to \mathbb{C}$ vom Grad n mit Koeffizienten $a_0, a_1, \ldots, a_n \in \mathbb{C}$ zerfällt über $\mathbb{C}$ in Linearfaktoren, das heißt es gilt*

$$p(z) = a_n (z - \lambda_1)^{\nu_1} \cdot \ldots \cdot (z - \lambda_k)^{\nu_k} \quad \text{für alle } z \in \mathbb{C}. \tag{5.7}$$

Dabei sind $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ die Nullstellen von p, und die $\nu_i \geq 1$ sind die Vielfachheiten dieser Nullstellen mit $\nu_1 + \ldots + \nu_k = n$.

Beweis: Das Restpolynom q in (5.6) hat keine Nullstelle in $\mathbb{C}$, also hat q nach dem Fundamentalsatz der Algebra den Grad $n = 0$, das heißt $q(z) = a_n$ für alle $z \in \mathbb{C}$. $\square$

Mithilfe des Fundamentalsatzes der Algebra können wir auch die Situation in $\mathbb{R}$ analysieren. Jedes reelle Polynom $p(x) = a_0 + a_1 x + \ldots + a_n x^n$, das heißt $a_i \in \mathbb{R}$, kann nämlich als komplexes Polynom aufgefasst werden, indem wir für x auch komplexe Zahlen z einsetzen. Damit wird die reelle Funktion $p : \mathbb{R} \to \mathbb{R}$ zu einer komplexen Funktion $p : \mathbb{C} \to \mathbb{C}$ fortgesetzt.

Lemma 5.8 *Sei $p(x) = a_0 + a_1 x + \ldots + a_n x^n$ mit $a_0, a_1, \ldots, a_n \in \mathbb{R}$ ein reelles Polynom vom Grad n. Dann besitzt p, aufgefasst als komplexes Polynom, eine Faktorisierung der Form*

$$p(z) = a_n \prod_{i=1}^{r} (z - \alpha_i)^{\mu_i} \prod_{j=1}^{k} (z - \lambda_j)^{\nu_j} (z - \overline{\lambda_j})^{\nu_j} \quad \text{für alle } z \in \mathbb{C}.$$

Dabei sind $\{\alpha_1, \ldots, \alpha_r\} \subset \mathbb{R}$ die reellen Nullstellen von p mit Vielfachheiten $\mu_i \geq 1$, und $\{\lambda_1, \overline{\lambda_1}, \ldots, \lambda_k, \overline{\lambda_k}\} \subset \mathbb{C}\backslash\mathbb{R}$ die nichtreellen Nullstellen. Diese treten also in Paaren $\lambda_j, \overline{\lambda_j}$ mit gleicher Vielfachheit ν_j auf. Es gilt $\mu_1 + \ldots + \mu_r + 2(\nu_1 + \ldots + \nu_k) = n$.

Beweis: Es gilt für $z \in \mathbb{C}$, da $a_i \in \mathbb{R}$, wegen Lemma 5.5(i)

$$\overline{p(z)} = \overline{a_0 + a_1 z + \ldots + a_n z^n} = a_0 + a_1 \overline{z} + \ldots + a_n \overline{z}^n = p(\overline{z}).$$

Insbesondere folgt aus $p(z) = 0$, dass auch $p(\overline{z}) = 0$. Ist daher $\lambda \in \mathbb{C}\backslash\mathbb{R}$ eine nichtreelle Nullstelle, so erhalten wir nach Satz 5.9 die Zerlegung

$$p(z) = (z - \lambda)(z - \overline{\lambda})\, q(z) = (z^2 - 2(\operatorname{Re}\lambda) z + |\lambda|^2)\, q(z), \tag{5.8}$$

mit $q(z) = b_0 + b_1 z + \ldots + b_{n-2} z^{n-2}$. Beachten Sie, dass wir Satz 5.9 hier über $\mathbb{C}$ anwenden, das heißt die b_i sind zunächst komplexe Zahlen. Um zu sehen, dass die b_i doch reell sind, setzen wir $z := x \in \mathbb{R}$ in (5.8) ein und bilden die Imaginärteile beider Seiten:

$$\begin{aligned} 0 &= \operatorname{Im} p(x) \quad \text{(da } p \text{ reelle Koeffizienten hat)} \\ &= \operatorname{Im} \Big((x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2)\, q(x) \Big) \\ &= (x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2) \operatorname{Im} q(x) \\ &= (x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2) \sum_{i=0}^{n-2} (\operatorname{Im} b_i) x^i. \end{aligned}$$

Da der linke Faktor in der letzten Zeile für $x \in \mathbb{R}$ niemals Null ist, folgt

$$\sum_{i=0}^{n-2} (\operatorname{Im} b_i) x^i = 0 \quad \text{für alle } x \in \mathbb{R}.$$

Nach Satz 5.10 muss $\operatorname{Im} b_i = 0$ für alle i gelten, das heißt $b_i \in \mathbb{R}$. Das Polynom q ist also wieder ein reelles Polynom, und durch Induktion erhalten wir nun die gewünschte Zerlegung. □

Folgerung 5.6 *Jedes Polynom $p : \mathbb{R} \to \mathbb{R}$ vom Grad n mit Koeffizienten $a_0, a_1, \ldots, a_n \in \mathbb{R}$ zerfällt über $\mathbb{R}$ in lineare und quadratische Faktoren. Genauer gilt*

$$p(x) = a_n \prod_{i=1}^{r} (x - \alpha_i)^{\mu_i} \prod_{j=1}^{k} (x^2 - 2(\operatorname{Re}\lambda_j)x + |\lambda_j|^2)^{\nu_j} \quad \textit{für alle } x \in \mathbb{R}, \tag{5.9}$$

wobei die $\alpha_i \in \mathbb{R}$ die reellen und die $\lambda_j, \overline{\lambda_j} \subset \mathbb{C}\backslash\mathbb{R}$ die nichtreellen Nullstellen von p sind.

Komplexe Zahlen traten in der italienischen Renaissance in der ersten Hälfte des 16. Jahrhunderts auf, bei der Lösung von polynomialen Gleichungen [3]. Damals wurden nur reelle Zahlen als Lösungen akzeptiert. Aber bei kubischen Gleichungen gibt es den sogenannten *casus irreducibilis*, wo tatsächlich reelle Lösungen existieren, ihre Berechnung mittels Umformungen und Wurzeln führt aber über komplexe Zahlen. Man sah diese als imaginär an, also nur für die Rechnung eingebildet und nicht real. Die Bezeichnung $i = \sqrt{-1}$ stammt von Euler (1777), und der Begriff *komplexe Zahl* taucht zuerst bei Gauß auf (1831). Der Fundamentalsatz der Algebra ist zuerst von Gauß in seiner Dissertation (Helmstedt 1799) bewiesen worden; wir werden den Satz im zweiten Semester zeigen. Trotz seines Namens erfordert der Beweis Methoden der Analysis oder Topologie. Es ist für die Algebraiker ein ewiges Ärgernis, dass sie ihren Fundamentalsatz nicht selber beweisen können.

[3] Hellmuth Gericke: Mathematik im Abendland, pp. 236-241, 3. Auflage, Fourier Verlag Wiesbaden 1994

6 Reihen

Unendliche Summen, sogenannte Reihen, spielen in der Analysis eine wichtige Rolle. Zum Beispiel ist die Eulersche Zahl als die Reihe $e = \sum_{k=0}^{\infty} 1/k!$ definiert. Die Exponentialfunktion und die trigonometrischen Funktionen können als Reihen definiert beziehungsweise dargestellt werden, das werden wir in den Kapiteln 11 und 12 noch sehen. Darum ist es wichtig, die Konvergenzfrage für Reihen zu untersuchen.

Definition 6.1 *Eine Folge $(S_n)_{n \in \mathbb{N}_0}$ heißt Reihe mit Gliedern $a_n \in \mathbb{C}$, falls gilt:*

$$S_n = \sum_{k=0}^{n} a_k \quad \textit{für alle } n \in \mathbb{N}_0.$$

Die Reihe heißt konvergent mit Wert $S \in \mathbb{C}$, wenn die Folge $(S_n)_{n \in \mathbb{N}_0}$ gegen S konvergiert:

$$\sum_{k=0}^{\infty} a_k := \lim_{n \to \infty} \sum_{k=0}^{n} a_n = \lim_{n \to \infty} S_n = S.$$

S_n wird allgemein als n-te Partialsumme der Reihe bezeichnet. Oft wird die Reihe kurz mit $a_0 + a_1 + a_2 + \ldots$ angegeben, also durch die ersten Glieder. Leider ist es auch üblich, die Reihe selbst mit dem Symbol $\sum_{k=0}^{\infty} a_k$ zu bezeichnen, egal ob sie konvergiert oder nicht.

Beispiel 6.1 Die Reihe $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \ldots$ wird auch mit $\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$ bezeichnet. Gemeint ist jeweils die Folge $(S_n)_{n \in \mathbb{N}}$ mit den Gliedern

$$S_1 = \frac{1}{1 \cdot 2} = \frac{1}{2}, \quad S_2 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} = \frac{2}{3}, \quad S_3 = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} = \frac{3}{4}, \ldots$$

Für die betrachtete Reihe gilt, für alle $n \in \mathbb{N}$,

$$S_n = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1} \right) = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=2}^{n+1} \frac{1}{k} = 1 - \frac{1}{n+1}.$$

Also ist die Reihe konvergent mit Wert

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1.$$

Wie hier im Beispiel kann eine Reihe statt bei $k = 0$ auch bei $k = 1$ oder einer anderen Zahl starten. Konvergente Reihen können addiert und mit komplexen Zahlen multipliziert werden. Genauer: sind $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ konvergente Reihen und $\lambda, \mu \in \mathbb{C}$, so folgt mit der Regel für Folgen, Satz 3.3(a),

$$\begin{aligned} \sum_{k=0}^{n} (\lambda a_k + \mu b_k) &= \lambda \sum_{k=0}^{n} a_k + \mu \sum_{k=0}^{n} b_k \to \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k, \quad \text{also} \\ \sum_{k=0}^{\infty} (\lambda a_k + \mu b_k) &= \lambda \sum_{k=0}^{\infty} a_k + \mu \sum_{k=0}^{\infty} b_k. \end{aligned}$$

Werden endlich viele Glieder einer Reihe geändert, hinzugefügt oder weggelassen, so wird dadurch die Konvergenz oder Divergenz nicht beeinflusst. Allerdings wird sich der Wert der Reihe natürlich entsprechend ändern. Zum Beispiel haben wir für $n \in \mathbb{N}_0$ und $m \geq n$

$$\sum_{k=n}^{m} a_k = \sum_{k=0}^{m} a_k - \sum_{k=0}^{n-1} a_k,$$

und mit $m \to \infty$ folgt, falls die Reihe $\sum_{k=0}^{\infty} a_k$ konvergiert,

$$\sum_{k=n}^{\infty} a_k = \sum_{k=0}^{\infty} a_k - \sum_{k=0}^{n-1} a_k. \tag{6.1}$$

Jede Folge $b_0, b_1, , \ldots$ komplexer Zahlen kann auch als Reihe $\sum_{k=0}^{n} a_k$ aufgefasst werden, denn mit $a_0 = b_0$ und $a_k = b_k - b_{k-1}$ für $k \geq 1$ gilt

$$\sum_{k=0}^{n} a_k = b_0 + \sum_{k=1}^{n} (b_k - b_{k-1}) = b_0 + \sum_{k=1}^{n} b_k - \sum_{k=0}^{n-1} b_k = b_n.$$

Der Unterschied zwischen Folgen und Reihen besteht in der Sichtweise, Reihen werden durch die Zuwächse a_n beschrieben. Es stellen sich zwei Fragen:

- Wie kann ich den Gliedern a_n ansehen, ob die Reihe konvergiert bzw. divergiert?
- Im Fall der Konvergenz: welchen Wert hat die Reihe?

Bei der zweiten Frage ist zum Beispiel gemeint, ob eine bereits definierte Zahl wie $\sqrt{2}$, e, π, ... als Grenzwert einer Reihe dargestellt werden kann. Im Folgenden steht aber die erste Frage im Zentrum des Interesses. Dabei ist das nächste Beispiel fundamental.

Beispiel 6.2 (Geometrische Reihe) Die geometrische Reihe $\sum_{k=0}^{\infty} z^k$ mit $z \in \mathbb{C}$ konvergiert genau für $|z| < 1$. Der Konvergenzbeweis ist wie in Beispiel 3.7, und zwar gilt

$$\sum_{k=0}^{n} z^k = \frac{1 - z^{n+1}}{1 - z} \to \frac{1}{1 - z} \quad \text{mit } n \to \infty.$$

Dagegen gilt für $|z| \geq 1$ mit $S_n = \sum_{k=0}^{n} z^k$

$$|S_{n+1} - S_n| = |z^{n+1}| = |z|^{n+1} \geq 1,$$

so dass $(S_n)_{n \in \mathbb{N}}$ keine Cauchyfolge sein kann.

Beispiel 6.3 (Unendliche Dezimalbrüche) Ist $(k_j)_{j \in \mathbb{N}_0}$ eine Folge von Ziffern $k_j \in \{0, 1, \ldots, 9\}$, so ist die Reihe $\sum_{j=0}^{\infty} k_j \, 10^{-j}$ konvergent, vgl. Beispiel 4.4.

Beispiel 6.4 (Harmonische Reihe) Die harmonische Reihe $\sum_{k=1}^{\infty} 1/k$ ist bestimmt divergent gegen $+\infty$. Dies zeigen wir, indem wir wie folgt Klammern setzen:

$$\underbrace{\left(\frac{1}{1}\right)}_{\geq 1 \cdot \frac{1}{2}} + \underbrace{\left(\frac{1}{2} + \frac{1}{3}\right)}_{\geq 2 \cdot \frac{1}{4}} + \underbrace{\left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right)}_{\geq 4 \cdot \frac{1}{8}} + \underbrace{\left(\frac{1}{8} + \frac{1}{9} + \ldots + \frac{1}{15}\right)}_{\geq 8 \cdot \frac{1}{16}} + \ldots$$

Allgemein enthält der Abschnitt $2^m \leq k < 2^{m+1}$, mit $m \in \mathbb{N}_0$, 2^m Summanden $1/k$, die jeweils größer sind als $2^{-(m+1)}$. Also ist der Beitrag dieses Abschnitts mindestens $2^m \cdot 2^{-(m+1)} = 1/2$, und die Reihe konvergiert gegen $+\infty$.

Nach diesen ersten Beispielen nun zur allgemeinen Konvergenzfrage für Reihen. Wir beginnen mit einem notwendigen Kriterium.

Satz 6.1 (Nullfolgentest) *Ist die Reihe $\sum_{k=0}^{\infty} a_k$ konvergent, so folgt $\lim_{k\to\infty} a_k = 0$.*

Beweis: Nach Voraussetzung ist die Folge der Partialsummen $S_n = \sum_{k=0}^{n} a_k$ konvergent mit Grenzwert $S \in \mathbb{C}$, also folgt $a_n = S_n - S_{n-1} \to S - S = 0$. □

Als nächstes formulieren wir unsere Konvergenzkriterien für Folgen neu in der Situation von Reihen. Die Cauchyfolgeneigenschaft sieht wie folgt aus.

Satz 6.2 (Konvergenzkriterium von Cauchy) *Eine Reihe $\sum_{k=0}^{\infty} a_k$ konvergiert dann und nur dann, wenn es zu jedem $\varepsilon > 0$ ein $N \in \mathbb{R}$ gibt, so dass gilt:*

$$\left|\sum_{k=n}^{m} a_k\right| < \varepsilon \quad \text{für alle } n, m \in \mathbb{N}_0 \text{ mit } m \geq n > N.$$

Beweis: Satz 4.2. □

Satz 6.3 (Reihen mit Gliedern $a_k \geq 0$) *Eine reelle Reihe $\sum_{k=0}^{\infty} a_k$ mit $a_k \geq 0$ für alle k konvergiert genau dann, wenn die Folge der Partialsummen $S_n = \sum_{k=0}^{n} a_k$ nach oben beschränkt ist.*

Beweis: Da $a_k \geq 0$, ist die Folge (S_n) der Partialsummen monoton wachsend. Die Behauptung folgt aus Satz 4.3 und Satz 3.2. □

Beispiel 6.5 Für $s > 1$ ist die Reihe $\sum_{k=1}^{\infty} k^{-s}$ konvergent[4]. Wir schätzen dazu die Abschnitte $2^m \leq k < 2^{m+1}$, $m \in \mathbb{N}_0$, nach oben ab. Die Funktion $f(x) = x^{-s}$ ist monoton fallend nach Satz 4.6, also gilt

$$\sum_{2^m \leq k < 2^{m+1}} k^{-s} \leq 2^m \cdot (2^m)^{-s} = (2^{1-s})^m.$$

Für die Summe der Abschnitte mit Nummern $m = 0, \ldots, M$ folgt

$$\sum_{1 \leq k < 2^{M+1}} k^{-s} = \sum_{m=0}^{M} \Big(\sum_{2^m \leq k < 2^{m+1}} k^{-s} \Big) \leq \sum_{m=0}^{M} (2^{1-s})^m \leq \frac{1}{1 - 2^{1-s}}.$$

Beachte dabei $2^{1-s} < 1$ für $s > 1$, so dass die geometrische Reihe rechts konvergiert, siehe Beispiel 3.7. Da die Folge der Partialsummen S_n monoton wachsend ist, gilt die Abschätzung für alle Partialsummen S_n, und die Reihe konvergiert nach Satz 6.3. Die Zahl $s \in (0, \infty)$ ist ein Parameter, wir haben also die Funktion

$$\zeta : (1, \infty) \to \mathbb{R},\ \zeta(s) = \sum_{k=1}^{\infty} k^{-s}.$$

Dies ist die Riemannsche Zetafunktion, sie spielt bei der Untersuchung der Verteilung der Primzahlen eine fundamentale Rolle.

[4] k^{-s} ist bisher nur definiert für $s \in \mathbb{Q}$, vgl. aber Definition 11.2

Die Reihe $\sum_{k=1}^{\infty}(-1)^{k-1}/k = 1 - 1/2 + 1/3 - + \ldots$ ist ein Beispiel für eine sogenannte alternierende Reihe. Während die entsprechende Reihe mit nur positiven Vorzeichen nicht konvergiert – es ist die harmonische Reihe aus Beispiel 6.4 – ist die Reihe mit dem Vorzeichenwechsel konvergent. Dies ergibt sich aus dem nächsten Satz, bei dessen Beweis wieder Monotonieargumente eine wesentliche Rolle spielen.

Satz 6.4 (Leibniz-Kriterium für alternierende Reihen) *Sei a_k, $k \in \mathbb{N}_0$, eine reelle, monoton fallende Nullfolge (also insbesondere $a_k \geq 0$). Dann ist die Reihe $\sum_{k=0}^{\infty}(-1)^k a_k$ konvergent. Außerdem gilt für alle $n \in \mathbb{N}_0$ die Abschätzung*

$$0 \leq (-1)^n \sum_{k=n}^{\infty}(-1)^k a_k \leq a_n. \tag{6.2}$$

Beweis: Wir betrachten die Partialsummen $S_n = \sum_{k=0}^{n}(-1)^k a_k$. Für alle $n \in \mathbb{N}_0$ gilt:

$$\begin{aligned} S_{2n+2} - S_{2n} &= (-1)^{2n+2}a_{2n+2} + (-1)^{2n+1}a_{2n+1} = a_{2n+2} - a_{2n+1} \leq 0, \\ S_{2n+3} - S_{2n+1} &= (-1)^{2n+3}a_{2n+3} + (-1)^{2n+2}a_{2n+2} = a_{2n+2} - a_{2n+3} \geq 0, \\ S_{2n+1} - S_{2n} &= (-1)^{2n+1}a_{2n+1} = -a_{2n+1} \leq 0. \end{aligned}$$

Also ist $(S_{2n})_{n\in\mathbb{N}_0}$ monoton fallend, und $(S_{2n+1})_{n\in\mathbb{N}_0}$ monoton wachsend. Weiter haben wir für alle $n \in \mathbb{N}_0$ die untere bzw. obere Schranke

$$S_{2n} \geq S_{2n+1} \geq S_1 \quad \text{und } S_{2n+1} \leq S_{2n} \leq S_0.$$

Nach Satz 4.3 sind die Folgen S_{2n} und S_{2n+1} konvergent. Aber $S_{2n+1} - S_{2n} = -a_{2n+1} \to 0$ nach Voraussetzung. Also konvergieren beide Folgen, und damit die gesamte Folge, gegen denselben Grenzwert $S \in \mathbb{R}$. Zusätzlich sehen wir

$$0 \leq a_1 - a_0 = S_1 \leq \lim_{n\to\infty} S_{2n} = S = \lim_{n\to\infty} S_{2n+1} \leq S_0 = a_0,$$

das heißt Abschätzung (6.2) gilt im Fall $n = 0$. Für beliebiges n wenden wir das an auf $b_l = a_{n+l}$. Das ist ebenfalls eine monoton fallende Nullfolge, also gilt

$$0 \leq \sum_{l=0}^{\infty}(-1)^l b_l \leq b_0 = a_n.$$

Abschätzung (6.2) folgt nun wegen

$$\sum_{l=0}^{\infty}(-1)^l b_l = (-1)^n \sum_{l=0}^{\infty}(-1)^{n+l} a_{n+l} = (-1)^n \sum_{k=n}^{\infty}(-1)^k a_k.$$

□

Bei der alternierenden Reihe $1 - 1/2 + 1/3 - + \ldots$ stößt man auf Merkwürdigkeiten, wenn man die Summationsreihenfolge ändert. Während die Ausgangsreihe konvergent ist, ist die durch Umordnung entstehende Reihe

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{5} - \frac{1}{4}\right) + \left(\frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13} - \frac{1}{6}\right) + \ldots.$$

bestimmt divergent gegen $+\infty$, denn die Summe der positiven Zahlen in der m-ten Klammer ist mindestens $2^{m-1} \cdot 2^{-(m+1)} = 1/4$. Dies ist ein interessantes Phänomen, jedoch sind wir in erster Linie an Reihen interessiert, deren Konvergenz stabiler ist. Dabei ist der folgende Begriff zentral.

Definition 6.2 *Die Reihe $\sum_{k=0}^{\infty} a_k$ mit $a_k \in \mathbb{C}$ heißt absolut konvergent, wenn die Reihe $\sum_{k=0}^{\infty} |a_k|$ konvergiert, das heißt $\sum_{k=0}^{\infty} |a_k| < \infty$.*

Satz 6.5 (absolut konvergent $\Rightarrow$ konvergent) *Wenn die Reihe $\sum_{k=0}^{\infty} a_k$ absolut konvergiert, so ist sie konvergent und es gilt*

$$\left|\sum_{k=0}^{\infty} a_k\right| \leq \sum_{k=0}^{\infty} |a_k|.$$

Beweis: Die Dreiecksungleichung besagt

$$\left|\sum_{k=n}^{m} a_k\right| \leq \sum_{k=n}^{m} |a_k| \quad \text{für } m \geq n \geq 0.$$

Aus dem Cauchykriterium für $\sum_{k=0}^{\infty} |a_k|$ folgt deshalb das Cauchykriterium für $\sum_{k=0}^{\infty} a_k$, das heißt die Reihe konvergiert nach Satz 6.2. Die Abschätzung folgt, indem wir in der Dreiecksungleichung $n = 0$ setzen und $m \to \infty$ gehen lassen. $\square$

Der folgende Satz fasst drei wichtige Kriterien für die absolute Konvergenz zusammen.

Satz 6.6 (Tests für absolute Konvergenz) *Sei $\sum_{k=0}^{\infty} a_k$ eine Reihe mit $a_k \in \mathbb{C}$. Ist eine der drei folgenden Bedingungen erfüllt, so ist $\sum_{k=0}^{\infty} a_k$ absolut konvergent:*

(a) Majorantenkriterium (M-Test): *Es gilt $|a_k| \leq c_k \in [0, \infty)$ mit $\sum_{k=0}^{\infty} c_k < \infty$.*

(b) Quotientenkriterium: *Es gibt ein $\theta \in [0, 1)$ und ein $n \in \mathbb{N}$ mit*

$$\frac{|a_{k+1}|}{|a_k|} \leq \theta \text{ für alle } k \geq n \quad (\text{wobei } a_k \neq 0 \text{ für } k \geq n).$$

(c) Wurzelkriterium: *Es gibt ein $\theta \in [0, 1)$ und ein $n \in \mathbb{N}$ mit $\sqrt[k]{|a_k|} \leq \theta$ für alle $k \geq n$.*

Umgekehrt ist die Reihe divergent, wenn für ein $n \in \mathbb{N}$ gilt:

$$\frac{|a_{k+1}|}{|a_k|} \geq 1 \quad \text{oder} \quad \sqrt[k]{|a_k|} \geq 1 \text{ für alle } k \geq n.$$

Für die harmonische Reihe $\sum_{k=1}^{\infty} 1/k$ haben wir

$$\frac{|a_{k+1}|}{|a_k|} = \frac{k}{k+1} < 1 \text{ für alle } k \quad \text{und} \quad \lim_{k\to\infty} \frac{|a_{k+1}|}{|a_k|} = 1.$$

Also reicht es *nicht*, in (b) nur die Ungleichung $|a_{k+1}|/|a_k| < 1$ vorauszusetzen. Andererseits folgt aus $\lim_{k\to\infty} |a_{k+1}|/|a_k| = 1$ auch nicht notwendig Divergenz: die Reihe $\sum_{k=1}^{\infty} 1/k^2$ hat $|a_{k+1}|/|a_k| = (k/k+1)^2 \to 1$, aber sie konvergiert nach Beispiel 6.5. Eine ganz analoge Diskussion gilt für das Wurzelkriterium. Die Bedingungen (b) beziehungsweise (c) sind äquivalent zu den folgenden:

$$\limsup_{k\to\infty} \frac{|a_{k+1}|}{|a_k|} < 1 \quad \text{bzw.} \quad \limsup_{k\to\infty} \sqrt[k]{|a_k|} < 1.$$

Beweis des Satzes: (a) folgt aus Satz 6.3, denn für alle $n \in \mathbb{N}$ hat man die obere Schranke

$$\sum_{k=0}^{n} |a_k| \leq \sum_{k=0}^{n} c_k \leq \sum_{k=0}^{\infty} c_k < \infty.$$

Voraussetzung (b) liefert per Induktion

$$|a_k| \leq \theta^{k-n} |a_n| \quad \text{für } k \geq n.$$

Da die geometrische Reihe wegen $0 \leq \theta < 1$ nach Beispiel 3.7 konvergiert, folgt die Behauptung aus (a) und wir erhalten außerdem die Abschätzung

$$\sum_{k=n}^{\infty} |a_k| \leq \frac{|a_n|}{1-\theta}. \tag{6.3}$$

Unter der Voraussetzung (c) gilt

$$|a_k| \leq \theta^k \quad \text{für } k \geq n.$$

Wieder folgt die Behauptung durch M-Test mit der geometrischen Reihe. Die Abschätzung lautet hier

$$\sum_{k=n}^{\infty} |a_k| \leq \frac{\theta^n}{1-\theta}. \tag{6.4}$$

Die Divergenzaussagen folgen unmittelbar aus dem Nullfolgentest, Satz 6.1. □

Die folgenden zwei Sätze werden in der Vorlesung nicht gebraucht, gehören aber zum Standardrepertoire der Analysis.

Erstens geht es um die Multiplikation von zwei Reihen $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$. Wir schreiben die Produkte $a_k b_l$ mit $k, l \in \mathbb{N}_0$ in einem Schema auf, so dass $a_k b_l$ in der k-ten Zeile und l-ten Spalte steht. Es ist dann naheliegend, erst die $n+1$ Produkte $a_k b_l$ in jeder Diagonale $k + l = n$ zu addieren, und dann die Konvergenz der resultierenden Reihe zu studieren. Im folgenden schreiben wir kurz $\sum_{k+l=n}$ für die Summe über alle Paare $k, l \in \mathbb{N}_0$ mit $k + l = n$.

Satz 6.7 (Cauchyprodukt) *Die Reihen $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ seien absolut konvergent. Dann ist auch die Reihe*

$$\sum_{k=0}^{\infty} c_n \quad \text{mit} \quad c_n = \sum_{k+l=n} a_k b_l = \sum_{k=0}^{n} a_k b_{n-k}$$

absolut konvergent, und es gilt

$$\sum_{n=0}^{\infty} c_n = \sum_{n=0}^{\infty} \left(\sum_{k+l=n} a_k \, b_l \right) = \left(\sum_{k=0}^{\infty} a_k \right) \cdot \left(\sum_{k=0}^{\infty} b_k \right).$$

Beweis: Wir setzen für $N \in \mathbb{N}_0$

$$A_N := \sum_{k=0}^{N} |a_k| \to \sum_{k=0}^{\infty} |a_k| =: A \quad \text{und} \quad B_N := \sum_{k=0}^{N} |b_k| \to \sum_{k=0}^{\infty} |b_k| =: B,$$

wobei nach Voraussetzung $A, B < \infty$. Es folgt

$$\sum_{n=0}^{N} |c_n| \le \sum_{k+l \le N} |a_k|\,|b_l| \le \sum_{k,l \le N} |a_k|\,|b_l| = A_N B_N \le AB < \infty.$$

Nach Satz 6.3 ist $\sum_{n=0}^{\infty} c_n$ absolut konvergent und insbesondere konvergent (Satz 6.5). Um den Grenzwert zu identifizieren, reicht es also die geraden Partialsummen $\sum_{n=0}^{2N} c_n$ zu betrachten. Wir berechnen

$$\left| \left(\sum_{k=0}^{2N} a_k \right) \cdot \left(\sum_{l=0}^{2N} b_l \right) - \sum_{n=0}^{2N} c_n \right| = \left| \sum_{k,l \le 2N} a_k b_l - \sum_{k+l \le 2N} a_k b_l \right| \le \sum_{k,l \le 2N, \max(k,l) > N} |a_k|\,|b_l|,$$

Die rechte Seite ist gleich $A_{2N}\, B_{2N} - A_N\, B_N$, konvergiert also gegen Null für $N \to \infty$. Für die letzte Abschätzung ist es hilfreich, die Indexbereiche zu skizzieren. Jedenfalls ist damit die gewünschte Formel für das Cauchyprodukt bewiesen. □

Zweitens kommen wir zu der Frage der Umordnung zurück und zeigen, dass absolut konvergente Reihen beliebig umgeordnet werden können.

Satz 6.8 (Umordnungssatz) *Sei $\tau : \mathbb{N} \to \mathbb{N}$ bijektiv. Falls $\sum_{k=1}^{\infty} a_k$ absolut konvergent ist, so konvergiert auch die Reihe $\sum_{i=1}^{\infty} a_{\tau(i)}$ absolut und hat denselben Grenzwert.*

Beweis: Für $K \ge \max\{\tau(i) : 1 \le i \le m\}$ ist

$$\sum_{i=1}^{m} |a_{\tau(i)}| \le \sum_{k=1}^{K} |a_k| \le \sum_{k=1}^{\infty} |a_k| < \infty.$$

Nach dem Majorantenkriterium konvergiert die umgeordnete Reihe absolut. Sei nun $S = \sum_{k=1}^{\infty} a_k$. Für gegebenes $K \in \mathbb{N}$ und $m \ge \max\{\tau^{-1}(k) : 1 \le k \le K\}$, bzw. äquivalent dazu $\{\tau(1), \dots, \tau(m)\} \supset \{1, \dots, K\}$, gilt

$$\left| S - \sum_{i=1}^{m} a_{\tau(i)} \right| \le \left| S - \sum_{k=1}^{K} a_k \right| + \left| \sum_{k=1}^{K} a_k - \sum_{i=1}^{m} a_{\tau(i)} \right| \le \left| S - \sum_{k=1}^{K} a_k \right| + \sum_{k=K+1}^{\infty} |a_k|.$$

Die Behauptung folgt, indem wir erst $m \to \infty$ und dann $K \to \infty$ gehen lassen. □

7 Stetigkeit

In diesem Kapitel beginnt das Studium von Funktionen mit den Methoden der Analysis. Eine Funktion auf einer Menge $D \subset \mathbb{R}^n$ mit Werten in $\mathbb{R}^m$ ist bekanntlich eine Abbildung $f : D \to \mathbb{R}^m$, $x \mapsto f(x)$. Im Fall $m = 1$ heißt die Funktion reellwertig.

Beispiel 7.1 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 $c(t) = (x(t), y(t), z(t))$.

Beispiel 7.2 Die Abstandsfunktion von einem Punkt $a \in \mathbb{R}^n$ ist

$$f : \mathbb{R}^n \to \mathbb{R},\ f(x) = \|x - a\| = \Big(\sum_{i=1}^{n} (x_i - a_i)^2\Big)^{\frac{1}{2}}.$$

Der Begriff der Stetigkeit ist ein fundamentales Konzept der Analysis. In der Schule wird oft die anschauliche Charakterisierung gegeben, dass sich der Graph der Funktion zeichnen lässt, ohne mit dem Stift abzusetzen. Das ist nicht schlecht, aber in mathematischen Argumenten können wir es nicht verwenden. Unsere Definition ist ähnlich zu der des Grenzwerts.

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

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

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 $f(D \cap B_\delta(x_0)) \subset B_\varepsilon(f(x_0))$.

Wir wollen einige einfache Beispiele betrachten.

Beispiel 7.3 Eine konstante Funktion $f : D \to \mathbb{R}^m$ ist in jedem $x_0 \in D$ stetig, denn

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

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

Beispiel 7.4 Für $f : \mathbb{R} \to \mathbb{R}$, $f(x) = ax + b$ mit $a \neq 0$, gilt die Abschätzung

$$|f(x) - f(x_0)| = |ax - ax_0| = |a|\,|x - x_0|.$$

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

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

Somit ist die Funktion auf ganz $\mathbb{R}$ stetig.

Beispiel 7.5 Die Funktion $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, ist auf ganz $\mathbb{R}$ stetig. Es gilt

$$\begin{aligned} |f(x) - f(x_0)| &= |x^2 - x_0^2| \\ &\leq (|x| + |x_0|)\,|x - x_0| \\ &\leq (1 + 2|x_0|)\,|x - x_0| \quad \text{für } |x - x_0| \leq 1. \end{aligned}$$

Hier wurde $|x| \leq |x - x_0| + |x_0|$ abgeschätzt. Zu $\varepsilon > 0$ wähle nun $\delta = \min(1, \frac{\varepsilon}{1+2|x_0|}) > 0$. Für $|x - x_0| < \delta$, insbesondere $|x - x_0| \leq 1$, folgt

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

Der Graph von f wird zunehmend steiler für x_0 groß, daher muss $\delta > 0$ von x_0 abhängen,

Beispiel 7.6 Die charakteristische Funktion (Indikatorfunktion) einer Menge $E \subset \mathbb{R}$ ist

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

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

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

Ist $\chi_E(x_0) = 1$, so folgt $\chi_E(x) = 1$ und somit $(x_0 - \delta, x_0 + \delta) \subset E$. Ist $\chi_E(x_0) = 0$, so folgt analog $(x_0 - \delta, x_0 + \delta) \subset \mathbb{R}^n \backslash E$. 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 1.6. Aber $\mathbb{R} \backslash \mathbb{Q}$ ist ebenfalls dicht, zum Beispiel ist $\sqrt{2} + \mathbb{Q}$ dicht in $\mathbb{R}$ und Teilmenge von $\mathbb{R} \backslash \mathbb{Q}$ nach Satz 4.1. 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.

Beispiel 7.7 Die Euklidische Norm auf $\mathbb{R}^n$ (Betragsfunktion im Fall $\mathbb{R}$ bzw. $\mathbb{C}$) ist

$$f : \mathbb{R}^n \to \mathbb{R},\ f(x) = \|x\| = (x_1^2 + \ldots + x_n^2)^{1/2}.$$

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 also $\delta = \varepsilon$ nehmen.

Ein sehr nützliches hinreichendes Kriterium für die Stetigkeit ist das folgende.

Beispiel 7.8 $f : D \to \mathbb{R}^m$ heißt Lipschitzstetig mit Konstante $L \in [0, \infty)$, falls gilt:

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

Zum Beispiel ist die Euklidische Norm Lipschitzstetig mit Konstante $L = 1$, vgl. Beispiel 7.7. 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 7.9 Die Abstandsfunktion $\mathrm{dist}_E : \mathbb{R}^n \to \mathbb{R}$ einer Menge $E \subset \mathbb{R}^n$ ist Lipschitzstetig mit Konstante Eins. Denn es gilt nach der 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 alle $y \in E$, 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 7.10 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, und schätzen mit der Dreiecksungleichung und Cauchy-Schwarz ab:

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

Also gilt die Abschätzung

$$\|Ax\| \leq \|A\|\,\|x\| \quad \text{für alle } x \in \mathbb{R}^n,$$

wobei die Euklidische Norm der Matrix A definiert ist durch

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

Es folgt weiter

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

Also ist A Lipschitzstetig mit Konstante $\|A\|$. Der Beweis verwendet, dass $\mathbb{R}^n$ eine endliche Basis besitzt. 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 3.3; 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 $(x_k)_{k\in\mathbb{N}}$ mit $x_k \in D$ und $\lim_{k\to\infty} x_k = x_0$ 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 (7.1) 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. □

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. □

Beispiel 7.11 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 $\min(f, g) = (f + g - |f - g|)/2$.

Lemma 7.1 *Eine Funktion $f : D \to \mathbb{R}^m$ ist genau dann stetig in $x_0 \in D$, wenn jede Koordinatenfunktion $f_i : D \to \mathbb{R}$ in x_0 stetig ist.*

Beweis: Nach Satz 5.6 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. □

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

Satz 7.3 (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: Die Funktionen sind grundsätzlich punktweise erklärt, das heißt für alle $x \in D$ gilt

$$(\lambda f + \mu g)(x) = \lambda f(x) + \mu g(x),\ (fg)(x) = f(x)g(x) \text{ und } (f/g)(x) = f(x)/g(x).$$

Wir führen die Aussagen auf die entsprechenden Regeln für Folgen zurück: sei $(x_n)_{n\in\mathbb{N}}$ eine beliebige Folge mit $x_n \in D$ und $\lim_{n\to\infty} x_n = x_0$. Nach Satz 7.1 gilt $f(x_n) \to f(x_0)$ und $g(x_n) \to g(x_0)$ mit $n \to \infty$. Aus Satz 3.3 folgt

$$\lambda f(x_n) + \mu g(x_n) \to \lambda f(x_0) + \mu g(x_0) \quad \text{sowie} \quad f(x_n)g(x_n) \to f(x_0)g(x_0).$$

Mit Satz 7.1 ergeben sich (1) und (2). Aussage (3) folgt ebenso, falls $g \neq 0$ auf D: dann ist $f/g : D \to \mathbb{R}$ definiert, und für jede Folge $x_n \to x_0$ gilt $f(x_n)/g(x_n) \to f(x_0)/g(x_0)$. Somit ist f/g stetig in x_0 nach Satz 7.1. Die Annahme $g \neq 0$ auf D ist aber zu stark, es wird ja nur eine lokale Aussage bei x_0 behauptet. Offenbar reicht $g \neq 0$ auf $D \cap B_\delta(x_0)$, wenn wir f/g nur auf $D \cap B_\delta(x_0)$ betrachten. Die Existenz dieser Umgebung wird im Anschluss bewiesen. □

Lemma 7.2 *Sei $g : D \to \mathbb{R}$ stetig in x_0 mit $g(x_0) \neq 0$. Dann gibt es ein $\delta > 0$ mit*

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

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

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

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

Eine offensichtliche Konsequenz von Satz 7.3(1), sowie von Lemma 7.1 im Fall $m \geq 2$, ist

Folgerung 7.1 (C^0-Räume) *Die stetigen Abbildungen $C^0(D, \mathbb{R}^m)$ bilden einen Untervektorraum des Raums aller Abbildungen $f : D \to \mathbb{R}^m$, mit der punktweisen Addition und Skalarmultiplikation.*

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

Definition 7.2 (Grenzwert für Funktionen) *Sei $D \subset \mathbb{R}^n$ und x_0 ein Häufungspunkt von D. 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* $x \to x_0$.

Für die Existenz und den Wert von $\lim_{x \to x_0} f(x)$ ist es egal, ob die Funktion f in x_0 definiert ist bzw. welchen Funktionswert sie dort hat. Die Beziehung zwischen Grenzwert und Stetigkeit ist wie folgt:

Lemma 7.3 (Stetigkeit und Grenzwert) *Sei $x_0 \in D$ ein Häufungspunkt von D. Für die Funktion $f : D \to \mathbb{R}^m$ 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:

(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. □

Für reellwertige Funktionen können wir den Konvergenzbegriff ausdehnen, indem wir (uneigentliche) Konvergenz gegen $\pm\infty$ zulassen.

Definition 7.3 (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*

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

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

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

Definition 7.4 (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*

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

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

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 7.4 (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. Einziger Unterschied: in Definition 7.2 kommen nur $x \in D$ vor mit $\|x - x_0\| > 0$. Entsprechend werden in (2) hier nur Folgen mit $x_k \neq x_0$ betrachtet. □

Für Grenzwerte von Funktionen gelten Rechenregeln analog zu Satz 7.3. Der Beweis wird den Lesern/innen überlassen.

Satz 7.5 (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*

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

Im Fall $D \subset \mathbb{R}$ und $x_0 = \pm\infty$ gelten analoge Aussagen.

(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 $a \geq 0$.*

(4) *Ist $f : D \to (0, \infty)$, so ist $\lim_{x\to x_0} f(x) = 0$ äquivalent zu $\displaystyle\lim_{x\to x_0} \frac{1}{f(x)} = +\infty$.*

Beispiel 7.12 (Rationale Funktionen) Seien p, q reelle Polynome vom Grad m bzw. n, also $p(x) = a_0 + a_1 x + \ldots + a_m x^m$ bzw. $q(x) = b_0 + b_1 x + \ldots + b_n x^n$ mit $a_i, b_j \in \mathbb{R}$ und $a_m, b_n \neq 0$. Setze $N = \{x \in \mathbb{R} : q(x) = 0\}$ und definiere

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

Wann hat f eine stetige Fortsetzung bei $x_0 \in N$? Sei dazu x_0 eine ν-fache Nullstelle von q mit $\nu \geq 1$, und eine μ-fache Nullstelle von p, evtl. $\mu = 0$ falls $p(x_0) \neq 0$. Also gilt

$$p(x) = (x - x_0)^{\mu}\, \tilde{p}(x) \quad \text{bzw.} \quad q(x) = (x - x_0)^{\nu} \tilde{q}(x)$$

für Polynome $\tilde{p}, \tilde{q}$ mit $\tilde{p}(x_0), \tilde{q}(x_0) \neq 0$. Es folgt

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

und wir erhalten aus den Rechenregeln für Grenzwerte

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

In diesen beiden Fällen ist die Funktion stetig fortsetzbar nach Lemma 7.3, dagegen hat f im Fall $\mu < \nu$ in x_0 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	$+\infty$	$-\infty$

Das Verhalten von f für $x \to \pm\infty$ wurde bereits in Beispiel 3.6 untersucht.

8 Zwischenwertsatz und monotone Funktionen

In diesem Abschnitt haben wir es mit reellen Funktionen zu tun, die auf einem Intervall definiert sind. Folgendes Kriterium ist dabei praktisch: eine Menge $I \subset \mathbb{R}$ ist genau dann ein Intervall mit Grenzen $a \leq b$, wenn gilt:

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

Unendliche Intervallgrenzen sind hier erlaubt, aber das Intervall soll stets Teilmenge von $\mathbb{R}$ sein, das heißt unendliche Intervallgrenzen sind offen. Der Durchschnitt von zwei Intervallen $I_{1,2}$ mit den Grenzen $a_{1,2}$ und $b_{1,2}$ ist wieder ein Intervall, und zwar mit den 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. Der folgende Beweis liefert die größte Lösung x_0 der Gleichung, denn für $x > x_0$ ist $f(x) > y_0$ (siehe unten).

Beweis: Sei oBdA $f(a) \leq y_0 \leq f(b)$. Dann ist die Menge $M = \{x \in [a,b] : f(x) \leq y_0\}$ nichtleer, da $a \in M$. Wir behaupten $f(x_0) = y_0$ für $x_0 = \sup M$. Für $n \in \mathbb{N}$ ist $x_0 - 1/n$ keine obere Schranke von M, also gibt es $x_n \in M$ mit $x_0 - 1/n < x_n \leq x_0$. Da f stetig, folgt $f(x_0) = \lim_{n\to\infty} f(x_n) \leq y_0$. Andererseits ist x_0 obere Schranke von M, also gilt $f(x) > y_0$ für $x_0 < x \leq b$. Ist $x_0 < b$, so folgt $f(x_0) = \lim_{x \searrow x_0} f(x) \geq y_0$. Im Fall $x_0 = b$ gilt $f(x_0) \geq y_0$ sowieso nach Voraussetzung. □

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 $y \in f(I)$. □

Jetzt zu monotonen Funktionen. Nach Satz 4.3 haben monotone, beschränkte Folgen einen Grenzwert; hier eine analoge Aussage für monotone Funktionen.

Lemma 8.1 (einseitige Grenzwerte) *Ist $f : I \to \mathbb{R}$ monoton wachsend, so existieren in jedem Punkt $x_0 \in I$ 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. $\Box$

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 (4.6) 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: von (1): Nach Folgerung 8.1 ist $f(I)$ das 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 außerdem $f(a) = \lim_{x \searrow a} f(x)$, da f stetig in a, also folgt insgesamt auch $\inf_I f \geq \lim_{x \searrow a} f(x)$. Für die rechte Intervallgrenze argumentieren wir entsprechend.

von (2): Sei $y_0 \in f(I)$ mit $y_0 > \alpha$. Wir zeigen: der linksseitige Grenzwert $x_- := \lim_{y \nearrow y_0} g(y)$ ist gleich dem Funktionswert $x_0 := g(y_0)$. Da g streng monoton wachsend ist, gilt zunächst $a < x_- \leq x_0$. Mit der Stetigkeit von f in x_- folgt nun

$$f(x_-) = \lim_{y \nearrow y_0} f(g(y)) = \lim_{y \nearrow y_0} y = y_0 = f(x_0).$$

Aber f ist streng monoton, also ist $x_- = x_0$ wie behauptet. Analog sehen wir, dass g in jedem Punkt $y_0 \in f(I)$ mit $y_0 < \beta$ rechtsseitig stetig ist, und damit insgesamt stetig.

von (3): folgt aus den Aussagen (1) und (2), angewandt auf g statt f. $\Box$

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 4.6). 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 5.6 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, genauer konnte er diese Gesetze alle 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 Satz 7.5(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 7.3. □

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 7.2. 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 Für $f_n(x) = x^n$ folgt aus Beispiel 9.2, also $f_1' = 1$, und der Produktregel

$$f_n'(x) = (f_1 f_{n-1})'(x) = f_1'(x) f_{n-1}(x) + f_1(x) f_{n-1}'(x) = x^{n-1} + x\, f_{n-1}'(x),$$

und damit per Induktion $f_n'(x) = nx^{n-1}$. Allgemeiner ergibt sich mit Satz 9.2(1) für Polynome $p(x) = \sum_{k=0}^n a_k x^k$ die Formel

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

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 betrachten wieder für $x \neq x_0$ den Differenzenquotienten:

$$\frac{g(f(x)) - g(f(x_0))}{x - x_0} = \begin{cases} \dfrac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} \dfrac{f(x) - f(x_0)}{x - x_0} & \text{falls } f(x) \neq f(x_0) \\ 0 & \text{falls } f(x) = f(x_0). \end{cases}$$

Wir erkennen hier schon die Kettenregel, nur gibt es die technische Schwierigkeit mit der Fallunterscheidung. Für $f(x) \neq f(x_0)$ gilt

$$\frac{g(f(x)) - g(f(x_0))}{f(x) - f(x_0)} = a(f(x)) \quad \text{mit } a : J\backslash\{f(x_0)\} \to \mathbb{R},\ a(y) = \frac{g(y) - g(f(x_0))}{y - f(x_0)}.$$

Da g an der Stelle $f(x_0)$ differenzierbar ist, gilt $a(y) \to g'(f(x_0))$ mit $y \to f(x_0)$. Wir erhalten also eine stetige Funktion $a : J \to \mathbb{R}$, wenn wir $a(f(x_0)) = g'(f(x_0))$ setzen. Nun gilt aber

$$\frac{g(f(x)) - g(f(x_0))}{x - x_0} = a(f(x)) \frac{f(x) - f(x_0)}{x - x_0} \quad \text{für alle } x \in I.$$

Im Fall $f(x) \neq f(x_0)$ hatten wir das schon gesehen, im Fall $f(x) = f(x_0)$ sind einfach beide Seiten Null. Mit $x \to x_0$ folgt

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

□

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 mit Zusatz (8.1) 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 $y \to y_0$. Für $y \neq y_0$ folgt

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

□

Bemerkung. Die Bedingung $f'(x_0) \neq 0$ ist nicht nur hinreichend (zusammen mit den sonstigen Annahmen), sondern auch notwendig. Denn sind f, g Umkehrfunktionen, die an den Stellen x_0 bzw. $y_0 = f(x_0)$ differenzierbar sind, so folgt aus der Kettenregel

$$g(f(x)) = x \quad \Rightarrow \quad g'(f(x_0)) f'(x_0) = 1.$$

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

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

Sind f, g insgesamt differenzierbar, so ergibt die Kettenregel (mit vertauschten Rollen)

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

Damit kann man die Ableitung $g'(y)$ herleiten, falls man die Formel vergessen hat.

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 $f'(x) = nx^{n-1} > 0$ für $x > 0$, siehe Beispiel 9.4. Die Umkehrfunktion ist $g : [0, \infty) \to \mathbb{R}$, $g(y) = y^{1/n}$. Nach Satz 9.4 ist g für $y > 0$ differenzierbar mit

$$g'(y) = \frac{1}{f'(y^{\frac{1}{n}})} = \frac{1}{ny^{1-\frac{1}{n}}} = \frac{1}{n} y^{\frac{1}{n}-1}.$$

Für $h(y) = y^\alpha$ mit $\alpha = m/n \in \mathbb{Q}$ gilt $h(y) = g(y)^m$, also folgt weiter aus der Kettenregel, siehe auch Beispiel 9.5,

$$h'(y) = mg(y)^{m-1} g'(y) = m\, y^{\frac{m}{n} - \frac{1}{n}} \frac{1}{n} y^{\frac{1}{n}-1} = \alpha y^{\alpha - 1}.$$

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

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

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

$$\begin{aligned} f = o(g) \;\text{ für } x \to x_0 \quad &\Leftrightarrow \quad \lim_{x \to x_0} \frac{|f(x)|}{|g(x)|} = 0, \\ f = \mathcal{O}(g) \;\text{ für } x \to x_0 \quad &\Leftrightarrow \quad \limsup_{x \to x_0} \frac{|f(x)|}{|g(x)|} < \infty. \end{aligned}$$

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. Da sich keine Unterschiede ergeben, betrachten wir dabei gleich reellwertige Funktionen auf einer Teilmenge des $\mathbb{R}^n$.

Wir benötigen einen Satz, der die Existenz von Extremalstellen für eine stetige Funktion $f : D \to \mathbb{R}$ 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 den Definitionsbereich D, wie das folgende Beispiel zeigt:

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

Hier wird weder das Supremum $\sup_{x \in D} f(x) = 1$ noch das Infimum $\inf_{x \in D} f(x) = 0$ durch die Funktion angenommen.

Satz 10.1 (Existenz von Extremalstellen) *Sei $D \subset \mathbb{R}^n$ nichtleer, abgeschlossen und beschränkt, und sei $f : D \to \mathbb{R}$ stetig. Dann ist f beschränkt und nimmt ihr Infimum und Supremum an, das heißt es gibt $x_0, x_1 \in D$ mit*

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

Beweis: Setze $\alpha = \inf_{x \in D} f(x) \in [-\infty, \infty)$. Wir zeigen die Existenz eines $x_0 \in D$ mit $f(x_0) = \alpha$, daraus folgt insbesondere $\alpha > -\infty$, das heißt f ist nach unten beschränkt. Für das Supremum kann analog argumentiert werden.

Nach Definition des Infimums gibt es eine Folge $x_k \in D$ mit $f(x_k) \to \alpha$, eine sogenannte *Minimalfolge*. Da D beschränkt, ist die Folge x_k beschränkt. Nach Bolzano-Weierstraß, siehe Satz 4.7 bzw. Folgerung 5.3, gibt es eine Teilfolge x_{k_j} und ein $x_0 \in \mathbb{R}^n$ mit $x_{k_j} \to x_0$ für $j \to \infty$. Aus D abgeschlossen, siehe Definition 5.11 folgt $x_0 \in D$. Aber dann gilt wegen der Stetigkeit von f

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

□

Im Beweis spielte die folgende Eigenschaft von D eine Rolle:

Definition 10.1 (Folgenkompaktheit) *Eine Menge $D \subset \mathbb{R}^n$ heißt folgenkompakt, wenn es zu jeder Folge $x_k \in D$ eine Teilfolge x_{k_j} gibt, die konvergiert mit Grenzwert in D:*

$$\lim_{j \to \infty} x_{k_j} = x_0 \in D.$$

Wir halten aus dem Beweis von Satz 10.1 folgende Beobachtung fest:

Satz 10.2 *Für $D \subset \mathbb{R}^n$ sind äquivalent:*

(1) *D ist folgenkompakt.*

(2) *D ist abgeschlossen und beschränkt.*

Beweis: Die Implikation (2) $\Rightarrow$ (1) wurde im Beweis von Satz 10.1 gezeigt. Jetzt setzen wir (1) voraus. Um die Abgeschlossenheit von D zu zeigen, sei $x_k \in D$ eine beliebige Folge mit $x_k \to x_0 \in \mathbb{R}^n$. Da D folgenkompakt, gibt es eine Teilfolge x_{k_j}, die gegen ein $x_0' \in D$ konvergiert. Aus der Eindeutigkeit des Grenzwerts folgt $x_0 = x_0' \in D$, also ist D abgeschlossen. Wäre D nicht beschränkt, so gibt es zu $k \in \mathbb{N}$ ein $x_k \in D$ mit $|x_k| \geq k$. Da D kompakt, existiert eine konvergente Teilfolge x_{k_j}, und diese ist nach Satz 3.2 beschränkt, ein Widerspruch. $\square$

Die Einführung des Begriffs *folgenkompakt* erscheint hier unnötig, man könnte es auch bei *abgeschlossen* und *beschränkt* lassen. Es gibt Optimierungsprobleme, bei denen statt eines Punkts im $\mathbb{R}^n$ eine ganze Funktion gesucht ist, zum Beispiel eine kürzeste Verbindungskurve auf einer Fläche. In Funktionenräumen sind abgeschlossene und beschränkte Mengen im allgemeinen aber nicht kompakt, daher der zusätzliche Begriff.

Ab jetzt geht es um Funktionen $f(x)$ einer reellen Variablen.

Definition 10.2 (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.

Satz 10.3 (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. Dann gibt es ein $\delta > 0$ mit $f(x) \geq f(x_0)$ für $x \in B_\delta(x_0)$, also gilt

$$\frac{f(x) - f(x_0)}{x - x_0} \quad \begin{cases} \geq 0 & \text{für } x \in (x_0, x_0 + \delta), \\ \leq 0 & \text{für } x \in (x_0 - \delta, x_0). \end{cases}$$

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

Der Beweis zeigt tatsächlich eine Aussage über *einseitige* lokale Minima, und zwar gilt

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

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

Satz 10.4 (Mittelwertsatz der Differentialrechnung) *Sei $f : [a,b] \to \mathbb{R}$ 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). Wir brauchen 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 10.3 und wir können $\xi = \xi_1$ wählen. Analog, wenn $\xi_2 \in (a,b)$. Im verbleibenden Fall $\xi_1, \xi_2 \in \{a,b\}$ folgt $\inf f = \sup f = 0$ bzw. $f(x) = 0$ für alle $x \in [a,b]$, und damit auch $f'(x) = 0$ für alle x. Seien nun $f(a), f(b)$ 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 $h(a) = h(b) = 0$. Also existiert ein $\xi \in (a,b)$ mit

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

□

Folgerung 10.1 (Monotoniekriterium) *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:

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

□

Folgerung 10.2 (Schrankensatz) *Ist $f : [a,b] \to \mathbb{R}$ stetig und differenzierbar auf (a,b), so gilt für $a \leq x_1 < x_2 \leq b$:*

$$\begin{aligned}
f'(x) \geq m \textit{ für alle } x \in (a,b) \quad &\Rightarrow \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \geq m \\
f'(x) \leq M \textit{ für alle } x \in (a,b) \quad &\Rightarrow \quad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq M.
\end{aligned}$$

Beweis: Wir zeigen die erste Aussage. Mit $g(x) = mx$ gilt $(f-g)' = f' - g' \geq m - m = 0$. Nach Folgerung 10.1 ist $f-g$ wachsend, ds heißt $f(x_2) - m\,x_2 \geq f(x_1) - m\,x_1$. Die Behauptung folgt. □

Der Mittelwertsatz gilt nicht für vektorwertige Funktionen $f : [a,b] \to \mathbb{R}^n$, $n \geq 2$. Ein ad hoc Beispiel ist $f(x) = (x^2, x^3)$ auf $[a,b] = [0,1]$. Angenommen es gibt ein $\xi \in (0,1)$ mit

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

Dann folgt $2\xi = 1$ und $3\xi^2 = 1$, ein Widerspruch. Das Problem ist, dass für die einzelnen Koordinaten im allgemeinen nicht dieselbe Stelle ξ gewählt werden kann. Dennoch können wir auch im vektorwertigen Fall eine Version des Schrankensatzes beweisen.

Folgerung 10.3 (f' beschränkt $\Rightarrow$ f Lipschitz) *Sei $f : [a,b] \to \mathbb{R}^n$ stetig und auf (a,b) differenzierbar. Falls $\|f'(x)\| \leq L$ für alle $x \in (a,b)$, so folgt*

$$\|f(x_2) - f(x_1)\| \leq L|x_2 - x_1| \quad \text{für alle } x_1, x_2 \in [a,b].$$

Beweis: Um die Aussage auf den reellwertigen Fall zu reduzieren, betrachten wir für einen beliebigen Vektor $v \in \mathbb{R}^n$ mit $\|v\| = 1$ die Funktion

$$\varphi : [a,b] \to \mathbb{R},\ \varphi(x) = \langle v, f(x) \rangle = \sum_{i=1}^{n} v_i f_i(x).$$

Aus der Ungleichung von Cauchy-Schwarz, siehe Satz 5.5, folgt

$$\varphi'(x) = \langle v, f'(x) \rangle \leq \|v\|\,\|f'(x)\| \leq L.$$

Für $x_1, x_2 \in [a,b]$ mit $x_1 < x_2$ erhalten wir aus Folgerung 10.2

$$\langle v, f(x_2) - f(x_1) \rangle = \varphi(x_2) - \varphi(x_1) \leq L(x_2 - x_1) = L|x_2 - x_1|.$$

Wir können $f(x_2) \neq f(x_1)$ annehmen, und wählen $v = (f(x_2) - f(x_1))/\|f(x_2) - f(x_1)\|$. Das ist ein Einheitsvektor, also folgt

$$\|f(x_2) - f(x_1)\| = \Big\langle \frac{f(x_2) - f(x_1)}{\|f(x_2) - f(x_1)\|}, f(x_2) - f(x_1) \Big\rangle = \langle v, f(x_2) - f(x_1) \rangle \leq L|x_1 - x_2|.$$

□

Wir kommen jetzt zu höheren Ableitungen.

Definition 10.3 *Die k-te Ableitung von $f : (a,b) \to \mathbb{R}$ in x_0 ist induktiv definiert durch*

$$f^{(k)}(x_0) = \big(f^{(k-1)}\big)'(x_0).$$

Damit $f^{(k)}(x_0)$ definiert ist, müssen also die Ableitungen bis Ordnung $k-1$ in einer Umgebung von x_0 definiert sein, und $f^{(k-1)}$ muss in x_0 differenzierbar sein.

Natürlich schreiben wir f' und f'' statt $f^{(1)}$ bzw. $f^{(2)}$. Mit der zweiten Ableitung gewinnen wir genauere Informationen über lokale Extrema.

Satz 10.5 *Sei $f : (a,b) \to \mathbb{R}$ differenzierbar. In $x_0 \in (a,b)$ gelte $f'(x_0) = 0$, und $f''(x_0)$ sei definiert. Dann gilt:*

(1) *Ist $f''(x_0) > 0$, so hat f in x_0 ein isoliertes, lokales Minimum.*

(2) *Hat f in x_0 ein lokales Minimum, so folgt $f''(x_0) \geq 0$.*

Analoge Aussagen gelten mit umgekehrten Ungleichungen für Maxima.

BEWEIS: Da $f'(x_0) = 0$ nach Voraussetzung, gilt

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

Unter der Voraussetzung (1) gibt es ein $\delta > 0$ mit $f'(x) < 0$ auf $(x_0 - \delta, x_0)$ und $f'(x) > 0$ auf $(x_0, x_0 + \delta)$. Nach Folgerung 10.1 ist f dann streng monoton fallend auf $(x_0 - \delta, x_0)$ und und streng monoton wachsend auf $(x_0, x_0 + \delta)$, hat also in x_0 ein isoliertes, lokales Minimum. Wäre $f''(x_0) < 0$ in (2), so hätte f in x_0 ein isoliertes lokales Maximum im Widerspruch zur Voraussetzung. □

Die Funktion $f(x) = x^4$ zeigt, dass in einem isolierten Minimum $f''(x_0) = 0$ gelten kann. Häufig ist man nicht wirklich an den lokalen Minima, sondern am globalen Minimum interessiert. Dafür ist der Begriff der Konvexität relevant.

Definition 10.4 *Sei $I \subset \mathbb{R}$ ein Intervall. $f : I \to \mathbb{R}$ heißt konvex, falls gilt:*

$$f\big((1-t)x_0 + tx_1\big) \leq (1-t)f(x_0) + tf(x_1) \quad \forall x_0, x_1 \in I,\ t \in [0,1]. \tag{10.1}$$

Gilt dies mit $\geq$ statt mit $\leq$, so heißt f konkav.

Die Sekante durch $\big(x_0, f(x_0)\big)$ und $\big(x_1, f(x_1)\big)$ hat die Geradengleichung

$$y = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) =: g(x).$$

An der Stelle $x(t) = (1-t)x_0 + tx_1$ gilt

$$g\big(x(t)\big) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}\, t(x_1 - x_0) = (1-t)f(x_0) + tf(x_1).$$

Die Konvexität bedeutet also, dass der Graph von f stets unterhalb der Sekanten liegt,

Satz 10.6 (Konvexitätskriterien) *Sei $f : (a,b) \to \mathbb{R}$ differenzierbar. Dann sind folgende Aussagen äquivalent:*

(1) *f' ist monoton wachsend auf (a,b).*

(2) *f ist konvex.*

(3) *$f(x_1) \geq f(x_0) + f'(x_0)(x_1 - x_0) \quad$ für alle $x_0, x_1 \in (a,b)$.*

Ist f zweimal differenzierbar auf (a,b), so ist außerdem äquivalent:

(4) *$f'' \geq 0$.*

Beweis: Wir zeigen (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (1). Sei (1) erfüllt. Angenommen f ist nicht konvex, das heißt es gibt $x_0, x_1 \in I$ und ein $t \in [0,1]$ mit

$$g(t) = f\big((1-t)x_0 + tx_1\big) - \big((1-t)f(x_0) + tf(x_1)\big) > 0.$$

Wir können $x_1 > x_0$ annehmen, sonst vertausche x_0, x_1 und ersetze t durch $1-t$. Die Funktion $g : [0,1] \to \mathbb{R}$ ist stetig mit $g(0) = g(1) = 0$. Nach Satz 10.1 nimmt g ihr strikt positives Maximum in einem $t_0 \in (a,b)$ an, und es gilt $g'(t_0) = 0$ nach Satz 10.3. Aber wegen f' wachsend ist auch g' wachsend:

$$g'(t) = f'\big(\underbrace{x_0 + t(x_1 - x_0)}_{\text{wachsend}}\big)\underbrace{(x_1 - x_0)}_{>0} - \big(\underbrace{f(x_1) - f(x_0)}_{\text{konstant}}\big).$$

Insbesondere folgt $g'(t) \geq 0$ für alle $t \geq t_0$, und daraus $g(1) \geq g(t_0) > 0$, Widerspruch.

Es gelte jetzt (2), also gilt für alle $t \in (0,1)$ und $x_0, x_1 \in (a,b)$ mit $x_0 \neq x_1$

$$(x_1 - x_0)\,\frac{f\big(x_0 + t(x_1 - x_0)\big) - f(x_0)}{t(x_1 - x_0)} \leq f(x_1) - f(x_0).$$

Mit $t \searrow 0$ folgt $(x_1 - x_0)f'(x_0) \leq f(x_1) - f(x_0)$, das ist Ungleichung (3). Schließlich folgt aus (3), indem wir dort x_0 und x_1 vertauschen und addieren,

$$\big(f'(x_1) - f'(x_0)\big)(x_1 - x_0) \geq 0,$$

womit wiederum (1) gezeigt ist. Schließlich: ist f zweimal differenzierbar, so impliziert $f'' \geq 0$ Bedingung (1) nach Folgerung 10.1, und umgekehrt folgt $f'' \geq 0$ aus (1) einfach durch Betrachtung des Differenzenquotienten. $\square$

Beispiel 10.1 (Youngsche Ungleichung) Für $x, y \geq 0$ gilt die Ungleichung

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \qquad \text{falls } p, q \in (1,\infty) \text{ mit } \frac{1}{p} + \frac{1}{q} = 1.$$

Dazu betrachten wir für festes $y > 0$ die Funktion

$$f : (0,\infty) \to \mathbb{R},\ f(x) = xy - \frac{x^p}{p}.$$

Es gilt $f'(x) = y - x^{p-1}$ und $f''(x) = -(p-1)x^{p-2} \leq 0$, das heißt f ist konkav nach Satz 10.6. Aber $f'(x_0) = 0$ für $x_0 = y^{1/(p-1)}$, also folgt mit Satz 10.6(3) wegen $p/(p-1) = q$

$$xy - \frac{x^p}{p} = f(x) \leq f(x_0) + f'(x_0)(x - x_0) = y^{1/(p-1)}y - \frac{y^{p/(p-1)}}{p} = \frac{y^q}{q}.$$

Definition 10.5 *Sei $I \subset \mathbb{R}$ ein offenes Intervall und $k \in \mathbb{N}_0$. Wir bezeichnen mit $C^k(I)$ den $\mathbb{R}$-Vektorraum der k mal stetig differenzierbaren Funktionen $f : I \to \mathbb{R}$, das heißt*

$$C^k(I) = \{f : I \to \mathbb{R} : f^{(i)} : I \to \mathbb{R} \text{ sind definiert und stetig für } i = 0, 1, \ldots, k\}.$$

Weiter definieren wir $C^\infty(I)$ als den $\mathbb{R}$-Vektorraum der unendlich oft differenzierbaren Funktionen, also ist

$$C^\infty(I) = \bigcap_{k \geq 0} C^k(I).$$

Der Umgang mit C^∞-Funktionen ist besonders angenehm, weil die Klasse im Gegensatz zu den Räumen $C^k(I)$ unter der Bildung von Ableitungen abgeschlossen ist. Es ist klar, dass Polynome unendlich oft differenzierbar sind; es gibt aber sehr viel mehr Funktionen in $C^\infty(I)$. Wir werden weitere im nächsten Abschnitt mithilfe der Exponentialfunktion konstruieren.

11 Die reelle Exponentialfunktion

Wachtstums- oder Zerfallsprozesse spielen in vielen Bereichen der Natur und der Anwendungen eine Rolle. Dabei wird die zukünftige (oder vergangene) Entwicklung einer Größe $K(t)$, $t \in I$, durch ein Wachstumsgesetz und einen Startwert $K(t_0) = K_0$ charakterisiert. Ein diskretes Modell ist der Zinseszins, bei dem das Intervall in Zeitschritte der Länge Δt unterteilt wird. In jedem Zeitschritt $[t, t + \Delta t]$ wird linear mit der Rate $\lambda \in \mathbb{R}$ verzinst. Das bedeutet

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

Um ein kontinuierliches Modell herzuleiten, lassen wir hier Δt gegen Null gehen. Formal erhalten wir das Wachstumsgesetz, und die Anfangsbedingung,

$$K'(t) = \lambda K(t) \text{ für } t \in I, \quad K(t_0) = K_0. \tag{11.1}$$

Ob dieses Modell sinnvoll ist, muss mathematisch untersucht werden. Zum einen ist die Existenz einer Lösungsfunktion $K(t)$ zu klären, sonst wäre alles Quatsch. Zweitens sollte die Lösung eindeutig sein, damit es Aussagen über die zukünftige oder vergangene Entwicklung erlaubt. Weitere Eigenschaften der Lösung wie Monotonie oder asymptotisches Verhalten für große Zeiten sind natürlich auch von Interesse. Die Aufgabe, zu einem Wachstumsgesetz und Startwert die zeitliche Entwicklung zu bestimmen, bezeichnet man allgemein als *Anfangswertproblem*. Wir kennen schon ein triviales aber fundamentales Beispiel: für die Gleichung $K' = 0$ ist die konstante Funktion $K(t) = K_0$ die eindeutige Lösung nach Folgerung 10.1. Enthält ein Wachstumsgesetz die Ableitung der Funktion, so spricht man von einer Differentialgleichung. Speziell ist $K' = \lambda K$ die Differentialgleichung des natürlichen Wachstums mit Rate $\lambda \in \mathbb{R}$. Wir studieren nun das Anfangswertproblem

$$f' = f \quad \text{auf } \mathbb{R}, \quad f(0) = 1. \tag{11.2}$$

Das ist der Spezialfall von (11.1) mit $I = \mathbb{R}$, $\lambda = 1$, $t_0 = 0$ und $K_0 = 1$. Wir werden die Lösung des allgemeinen Problems im Anschluss durch Skalierung erhalten.

Lemma 11.1 *Eine Lösung $f : \mathbb{R} \to \mathbb{R}$ von (11.2) hat folgende Eigenschaften:*

(a) $f \in C^0(\mathbb{R})$,

(b) $f(x)f(-x) = 1$ *für alle* $x \in \mathbb{R}$,

(c) f *ist strikt positiv.*

Beweis: Nach Voraussetzung ist f differenzierbar, also stetig mit Satz 9.1. Weiter gilt

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

Also ist $f(x)f(-x) = f(0)^2 = 1$, insbesondere $f(x) \neq 0$, für alle $x \in \mathbb{R}$. Nun ist $f(0) = 1$. Wäre $f(x) < 0$ für ein $x \in \mathbb{R}$, so hätte f eine Nullstelle nach dem Zwischenwertsatz, Widerspruch. $\square$

Satz 11.1 (Eindeutigkeit) *Es gibt höchstens ein $f : \mathbb{R} \to \mathbb{R}$ mit $f' = f$ auf $\mathbb{R}$ und $f(0) = 1$.*

Beweis: Seien f, g zwei Lösungen von (11.2). Dann ist $g > 0$ nach Lemma 11.1,und es folgt

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

Also ist f/g konstant, mit $f(0)/g(0) = 1$ also $f = g$. □

Zur Existenz machen wir mit Cauchy den Reihenansatz $f(x) = \sum_{k=0}^{\infty} a_k x^k$, und berechnen

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

Natürlich ist die Vertauschung der Grenzprozesse hier nicht begründet, und überhaupt ist unklar ob es eine Lösung von (11.2) gibt, die so als Reihe geschrieben werden kann. Aber wir wischen diese Einwände beiseite, stattdessen leiten wir eine potentielle Formel für die Lösung her: Koeffizientenvergleich in der Gleichung $f' = f$ liefert $a_{k+1} = a_k/(k+1)$. Mit $a_0 = f(0) = 1$ folgt per Induktion $a_k = 1/k!$, und wir kommen zu dem *educated guess*

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

Satz 11.2 (Existenz) *$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ konvergiert für alle $x \in \mathbb{R}$ und ist Lösung des Anfangswertproblems (11.2), das heißt es gilt $f' = f$ auf $\mathbb{R}$, $f(0) = 1$.*

Beweis: Offensichtlich ist $f(0) = 1$. Für $x \neq 0$ gilt

$$\frac{|x|^{k+1}/(k+1)!}{|x|^k/k!} = \frac{|x|}{k+1} \to 0 \quad \text{mit } k \to \infty.$$

Die Reihe konvergiert absolut nach dem Quotientenkriterium, also existiert der Grenzwert

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

Wir schätzen die f_n unabhängig von n ab. Für $|x| \leq R$ und alle $n \in \mathbb{N}_0$ gilt

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

Aber $f_n' = f_{n-1}$ für $n \geq 1$. Es folgt $|f_n'(x)| \leq C$ für $|x| \leq R$, also mit dem Schrankensatz

$$|f_n(x_1) - f_n(x_2)| \leq C\,|x_1 - x_2| \quad \text{für alle } n \in \mathbb{N}_0,\ |x_1|, |x_2| \leq R.$$

Weiter ist dann $|f_n'(x_1) - f_n'(x_2)| = |f_{n-1}(x_1) - f_{n-1}(x_2)| \leq C|x_1 - x_2|$. Für $|x_0|, |x| \leq R$ folgt aus dem Mittelwertsatz, für ein ξ zwischen x_0 und x,

$$\left|\frac{f_n(x) - f_n(x_0)}{x - x_0} - f_n'(x_0)\right| = |f_n'(\xi) - f_n'(x_0)| \leq C|\xi - x_0| \leq C|x - x_0|.$$

Jetzt lassen wir $n \to \infty$ gehen. Da $f_n'(x_0) = f_{n-1}(x_0) \to f(x_0)$, erhalten wir

$$\left|\frac{f(x) - f(x_0)}{x - x_0} - f(x_0)\right| \leq C|x - x_0| \quad \text{für } |x_0|, |x| \leq R.$$

Es folgt $f'(x_0) = f(x_0)$ für $x_0 \in (-R, R)$. Da $R > 0$ beliebig ist, folgt die Behauptung. □

Definition 11.1 (Exponentialfunktion) *Die Funktion*

$$\exp : \mathbb{R} \to \mathbb{R},\ \exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{11.3}$$

heißt Exponentialfunktion. Sie ist die eindeutige Lösung des Anfangswertproblems (11.2).

Wie angekündigt erhalten wir die Lösung von (11.1) nun durch Skalierung.

Folgerung 11.1 *Die eindeutige Lösung von (11.1) ist*

$$K(t) = K_0 \exp\big(\lambda(t-t_0)\big) \quad \text{für } t \in I.$$

Beweis: Dass die Formel eine Lösung liefert ist klar. Für eine beliebige Lösung $K(t)$ berechnen wir, vgl. den Beweis von Lemma 11.1,

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

Einsetzen von $t = t_0$ ergibt $\exp(-\lambda(t-t_0))K(t) = \exp(0)K(t_0) = K_0$. Multiplikation mit $\exp(\lambda(t-t_0))$ liefert nun $K(t) = K_0 \exp(\lambda(t-t_0))$, wobei Lemma 11.1 (b) benutzt wurde. $\square$

Jetzt interessieren wir uns für weitere Eigenschaften der Exponentialfunktion. Dazu könnten wir die Reihendarstellung benutzen, es ist aber allemal einfacher mit der Differentialgleichung zu argumentieren.

Satz 11.3 (Eigenschaften der Exponentialfunktion) *Es gelten folgende Aussagen:*

(a) $\exp(1) = e = 2,71828\ldots$.

(b) $\exp$ *ist in* $C^\infty(\mathbb{R})$, *strikt positiv und streng monoton wachsend mit Grenzwerten*

$$\lim_{x \nearrow \infty} \exp(x) = +\infty \quad \text{und} \quad \lim_{x \searrow -\infty} \exp(x) = 0.$$

(c) *Es gilt die Funktionalgleichung*

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

insbesondere $\exp(x)\exp(-x) = 1$.

Beweis: (a) Nach den Definitionen 11.3 und 4.3 gilt $\exp(1) = \sum_{k=0}^{\infty} \frac{1}{k!} = e$.

(b) In Lemma 11.1 (a) wurde schon $f \in C^0(\mathbb{R})$ festgestellt. Aus $f' = f$ folgt dann leicht mit Induktion $f \in C^k(\mathbb{R})$ und $f^{(k)} = f$ für alle $k \in \mathbb{N}_0$. Weiter wissen wir $\exp > 0$ nach Lemma 11.1 (c). Damit ist auch $\exp' = \exp > 0$, also exp streng monoton wachsend. Für $x \geq 0$ ist somit $\exp'(x) = \exp(x) \geq \exp(0) = 1$, also nach Schrankensatz

$$\exp(x) \geq 1 + x \to +\infty \quad \text{für } x \nearrow \infty. \tag{11.4}$$

Schließlich gilt $\exp(x)\exp(-x) = 1$ nach Lemma 11.1, also $\exp(x) \to 0$ mit $x \to -\infty$.

(c) Für $g(x) = \exp(x+y)$, mit $y \in \mathbb{R}$ fest, gilt $g' = g$ und $g(0) = \exp(y)$, also mit Folgerung 11.1

$$\exp(x+y) = g(x) = g(0)\exp(x) = \exp(x)\exp(y).$$

$\square$

In der Sprache der Algebra besagt die Funktionalgleichung, dass $\exp : (\mathbb{R},+) \to (\mathbb{R}^+,\cdot)$ ein Gruppenhomomorphismus ist, wobei $\mathbb{R}^+ = \{r \in \mathbb{R} : r > 0\}$. Genauer ist exp ein Isomorphismus, wie wir gleich feststellen, der inverse Gruppenisomorphismus ist der Logarithmus.

Satz 11.4 (Definition des Logarithmus) $\exp : (-\infty,\infty) \to (0,\infty)$ *ist streng monoton wachsend und bijektiv. Die Umkehrfunktion* $\log : (0,\infty) \to (-\infty,\infty)$ *heißt (natürlicher) Logarithmus. Sie ist streng monoton wachsend und bijektiv, und hat die Grenzwerte*

$$\lim_{r\searrow 0} \log(r) = -\infty \quad \textit{und} \quad \lim_{r\to\infty} \log(r) = \infty. \tag{11.5}$$

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

$$\log'(r) = \frac{1}{r} \quad \textit{für alle } r \in (0,\infty). \tag{11.6}$$

Schließlich gilt die Funktionalgleichung

$$\log(rs) = \log(r) + \log(s) \quad \textit{für alle } r,s > 0. \tag{11.7}$$

Beweis: Nach Satz 8.2 zur Umkehrfunktion ist exp bijektiv, die Umkehrfunktion log ist ebenfalls streng monoton wachsend. Die Grenzwerte (11.5) folgen aus der dortigen Aussage (8.1). Die Funktionswerte ergebn sich direkt aus $\exp(0) = 1$ bzw. $\exp(1) = e$. Da $\exp'(x) = \exp(x) > 0$, ist die Umkehrfunktion differenzierbar und hat die Ableitung, siehe Satz 9.4,

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

Die Funktionalgleichung (11.7) ergibt sich schließlich aus $\exp(x)\exp(y) = \exp(x+y)$, indem wir $x = \log r$, $y = \log(s)$ einsetzen und dann den Logarithmus nehmen. $\square$

Für $a > 0$ hatten wir die Potenz a^x bisher nur im Fall $x \in \mathbb{Q}$ definiert. Diese Einschränkung können wir mit dem Logarithmus nun abschaffen.

Definition 11.2 (Potenz mit reellen Exponenten) *Für* $a > 0$, $x \in \mathbb{R}$ *setzen wir*

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

Mit den Funktionalgleichungen von log und exp ergeben sich die Regeln

$$a^x\,a^y = a^{x+y} \qquad (a^x)^y = a^{xy}$$
$$\left(\frac{1}{a}\right)^x = a^{-x} \qquad a^x\,b^x = (ab)^x.$$

Induktion liefert $a^{nx} = (a^x)^n$. Mit $x = \frac{1}{n}$ folgt $(a^{\frac{1}{n}})^n = a^{n\frac{1}{n}} = a$, das heißt $a^{\frac{1}{n}} = \sqrt[n]{a}$. Damit stimmt für $x \in \mathbb{Q}$ die neue mit der bisherigen Definition der Potenz überein. Weiter gelten für $x \in \mathbb{R}$ die Ableitungsregeln

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

Satz 11.5 (Wachstum von exp **und** log**)** *Es gelten folgende Aussagen:*

$$\lim_{x\to\infty} x^{-s}e^{x} = \infty \quad \textit{und} \quad \lim_{x\to\infty} x^{s}e^{-x} = 0 \quad \textit{für jedes } s > 0, \tag{11.8}$$
$$\lim_{y\to\infty} y^{-s}\log y = 0 \quad \textit{und} \quad \lim_{y\searrow 0} y^{s}(-\log y) = 0 \quad \textit{für jedes } s > 0. \tag{11.9}$$

Beweis: Zu $s > 0$ wähle $n \in \mathbb{N}$ mit $n > s$. Für $x > 0$ folgt dann

$$x^{-s}e^{x} \geq \frac{1}{n!}x^{n-s} = \frac{1}{n!}e^{(n-s)\log x} \to \infty \quad \text{mit } x \to \infty.$$

Die zweite Aussage in (11.8) folgt durch Übergang zum Kehrwert. Für die zweite Aussage in (11.9) substituieren wir $y = e^{-x/s}$. Mit $y \searrow 0$ geht $x = -s\log y \to \infty$, also folgt aus (11.8)

$$y^{s}(-\log y) = e^{-x}\,\frac{x}{s}\Big|_{x=-s\log y} \to 0 \quad \text{mit } y \searrow 0.$$

Setzen wir $y = 1/z$ mit $z \to \infty$, so folgt auch der linke Grenzwert in (11.9). □

Folgerung 11.2 *Die Funktion*

$$f(x) = \begin{cases} e^{-1/x} & \textit{für } x > 0 \\ 0 & \textit{für } x \leq 0 \end{cases}$$

ist unendlich oft differenzierbar.

Beweis: Wir zeigen durch Induktion, dass es Polynome p_n gibt mit

$$f^{(n)}(x) = \begin{cases} p_n(\frac{1}{x})\, e^{-1/x} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0. \end{cases}$$

Für $n = 0$ ist das richtig mit $p_0(s) \equiv 1$. Ist die Aussage für ein $n \in \mathbb{N}_0$ gezeigt, so folgt:

$$f^{(n+1)}(x) = \begin{cases} \left(\frac{1}{x^2}\, p_n(\frac{1}{x}) - \frac{1}{x^2}\, p_n'(\frac{1}{x})\right)\, e^{-1/x} & \text{für } x > 0, \\ 0 & \text{für } x < 0. \end{cases}$$

Für $x \neq 0$ gilt der Induktionsschluss also mit $p_{n+1}(s) = s^2\,(p_n(s) - p_n'(s))$. Zu zeigen bleibt $f^{(n+1)}(0) = 0$. Für den linksseitigen Differenzenquotienten ist das klar, für $x > 0$ berechnen wir mit (11.8)

$$\frac{f^{(n)}(x) - f^{(n)}(0)}{x} = \frac{1}{x}\,p_n\Big(\frac{1}{x}\Big)\,e^{-1/x} \to 0 \quad \text{mit } x \searrow 0.$$

□

Wir haben die Differentialgleichung $f' = f$ als Grenzwert des diskreten Zinsmodells motiviert, wenn der Zeitschritt gegen Null geht. Die diskreten Lösungen sollten dann die kontinuierliche Lösung approximieren. Dies sollte zu einer alternativen Formel für die Exponentialfunktion führen. Unterteilen wir das Zeitintervall $[0,t]$ in n Abschnitte der Länge $\Delta t = t/n$, so ergibt sich bei Zinsrate Eins und Anfangswert $f_n(0) = 1$ mittels Induktion

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

Satz 11.6 (Eulerapproximation) *Für alle $x \in \mathbb{R}$ gilt*

$$\exp(x) = \lim_{n\to\infty} \Big(1 + \frac{x}{n}\Big)^n.$$

Beweis: Wir zeigen $n \log\big(1 + \frac{x}{n}\big) \to x$, die Behauptung folgt dann mit der Stetigkeit von exp im Punkt x. Wir berechnen

$$\Big|\frac{d}{dx}\Big(x - n\log\big(1 + \frac{x}{n}\big)\Big)\Big| = \Big|1 - \frac{1}{1+\frac{x}{n}}\Big| = \frac{1}{|1+\frac{x}{n}|}\,\frac{|x|}{n} \le \frac{2|x|}{n} \qquad \text{für } n \ge 2|x|.$$

Diese Abschätzung gilt für alle $x' \in \mathbb{R}$, $|x'| \le |x|$, also liefert der Schrankensatz

$$\Big|x - n\log\big(1 + \frac{x}{n}\big)\Big| \le \frac{2|x|^2}{n} \to 0 \qquad \text{mit } n \to \infty.$$

$\square$

Wir können die Funktionalgleichung anschaulich aus dem Zins-Modell verstehen: bei konstantem Zinssatz sollte es egal sein, ob das Geld erst für einen Zeitraum t_1 angelegt wird, und dann die ersparte Summe wieder für einen Zeitraum t_2, oder eben gleich für die Gesamtdauer $t_1 + t_2$. Bei Zinssatz $x = 1$ bedeutet das genau

$$\exp(t_1 + t_2) = \exp(t_1)\exp(t_2).$$

12 Die trigonometrischen Funktionen

Bei einer harmonischen Schwingung wird die Auslenkung eines schwingungsfähigen Systems aus seiner Ruhelage durch eine Funktion $u : I \to \mathbb{R}$ beschrieben. Ein klassisches Beispiel ist die Schwingung einer Masse $m > 0$ an einer Feder. Nach dem Hookeschen Gesetz (1676) ist die elastische Rückstellkraft der Feder proportional zur Auslenkung, also $F = -ku$ mit $k > 0$. Die Kraft bewirkt eine Beschleunigung der Masse, nach dem Newtonschen Gesetz (1687) ist $F = mu''$. Geben wir noch die Auslenkung und Geschwindigkeit zu einer Zeit $t_0 \in I$ vor und setzen $\omega = \sqrt{k/m}$, so erhalten wir das Anfangswertproblem

$$u''(t) + \omega^2 u(t) = 0 \text{ für } t \in I, \quad u(t_0) = x_0,\, u'(t_0) = y_0. \tag{12.1}$$

Bei einer Schwingung findet ein ständiger Austausch zwischen kinetischer und potentieller Energie statt. Bis auf Konstanten ist die kinetische Energie das Quadrat der Geschwindigkeit. Die potentielle Energie ist das Quadrat der Auslenkung, das folgt aus dem Hookeschen Gesetz. Die Gesamtenergie bleibt konstant. Wir wollen das mathematisch begründen, wobei wir $\omega = 1$ annehmen.

Lemma 12.1 *Für Funktionen $u, v : I \to \mathbb{R}$ sind folgende Aussagen* (1) *und* (2) *äquivalent:*

(1) *u ist Lösung von $u'' + u = 0$, und $v = -u'$.*

(2) *$c = u + iv : I \to \mathbb{C}$ ist Lösung von $c' = ic$, wobei $i = \sqrt{-1}$.*

Weiter: Für jede Lösung von (2) *ist $|c(t)|^2 = u(t)^2 + v(t)^2$ konstant.*

Beweis: Die Gleichung (2), also $c' = ic$, lautet in Koordinaten $c = u + iv$

$$u' + iv' = i(u + iv) \quad \Leftrightarrow \quad (u', v') = (-v, u). \tag{12.2}$$

Sind u, v Lösungen von (1), so folgt aus der hinteren Gleichung $u' = -v$ und dann aus der ersten $v' = -u'' = u$, also gilt (2). Sei andererseits (2) erfüllt, also $u' = -v$ und $v' = u$. Dann folgt $u'' = -u$, also gilt (1). Schließlich berechnen wir mit (2)

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

□

Es gilt also $|c(t)|^2 = u(t)^2 + v(t)^2 = R^2$ für einen festen Radius $R \geq 0$. Physikalisch ist das der Energieerhaltungssatz, denn u ist die Auslenkung und v ist (minus) die Geschwindigkeit. Wir bezeichnen $c' = ic$ als Differentialgleichung der Kreisbewegung. Der Geschwindigkeitsvektor $c' = (-v, u)$ steht senkrecht auf $c = (u, v)$ und hat die Länge $|c'| = R$. Dadurch ist c' bis auf Richtung eindeutig festgelegt, diese ergibt sich aus

$$\det(c, c') = \det \begin{pmatrix} u & u' \\ v & v' \end{pmatrix} = uv' - u'v = u^2 + v^2 > 0 \quad (\text{außer wenn } R = 0).$$

Anschaulich liegt die Kreisscheibe bei dieser Fahrtrichtung auf der linken Seite. Man sagt: der Kreis wird im mathematisch positiven Sinn durchlaufen, also gegen den Uhrzeigersinn. Wir zeigen jetzt, dass die Kreisbewegung durch die Anfangsposition eindeutig festgelegt ist, und folgern die Eindeutigkeit für das Anfagswertproblem der harmonischen Schwingung.

Satz 12.1 (Eindeutigkeit für Kreisbewegung/Schwingung) *Sei $t_0 \in I = (a,b)$.*

(1) *Zu $z_0 \in \mathbb{R}^2$ gibt es höchstens ein $c : I \to \mathbb{C}$ mit*

$$c' = ic \quad \text{auf } I, \quad c(t_0) = z_0. \tag{12.3}$$

(2) *Zu $x_0, y_0 \in \mathbb{R}$ gibt es höchstens ein $u : I \to \mathbb{R}$ mit*

$$u'' + u = 0 \quad \text{auf } I, \quad u(t_0) = x_0,\, u'(t_0) = y_0, \tag{12.4}$$

Beweis: Seien $c_{1,2} : I \to \mathbb{C}$ zwei Lösungen von (12.3). Dann gilt für $c = c_1 - c_2$

$$c' = (c_1 - c_2)' = (c_1)' - (c_2)' = i(c_1 - c_2) = ic, \quad c(t_0) = c_1(t_0) - c_2(t_0) = 0.$$

Aus Lemma 12.1 folgt $|c(t)|^2 = |c(t_0)|^2 = 0$ für alle t, also $c_1 = c_2$. Für (2) seien $u_k : I \to \mathbb{R}$, $k = 1,2$, Lösungen von (12.4). Nach Lemma 12.1 sind dann $c_k = u_k - iu_k'$ Lösungen von $c_k' = ic_k$. Da $c_k(t_0) = u_k(t_0) - iu_k'(t_0) = x_0 - iy_0$, folgt $c_1 = c_2$ aus (1), also $u_1 = u_2$. □

Die Gleichung $c' = ic$ ist formal identisch zur reellen Gleichung $f' = \lambda f$ aus dem vorigen Kapitel, nur dass c Werte in $\mathbb{C}$ hat und auf der rechten Seite die komplexe Multiplikation mit $i = \sqrt{-1}$ steht. Es ist deshalb naheliegend, die Definition der Exponentialfunktion auf $\mathbb{C}$ zu verallgemeinern und damit eine Lösung zu konstruieren.

Definition 12.1 (Exponentialfunktion) *Die komplexe Exponentialfunktion ist*

$$\exp : \mathbb{C} \to \mathbb{C}, \quad \exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}. \tag{12.5}$$

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

Die absolute Konvergenz der Reihe folgt wie in Satz 11.2 aus dem Quotientenkriterium, also ist die Funktion überall definiert.

Satz 12.2 (Existenz der Kreisbewegung) *Die Funktion $c : \mathbb{R} \to \mathbb{C}$, $c(t) = \exp(it)$, ist die eindeutige Lösung des Anfangswertproblems*

$$c' = ic \quad \text{auf } \mathbb{R}, \quad c(0) = 1. \tag{12.6}$$

Beweis: Es folgt direkt $c(0) = \exp(0) = 1$. Betrachte die Approximationen

$$c_n : \mathbb{R} \to \mathbb{C},\, c_n(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} = \sum_{k=0}^{n} i^k \frac{t^k}{k!}.$$

Es gilt $c_n(t) \to c(t)$ mit $n \to \infty$. Wir berechnen

$$c_n'(t) = \sum_{k=1}^{n} i^k \frac{t^{k-1}}{(k-1)!} = i \sum_{k=0}^{n-1} \frac{(it)^k}{k!} = ic_{n-1}(t).$$

Jetzt argumentieren wir wie in Satz 11.2. Für $|t| \leq R$ gilt

$$|c_n(t)| \leq \sum_{k=0}^{n} \frac{|t|^k}{k!} \leq \sum_{k=0}^{\infty} \frac{R^k}{k!} =: C < \infty.$$

Es folgt $|c_n'(t)| = |c_{n-1}(t)| \leq C$, also nach dem (vektorwertigen) Schrankensatz

$$|c_n(t_1) - c_n(t_2)| \leq C\,|t_1 - t_2| \quad \text{für } |t_1|, |t_2| \leq R.$$

Setze nun $f_n(t) = c_n(t) - c_n(t_0) - c_n'(t_0)(t - t_0)$ für $|t_0| < R$. Dann ist $f_n(t_0) = 0$, und

$$|f_n'(t)| = |c_n'(t) - c_n'(t_0)| = |c_{n-1}(t) - c_{n-1}(t_0)| \leq C\,|t - t_0| \quad \text{für } |t| \leq R.$$

Für $s \in [t_0, t]$ bzw. $s \in [t, t_0]$ folgt $|f_n'(s)| \leq C\,|s - t_0| \leq C\,|t - t_0|$. Also gilt

$$|f_n(t)| = |f_n(t) - f_n(t_0)| \leq C\,|t - t_0|^2 \quad \text{für } |t| \leq R,$$

wieder mit dem (vektorwertigen) Schrankensatz. Wir erhalten

$$\Big|\frac{c_n(t) - c_n(t_0)}{t - t_0} - c_n'(t_0)\Big| = \frac{|c_n(t) - c_n(t_0) - c_n'(t_0)(t - t_0)|}{|t - t_0|} \leq C\,|t - t_0|.$$

Da $c_n'(t_0) = ic_{n-1}(t_0) \to ic(t_0)$ mit $n \to \infty$, folgt

$$\Big|\frac{c(t) - c(t_0)}{t - t_0} - ic(t_0)\Big| \leq C\,|t - t_0|.$$

Also ist $c'(t_0) = ic(t_0)$, der Satz ist bewiesen. □

Satz 12.3 (Eigenschaften der Kreisbewegung) *Die Kreisbewegung $c(t) = e^{it}$ ist in $C^\infty(\mathbb{R}, \mathbb{C})$ und hat folgende Eigenschaften:*

(a) $|e^{it}| = 1$ *für alle $t \in \mathbb{R}$.*

(b) $e^{i(s+t)} = e^{is}e^{it}$ *für alle $s, t \in \mathbb{R}$ (Funktionalgleichung).*

(c) $\overline{e^{it}} = e^{-it}$.

Beweis: Nach Satz 12.2 gilt $c' = ic$ auf $\mathbb{R}$ sowie $c(0) = 1$. Daraus folgt mit Satz 9.1 die Stetigkeit von c. Durch Induktion sieht man nun $c \in C^k(\mathbb{R}, \mathbb{C})$ und $c^{(k)} = i^k c$ für alle $k \in \mathbb{N}_0$, analog zum reellen Fall. Aussage (a) folgt aus Lemma 12.1, es ist

$$|c(t)| = |c(0)| = 1 \quad \text{für alle } t \in \mathbb{R}.$$

Für die Funktionalgleichung (b) berechne

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

Aus der Eindeutigkeit, Satz 12.1, folgt $c(s+t) = c(s)c(t)$ wie behauptet. Insbesondere ist $c(t)c(-t) = c(0) = 1$. Andererseits ist auch $c(t)\overline{c(t)} = |c(t)|^2 = 1$ nach (a), also $\overline{c(t)} = c(-t)$. □

Wir kommen nun zu einer weiteren wichtigen Eigenschaft, der Periodizität. Eine Zahl $p \in \mathbb{R}$ heißt Periode einer Funktion $c : \mathbb{R} \to \mathbb{C}$, wenn gilt:

$$c(t + p) = c(t) \quad \text{für alle } t \in \mathbb{R}.$$

Jede Funktion c hat die triviale Periode $p = 0$. Generell bilden die Perioden von c eine Untergruppe von $(\mathbb{R}, +)$, denn für zwei Perioden $p, q \in \mathbb{R}$ gilt

$$c(t + p - q) = c(t + p - q + q) = c(t + p) = c(t).$$

Eine Periode $p > 0$ heißt kleinste Periode, wenn im Intervall $(0, p)$ keine Perioden von c liegen. Die Menge aller Perioden ist dann das Gitter $\mathbb{Z} \cdot p$. Denn zu $q \notin \mathbb{Z} \cdot p$ gibt es ein $k \in \mathbb{Z}$ mit $kp < q < (k + 1)p$. Wäre q eine Periode, so auch $q - kp \in (0, p)$, ein Widerspruch.

Lemma 12.2 *Die Funktion $c(t) = e^{it}$ hat eine Periode $p > 0$.*

Beweis: Es reicht zu zeigen dass ein $\tau > 0$ existiert mit $c(\tau) = i$. Denn dann gilt

$$\begin{aligned}
\frac{d}{dt} c(t + \tau) &= c'(t + \tau) = ic(t + \tau) \quad \text{und} \quad c(t + \tau)|_{t=0} = i, \\
\frac{d}{dt} ic(t) &= i(ic'(t)) \quad \text{und} \quad ic(t)|_{t=0} = i.
\end{aligned}$$

Mit der Eindeutigkeit, Satz 12.1, folgt $c(t + \tau) = ic(t)$ für alle $t \in \mathbb{R}$, und dann

$$c(t + 4\tau) = i^4 c(t) = c(t) \quad \text{für alle } t \in \mathbb{R}.$$

Also ist $p = 4\tau > 0$ eine Periode von c. Schreibe nun $c(t) = u(t) + iv(t)$, und setze

$$\tau = \sup\{t > 0 : u, v > 0 \text{ auf } (0, t)\}.$$

Es gilt $u(0) = 1$, $v(0) = 0$ und $v'(0) = 1$, also gibt es ein $\delta > 0$ mit $u, v > 0$ auf $(0, \delta]$, insbesondere ist $\tau \geq \delta$. Wir behaupten dass τ endlich ist. Es gilt $u, v > 0$ auf $(0, \tau)$. Wegen $u' = -v$ und $v' = u$ ist u streng monoton fallend und v streng monoton wachsend auf $(0, \tau)$. Mit $\varepsilon = v(\delta) > 0$ ist auf $[\delta, \tau)$ sogar $u' = -v \leq -\varepsilon$, und der Schrankensatz liefert

$$u(t) \leq u(\delta) - \varepsilon(t - \delta), \quad \text{das heißt } \tau \leq \delta + \frac{u(\delta)}{\varepsilon} < \infty.$$

Mit $t \nearrow \tau$ ergibt sich $u(\tau) \geq 0$ und $v(\tau) > 0$, außerdem ist $u(\tau)^2 + v(\tau)^2 = 1$. Wäre $u(\tau) > 0$, so hätten wir $u, v > 0$ auf einem größeren Intervall, Widerspruch. Also ist $u(\tau) = 0$ und $v(\tau) = 1$ bzw. $c(\tau) = i$. $\square$

Wir wollen nun den Bezug zur Geometrie herstellen. Dazu brauchen wir den Begriff der Länge einer Kurve. Für eine Kurve $c : [a, b] \to \mathbb{R}^n$ betrachten wir die äquidistante Unterteilung $t_k = a + k\frac{b-a}{n}$, $k = 0, \ldots, n$, und setzen

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

Das ist die Länge des Streckenzugs mit den Punkten $c(t_k)$.

Definition 12.2 (Bogenlänge) *Die Länge von $c \in C^1([a, b], \mathbb{R}^n)$ ist $L(c) = \lim_{n \to \infty} L_n(c)$.*

Man kann zeigen, dass der Grenzwert für jede C^1-Kurve $c : [a, b] \to \mathbb{R}$ existiert, und dass er einen sinnvollen Begriff der Länge liefert. Wir beschränken uns hier auf den Fall $c(t) = e^{it}$.

Lemma 12.3 *Die Kurve $c : \mathbb{R} \to \mathbb{C}$, $c(t) = e^{it}$, bildet längentreu ab: für $I = [a, b]$ gilt*

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

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

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

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

$$L(c|_I) = \lim_{n\to\infty} L_n(c|_I) = |I|. \tag{12.7}$$

□

Auf dem Intervall $[0, \tau]$ durchläuft $c(t) = e^{it}$ den Viertelkreis im ersten Quadranten streng monoton von 1 bis i. Das motiviert die Definition

$$\frac{\pi}{2} := L(c|_{[0,\tau]}) = 1,5707\ldots. \tag{12.8}$$

Nach Lemma 12.3 ist demnach $\tau = \pi/2$ bzw. $4\tau = 2\pi$.

Satz 12.4 (Periodizität von e^{it}) *Die Kreisbewegung $c(t) = e^{it}$ hat die kleinste Periode 2π. Auf $[0, 2\pi]$ durchläuft $c(t)$ den Einheitskreis $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ genau einmal entgegen dem Uhrzeigersinn.*

BEWEIS: Nach Lemma 12.2 ist 4τ eine Periode von c. Wegen $c(t+\tau) = ic(t)$ gehen die Zeiten $k\tau$ für $k = 0, 1, 2, 3, 4$ auf die Punkte $1, i, -1, -i, 1$. Die Intervalle dazwischen werden der Reihe nach bijektiv auf die Viertelkreise in den vier offenen Quadranten abgebildet. Somit ist 4τ die kleinste Periode, und $c : [0, 4\tau] \to \mathbb{S}^1$ ist bijektiv bis auf $c(0) = c(4\tau)$. □

Definition 12.3 (Cosinus und Sinus) *Die Funktionen* $\cos t$ *und* $\sin t$ *sind definiert durch*

$$\cos t = \operatorname{Re}(e^{it}) = \frac{e^{it} + e^{-it}}{2}, \quad \sin t = \operatorname{Im}(e^{it}) = \frac{e^{it} - e^{-it}}{2i}. \tag{12.9}$$

Mit anderen Worten gilt die Eulersche Formel

$$e^{it} = \cos t + i \sin t \quad \textit{für alle } t \in \mathbb{R}. \tag{12.10}$$

Satz 12.5 (Eigenschaften von $\cos$ und $\sin$) *Es gelten folgende Eigenschaften:*

(1) $\cos$ *und* $\sin$ *sind in $C^\infty(\mathbb{R})$, und es gelten die Reihendarstellungen*

$$\cos t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \quad \textit{und} \quad \sin t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}. \tag{12.11}$$

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

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

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

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

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

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

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

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

Beweis: Nach Satz 12.3 ist $c(t) = e^{it}$ in $C^\infty(\mathbb{R}, \mathbb{C})$, also gilt das auch für die Komponenten $\cos t$ und $\sin t$. Aus der Definition der Exponentialfunktion folgt

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

Es gilt $c' = ic$ nach Satz 12.2, die Gleichungen (2) folgen daraus durch Komponentenvergleich, vgl. (12.2). In (3) ergeben sich die Anfangswerte aus $c(0) = 1$ und $c'(0) = i$. Die Differentialgleichungen folgen aus (2) durch Differentiation, vgl. Lemma 12.1. Die Identität (4) ist gleichbedeutend mit $|e^{it}| = 1$, das wurde in Aussage (a) von Satz 12.3 bewiesen. Mit Aussage (c) dieses Satz bekommen wir

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

Also gilt Behauptung (5). Berechne für (6) mit der Funktionalgleichung, siehe Satz 12.3(b),

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

Nach Definition ist klar, dass jede Periode von e^{it} auch Periode von cos und sin ist. Es gilt aber auch die Umkehrung: ist zum Beispiel $\cos(t+p) = \cos t$ für alle $t \in \mathbb{R}$, so folgt durch Ableiten $\sin(t+p) = \sin t$ für alle $t \in \mathbb{R}$, und damit $\exp i(t+p) = \exp it$. Nach Satz 12.4 ist 2π die kleinste Periode von $c(t) = e^{it}$, also gilt das auch für $\cos t$ und $\sin t$. □

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

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

Die Nullstellen von Cosinus und Sinus sind wie folgt.

$$\begin{aligned} e^{it} = 1 \quad &\Leftrightarrow \quad t = 2k\pi \quad \text{ mit } k \in \mathbb{Z}, \\ \cos t = 0 \quad &\Leftrightarrow \quad t = \pi/2 + k\pi \quad \text{ mit } k \in \mathbb{Z}, \\ \sin t = 0 \quad &\Leftrightarrow \quad t = k\pi \quad \text{ mit } k \in \mathbb{Z}. \end{aligned} \tag{12.12}$$

Die Funktionen cos und sin sind periodisch und damit nicht injektiv. Wir können sie aber auf ein Intervall einschränken, auf dem sie streng monoton sind und damit eine Umkehrfunktion *arcus cosinus* bzw. *arcus sinus* [5] haben. Man spricht genauer von einem Zweig dieser Umkehrfunktionen, denn Einschränkung auf ein anderes Monotonieintervall führt zu einer anderen Umkehrfunktion. Unsere Wahl hier wird oft als Hauptzweig bezeichnet.

Satz 12.6 (Arcusfunktionen) $\cos : [0,\pi] \to [-1,1]$ *und* $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1,1]$ *haben stetige Umkehrfunktionen, die wie folgt bezeichnet werden:*

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

$$\arccos'(x) = -\frac{1}{\sqrt{1-x^2}} \quad \textit{für } x \in (-1,1). \tag{12.13}$$

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

$$\arcsin'(x) = \frac{1}{\sqrt{1-x^2}} \quad \textit{für } x \in (-1,1). \tag{12.14}$$

Beweis: Es gilt $\cos' = -\sin < 0$ auf $(0,\pi)$, und $\sin' = \cos > 0$ auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Also ist cos auf $[0,\pi]$ streng monoton fallend, und sin auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$ streng monoton wachsend. Die Existenz und Stetigkeit der Umkehrfunktionen gilt nach Satz 8.2. Die Arcusfunktionen sind differenzierbar auf $(-1,1)$ nach Satz 9.4. Wir berechnen, wobei wir $\sin(\arccos x) > 0$ haben wegen $\arccos x \in (0,\pi)$:

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

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

Mithilfe der Funktion arccos können wir den Winkel zwischen Einheitsvektoren $v, w \in \mathbb{R}^n$ definieren. Betrachte erst im $\mathbb{R}^2$ die Vektoren $(1,0)$ und (x,y). Wähle $t \in [0,2\pi]$ mit $(x,y) = (\cos t, \sin t)$. Dann gilt

$$\arccos x = \arccos(\cos t) = \begin{cases} t & \text{falls } t \in [0,\pi], \text{ also } y \geq 0, \\ 2\pi - t & \text{falls } t \in [\pi, 2\pi], \text{ also } y \leq 0. \end{cases}$$

Die Punkte $(1,0)$ und (x,y) zerlegen den Vollkreis in zwei Winkel, $\arccos x$ ist die Bogenlänge des kleineren Winkels. Sei nun E die von $v, w \in \mathbb{R}^n$ aufgespannte Ebene. Wähle eine Orhonormalbasis e_1, e_2 von E mit $v = e_1$. Mit $w = xe_1 + ye_2$ gilt dann

$$\langle v, w \rangle = \langle e_1, xe_1 + ye_2 \rangle = x.$$

[5] Arcus = Bogen (Latein)

Definition 12.4 (Winkel) *Der Winkel zwischen $v, w \in \mathbb{R}^n$, $v, w \neq 0$, ist*

$$\angle(v,w) = \arccos \left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\rangle \in [0,\pi].$$

Das Skalarprodukt liegt in $[-1,1]$ nach der Ungleichung von Cauchy-Schwarz, Satz 5.5. In den Grenzfällen sind v, w parallel bzw. antiparallel, der Winkel ist dann $\arccos 1 = 0$ bzw. $\arccos -1 = \pi$.

In der Schule werden die trigonometrischen Funktionen elementargeometrisch erklärt. Um für ein gegebenes $\alpha \in \mathbb{R}$ die Werte $\cos\alpha$ und $\sin\alpha$ zu definieren, wählt man ein rechtwinkliges Dreieck mit Winkel α und setzt

$$\cos\alpha = \frac{\text{Ankathete}}{\text{Hypothenuse}} \quad \text{und} \quad \sin\alpha = \frac{\text{Gegenkathete}}{\text{Hypothenuse}}.$$

Durch Einzeichnen der Winkelhalbierenden im gleichseitigen Dreieck und im Quadrat, und mit dem Satz des Pythagoras, ergeben sich die speziellen Werte

α	0	$\pi/6$	$\pi/4$	$\pi/3$	$\pi/2$
$\cos\alpha$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
$\sin\alpha$	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1

Allgemein beruft sich die Winkelmessung in der Schule aber auf die Anschauung, der Grenzwert der Bogenlänge wird nicht behandelt. Für allgemeine $\alpha \in \mathbb{R}$ ist die Konstruktion eines rechtwinkligen Dreiecks mit Winkel α exakt nicht möglich, sondern höchstens praktisch mit dem Geodreieck.

Folgerung 12.1 (Polarkoordinaten) *Zu jedem $z \in \mathbb{C}\backslash\{0\}$ gibt es eindeutig bestimmte $r > 0$, $\vartheta \in [0, 2\pi)$ mit $z = re^{i\vartheta}$.*

Beweis: Sei zunächst $|z| = 1$. Wir definieren

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

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

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

Für $z \in \mathbb{C}\backslash\{0\}$ beliebig setzen wir $r = |z| > 0$ und definieren $\vartheta \in [0, 2\pi]$ mit $z/r = e^{i\vartheta}$. Die Eindeutigkeit von r ist klar, denn $r = |z|$, für den Polarwinkel ϑ folgt sie aus Satz 12.4. $\square$

Allgemein ist für Punkte $v, w \in \mathbb{S}^1$ ein orientierter Winkel in $[0, 2\pi)$ definiert, und zwar als Länge des Bogens, der bei Start in v im positiven Sinn zuerst durchlaufen wird. Für Vektoren v, w im $\mathbb{R}^n$ kann der Winkel nur in $[0,\pi]$ definiert werden, weil die aufgespannte Ebene a priori nicht orientiert ist.

Zum Schluss des Kapitels diskutieren wir die wesentlichen Eigenschaften der Exponentialfunktion für beliebige komplexe Argumente $z \in \mathbb{C}$.

Satz 12.7 (komplexe Exponentialfunktion) *Die Funktion*

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

hat folgende Eigenschaften:

(1) *Für* $\lambda \in \mathbb{C}$ *ist* $z(t) = e^{\lambda t}$ *die eindeutige Lösung des Anfangswertproblems*

$$z'(t) = \lambda z(t) \text{ für } t \in \mathbb{R}, \quad z(0) = 1.$$

(2) $e^{z_1+z_2} = e^{z_1} e^{z_2}$ *für alle* $z_1, z_2 \in \mathbb{C}$ *(Funktionalgleichung).*

(3) *Es gilt* $\exp(\mathbb{C}) = \mathbb{C}\backslash\{0\}$*, und* $e^{z_1} = e^{z_2}$ *genau wenn* $z_1 - z_2 \in 2\pi i\mathbb{Z}$.

Beweis:
zu (1): Dass $e^{\lambda t}$ eine Lösung ist, wurde im Fall $\lambda = i$ in Satz 12.2 gezeigt. Das dortige Argument lässt sich wörtlich übertragen, man muss nur die Zahl i durch λ ersetzen. Ist $z(t)$ irgendeine Lösung des Anfangswertproblems, so folgt

$$\frac{d}{dt}\big(e^{-\lambda t} z(t)\big) = -\lambda e^{-\lambda t} z(t) + e^{-\lambda t} \lambda z(t) = 0,$$

also $e^{-\lambda t} z(t) = e^{-\lambda t} z(t)|_{t=0} = 1$. Wir können in der Rechnung auch $e^{\lambda t}$ statt $z(t)$ nehmen, damit folgt $z(t) = 1/e^{-\lambda t} = e^{\lambda t}$.

zu (2): Wir berechnen mit (1) und der Produktregel

$$\frac{d}{dt}\big(e^{tz_1} e^{tz_2}\big) = z_1 e^{tz_1} e^{tz_2} + e^{tz_1} z_2 e^{tz_2} = (z_1 + z_2) e^{tz_1} e^{tz_2} \quad \text{und} \quad e^{tz_1} e^{tz_2}|_{t=0} = 1.$$

Aus (1) folgt $e^{tz_1} e^{tz_2} = e^{t(z_1+z_2)}$, und mit $t = 1$ die Behauptung.

zu (3): Für $z = x + iy$ gilt $e^z = e^x e^{iy}$, also $|e^z| = e^x > 0$. Sei $w \in \mathbb{C}\backslash\{0\}$ gegeben. Nach Folgerung 12.1 gibt es ein $y \in [0, 2\pi)$ mit $e^{iy} = w/|w|$. Setzen wir $x = \log|w|$, so folgt

$$w = |w| \frac{w}{|w|} = e^x e^{iy} = e^z \quad \text{mit } z = x + iy.$$

Weiter gilt, da 2π kleinste Periode von e^{iy} ist,

$$e^z = e^x e^{iy} = 1 \quad \Leftrightarrow \quad e^x = e^{iy} = 1 \quad \Leftrightarrow \quad x = 0,\ y \in 2\pi\mathbb{Z}.$$

Daraus folgt mit der Funktionalgleichung

$$e^{z_1} = e^{z_2} \quad \Leftrightarrow \quad e^{z_1 - z_2} = 1 \quad \Leftrightarrow \quad z_1 - z_2 \in 2\pi i\mathbb{Z}.$$

Die Exponentialabbildung hat also die Periode $2\pi i$, der Streifen $0 \le \operatorname{Im} z < 2\pi$ wird bijektiv auf $\mathbb{C}\backslash\{0\}$ abgebildet. Damit ist (3) gezeigt. □

13 Das Riemannsche Integral

In diesem Kapitel wird das Integral für Funktionen $f : I = [a, b] \to \mathbb{R}$ einer Variablen eingeführt. Anschaulich ist das Integral, für $f \geq 0$, der Flächeninhalt der Menge

$$\{(x, y) \in \mathbb{R}^2 : x \in I,\, 0 < y < f(x)\}.$$

Allerdings haben wir den Begriff des Flächeninhalts noch gar nicht behandelt, so dass diese Vorstellung nicht als Definition dienen kann.

Für Treppenfunktionen, also stückweise konstante Funktionen, kann das Integral elementar als Rechtecksumme definiert werden. Unsere Strategie ist, dies auf allgemeinere Funktionen f durch Approximation mit Treppenfunktionen f_k zu erweitern. Wir brauchen dazu ein geeignetes Konzept für die Konvergenz $f_k \to f$ von Funktionen. Um die Definition des Integrals übersichtlich zu halten, erledigen wir das vorab. Ein naheliegender Begriff ist die punktweise Konvergenz: für Funktionen $f_k, f : D \to \mathbb{R}$ auf einer Menge D definieren wir

$$f_k \to f \text{ punktweise auf } D \quad \Leftrightarrow \quad f_k(x) \to f(x) \text{ für alle } x \in D. \tag{13.1}$$

Beispiel 13.1 Für $c_k > 0$ betrachte $f_k : I = [0, 1] \to \mathbb{R}$ mit

$$f_k(x) = \begin{cases} c_k & \text{für } 0 < x \leq \frac{1}{k}, \\ 0 & \text{sonst.} \end{cases}$$

Die Folge konvergiert punktweise gegen $f = 0$, unabhängig von der Wahl der c_k. Denn es ist $f_k(0) = 0$ für alle k, und für $x > 0$ ist $f_k(x) = 0$ für $k > \frac{1}{x}$. Der Inhalt des Rechtecks $[0, 1/k] \times (0, c_k)$ ist c_k/k, diese Folge kann gegen jede Zahl in $[0, \infty]$ gehen je nach Wahl der c_k. Dagegen ist das Integral der Grenzfunktion $f = 0$ natürlich Null.

Die punktweise Konvergenz reicht also nicht aus, um die Konvergenz der Integrale zu garantieren; wir brauchen mehr Kontrolle.

Definition 13.1 (Supremumsnorm) *Sei D eine Menge. Für $f : D \to \mathbb{R}$ definieren wir*

$$\|f\|_D = \sup\{|f(x)| : x \in D\} \in [0, \infty].$$

Definition 13.2 (Gleichmäßige Konvergenz) *Eine Folge von Funktionen $f_k : D \to \mathbb{R}$ konvergiert gleichmäßig gegen $f : D \to \mathbb{R}$, falls gilt:*

$$\|f - f_k\|_D \to 0 \quad \textit{mit } k \to \infty.$$

Beispiel 13.2 Für die f_k in Beispiel 13.1 ist $\|f_k\|_I = c_k$, also gilt mit $f = 0$

$$\|f_k - f\|_I \to 0 \quad \Rightarrow \quad c_k \to 0 \quad \Rightarrow \quad \frac{c_k}{k} \to 0.$$

Das sieht, jedenfalls was das Beispiel angeht, besser aus. Der Unterschied zwischen den beiden Konvergenzbegriffen wird auch in folgender Formulierung deutlich:

punktweise:	$\forall \varepsilon > 0$	$\forall x \in D$	$\exists K :$	$\lvert f_k(x) - f(x)\rvert < \varepsilon$ für alle $k > K$,
gleichmäßig:	$\forall \varepsilon > 0$	$\exists K$	$\forall x \in D :$	$\lvert f_k(x) - f(x)\rvert < \varepsilon$ für alle $k > K$.

Punktweise Konvergenz erlaubt dass K von x abhängt, also $K = K(\varepsilon, x)$, dagegen verlangt gleichmäßige Konvergenz eine gemeinsame Schranke $K = K(\varepsilon)$ für alle $x \in D$. Kontraposition der Eigenschaft von $K = K(\varepsilon, x)$ ergibt die Implikation

$$|f_k(x) - f(x)| \geq \varepsilon \quad \Rightarrow \quad K(\varepsilon, x) \geq k.$$

Beispiel 13.3 Wähle $c_k = 1$ in Beispiel 13.1. Für $0 < \varepsilon \leq 1$ gilt mit $x_n = \frac{1}{n}$

$$|f_n(x_n) - f(x_n)| = |f_n(x_n)| = 1 \geq \varepsilon, \quad \text{also} \quad K(\varepsilon, x_n) \geq n.$$

Es gibt somit keine gemeinsame Schranke $K = K(\varepsilon)$ für alle $x \in [0,1]$.

Wegen $|f_k(x) - f(x)| \leq \|f_k - f\|_D$ ist klar: aus gleichmäßig folgt punktweise. Um eine Folge f_k auf gleichmäßige Konvergenz zu testen, kann man also in zwei Schritten vorgehen:

- Konvergiert die Folge punktweise? Wenn nicht, so erst recht nicht gleichmäßig. Wenn ja, so ist die punktweise Grenzfunktion die einzig mögliche Kandidatin für den gleichmäßigen Grenzwert.
- Nun bestimme $\|f_k - f\|_D$ bzw. schätze diese Norm ab. Gilt $\|f_k - f\|_D \to 0$ mit $k \to \infty$, so ist f_k gleichmäßig konvergent gegen f. Wenn nicht, so ist f_k zwar punktweise, aber nicht gleichmäßig konvergent.

Wir bezeichnen mit $\mathcal{B}(D)$ den Raum der beschränkten Funktionen auf D, also

$$\mathcal{B}(D) = \{f : D \to \mathbb{R} : \|f\|_D < \infty\}. \tag{13.2}$$

Dies ist ein Untervektorraum des Raums aller Funktionen auf D. Die Supremumsnorm ist eine Norm auf $\mathcal{B}(D)$, das heißt sie hat folgende Eigenschaften.

Positivität: $\|f\|_D \geq 0$ mit GLeichheit genau wenn $f = 0$,

Halblinearität: $\|\lambda f\|_D = |\lambda| \, \|f\|_D$ für $\lambda \in \mathbb{R}$.

Dreiecksungleichung: $\|f + g\|_D \leq \|f\|_D + \|g\|_D$.

Damit kommen wir zurück zum eigentlichen Thema, nämlich der Konstruktion des Integrals.

Definition 13.3 (Zerlegung) *Eine Zerlegung Z des Intervalls $I = [a,b]$ ist eine geordnete Menge von Punkten $a = x_0 \leq x_1 \leq \ldots \leq x_N = b$. Wir setzen $I_k = [x_{k-1}, x_k]$ sowie $\Delta x_k = x_k - x_{k-1}$ für $k = 1, \ldots, N$, und definieren die Feinheit von Z durch*

$$\Delta(Z) = \max_{1 \leq k \leq N} \Delta x_k. \tag{13.3}$$

Definition 13.4 (Riemannsche Summe) *Sei $f : I = [a,b] \to \mathbb{R}$. Die Riemannsche Summe von f zur Zerlegung Z und den Stützstellen $\xi_k \in I_k$ ist*

$$S_{Z,\xi}(f) = \sum_{k=1}^{N} f(\xi_k) \, \Delta x_k \in \mathbb{R}.$$

Die Riemannsche Summe ist ein Näherungswert für das noch zu definierende Integral. Eine konkrete Wahl der Zerlegung und der Stützstellen, zum Beispiel äquidistante Zerlegung und Intervallmittelpunkte, würde ein numerisches Verfahren zur Approximation des Integrals ergeben. Aber um das Integral stabil zu definieren, brauchen wir beliebige Approximationen, siehe Beispiel 13.5. Jedenfalls sollte bei Verfeinerung einer Zerlegung der Approximationsfehler kleiner werden. Dies führt auf B. Riemanns Begriff der Integrierbarkeit. [6]

Definition 13.5 (Riemann-Integral) *Eine beschränkte Funktion $f : I = [a,b] \to \mathbb{R}$ heißt (Riemann-)integrierbar mit Integral $S \in \mathbb{R}$, falls es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass für jede Zerlegung Z und jede Wahl ξ der Stützstellen gilt:*

$$\Delta(Z) < \delta \quad \Rightarrow \quad |S_{Z,\xi}(f) - S| < \varepsilon.$$

Wir nennen dann S das (bestimmte) Integral von f auf $[a,b]$ und schreiben

$$S = \int_I f \in \mathbb{R}.$$

Wir bezeichnen mit $\mathcal{R}(I)$ die Klasse der Riemann-integrierbaren Funktionen auf I.

Bemerkung Die Zahl S ist durch die Definition eindeutig bestimmt. Wähle dazu eine Folge von Riemannschen Summen $S_n(f)$ zu Zerlegungen Z_n mit $\Delta(Z_n) \to 0$. Aus Definition 13.5 folgt $S = \lim_{n\to\infty} S_n(f)$, dadurch ist S festgelegt.

Die Annahme, dass f beschränkt ist, kann in der Definition weggelassen werden; die Beschränktheit ergibt sich dann als Folgerung. Wir verzichten auf den Beweis.

Beispiel 13.4 Die konstante Funktion $f : I = [a,b] \to \mathbb{R}$, $f(x) = c$, ist integrierbar mit

$$\int_I f = c\,(b-a).$$

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

$$S_{Z,\xi}(f) = \sum_{k=1}^{N} c\,\Delta x_k = c\,\sum_{k=1}^{N}(x_k - x_{k-1}) = c\,(b-a).$$

Beispiel 13.5 Die Dirichletfunktion

$$\chi_{\mathbb{Q}} : [0,1] \to \mathbb{R},\ \chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{für } x \in \mathbb{Q}, \\ 0 & \text{für } x \notin \mathbb{Q} \end{cases}$$

ist nicht (Riemann-) integrierbar. Sei zum Beispiel Z die Unterteilung mit $I_k = [\frac{k-1}{N}, \frac{k}{N}]$, also $\Delta(Z) = \frac{1}{N} \to 0$. Wählt man rationale Stützstellen ξ_k so ist $S_{Z,\xi}(\chi_{\mathbb{Q}}) = 1$, für ξ_k irrational ist dagegen $S_{Z,\xi}(\chi_{\mathbb{Q}}) = 0$.

Satz 13.1 (Linearität des Integrals) *$\mathcal{R}(I)$ ist ein Untervektorraum von $\mathcal{B}(I)$, und das Integral $\mathcal{R}(I) \to \mathbb{R}$, $f \mapsto \int_I f$, ist ein lineares Funktional. Es gilt also für $\lambda, \mu \in \mathbb{R}$*

$$\int_I (\lambda\, f + \mu\, g) = \lambda \int_I f + \mu \int_I g.$$

[6] ein anderer Zugang zum Integral stammt von H. Lebesgue, dieser wird in Analysis 3 behandelt.

Beweis: Seien $f, g \in \mathcal{R}(I)$. Dann ist $\lambda f + \mu g$ beschränkt, und es gilt

$$S_{Z,\xi}(\lambda f + \mu g) = \lambda S_{Z,\xi}(f) + \mu S_{Z,\xi}(g).$$

Also folgt für $\Delta(Z)$ hinreichend klein

$$\Big|S_{Z,\xi}(\lambda f + \mu g) - \Big(\lambda \int_I f + \mu \int_I g\Big)\Big| \leq |\lambda| \, \Big|S_{Z,\xi}(f) - \int_I f\Big| + |\mu| \, \Big|S_{Z,\xi}(g) - \int_I g\Big| < \varepsilon.$$

□

Integrierbarkeit ist eine globale Eigenschaft, es kommt auf die Funktion als Ganzes an. Deshalb ist nicht direkt klar, wie groß der Raum $\mathcal{R}(I)$ ist. Wir zeigen in diesem Kapitel dass stückweise stetige Funktionen integrierbar sind. Unsere Strategie ist wie folgt:

- Treppenfunktionen sind in $\mathcal{R}(I)$.
- $\mathcal{R}(I)$ ist abgeschlossen unter gleichmäßiger Konvergenz.
- Stetige Funktionen sind durch Treppenfunktionen approximierbar.

Lemma 13.1 *Seien $f, \tilde{f} : I \to \mathbb{R}$ mit $\tilde{f}(x) = f(x)$ für alle $x \in I \backslash \{p_1, \ldots, p_r\}$. Mit $f \in \mathcal{R}(I)$ ist dann auch $\tilde{f} \in \mathcal{R}(I)$ und es gilt $\int_I \tilde{f} = \int_I f$.*

Beweis: Wir zeigen die Aussage für einen Ausnahmepunkt $p \in I$; der allgemeine Fall folgt daraus per Induktion. Es gilt für jede Zerlegung Z mit Stützstellen ξ_k

$$S_{Z,\xi}(\tilde{f}) - S_{Z,\xi}(f) = \sum_{k=1}^{N} (\tilde{f}(\xi_k) - f(\xi_k)) \, \Delta x_k = (\tilde{f}(p) - f(p)) \sum_{\{k : \xi_k = p\}} \Delta x_k.$$

Es gibt höchstens zwei k mit $p \in I_k$ und $\Delta x_k > 0$. Damit schätzen wir ab

$$\begin{aligned} \Big|S_{Z,\xi}(\tilde{f}) - \int_I f\Big| &\leq \Big|S_{Z,\xi}(\tilde{f}) - S_{Z,\xi}(f)\Big| + \Big|S_{Z,\xi}(f) - \int_I f\Big| \\ &\leq 2 \, \Big|\tilde{f}(p) - f(p)\Big| \, \Delta(Z) + \Big|S_{Z,\xi}(f) - \int_I f\Big|, \end{aligned}$$

und die rechte Seite geht gegen Null mit $\Delta(Z) \to 0$. □

Wie Beispiel 13.5 zeigt, ist Lemma 13.1 nicht richtig für eine abzählbare Ausnahmemenge.

Lemma 13.2 *Sei $I = I_1 \cup \ldots \cup I_n$ eine Zerlegung von $I = [a, b]$ in Intervalle $I_k = [a_{k-1}, a_k]$. Ist $f : I \to \mathbb{R}$ auf jedem der Teilintervalle I_k integrierbar, so folgt $f \in \mathcal{R}(I)$ und*

$$\int_I f = \sum_{k=1}^{n} \int_{I_k} f.$$

Beweis: Es reicht den Fall $n = 2$ zu betrachten. Sei $I = I' \cup I''$ mit $I' = [a, p]$ und $I'' = [p, b]$, und sei $f : I \to \mathbb{R}$ integrierbar auf I' und I'', insbesondere $\|f\|_I = \max(\|f\|_{I'}, \|f\|_{I''}) < \infty$. Sei Z eine Zerlegung von I mit Punkten $a = x_0 \leq \ldots \leq x_N = b$ sowie Stützstellen $\xi_1, \ldots, \xi_N$.

Wähle $r \in \{1, \ldots, N\}$ mit $p \in [x_{r-1}, x_r]$, und definiere Zerlegungen Z',Z'' von I', I'' mit Stützstellen ξ', ξ'' wie folgt:

$$\begin{aligned} Z' &= \{a = x_0 \leq \ldots \leq x_{r-1} \leq p\} \quad & \xi' &= \{\xi_1, \ldots, \xi_{r-1}, p\}, \\ Z'' &= \{p \leq x_r \leq \ldots \leq x_N = b\} \quad & \xi'' &= \{p, \xi_{r+1}, \ldots, \xi_N\}. \end{aligned}$$

Offenbar gilt $\Delta(Z'), \Delta(Z'') \leq \Delta(Z)$. Die Bilanz der Riemannschen Summen lautet

$$|S_{Z,\xi}(f) - \big(S_{Z',\xi'}(f) + S_{Z'',\xi''}(f)\big)| = |(f(\xi_r) - f(p))\,\Delta x_r| \leq 2\,\|f\|_I\,\Delta(Z).$$

Es folgt mit der Dreiecksungleichung

$$\Big|S_{Z,\xi}(f) - \Big(\int_{I'} f + \int_{I''} f\Big)\Big| \leq 2\,\|f\|_I\,\Delta(Z) + \Big|S_{Z',\xi'}(f) - \int_{I'} f\Big| + \Big|S_{Z'',\xi''}(f) - \int_{I''} f\Big|.$$

Die rechte Seite geht mit $\Delta(Z) \to 0$ gegen Null. □

Folgerung 13.1 *Sei $f : I \to \mathbb{R}$ eine (Riemannsche) Treppenfunktion, d. h. es gibt eine Unterteilung $a = a_0 < a_1 < \ldots < a_n = b$ und $c_i \in \mathbb{R}$ $(i = 1, \ldots, n)$ mit $f(x) = c_i$ für alle $x \in (a_{i-1}, a_i)$. Dann ist f integrierbar und es gilt*

$$\int_I f = \sum_{i=1}^{n} (a_i - a_{i-1}) c_i.$$

Beweis: Nach Lemma 13.1 ist $f : [a_{i-1}, a_i] \to \mathbb{R}$ integrierbar für alle $i = 1, \ldots, n$. Aus Lemma 13.2 folgt die Behauptung. □

Im zweiten Schritt geht es um das Verhalten des Integrals unter Approximationen. Hier ein weiteres Beispiel dafür, dass punktweise Konvergenz nicht ausreicht.

Beispiel 13.6 Sei $q_1, q_2, \ldots$ eine Abzählung der rationalen Zahlen in $[0, 1]$. Betrachte

$$\chi_n : [0,1] \to \mathbb{R},\ \chi_n(x) = \begin{cases} 1 & \text{für } x \in \{q_1, \ldots, q_n\} \\ 0 & \text{sonst .} \end{cases}$$

Die Folge χ_n konvergiert punktweise gegen $\chi_{\mathbb{Q}}$, die nach Beispiel 13.5 nicht integrierbar ist.

Wir betrachten deshalb nun die Supremumsnorm.

Satz 13.2 (Abschätzung des Integrals) *Für $f \in \mathcal{R}(I)$ gilt*

$$\Big|\int_I f\Big| \leq \|f\|_I\,|I|.$$

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

$$\begin{aligned} |S_{Z,\xi}(f)| &\leq \Big|\sum_{k=1}^{N} f(\xi_k)\,\Delta x_k\Big| \\ &\leq \sum_{k=1}^{N} |f(\xi_k)|\,\Delta x_k \quad \big(= S_{Z,\xi}(|f|)\big) \\ &\leq \|f\|_I \sum_{k=1}^{N} \Delta x_k = \|f\|_I\,|I|. \end{aligned}$$

Die Abschätzung für das Integral folgt. □

Bemerkung. Genauer zeigt der Beweis, falls zusätzlich $|f| \in \mathcal{R}(I)$, die Ungleichungen

$$\Big| \int_I f \Big| \leq \int_I |f| \leq \|f\|\,|I|. \tag{13.4}$$

Tatsächlich ist für $f \in \mathcal{R}(I)$ immer auch $|f| \in \mathcal{R}(I)$, was wir aber hier nicht beweisen.

Für $f, f_k \in \mathcal{R}(I)$ haben wir jedenfalls nun

$$\Big| \int_I f_k - \int_I f \Big| = \Big| \int_I (f_k - f) \Big| \leq \|f_k - f\|_I\,|I|.$$

Damit folgt aus $\|f_k - f\|_I \to 0$ auch die Konvergenz der Integrale.

Satz 13.3 (Integral und gleichmäßige Konvergenz) *Konvergiert die Folge $f_k \in \mathcal{R}(I)$ gleichmäßig gegen $f : I \to \mathbb{R}$, so ist $f \in \mathcal{R}(I)$ und es gilt*

$$\int_I f = \lim_{k\to\infty} \int_I f_k.$$

Beweis: Die Funktion f ist beschränkt wegen $\|f\|_I \leq \|f - f_k\|_I + \|f_k\|_I < \infty$. Mit $S_k = \int_I f_k$ gilt für k, l hinreichend groß nach Satz 13.2

$$|S_k - S_l| \leq \|f_k - f_l\|_I\,|I| \leq \left(\|f_k - f\|_I + \|f - f_l\|_I\right)|I| < \varepsilon.$$

Wir setzen $S = \lim_{k\to\infty} S_k$ und zeigen, dass f integrierbar ist mit $\int_I f = S$. Für jede Zerlegung Z mit Stützstellen ξ_j und jedes $k \in \mathbb{N}$ haben wir mit (13.4)

$$\begin{aligned} |S_{Z,\xi}(f) - S| &\leq |S_{Z,\xi}(f) - S_{Z,\xi}(f_k)| + |S_{Z,\xi}(f_k) - S_k| + |S_k - S| \\ &\leq \|f - f_k\|_I\,|I| + |S_{Z,\xi}(f_k) - S_k| + |S_k - S|. \end{aligned}$$

Zu $\varepsilon > 0$ wählen wir erst $k \in \mathbb{N}$ mit $\|f - f_k\|_I\,|I| < \varepsilon/3$ und $|S_k - S| < \varepsilon/3$. Für $\Delta(Z) < \delta$ ist dann auch $|S_{Z,\xi}(f_k) - S_k| < \varepsilon/3$, da f_k integrierbar ist. Damit ist der Satz bewiesen. □

Im dritten Schritt geht es um die Approximation von stetigen Funktionen durch Treppenfunktionen. Um die Stetigkeit einer Funktion $f : D \to \mathbb{R}$ mit $D \subset \mathbb{R}$ zu quantifizieren, betrachten wir die Oszillation (oder Schwankung) auf Skala $\delta > 0$:

$$\omega_f(\delta) = \sup_{x,x' \in D,\, |x-x'|<\delta} |f(x) - f(x')| \in [0,\infty]. \tag{13.5}$$

Wird $\delta > 0$ verkleinert, so sind am Supremum weniger Punkte x, x' beteiligt, der Funktionswert wird also höchstens kleiner. Insbesondere existiert nach Lemma 8.1 der Grenzwert $\lim_{\delta\searrow 0} \omega_f(\delta)$. Für f stetig sollte dieser Grenzwert Null sein, da f auf kleinen Skalen wenig schwankt. Das ist aber so nicht richtig, wie das folgende Beispiel zeigt.

Beispiel 13.7 Sei $f : I = (0,1] \to \mathbb{R}$, $f(x) = \sin \frac{1}{x}$. Dann ist f stetig, und $|f(x) - f(x')| \leq 2$ für alle $x, x' \in (0,1]$. Aber es gilt

$$|f(x_n^+) - f(x_n^-)| = 2 \quad \text{für} \quad x_n^{\pm} = \frac{1}{n\pi \pm \pi/2} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Es folgt $\omega_f(\delta) = 2$ für alle $\delta > 0$. Die Funktion kann nicht auf $[0,1]$ stetig fortgesetzt werden, da sie in jeder Umgebung von $x = 0$ zwischen den Werten ± 1 oszilliert.

Satz 13.4 *Ist $f : D \to \mathbb{R}$ stetig mit $D \subset \mathbb{R}$ folgenkompakt, so gilt $\lim_{\delta \searrow 0} \omega_f(\delta) = 0$.*

Beweis. Andernfalls gibt es ein $\varepsilon > 0$ und eine Folge $\delta_n \searrow 0$ mit $\omega_f(\delta_n) \geq \varepsilon > 0$. Dann existieren $x_n, x'_n \in D$ mit $|x_n - x'_n| < \delta_n$, so dass $|f(x_n) - f(x'_n)| \geq \frac{\varepsilon}{2}$. Nach Übergang zu einer Teilfolge konvergiert x_n gegen ein $x_0 \in D$. Aber dann gilt auch $x'_n \to x_0$, und wegen f stetig in x_0 folgt

$$\frac{\varepsilon}{2} \leq \lim_{n \to \infty} |f(x_n) - f(x'_n)| = |f(x_0) - f(x_0)| = 0, \text{ Widerspruch.}$$

□

Definition 13.6 *$f : D \to \mathbb{R}$ heißt gleichmäßig stetig wenn $\lim_{\delta \searrow 0} \omega_f(\delta) = 0$.*

Der Satz wird daher oft so zitiert: eine stetige Funktion f auf einer kompakten Menge D ist automatisch gleichmäßig stetig. Wir bemerken ohne Beweis, dass die gleichmäßige Stetigkeit auch wie folgt charakterisiert werden kann:

$$\forall \varepsilon > 0 \quad \exists \delta > 0 : \quad |f(x) - f(x')| < \varepsilon \text{ für alle } x, x' \in D \text{ mit } |x - x'| < \delta.$$

Zur Erinnerung: die Stetigkeit von f auf D bedeutet

$$\forall x_0 \in D \; \forall \varepsilon > 0 \; \exists \delta > 0 : \quad |f(x) - f(x_0)| < \varepsilon \text{ für alle } x \in D \text{ mit } |x - x_0| < \delta.$$

Im allgemeinen kann keine gemeinsame Schranke $\delta > 0$ für alle $x_0 \in D$ gewählt werden, siehe Beispiel 13.7. Genau das ist aber möglich wenn f gleichmäßig stetig ist.

Lemma 13.3 *Jede stetige Funktion $f : I = [a,b] \to \mathbb{R}$ kann gleichmäßig durch Treppenfunktionen approximiert werden.*

Beweis: Sei $x_k = a + k\frac{b-a}{N}$ für $k = 0, \ldots, N$, und setze

$$\varphi_N(x) = \begin{cases} f(x_0) & \text{für } x = x_0, \\ f(x_k) & \text{für } x \in (x_{k-1}, x_k] \text{ mit } k \in \{1, \ldots, N\}, \end{cases}$$

Für $x \in (x_{k-1}, x_k]$ ist $|x - x_k| \leq \frac{1}{N}$, also

$$|\varphi_N(x) - f(x)| = |f(x_k) - f(x)| \leq \omega_f\Big(\frac{1}{N}\Big).$$

Da $\varphi_N(x_0) = f(x_0)$, folgt mit Satz 13.4

$$\|\varphi_N - f\|_I \leq \omega_f\Big(\frac{1}{N}\Big) \to 0 \quad \text{ für } N \to \infty.$$

□

Definition 13.7 (stückweise stetig) *Eine Funktion $f : [a,b] \to \mathbb{R}$ heißt stückweise stetig, wenn es eine Zerlegung $a = a_0 < \ldots < a_N = b$ gibt, so dass f auf jedem Teilintervall $[a_{k-1}, a_k]$ stetig ist, nach eventueller Abänderung in den Endpunkten a_{k-1} und a_k.*

Satz 13.5 (stückweise stetig ⇒ integrierbar) *Ist $f : I = [a,b] \to \mathbb{R}$ stückweise stetig, so ist f Riemann-integrierbar.*

Beweis: Sei zunächst f stetig. Dann ist f gleichmäßiger Limes von Treppenfunktionen nach Lemma 13.3. Nach Folgerung 13.1 sind Treppenfunktionen integrierbar, also auch f wegen Satz 13.3. Ist f nur stückweise stetig, so ist f auf jedem der Teilintervalle stetig nach eventueller Abänderung in den zwei Endpunkten. Wie bewiesen ist die stetige Funktion auf dem Teilintervall integrierbar, also auch f nach Lemma 13.1. Schließlich folgt aus Lemma 13.2, dass f auf ganz I integrierbar ist. $\square$

Unsere Definition des Integrals gilt ohne Änderungen für Funktionen $f : I \to \mathbb{R}^n$. Es kann dann komponentenweise berechnet werden: $f \in \mathcal{B}(I, \mathbb{R}^n)$ ist genau dann integrierbar, wenn alle Komponenten $f_i \in \mathcal{B}(I)$, $i = 1, \ldots, n$, integrierbar sind, und

$$\int_I f = \sum_{i=1}^n \Big(\int_I f_i \Big) e_i \in \mathbb{R}^n.$$

Denn für die Riemannschen Summen gilt offensichtlich

$$S_{Z,\xi}(f) = \sum_{i=1}^n S_{Z,\xi}(f_i) e_i,$$

dies überträgt sich auf die Integrale. Allgemeiner zeigt man, dass das vektorwertige Integral mit linearen Abbildungen $A : \mathbb{R}^n \to \mathbb{R}^m$ vertauscht, das heißt $A(\int_I f) = \int_I (Af)$. Doch nun wieder zu Funktionen $f : I \to \mathbb{R}$, wobei jetzt die Anordnung von $\mathbb{R}$ eine Rolle spielt.

Satz 13.6 (Monotonie des Integrals) *Sind $f, g \in \mathcal{R}(I)$, so gilt:*

$$f \le g \quad \Rightarrow \quad \int_I f \le \int_I g.$$

Beweis: Für jede Zerlegung Z mit Stützstellen ξ_k gilt

$$S_{Z,\xi}(f) = \sum_{k=1}^N f(\xi_k) \Delta x_k \le \sum_{k=1}^N g(\xi_k) \Delta x_k = S_{Z,\xi}(g).$$

$\square$

Folgerung 13.2 (Mittelwertsatz der Integralrechnung) *Sei $\varphi : I = [a, b] \to [0, \infty)$ stetig mit $\int_I \varphi > 0$. Dann gibt es für jede Funktion $f \in C^0(I)$ ein $\xi \in I$ mit*

$$f(\xi) = \frac{1}{\int_I \varphi} \int_I f\varphi.$$

Im Spezialfall $\varphi = 1$ gilt also $f(\xi) = \frac{1}{|I|} \int_I f$.

Beweis: Setze $m = \min_{x \in I} f(x)$ und $M = \max_{x \in I} f(x)$. Wegen $m\varphi \le f\varphi \le M\varphi$ gilt

$$m = \frac{1}{\int_I \varphi} \int_I m\varphi \le \frac{1}{\int_I \varphi} \int_I f\varphi \le \frac{1}{\int_I \varphi} \int_I M\varphi = M.$$

Die Existenz von ξ folgt aus dem Zwischenwertsatz. $\square$

Der Vollständigkeit halber wollen wir noch eine alternative Definition des Riemann-Integrals erklären. Sei $f : I = [a, b] \to \mathbb{R}$ beschränkt und Z eine Zerlegung von I in Teilintervalle I_k der Länge Δx_k. Dann sind Ober- und Untersumme von f bzgl. Z definiert durch

$$\overline{S}_Z(f) = \sum_{k=1}^N (\sup_{I_k} f) \, \Delta x_k \quad \text{und} \quad \underline{S}_Z(f) = \sum_{k=1}^N (\inf_{I_k} f) \, \Delta x_k.$$

Nach Definition gilt $\underline{S}_Z(f) \leq \overline{S}_Z(f)$. Nehmen wir zu Z den Unterteilungspunkt $\xi \in [x_{k-1}, x_k]$ hinzu, so folgt mit $I'_k = [x_{k-1}, \xi]$ und $I''_k = [\xi, x_k]$

$$\begin{aligned} \underline{S}_{Z \cup \{\xi\}}(f) - \underline{S}_Z(f) &= (\inf_{I'_k} f)\,(\xi - x_{k-1}) + (\inf_{I''_k} f)\,(x_k - \xi) - (\inf_{I_k} f)\,(x_k - x_{k-1}) \\ &= (\inf_{I'_k} f - \inf_{I_k} f)\,(\xi - x_{k-1}) + (\inf_{I''_k} f - \inf_{I_k} f)\,(x_k - \xi). \end{aligned}$$

Mit $\inf_{I'_k} f, \inf_{I''_k} f, -\inf_{I_k} f \leq \|f\|_I$ sowie $0 \leq x_k - x_{k-1} \leq \Delta(Z)$ erhalten wir

$$0 \leq \underline{S}_{Z \cup \{\xi\}}(f) - \underline{S}_Z(f) \leq 2\|f\|_I\, \Delta(Z),$$

und analog für die Obersummen

$$0 \geq \overline{S}_{Z \cup \{\xi\}}(f) - \overline{S}_Z(f) \geq -2\|f\|_I\, \Delta(Z).$$

Für beliebige Zerlegungen Z, Z' ergibt sich per Induktion

$$\underline{S}_Z(f) \leq \underline{S}_{Z \cup Z'}(f) \leq \overline{S}_{Z \cup Z'}(f) \leq \overline{S}_{Z'}(f). \tag{13.6}$$

Bezeichnet N die Zahl der Teilintervalle von Z', so folgt ebenfalls induktiv

$$\underline{S}_{Z \cup Z'}(f) - 2N\,\|f\|_I\, \Delta(Z) \leq \underline{S}_Z(f) \leq \overline{S}_Z(f) \leq \overline{S}_{Z \cup Z'}(f) + 2N\|f\|_I\, \Delta(Z). \tag{13.7}$$

Wir definieren nun das Ober- bzw. Unterintegral von f durch

$$\begin{aligned} \overline{S}(f) &= \inf\{\overline{S}_Z(f) : Z \text{ ist Zerlegung von } I\}, \\ \underline{S}(f) &= \sup\{\underline{S}_Z(f) : Z \text{ ist Zerlegung von } I\}. \end{aligned}$$

Aus (13.6) folgt $\underline{S}(f) \leq \overline{S}(f)$.

Satz 13.7 *Eine beschränkte Funktion $f : I \to \mathbb{R}$ ist genau dann (Riemann-) integrierbar, wenn ihr Ober- und Unterintegral übereinstimmen, und es gilt dann*

$$\int_I f = \underline{S}(f) = \overline{S}(f).$$

Beweis: Ist f Riemannintegrierbar, so folgt für $\Delta(Z) < \delta$

$$\int_I f - \varepsilon \leq \inf_\xi S_{Z,\xi}(f) = \underline{S}_Z(f) \leq \underline{S}(f) \leq \overline{S}(f) \leq \overline{S}_Z(f) = \sup_\xi S_{Z,\xi}(f) \leq \int_I f + \varepsilon.$$

Umgekehrt sei Z' eine Zerlegung mit $\underline{S}_{Z'}(f) > \underline{S}(f) - \varepsilon/2$, und N sei die Zahl der Teilintervalle von Z'. Für jede Zerlegung Z gilt $\underline{S}_{Z \cup Z'}(f) \geq \underline{S}_{Z'}(f)$, also folgt mit (13.7)

$$S_{Z,\xi}(f) \geq \underline{S}_Z(f) \geq \underline{S}_{Z \cup Z'}(f) - 2N\|f\|_I\, \Delta(Z) > \underline{S}(f) - \varepsilon/2 - 2N\|f\|_I\, \Delta(Z).$$

Für $\Delta(Z) < \delta$ folgt $S_{Z,\xi}(f) > \underline{S}(f) - \varepsilon$. Die Abschätzung nach oben ist analog. $\square$

14 Ableitung und Integral

Wir kommen nun zu dem zentralen, von Newton und Leibniz studierten Zusammenhang zwischen Differentiation und Integration.

Definition 14.1 *Sei $f : (a,b) \to \mathbb{R}$. Eine Funktion $F : (a,b) \to \mathbb{R}$ heißt Stammfunktion von f, wenn gilt:*

$$F' = f \quad \Leftrightarrow \quad F'(x) = f(x) \quad \textit{für alle } x \in (a,b).$$

Satz 14.1 *Ist F eine Stammfunktion von f auf (a,b), so ist jede Stammfunktion von f auf (a,b) von der Form $F + c$, für eine Konstante $c \in \mathbb{R}$.*

Beweis: Sei G auch Stammfunktion von f auf (a,b). Es folgt

$$(G - F)' = G' - F' = f - f = 0.$$

Nach Folgerung 10.1 gibt es eine Konstante $c \in \mathbb{R}$ mit $G - F = c$, also $G = F + c$. □

Die Gleichung $F' = f$ ist ein weiteres Beispiel für eine Differentialgleichung. Satz 14.1 sagt aus, dass eine Lösung F der Gleichung bis auf eine additive Konstante $c \in \mathbb{R}$ eindeutig bestimmt ist. Es stellt sich nun das Problem der Existenz. Die Gleichung enthält auf der rechten Seite eine willkürliche Funktion, darum lautet hier die Frage:

Für welche f ist die Differentialgleichung $F' = f$ auf dem Intervall (a,b) lösbar?

Zum Beispiel lässt sich eine Lösung direkt hinschreiben, wenn die rechte Seite ein Polynom $f(x) = \sum_{k=0}^n a_k x^k$ ist, und zwar $F(x) = \sum_{k=0}^n a_k \frac{x^{k+1}}{k+1}$. Das wäre aber zu speziell. Mithilfe des Integrals werden wir eine Lösung für jede stetige rechte Seite konstruieren. Setze dazu

$$\int_a^b f(x)\,dx = \begin{cases} \int_I f & \text{falls } a \le b, \\ -\int_I f & \text{falls } a > b. \end{cases}$$

Es gilt dann für beliebige $a,b,c \in \mathbb{R}$, sofern f auf allen Intervallen Riemann-integrierbar ist,

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx + \int_c^a f(x)\,dx = 0. \tag{14.1}$$

Für $a \le b \le c$ ist das Lemma 13.2. Die linke Seite bleibt nun gleich bei zyklischen Vertauschungen von a,b,c, und ändert nur das Vorzeichen bei antizyklischen Vertauschungen. Daher gilt (14.1) allgemein. Die Notation $f(x)\,dx$ ist rein formal, ein dx wurde in der Vorlesung nicht definiert. Aber sie ist nützlich, zum Beispiel wenn f noch von weiteren Variablen $y,z,a,\ldots$ abhängt: es wird spezifiziert, bezüglich welcher Variablen integriert werden soll. Außerdem erinnert die Notation an die Riemannschen Summen, mit denen das Integral definiert wurde.

Satz 14.2 (Hauptsatz der Differential- und Integralrechnung) *Sei $f : I = [a,b] \to \mathbb{R}$ stetig. Dann ist für jedes $x_0 \in I$ die Funktion*

$$F : I \to \mathbb{R}, \quad F(x) = \int_{x_0}^x f(\xi)\,d\xi$$

eine Stammfunktion von f, das heißt es gilt $F'(x) = f(x)$ für alle $x \in I$.

Bemerkung. In den Endpunkten des Intervalls ist dies im Sinne der einseitigen Ableitungen $F'_+(a) = f(a)$ bzw. $F'_-(b) = f(b)$ zu verstehen.

Beweis: Die Funktion F ist definiert nach Satz 13.5. Wir berechnen für $x, x+h \in I$, $h \neq 0$, mit (14.1) und der Abschätzung aus Satz Satz 13.2 [7]

$$\begin{aligned} \left| \frac{F(x+h) - F(x)}{h} - f(x) \right| &= \frac{1}{|h|} \left| \int_{x_0}^{x+h} f(\xi)\, d\xi - \int_{x_0}^{x} f(\xi)\, d\xi - hf(x) \right| \\ &= \frac{1}{|h|} \left| \int_{x}^{x+h} \big(f(\xi) - f(x)\big)\, d\xi \right| \\ &\leq \sup_{\xi \in [x, x+h]} |f(\xi) - f(x)|. \end{aligned}$$

Da f stetig in x ist, geht die rechte Seite mit $h \to 0$ gegen Null. □

Folgerung 14.1 *Sei $f \in C^0([a,b])$. Ist $F \in C^0([a,b])$ Stammfunktion von f auf (a,b), so gilt für beliebiges $x_0 \in [a,b]$*

$$F(x) = F(x_0) + \int_{x_0}^{x} f(\xi)\, d\xi \quad \textit{für alle } x \in [a,b]. \tag{14.2}$$

Beweis: Nach Satz 14.1 und Satz 14.2 gibt es ein $c \in \mathbb{R}$ mit

$$F(x) = c + \int_{x_0}^{x} f(\xi)\, d\xi \quad \text{für alle } x \in (a,b).$$

Durch Grenzübergang folgt das auch für $x = a, b$, denn man hat zum Beispiel für $x = b$

$$F(b) = \lim_{x \nearrow b} F(x) \quad \text{und} \quad \int_{x_0}^{b} f(\xi)\, d\xi = \lim_{x \nearrow b} \int_{x_0}^{x} f(\xi)\, d\xi.$$

Das erste gilt wegen $F \in C^0(I)$, das zweite folgt mit

$$\left| \int_{x_0}^{b} f(\xi)\, d\xi - \int_{x_0}^{x} f(\xi)\, d\xi \right| = \left| \int_{x}^{b} f(\xi)\, d\xi \right| \leq |b - x|\, \|f\|_I \to 0.$$

Setze nun in die Gleichheit $x = x_0$ ein, es folgt $c = F(x_0)$. □

Folgerung 14.2 (Berechnung von Integralen mit Stammfunktionen) *Sei $f \in C^0([a,b])$. Ist $F \in C^0([a,b])$ Stammfunktion von f auf (a,b), so gilt*

$$\int_a^b f(x)\, dx = F(b) - F(a) =: \big[F(x)\big]_{x=a}^{x=b}$$

Beweis: Folgt aus (14.2) mit $x_0 = a$, $x = b$. □

Mit dem Hauptsatz bzw. mit Folgerung 14.2 lassen sich die Differentiationsregeln aus Kapitel 9 in Integrationsregeln übersetzen. Die resultierenden Integrationsregeln sind sowohl theoretisch wichtig als auch zur Berechnung konkreter Integrale.

[7] für $h < 0$ setze $[x, x+h] := [x+h, x]$.

Satz 14.3 (Partielle Integration) *Seien $f, g \in C^1(I)$ mit $I = [a,b]$. Dann gilt*

$$\int_a^b fg' = \big[f(x)\,g(x)\big]_{x=a}^{x=b} - \int_a^b f'g.$$

Beweis: Es gilt nach der Produktregel

$$(fg)' = f'g + fg' \quad \text{auf } I = [a,b].$$

Folgerung 14.2 liefert die Behauptung. □

Satz 14.4 (Substitutions- oder Transformationsregel) *Sei $I = [a,b]$, $I^* = [\alpha,\beta]$ und $\varphi \in C^1(I)$ mit $\varphi(I) \subset I^*$. Dann gilt für $f \in C^0(I^*)$*

$$\int_{\varphi(a)}^{\varphi(b)} f(y)\,dy = \int_a^b f\big(\varphi(x)\big)\varphi'(x)\,dx.$$

Beweis: Wähle nach Satz 14.2 eine Stammfunktion $F \in C^1(I^*)$ von f. Dann gilt

$$\begin{aligned}
\int_{\varphi(a)}^{\varphi(b)} f(y)\,dy &= \big[F(y)\big]_{y=\varphi(a)}^{y=\varphi(b)} \quad \text{(Folgerung 14.2)}\\
&= \big[F\big(\varphi(x)\big)\big]_{x=a}^{x=b}\\
&= \int_a^b (F\circ\varphi)'(x)\,dx \quad \text{(Folgerung 14.2)}\\
&= \int_a^b F'\big(\varphi(x)\big)\varphi'(x)\,dx \quad \text{(Kettenregel)}\\
&= \int_a^b f\big(\varphi(x)\big)\,\varphi'(x)\,dx.
\end{aligned}$$

□

Für die Anwendung der Substitutionsregel ist folgendes *Kochrezept* nützlich: bei einem gegebenen Integral $\int_\alpha^\beta f(y)\,dy$ möchten wir $y = y(x)$ substituieren. Berechne dazu

$$y = y(x) \quad \Rightarrow \quad dy = y'(x)\,dx.$$

Bestimme die neuen Intervallgrenzen a, b durch Auflösen der Gleichungen $\alpha = y(a)$, $\beta = y(b)$. Es ergibt sich die Substitutionsformel

$$\int_\alpha^\beta f(y)\,dy = \int_a^b f\big(y(x)\big)\,y'(x)\,dx.$$

Wir zeigen nun exemplarisch, wie die Integrationsregeln angewandt werden. Zunächst ergeben sich direkt Integrationsformeln, wenn die Stammfunktion bekannt ist.

Beispiel 14.1

$$\begin{aligned}
\int_a^b x^\alpha\,dx &= \left[\frac{x^{\alpha+1}}{\alpha+1}\right]_{x=a}^{x=b} \quad (\alpha \in \mathbb{R}\backslash\{-1\},\ a,b>0),\\
\int_a^b \frac{dx}{x} &= \big[\log x\big]_{x=a}^{x=b} \quad (a,b>0),\\
\int_a^b \frac{dx}{\sqrt{1-x^2}} &= \big[\arcsin x\big]_{x=a}^{x=b}, \quad (-1<a,b<1).
\end{aligned}$$

Ein Beispiel für die Anwendung der partiellen Integration ist

Beispiel 14.2

$$\int_1^x \log u\,du = \int_1^x 1\cdot \log u\,du = \big[u\log u\big]_{u=1}^{u=x} - \int_1^x u\frac{1}{u}\,du = x\log x - (x-1).$$

Eine schöne Anwendung von Satz 14.3 ist das

Beispiel 14.3 (Wallis-Produkt) Wir berechnen hier $A_n = \int_0^{\pi/2} \sin^n x\,dx$. Offenbar gilt

$$A_0 = \pi/2 \quad \text{und } A_1 = \big[-\cos x\big]_{x=0}^{x=\pi/2} = 1.$$

Für $n \geq 1$ leiten wir durch partielle Integration eine Rekursionsformel her:

$$\begin{aligned} A_{n+1} &= \int_0^{\pi/2} \sin x\,\sin^n x\,dx \\ &= \big[\underbrace{-\cos x\,\sin^n x}_{=0}\big]_{x=0}^{x=\pi/2} + n\int_0^{\pi/2} \cos^2 x\,\sin^{n-1} x\,dx \\ &= n\int_0^{\pi/2} \sin^{n-1} x\,dx - n\int_0^{\pi/2} \sin^{n+1} x\,dx, \end{aligned}$$

wobei wir $\cos^2 x = 1 - \sin^2 x$ benutzt haben. Es folgt

$$A_{n+1} = \frac{n}{n+1} A_{n-1} \quad (n \geq 1).$$

Durch Induktion erhalten wir

$$\begin{aligned} A_{2n} &= \frac{2n-1}{2n}\cdot\frac{2n-3}{2n-2}\cdot\ldots\cdot\frac{1}{2}\cdot A_0 = \left(\prod_{k=1}^n \frac{2k-1}{2k}\right)\frac{\pi}{2}, \\ A_{2n+1} &= \frac{2n}{2n+1}\cdot\frac{2n-2}{2n-1}\cdot\ldots\cdot\frac{2}{3}\cdot A_1 = \left(\prod_{k=1}^n \frac{2k}{2k+1}\right)1. \end{aligned}$$

Es folgt weiter

$$\frac{A_{2n}}{A_{2n+1}} = \frac{\pi}{2}\cdot\prod_{k=1}^n \frac{4k^2-1}{4k^2}.$$

Nun gilt $A_{n+1} = \int_0^{\pi/2} \sin^{n+1} x\,dx \leq \int_0^{\pi/2} \sin^n x\,dx = A_n$, und es folgt

$$1 \leq \frac{A_n}{A_{n+1}} \leq \frac{A_{n-1}}{A_{n+1}} = \frac{n+1}{n} \longrightarrow 1.$$

Daraus ergibt sich die Produktdarstellung von Wallis

$$\frac{\pi}{2} = \prod_{k=1}^\infty \frac{4k^2}{4k^2-1} = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\ldots$$

Als nächstes behandeln wir Beispiele zur Substitutionsregel.

Beispiel 14.4 (Lineare Parameterwechsel) Mit der Substitution $y = (x - x_0)/m$ haben wir $dy = 1/m\, dx$ und $x = x_0 + my$, also neue Grenzen $a = x_0 + m\alpha$ und $b = x_0 + m\beta$. Es folgt

$$\int_\alpha^\beta f(y)\, dy = \frac{1}{m} \int_{x_0+m\alpha}^{x_0+m\beta} f\Big(\frac{x - x_0}{m}\Big)\, dx.$$

Beispiel 14.5 (Integration von Ableitungen)

$$\begin{aligned}
\int_a^b \frac{f'(x)}{f(x)}\, dx &= \int_a^b (\log f)'(x)\, dx = \big[\log f(x)\big]_{x=a}^{x=b} \quad (f > 0),\\
\int_a^b u\sqrt{1+u^2}\, du &= \frac{1}{3}\int_a^b \big((1+u^2)^{\frac{3}{2}}\big)'\, du = \Big[\frac{1}{3}(1+u^2)^{\frac{3}{2}}\Big]_{u=a}^{u=b}.\\
\int_a^b F'\big(f(x)\big) f'(x)\, dx &= \big[F\big(f(x)\big)\big]_{x=a}^{x=b}.
\end{aligned}$$

Beispiel 14.6 (Flächeninhalt unter Hyperbel) Zu berechnen ist das Integral

$$A(x) = \int_1^x \sqrt{u^2 - 1}\, du \qquad \text{für } x \geq 1.$$

Hierfür brauchen wir die hyperbolischen Funktionen

$$\cosh(t) = \frac{e^t + e^{-t}}{2} \qquad \text{und} \qquad \sinh(t) = \frac{e^t - e^{-t}}{2}.$$

Aus den Definitionen folgt leicht

$$\cosh'(t) = \sinh(t) \quad \sinh'(t) = \cosh(t) \quad \cosh^2(t) - \sinh^2(t) = 1.$$

Man sieht leicht $\sinh(t) > 0$ für $t > 0$, also ist $\cosh(t)$ streng monoton wachsend auf $[0, \infty)$ und hat die Umkehrfunktion Arcosh $: [1, \infty) \to [0, \infty)$. Wir substituieren nun $u = \cosh t$, also $du = \sinh t\, dt$, und erhalten

$$\begin{aligned}
A(x) &= \int_0^{\text{Arcosh } x} \sinh^2 t\, dt\\
&= \frac{1}{4}\int_0^{\text{Arcosh } x} (e^{2t} + e^{-2t} - 2)\, dt\\
&= \Big[\frac{1}{8}(e^{2t} - e^{-2t})\Big]_{t=0}^{t\,=\,\text{Arcosh } x} - \frac{1}{2}\text{Arcosh } x\\
&= \frac{1}{2}\Bigg(\Big[\frac{e^t + e^{-t}}{2}\,\frac{e^t - e^{-t}}{2}\Big]_{t=0}^{t\,=\,\text{Arcosh } x} - \text{Arcosh } x\Bigg)\\
&= \frac{1}{2}\big(x\sqrt{x^2 - 1} - \text{Arcosh } x\big).
\end{aligned}$$

Für rationale Funktionen, also Quotienten von Polynomen, hat man ein spezielles Integrationsverfahren, die Partialbruchzerlegung, die wir hier nur an einem Beispiel vorführen:

Beispiel 14.7 (Partialbruchzerlegung) Um das Integral $\int_{-1/2}^{1/2} \frac{dx}{1-x^2}$ zu berechnen, machen wir den Ansatz

$$\frac{1}{1-x^2} = \frac{1}{(1+x)(1-x)} \overset{!}{=} \frac{A}{1-x} + \frac{B}{1+x} = \frac{(A-B)x + (A+B)}{1-x^2}.$$

Der Koeffizientenvergleich ergibt $A = B = 1/2$, also folgt

$$\int_{-1/2}^{1/2} \frac{dx}{1-x^2} = \int_{-1/2}^{1/2} \left(\frac{1/2}{1-x} + \frac{1/2}{1+x} \right) dx = \frac{1}{2} \left[\log \frac{1+x}{1-x} \right]_{x=-\frac{1}{2}}^{x=\frac{1}{2}} = \frac{1}{2}(\log 3 - \log 1/3) = \log 3.$$

Bei der Definition des Riemannschen Integrals ist das Definitionsintervall $I = [a,b]$ nach Voraussetzung kompakt. Wir wollen ganz kurz erläutern, wie auch unendliche Integrationsintervalle $-\infty \leq a < b \leq \infty$ und in den Intervallgrenzen unbeschränkte Funktionen im Rahmen des Riemann-Integrals behandelt werden können.

Definition 14.2 (uneigentliches Riemann-Integral) *Sei $I = [a,b)$ mit $a < b \leq \infty$. Die Funktion $f : I \to \mathbb{R}$ sei Riemann-integrierbar auf $[a,b']$ für alle $b' < b$. Falls $\lim_{x \nearrow b} \int_a^x f(\xi)\, d\xi$ existiert, so heißt das Integral $\int_a^b f(\xi)\, d\xi$ konvergent (oder existent) und wir setzen*

$$\int_a^b f(\xi)\, d\xi = \lim_{x \nearrow b} \int_a^x f(\xi)\, d\xi.$$

Die Konvergenzaussagen für das uneigentliche Integral sind analog zu den Konvergenzkriterien für Reihen. Die folgende Aussage entspricht dem Cauchykriterium.

Satz 14.5 *In der Situation von Definition 14.2 ist $\int_a^b f(x)\, dx$ genau dann konvergent, wenn es zu jedem $\varepsilon > 0$ ein $b' < b$ gibt mit*

$$\left| \int_{x_1}^{x_2} f(x)\, dx \right| < \varepsilon \quad \text{für alle } x_1, x_2 \in (b', b). \tag{14.3}$$

Beweis: Setze $F(x) = \int_a^x f(\xi)\, d\xi$ für $x \in [a,b)$. Existiert das uneigentliche Integral, das heißt $F(x) \to S$ mit $x \nearrow b$, so gibt es ein $b' < b$ mit $|F(x) - S| < \varepsilon/2$ für alle $x \in (b', b)$. Es folgt

$$|F(x_1) - F(x_2)| \leq |F(x_1) - S| + |F(x_2) - S| < \varepsilon \quad \text{für } x_{1,2} \in (b', b).$$

Jetzt folgern wir umgekehrt aus (14.3) die Existenz des Grenzwerts $\lim_{x \nearrow b} F(x)$. Sei $x_k \in [a,b)$, $x_k \to b$, eine Folge. Zu $\varepsilon > 0$ sei $b' < b$ wie in (14.3) gewählt. Dann folgt

$$|F(x_l) - F(x_k)| = \left| \int_{x_k}^{x_l} f(x)\, dx \right| < \varepsilon \quad \text{für } x_k, x_l \in (b', b).$$

Da die Folge beliebig war, existiert der Grenzwert $\lim_{x \nearrow b} F(x)$. □

Hier sind einige Beispiele von uneigentlichen Riemann-Integralen.

Beispiel 14.8

$$\int_1^\infty x^\alpha\,dx \quad = \quad \begin{cases} \lim_{R\nearrow\infty}\left[\dfrac{x^{\alpha+1}}{\alpha+1}\right]_{x=1}^{x=R} = -\dfrac{1}{\alpha+1} & \text{für } \alpha < -1, \\ \text{divergent} & \text{für } \alpha \geq -1. \end{cases}$$

$$\int_0^1 x^\alpha\,dx \quad = \quad \begin{cases} \text{divergent} & \text{für } \alpha \leq -1, \\ \lim_{\varepsilon\searrow 0}\left[\dfrac{x^{\alpha+1}}{\alpha+1}\right]_{x=\varepsilon}^{x=1} = \dfrac{1}{\alpha+1} & \text{für } \alpha > -1. \end{cases}$$

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} \quad = \quad \lim_{x\nearrow 1}\int_0^x \frac{d\xi}{\sqrt{1-\xi^2}} = \lim_{x\nearrow 1}\arcsin x = \pi/2.$$

In den vorangegangenen Beispielen sind die Integranden positiv. Das uneigentliche Integral $\int_a^b f(x)\,dx$ heißt absolut konvergent, wenn $\int_a^b |f(x)|\,dx$ konvergiert. Aus Satz 14.5 folgt sofort, dass ein absolut konvergentes Integral konvergiert. Wie bei Reihen impliziert die Konvergenz aber nicht umgekehrt die absolute Konvergenz. Ein simples Beispiel ist das uneigentliche Integral $\int_1^\infty f(x)\,dx$, wobei

$$f(x) = \frac{(-1)^{k-1}}{k} \text{ für } k \leq x < k+1.$$

Für die Existenz des uneigentlichen Integrals auf einem beidseitig offenen Intervall (a,b) wählt man einen Zwischenpunkt $c \in (a,b)$ und verlangt die Existenz der Integrale auf $(a,c]$ und $[c,b)$. Das Integral über (a,b) ergibt sich dann als Summe. Es ist leicht zu sehen, dass diese Definition nicht von der Wahl des Zwischenpunkts c abhängt: für $a < c_1 < c_2 < b$ gilt

$$\int_{c_1}^x f(\xi)\,d\xi = \int_{c_1}^{c_2} f(\xi)\,d\xi + \int_{c_2}^x f(\xi)\,d\xi.$$

Mit $x \nearrow b$ ergibt sich, wobei die Konvergenz auf einer Seite die auf der anderen impliziert,

$$\int_{c_1}^b f(\xi)\,d\xi = \int_{c_1}^{c_2} f(\xi)\,d\xi + \int_{c_2}^b f(\xi)\,d\xi,$$

Analog sieht man

$$\int_a^{c_2} f(\xi)\,d\xi = \int_a^{c_1} f(\xi)\,d\xi + \int_{c_1}^{c_2} f(\xi)\,d\xi,$$

und durch Subtraktion folgt

$$\int_a^{c_1} f(\xi)\,d\xi + \int_{c_1}^b f(\xi)\,d\xi = \int_a^{c_2} f(\xi)\,d\xi + \int_{c_2}^b f(\xi)\,d\xi.$$

Ein Beispiel ist, mit $c = 0$,

$$\int_{-\infty}^\infty \frac{dx}{1+x^2} = 2\lim_{b\nearrow\infty}\int_0^b \frac{dx}{1+x^2} = 2\lim_{b\nearrow\infty}\big[\arctan x\big]_{x=0}^{x=b} = \pi.$$