Analysis

Tammo tom Dieck

Mathematisches Institut
Georg-August-Universität

Version vom 4. Mai 2004

Inhaltsverzeichnis

1 Die reellen Zahlen

Die reellen Zahlen sind die Grundlage der Analysis. Der praktische Umgang mit den Zahlen ist bekannt. Um eine begriffliche Grundlage für die theoretische Entwicklung zu haben, wird das Rechensystem der reellen Zahlen durch Grundeigenschaften (Axiome) festgelegt. Es gibt drei Sorten von Axiomen: Die algebraischen Körperaxiome über das Rechnen; die Anordnungsaxiome über den Größenvergleich der Zahlen; das Vollständigkeitsaxiom. Das eigentliche Herzstück der Analysis ist das Vollständigkeitsaxiom. Es dient zur Entwicklung der Grenzwerttheorie. Ohne die Grenzwerttheorie liefert die Analysis keine praktisch verwertbaren Resultate. Wie in jedem wissenschaftlichen Gedankengebäude sind die konkreten Ergebnisse das Endresultat der Theorie.

Das Rechensystem der reellen Zahlen wird axiomatisch definiert als ein *vollständig angeordneter Körper*. Die Axiome werden für eine Menge $\mathbb{R}$ mit drei *Strukturdaten* $\alpha, \mu, \mathbb{P}$ formuliert:

$$\begin{array}{lll}
\alpha\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}, & (a,b)\mapsto a+b & \text{Addition}\\
\mu\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}, & (a,b)\mapsto ab = a\cdot b & \text{Multiplikation}\\
\mathbb{P}\subset\mathbb{R} & & \text{Positivbereich}
\end{array}$$

1 Die Körperaxiome

Eine Menge K zusammen mit zwei Strukturdaten $\alpha\colon (a,b)\mapsto a+b$ und $\mu\colon (a,b)\mapsto ab$ wie oben für $\mathbb{R}$ heißt *Körper*, wenn die folgenden *Körperaxiome* erfüllt sind.

(1)	$a+(b+c)=(a+b)+c$	(5)	$a(bc)=(ab)c$
(2)	$a+b=b+a$	(6)	$ab=ba$
(3)	Es gibt 0 mit $0+a=a$	(7)	Es gibt $1\neq 0$ mit $1\cdot a=a$
(4)	Zu a gibt es b mit $b+a=0$	(8)	Zu $a\neq 0$ gibt es b mit $ba=1$
(9)	$a(b+c)=ab+ac$		

(1) und (5) heißen Assoziativgesetz, (2) und (6) Kommutativgesetz, (9) Distributivgesetz. In (4) schreibt man $b=-a$. In (8) schreibt man $b=a^{-1}=\frac{1}{a}=1/a$. $\diamond$

Der Umgang mit dem Minuszeichen, mit Klammern, mit den Regeln der Bruchrechnung, Punktrechnung vor Strichrechnung und dergleichen ist bekannt. Üblicherweise schreibt man für

$$a+(-b),\ a+(b+c),\ a\cdot b^{-1},\ a(bc),\ 1+1,\ a+a,\ aa,\ \dots$$

der Reihe nach

$$a-b,\ a+b+c,\ a/b=\frac{a}{b},\ abc,\ 2,\ 2a,\ a^2,\ \dots$$

Ohne weitere Axiome kann man bekanntlich Vorkommnisse wie $2=0$ oder $1+1+1=3=0$ nicht ausschließen.

2 Anordnung

Sei (K, α, μ) ein Körper und $\mathbb{P} \subset K$. Wir nennen $(K, \alpha, \mu, \mathbb{P})$ einen *geordneten Körper*, wenn die folgenden *Anordnungsaxiome* erfüllt sind.

(1) $\mathbb{P} + \mathbb{P} \subset \mathbb{P}$.

(2) $\mathbb{P} \cdot \mathbb{P} \subset \mathbb{P}$.

(3) $K = \mathbb{P} \cup \{0\} \cup -\mathbb{P}$ ist eine disjunkte Zerlegung.

Darin haben wir für eine Menge $P \subset K$ gesetzt $P + P = \{a + b \mid a, b \in P\}$, $P \cdot P = \{ab \mid a, b \in P\}$, $-P = \{-a \mid a \in P\}$. ◇

Mit diesen Daten definieren wir $a < b$ als $b - a \in \mathbb{P}$ und $a \leq b$ als $b - a \in \mathbb{P} \cup \{0\}$. Die Menge K ist *geordnet*, das heißt, die Relation $<$ auf K genügt den Regeln:

(1) Sind x und y aus K, so gilt genau eine der Aussagen $x < y$, $x = y$, $y < x$.

(2) Aus $x < y$ und $y < z$ folgt $x < z$.

3 Anordnung und Rechnen

Sei K ein geordneter Körper. Aus den Körper- und Anordnungsaxiomen leitet man eine Reihe von gebräuchlichen Regeln her:

(3.1) Rechenregeln der Anordnung. *Für die Anordnungszeichen $<$ und $>$ sowie $\leq$ und $\geq$ gilt:*

$$
\begin{array}{rcl@{\qquad\qquad}rcl}
a < b,\ b < c & \Rightarrow & a < c & a \leq b,\ b \leq c & \Rightarrow & a \leq c \\
a < b & \Rightarrow & a + c < b + c & a \leq b & \Rightarrow & a + c \leq b + c \\
a < b,\ c > 0 & \Rightarrow & ac < bc & a \leq b,\ c \geq 0 & \Rightarrow & ac \leq bc \\
a < b,\ c < 0 & \Rightarrow & ac > bc & a \leq b,\ c \leq 0 & \Rightarrow & ac \geq bc
\end{array}
$$

(Wir benutzen hier und weiterhin $\Rightarrow$ als Symbol für die logische Folgerung.) □

Im Zusammenhang mit den Zeichen $=$, $<$, $\leq$ hat sich eine Sprechweise eingebürgert: Eine irgendwie geartete Relation der Form $x = y$ heißt *Gleichung*, eine der Form $x < y$ oder $x \leq y$ *Ungleichung*. Anwendung der eben aufgeführten Rechenregeln bezeichnet man dann als *Rechnen mit Ungleichungen*.

(3.2) Notiz. *Aus den Anordnungsregeln folgt:*

(1) *Sei $ab > 0$. Dann gilt: $a < b \Leftrightarrow a^{-1} > b^{-1}$.*

(2) *Sei $x \neq 0$. Dann ist $x^2 > 0$. Insbesondere ist $0 < 1$.*

(3) *Seien a und b positiv. Für jede natürliche Zahl n gilt: $a < b \Leftrightarrow a^n < b^n$.* □

Sind $a_1, \dots, a_n$ reelle Zahlen, so wird der Zahlenwert der größten darunter das *Maximum* $\max(a_1, \dots, a_n)$, der Wert der kleinsten darunter das *Minimum* $\min(a_1, \dots, a_n)$ von $(a_1, \dots, a_n)$ genannt.

4 Der Betrag

Sei K ein angeordneter Körper. Der *(absolute) Betrag* $|a|$ einer Zahl $a \in K$ wird durch $|a| = \max(a, -a)$ erklärt. Aus der Definition verifiziert man:

(4.1) Regeln über den Betrag.

(1) $|a| \geq 0$. $|a| = 0 \Leftrightarrow a = 0$.

(2) $|a + b| \leq |a| + |b|$.

(3) $|ab| = |a||b|$.

(4) $||b| - |a|| \leq |b - a|$.

(5) $|a| < \varepsilon \quad \Leftrightarrow \quad -\varepsilon < a < \varepsilon$.

(6) $\max(x, y) = \frac{1}{2}(x + y + |x - y|)$, $\qquad \min(x, y) = \frac{1}{2}(x + y - |x - y|)$.

Die Regel (2) heißt *Dreiecksungleichung*. Sie läßt sich durch vollständige Induktion verallgemeinern $|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n|$. Man nennt $|a - b|$ den *Abstand* der Zahlen (der Punkte) a und b. ◇

5 Natürliche, ganze und rationale Zahlen

In einem geordneten Körper K kann man ein Modell der natürlichen Zahlen finden. Eine Menge $M \subset K$ heiße *induktiv*, wenn gilt:

(1) $1 \in M$.

(2) $x \in M \;\Rightarrow\; x + 1 \in M$.

Es gibt induktive Teilmengen, zum Beispiel K oder $\mathbb{P}$. Als die Menge $\mathbb{N}$ der *natürlichen Zahlen* in K definieren wir den Durchschnitt aller induktiven Teilmengen. Das ist selbst eine induktive Teilmenge. Daraus definieren wir die Menge $\mathbb{Z}$ der *ganzen Zahlen* in K als $\mathbb{Z} = \{m - n \mid m, n \in \mathbb{N}\}$ und die Menge $\mathbb{Q}$ der *rationalen Zahlen* in K als $\mathbb{Q} = \{p/q \mid p \in \mathbb{Z}, q \in \mathbb{Z} \setminus \{0\}\}$.

6 Archimedisch geordnete Körper

Sei K ein geordneter Körper. Er heißt *archimedisch geordnet*, wenn zusätzlich das folgende Axiom gilt: Zu je zwei Elementen $a, b \in \mathbb{P}$ gibt es ein $n \in \mathbb{N}$ mit $b < na$. ◇

(6.1) Satz. *Der Körper K ist genau dann archimedisch geordnet, wenn $\mathbb{Q}$ dicht in K liegt, das heißt, zu je zwei verschiedenen Elementen in K gibt es eine rationale Zahl, die dazwischen liegt.* □

Sei im weiteren der Körper archimedisch geordnet. Einige einfache Konsequenzen der Axiome sind:

(6.2) Weitere Eigenschaften der Anordnung.

(1) Zu jeder positiven Zahl ε gibt es eine natürliche Zahl n, so daß $n^{-1} < \varepsilon$ ist.

(2) Zu jeder Zahl b gibt es genau eine ganze Zahl n, die den Ungleichungen $n \leq b < n + 1$ genügt. □

(6.3) Bernoullische Ungleichung. *Sei $x \geq -1$. Dann gilt für jede natürliche Zahl n die Ungleichung $(1+x)^n \geq 1+nx$.*

Beweis. Induktion nach n. Der Induktionsanfang ist klar. Der Induktionsschritt:

$$(1+x)^{n+1} = (1+x)^n(1+x) \geq (1+nx)(1+x) = 1+(n+1)x+x^2 \geq 1+(n+1)x.$$

In dieser Rechnung haben wir die Induktionsvoraussetzung benutzt, sowie die Ungleichungen $1+x \geq 0$ und $x^2 \geq 0$. □

(6.4) Wachstum der Potenzen.

(1) *Ist $b > 1$, so gibt es zu jeder reellen Zahl K eine natürliche Zahl n, so daß $b^n > K$ ist.*

(2) *Ist $0 < q < 1$, so gibt es zu jeder reellen Zahl $\varepsilon > 0$ eine natürliche Zahl n, so daß $q^n < \varepsilon$ ist.*

(3) *Sei $a > 1$ und $x > 0$. Dann gibt es genau eine ganze Zahl k, so daß $a^k \leq x < a^{k+1}$.*

Beweis. (1) Wir schreiben $b = 1+x$. Dann ist nach Voraussetzung $x > 0$. Die Bernoullische Ungleichung liefert $b^n \geq 1+nx$. Nach dem Archimedischen Prinzip gibt es ein n, so daß $n > K/x$, also $nx > K$ ist. Insgesamt folgt $b^n \geq 1+nx > nx > K$.

(2) folgt aus (1); man setze dazu $b = q^{-1}$ und $K = \varepsilon^{-1}$.

(3) folgt aus (1) und (2), weil $0 < a^{-1} < 1 < a$ ist. □

7 Das Vollständigkeitsaxiom

Die für die gesamte Analysis und insbesondere die Grenzwerttheorie fundamentale Eigenschaft der Anordnung wird hier als Grundeigenschaft, als Axiom vorausgesetzt. Sei K eine geordnete Menge. Ein Element $M \in K$ heißt *obere Schranke* der Teilmenge $S \subset K$, wenn für alle $s \in S$ die Ungleichung $s \leq M$ gilt. Ein Element m heißt *untere Schranke* von S, wenn für alle $s \in S$ die Ungleichung $m \leq s$ gilt. Wir sagen, S ist nach unten (oben) beschränkt, wenn es eine untere (obere) Schranke für S gibt. Gibt es beide Schranken, so nennen wir S *beschränkt*.

Ist M obere Schranke von S, so auch jedes größere Element. Deshalb ist es nur interessant, nach möglichst kleinen oberen Schranken zu suchen. Wir nennen M *kleinste obere Schranke* oder *Supremum* von S, wenn M obere Schranke ist und wenn für jede andere obere Schranke M' von S gilt $M \leq M'$. Analog wird der Begriff *größte untere Schranke* oder *Infimum* definiert. Wir verwenden die Bezeichnungen

$$M = \sup(S) \quad \text{bzw.} \quad m = \inf(S)$$

für das Supremum bzw. das Infimum von S. Gibt es $M \in S$, so daß für alle $s \in S$ die Ungleichung $s \leq M$ gilt, so heißt M *Maximum* von S, in Zeichen $M = \max(S)$. Gibt es $m \in S$, so daß für alle $s \in S$ die Ungleichung $m \leq s$

gilt, so heißt m *Minimum* von S, in Zeichen $m = \min(S)$. Unmittelbar aus den Definitionen verifiziert man:

(7.1) Notiz. *Hat S ein Maximum M (Minimum m), so ist $M = \sup(S)$ ($m = \inf(S)$). Ein Element M (m) ist genau dann Supremum (Infimum) von S, wenn M obere Schranke (m untere Schranke) von S ist und wenn für jedes $N < M$ ein $s \in S$ mit $N < s$ (wenn für jedes $m < n$ ein $s \in S$ mit $n < s$) existiert. Maximum, Minimum, Supremum, Infimum sind, wenn sie existieren, eindeutig durch die Menge bestimmt.* $\diamond$

(7.2) Vollständigkeitsaxiom. Jede nichtleere, nach oben beschränkte Menge $S \subset K$ hat ein Supremum. $\diamond$

Äquivalent dazu könnte man ein analoges Axiom über das Infimum postulieren. Wir zeigen:

(7.3) Notiz. *Erfüllt K das Supremumsaxiom* (7.2)*, so hat jede nach unten beschränkte nichtleere Menge $T \subset K$ ein Infimum.*

Beweis. Sei S die Menge der unteren Schranken von T. Jedes Element von T ist obere Schranke von S. Sei $m = \sup S$. Da m die kleinste obere Schranke von S ist, gilt $m \leq t$ für alle $t \in T$. Also ist m eine untere Schranke von T. Eine größere kann es aber nach Definition von S nicht geben. $\square$

(7.4) Satz. *Gilt für die Ordnung eines geordneten Körpers das Supremumsaxiom, so gilt auch das archimedische Axiom.*

Beweis. Seien x und y positive Zahlen. Angenommen, die Zahlen nx, $n\mathbb{N}$ seien alle kleinergleich y. Dann ist y eine obere Schranke dieser Menge. Sei s deren Supremum. Da $s - 1$ kleiner als s ist und also keine obere Schranke, gibt es $n \in \mathbb{N}$, so daß $s - 1 < nx$ ist. Es folgt $s < (n + 1)x$ im Widerspruch zur Definition von s. $\square$

2 Die Topologie der reellen Zahlen

Wir legen einen vollständig geordneten Körper zugrunde. Er heiße der Körper der reellen Zahlen $\mathbb{R}$. Die folgenden Definitionen können aber, soweit sie nur von der Ordnungsrelation Gebrauch machen, in jedem geordneten Körper oder sogar in jeder geordneten Menge gegeben werden.

1 Intervalle

In vielen Anwendungen und Formulierungen der Analysis ist es zweckmäßig, neben den reellen Zahlen noch die Symbole $\pm\infty$ zur Verfügung zu haben. Wir erweitern deshalb die Menge $\mathbb{R}$ durch zwei neue Elemente $\overline{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ und legen die Anordnungsregeln $-\infty < a < \infty$ für jede reelle Zahl a fest. Damit wird $\overline{\mathbb{R}}$ eine geordnete Menge. Zunächst rechnen wir mit den Symbolen $\pm\infty$ jedoch nicht. Für $\overline{\mathbb{R}}$ liefert das Vollständigkeitsaxiom: Jede nichtleere Menge $S \subset \overline{\mathbb{R}}$ hat ein Supremum und ein Infimum.

Ein *Intervall* ist eine Teilmenge von $\mathbb{R}$ oder $\overline{\mathbb{R}}$, die mit je zwei Elementen auch alle dazwischenliegenden enthält. Für Intervalle verwenden wir die folgenden Bezeichnungen:

$$\begin{aligned}
[a,b] &= \{x \mid a \leq x \leq b\} \\
]a,b[&= \{x \mid a < x < b\} \\
[a,b[&= \{x \mid a \leq x < b\} \\
]a,b] &= \{x \mid a < x \leq b\}
\end{aligned}$$

Die *Intervallgrenzen* a und b dürfen dabei auch $\pm\infty$ sein. So ist $[a,\infty[= \{x \mid a \leq x\}$ und $]-\infty,\infty[= \mathbb{R}$. Ein Intervall der Form $[a,b]$ heißt *abgeschlossenes* Intervall von a nach b, eines der Form $]a,b[$ ein *offenes* Intervall von a nach b. Der Fall $a = b$ ist nicht besonders interessant und werde stillschweigend ausgeschlossen (entartetes Intervall). Die Punkte a und b heißen auch die *Randpunkte* des Intervalls $[a,b]$; analog in den anderen Fällen. Die Randpunkte können also im Intervall liegen oder auch nicht.

2 Umgebungen

Mengen der Form $\{x \mid x < a\}$, $\{x \mid x > a\}$ und $\{x \mid a < x < b\}$ heißen *Basismengen* von $\overline{\mathbb{R}}$. Ein Durchschnitt von endlich vielen Basismengen ist wieder eine Basismenge. Wir nennen U eine *Umgebung* von x, wenn es eine Basismenge B gibt, die $x \in B \subset U$ erfüllt. Jede Obermenge einer Umgebung von x ist wieder eine Umgebung von x, und der Schnitt endlich vieler Umgebungen von x ist wieder eine Umgebung von x. Sei ε eine positive reelle Zahl. Die Menge

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

heißt ε-Umgebung von $a \in \mathbb{R}$. Genau dann ist U eine Umgebung von $a \in \mathbb{R}$, wenn es ein $\varepsilon > 0$ gibt, so daß $U_\varepsilon(a) \subset U$ ist. Die Mengen

$$U_M(\infty) = \{x \in \overline{\mathbb{R}} \mid x > M\} \quad \text{bzw.} \quad U_M(-\infty) = \{x \in \overline{\mathbb{R}} \mid x < M\}$$

heißen M-Umgebungen von ∞ bzw. $-\infty$. Genau dann ist U eine Umgebung von $\pm\infty$, wenn es ein $M \in \mathbb{R}$ so gibt, daß die zugehörige M-Umgebung in U liegt.

3 Offene und abgeschlossene Mengen

Eine Teilmenge $W \subset \overline{\mathbb{R}}$ heißt *offen*, wenn sie Umgebung jedes ihrer Punkte ist. Eine Basismenge ist offen. Wir nennen auch die leere Menge offen. Die Menge

$\mathcal{O}(\overline{\mathbb{R}})$ aller offenen Mengen von $\overline{\mathbb{R}}$ heißt die *Topologie* von $\overline{\mathbb{R}}$. Ist X eine beliebige Teilmenge von $\overline{\mathbb{R}}$, so nennen wir

$$\mathcal{O}(X) = \{U \subset X \mid \exists V \in \mathcal{O}(\overline{\mathbb{R}}) \text{ mit } U = V \cap X\}$$

die *Topologie* auf X oder genauer die von $\mathcal{O}(\overline{\mathbb{R}})$ auf X induzierte *Relativtopologie*. Die Mengen in $\mathcal{O}(X)$ heißen dann die offenen Mengen von X. Eine Menge $U \subset \mathbb{R}$ ist genau dann offen, wenn es zu jedem $u \in U$ ein $\varepsilon > 0$ gibt, so daß $U_\varepsilon(u) \subset U$ ist.

Mit dem Begriff der offenen Menge definieren wir den Begriff einer Umgebung für Punkte in beliebigen Teilmengen X von $\overline{\mathbb{R}}$. Wir nennen $U \subset X$ eine Umgebung von $x \in X$, wenn es eine offene Menge V von X mit $x \in V \subset U$ gibt.

(3.1) Notiz. *Das System der offenen Mengen von $X \subset \overline{\mathbb{R}}$ hat die folgende Eigenschaften:*

(1) *$\emptyset$ und X sind offen.*

(2) *Ist $(U_j \mid j \in J)$ eine Familie offener Mengen U_j, so ist auch deren Vereinigung offen.*

(3) *Ein Schnitt endlich vieler offener Mengen ist offen.* □

Eine Menge $A \subset X$ heißt *abgeschlossen* in X, wenn ihr Komplement $X \backslash A$ offen in X ist. Durch mengentheoretische Dualität erhalten wir aus den Eigenschaften offener Mengen:

(3.2) Notiz. *Das System der abgeschlossenen Mengen von X hat die folgenden Eigenschaften:*

(1) *$\emptyset$ und X sind abgeschlossen.*

(2) *Ist $(A_j \mid j \in J)$ eine Familie abgeschlossener Mengen A_j, so ist auch ihr Durchschnitt abgeschlossen.*

(3) *Eine Vereinigung endlich vieler abgeschlossener Menge ist abgeschlossen.* □

4 Kompakte Mengen

Eine Familie offener Menge $(U_j \mid j \in J)$ heißt offene *Überdeckung* von A, wenn A in der Vereinigung der U_j enthalten ist. Wir nennen A *kompakt*, wenn zu jeder offenen Überdeckung $(U_j \mid j \in J)$ ein endliches Teilsystem E der Indexmenge J existiert, so daß $(U_j \mid j \in E)$ immer noch eine Überdeckung von A ist. Kompaktheit ist das Äußerste an „Endlichkeit“, will sagen „Menschlichkeit“, was man der exorbitanten Unendlichkeit der Zahlen abgewinnen kann.

Die fundamentale Konsequenz aus dem Vollständigkeitsaxiom ist der folgende *Überdeckungssatz von Heine-Borel*:

(4.1) Satz. *Ein abgeschlossenes Intervall aus $\overline{\mathbb{R}}$ ist kompakt.*

Beweis. Sei $(U_j \mid j \in J)$ eine offene Überdeckung des Intervalls. Zu jedem x sei $U(x)$ eine Überdeckungsmenge, die x enthält. Sei $S \subset [a, b]$ die Menge der s, für die $[a, s]$ in der Vereinigung endlich vieler $U(x)$ enthalten ist. Wegen $a \in S$ ist S nichtleer. Sei $z = \sup(S)$. Wir behaupten $z = b \in S$. Wir wählen eine Zahl t in $]a, z[\cap U(z)$. Dann ist $a < t < z$, und deshalb ist, nach Definition von z, $[a, t]$ in der Vereinigung endlich vieler Mengen der Form $U(x)$ enthalten. Dann ist aber auch $[a, t] \cup U(z)$ in endlich vielen enthalten. Also liegt z in S. Ist $z < b$, so widerspricht das der Supremumseigenschaft von z. □

Wir betonen, daß in dem letzten Satz auch die Werte $\pm\infty$ als Intervallgrenzen zugelassen sind. Durch einfache mengentheoretische Umformungen gewinnen wir weitere kompakte Mengen.

(4.2) Satz. *Sei K kompakt und A abgeschlossen. Dann ist auch $A \cap K$ kompakt.*

Beweis. Sei $(U_j \mid j \in J)$ eine offene Überdeckung von $A \cap K$. Nehmen wir noch das Komplement U von A hinzu, so erhalten wir eine offene Überdeckung von K. Da K mit endlich vielen überdeckt wird, so auch $A \cap K$. □

(4.3) Satz. *Eine Menge $A \subset \mathbb{R}$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.*

Beweis. Ist A beschränkt, so liegt A in einem kompakten Intervall K. Da A abgeschlossen ist, so ist nach dem vorigen Satz $A = A \cap K$ kompakt.

Sei A kompakt. Die Mengen $]-x, x[$, $x \in \mathbb{R}$ bilden eine offene Überdeckung. Da A kompakt ist, so ist A in endlich vielen davon enthalten, also beschränkt. Wir zeigen ferner, daß das Komplement U von A offen ist. Sei $u \in U$. Zu jedem $x \in A$ wählen wir offene Mengen $U(x)$ und $V(x)$, so daß $u \in U(x)$, $x \in V(x)$ und $U(x) \cap V(x) = \emptyset$. Da A kompakt ist, wird A von endlich vielen der V-Umgebungen überdeckt, etwa von $V(x_1), \ldots, V_(x_n)$. Sei $U_x = \cap_{i=1}^n U(x_i)$. Dann ist U_x disjunkt zur Vereinigung der $V(x_j)$ und damit in U enthalten. Zu jedem Punkt von U gibt es also eine in U enthaltene Umgebung, das heißt, U ist offen. □

3 Stetigkeit. Grenzwert

Der Begriff der Stetigkeit ist grundlegend für die gesamte Analysis. Wir betrachten Funktionen $f\colon A \to \overline{\mathbb{R}}$ mit einem Definitionsbereich $A \subset \overline{\mathbb{R}}$. Etliche Hauptsätze über stetige Funktionen haben nichts mit dem Zahlenrechnen zu tun, sondern benutzen nur die Anordnung.

1 Die Hauptsätze über stetige Funktionen

Eine Funktion $f\colon A \to \overline{\mathbb{R}}$ heißt *stetig an der Stelle* $a \in A$ oder *bei* $a \in A$, wenn das Urbild jeder Umgebung von $f(a)$ bei f eine Umgebung von a in A ist. Sie heißt *stetig*, wenn sie an jedem Punkt ihres Definitionsbereichs stetig ist.

Durch logische Verneinung ergibt diese Definition: Eine Funktion $f\colon A \to \overline{\mathbb{R}}$ ist *unstetig* bei a, wenn es eine Umgebung U von $f(a)$ derart gibt, daß in jeder Umgebung V von a ein x mit $f(x) \notin U$ existiert.

Da jede Obermenge einer Umgebung wieder eine Umgebung ist, sind die folgenden Aussagen aus mengentheoretischen Gründen gleichwertig zur Definition der Stetigkeit im Punkt a:

(1) Das Urbild $f^{-1}(U)$ jeder Umgebung von $f(a)$ enthält eine Umgebung von a in A.

(2) Das Urbild jeder offenen Umgebung von $f(a)$ enthält eine offene Umgebung von a in A.

(3) Falls $f(a) \in \mathbb{R}$ und $a \in \mathbb{R}$: Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß für $|x - a| < \delta$ die Ungleichung $|f(x) - f(a)| < \varepsilon$ gilt. Wir nennen dann (ε, δ) ein *Stetigkeitspaar* (= SP) für f bei a.

Ist (ε, δ) ein Stetigkeitspaar für f bei a so auch (ε', δ) und (ε, δ'), wenn $\varepsilon' \geq \varepsilon$ und $\delta \geq \delta'$. Für den Nachweis der Stetigkeit spielen deshalb nur kleine ε eine Rolle; man kann also in den Untersuchungen ε von vornherein gewissen Beschränkungen unterwerfen. In den Anwendungen schreiben wir x manchmal in der Form $x = a + h$; im Fall (3) ist dann (ε, δ) ein SP, wenn aus $|h| < \delta$ immer $|f(a + h) - f(a)| < \varepsilon$ folgt.

Die Definition liefert unmittelbar: Ist $f\colon B \to \overline{\mathbb{R}}$ bei $a \in B$ stetig und gilt $a \in A \subset B$, so ist auch die Einschränkung $f|A$ von f auf A bei a stetig. Ist $A \subset B$ und ist $i\colon A \to B, x \mapsto x$ die Inklusion, aufgefaßt als Abbildung, so ist i stetig.

(1.1) Satz. *Eine Funktion ist genau dann stetig, wenn das Urbild jeder offenen Menge offen ist.*

Beweis. Sei die Funktion $f\colon A \to B$ stetig. Sei $U \subset B$ offen. Dann ist U eine Umgebung aller $u \in U$ und folglich, nach Definition der Stetigkeit, $V = f^{-1}(U)$ eine Umgebung aller Punkte $v \in V$. Die Umkehrung ist nach der Äquivalenz (2) oben klar. □

(1.2) Satz. *Sei $f\colon A \to B$ bei a und $g\colon B \to C$ bei $b = f(a)$ stetig. Dann ist $g \circ f\colon A \to C$ bei a stetig.*

Beweis. Sei U eine Umgebung von $(g \circ f)(a) = g(f(a))$. Dann ist $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$ eine Umgebung von a. □

(1.3) Schrankensatz. *Sei $f\colon A \to \overline{\mathbb{R}}$ bei a stetig. Sei $c \in \overline{\mathbb{R}}$. Gibt es in jeder Umgebung U von a ein $u \in U$ mit $c \leq f(u)$, so ist $c \leq f(a)$. Analog, wenn man $\leq$ beidemal durch $\geq$ oder durch $=$ ersetzt.*

Beweis. Angenommen $f(a) < c$. Dann ist $f^{-1}([-\infty, c[\,)$ eine Umgebung von a, in der es kein u mit $c \leq f(u)$ gibt. Widerspruch. □

(1.4) Eindeutigkeitsatz. *Sei $f\colon A \to \overline{\mathbb{R}}$ bei p stetig. Gibt es in jeder Umgebung U von p Stellen $u, v \in U$ mit $f(u) \leq c \leq f(v)$, so ist $f(p) = c$. Der Funktionswert* *an der Stelle p einer bei p stetigen Funktion ist also durch die Einschränkung $f|A \setminus \{p\}$ bestimmt.*

Beweis. Das ist eine direkte Folgerung aus (1.3). □

(1.5) Zwischenwertsatz. *Sei $f\colon [a,b] \to \overline{\mathbb{R}}$ stetig. Dann nimmt f jeden Wert zwischen $f(a)$ und $f(b)$ an.*

Beweis. Sei $f(a) < c < f(b)$ und sei $p \in [a,b]$ das Supremum der $x \in [a,b]$ mit $f(x) \le c$. Nach (1.4) ist $f(p) = c$. □

(1.6) Minimaxsatz. *Eine stetige Funktion $f\colon [a,b] \to \overline{\mathbb{R}}$ hat einen minimalen und einen maximalen Funktionswert.*

Beweis. Sei $M = \sup f([a,b])$ das Supremum der Menge der Funktionswerte. Ist M kein Funktionswert, so bilden die Mengen $f^{-1}([-\infty, c[$ für $c < M$ nach (1.1) eine offene Überdeckung von $[a,b]$, die keine endliche Teilüberdeckung hat, im Widerspruch zum Überdeckungssatz von Heine-Borel. □

(1.7) Intervallsatz. *Das Bild eines Intervalles bei einer stetigen Funktion ist wieder ein Intervall. (Dabei lassen wir auch $\pm\infty$ als Intervallgrenzen zu.) Ist das Intervall abgeschlossen, so auch das Bild.*

Beweis. Ein Intervall ist eine Teilmenge, die mit jedem Punkt auch alle dazwischenliegenden enthält. Die Behauptung folgt also direkt aus dem Zwischenwertsatz. Die Aussage über die Abgeschlossenheit benutzt außerdem den Minimaxsatz. □

Eine Abbildung $f\colon A \to B$ heißt *injektiv*, wenn aus $a_1 \ne a_2$ immer $f(a_1) \ne f(a_2)$ folgt, und *surjektiv*, wenn zu jedem $b \in B$ ein $a \in A$ mit $b = f(a)$ existiert. Sie heißt *bijektiv*, wenn sie sowohl surjektiv als auch injektiv ist.

Sei $f\colon X \to Y$ eine Funktion. Eine *Umkehrfunktion* von f ist eine Funktion $g\colon Y \to X$ (in der „umgekehrten Richtung“), so daß für alle $x \in X$ und alle $y \in Y$ die Gleichungen $g(f(x)) = x$ und $f(g(y)) = y$ gelten. Eine Funktion $f\colon X \to Y$ hat genau dann eine Umkehrfunktion, wenn sie bijektiv ist. Ist g ein *Linksinverses* von f, das heißt gilt nur $g(f(x)) = x$ für alle x, so ist f injektiv und g surjektiv.

Sei $X \subset \overline{\mathbb{R}}$. Eine Funktion $f\colon X \to \overline{\mathbb{R}}$ heißt *streng monoton wachsend* (bzw. *streng monoton fallend*), wenn aus $a < b$ immer $f(a) < f(b)$ (bzw. $f(a) > f(b)$) folgt. Eine streng monoton wachsende oder fallende Funktion ist also injektiv. Wählen wir für eine solche Funktion Y als die Menge ihrer Funktionswerte, so hat $f\colon X \to Y$ eine Umkehrfunktion. Hat eine streng monoton wachsende Funktion f eine Umkehrfunktion g, so ist auch g streng monoton wachsend. Impliziert $a < b$ nur $f(a) \le f(b)$, so heißt f monoton wachsend.

(1.8) Monotoniesatz. *Sei $A \subset \overline{\mathbb{R}}$ ein Intervall. Sei $f\colon A \to \overline{\mathbb{R}}$ injektiv und stetig. Dann ist f streng monoton.*

Beweis. Seien x, y aus A und sei $x < y$. Da f injektiv ist, gilt entweder $f(x) < f(y)$ oder $f(x) > f(y)$. Angenommen $f(x) < f(y)$. Liege z zwischen x und y. Da

f injektiv ist, so ist $f(z)$ von $f(x)$ und $f(y)$ verschieden. Wäre $f(z) > f(y)$, so würde nach dem Zwischenwertsatz für ein $x < c < z$ die Gleichung $f(c) = f(y)$ bestehen, im Widerspruch zur Injektivität. Also ist $f(z) < f(y)$; ebenso schließt man auf $f(x) < f(z)$. Für $y < z$ muß $f(y) < f(z)$ sein, da nach dem schon Bewiesenen $f(x) < f(y) < f(z)$ gelten muß. Für $z < x$ folgt ebenso $f(z) < f(x)$. Sei nun a, b aus A ein beliebiges Paar mit $a < b$. Indem wir das schon Bewiesene auf die noch möglichen Fälle $a < b < x$, $a < x < b$ und $x < a < b$ geeignet anwenden, folgt $f(a) < f(b)$. □

(1.9) Umkehrsatz. *Sei $f\colon A \to \overline{\mathbb{R}}$ eine stetige streng monotone Funktion, die auf dem Intervall $A \subset \overline{\mathbb{R}}$ definiert ist. Sei B das Bildintervall von f. Dann ist die Umkehrabbildung $g\colon B \to A$ stetig.*

Beweis. Wir bemerken, daß eine streng monotone Funktion eine bijektive Abbildung auf ihr Bild liefert. Sei A offen. Dann ist auch B offen, da f streng monoton ist. Für den Nachweis der Stetigkeit von g bei $b = f(a)$ haben wir zu zeigen: Sei $W \subset A$ ein offenes Intervall um $a = g(b)$; dann enthält $g^{-1}(W)$ ein offenes Intervall um b. Nun ist aber $g^{-1}(W) = f(W)$ ein offenes Intervall, weil f stetig und streng monoton ist. Falls A nicht offen ist, so muß man die Randpunkte nach demselben Verfahren behandeln. □

Der letzte unserer Hauptsätze benutzt den Zusammenhang zwischen den Anordnungs- und den Körperaxiomen, insbesondere den Betrag und die Dreiecksungleichung.

Sei $A \subset \mathbb{R}$. Eine Funktion $f\colon A \to \mathbb{R}$ heißt *gleichmäßig stetig*, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß aus $|x - y| < \delta$, $x, y \in A$ immer die Abschätzung $|f(x) - f(y)| < \varepsilon$ folgt.

(1.10) Gleichmäßigkeitssatz. *Eine stetige Funktion $f\colon [a, b] \to \mathbb{R}$ auf einem abgeschlossenen Intervall ist gleichmäßig stetig.*

Beweis. Sei $\varepsilon > 0$ gegeben. Zu jedem $c \in [a, b]$ wählen wir ein Stetigkeitspaar der Form $(\varepsilon/2, 2\delta(c))$ für f bei c. Nach dem Satz von Heine-Borel ist $[a, b]$ in endlich vielen Mengen $U(c) = U_{\delta(c)}(c)$ enthalten, etwa in $U(c_1), \ldots, U(c_n)$. Sei δ das Minimum der $\delta(c_j)$. Sei $|x - y| < \delta$. Da $x \in [a, b] \subset \bigcup_{i=1}^n U(c_i)$, gibt es ein k mit $x \in U(c_k)$, das heißt $|x - c_k| < \delta(c_k)$. Es folgt $|y - c_k| \le |y - x| + |x - c_k| < \delta + \delta(c_k) \le 2\delta(c_k)$. Es folgt

$$|f(x) - f(y)| \le |f(x) - f(c_k)| + |f(y) - f(c_k)| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

Damit ist die gleichmäßige Stetigkeit gezeigt. □

2 Grenzwerte

Ein Element $b \in \overline{\mathbb{R}}$ heißt *Häufungspunkt* (= HP) von $A \subset \overline{\mathbb{R}}$, wenn jede Umgebung von b ein von b verschiedenes Element von A enthält. Ein Punkt $a \in A$, der nicht HP von A ist, heißt *isolierter* Punkt von A.

(2.1) Beispiele. Jeder Punkt eines nichtentarteten Intervalls A ist Häufungspunkt von A. Jede reelle Zahl ist Häufungspunkt von $\mathbb{Q}$, siehe I(5.4.3). Die Elemente $\pm\infty$ sind HPe von $\mathbb{R}$ und $\mathbb{Z}$. Eine Funktion $f\colon A \to \overline{\mathbb{R}}$ ist in einem isolierten Punkt von A immer stetig. ◇

Sei $f\colon A \to \overline{\mathbb{R}}$ eine Funktion und a Häufungspunkt von A. Das Symbol

$$\lim_{x\to a} f(x) = c$$

ist eine Abkürzung für die folgende Aussage: Zu jeder Umgebung U von c gibt es eine Umgebung V von a, so daß für alle $x \in V \cap (A \setminus a)$ die Relation $f(x) \in U$ gilt. Das Symbol wird gelesen: c ist *Limes* (*Grenzwert*) von f für (die Annäherung von) x gegen a. Wir sagen, $\lim_{x\to a} f(x)$ *existiert* (oder: der Limes von $f(x)$ für x gegen a existiert; oder: $f(x) \to c$ für $x \to a$), wenn es ein c gibt, so daß $\lim_{x\to a} f(x) = c$ gilt. Nach unseren Vereinbarungen ist übrigens auch $a = \pm\infty$ und $c = \pm\infty$ in dieser Definition zugelassen. Aus (1.5) und (2.3) folgt, daß der Limes c, falls er existiert, durch f eindeutig bestimmt ist.

Es ist also $\lim_{x\to a} f(x) = c$ zunächst keine Gleichung zwischen Zahlen, sondern eine Kurzform für eine Aussage. Falls allerdings die Aussage richtig ist und $c \in \mathbb{R}$, so betrachten wir auch $\lim_{x\to a} f(x)$ als die Zahl c selbst.

Wir definieren aus f eine Hilfsfunktion $f^{\#}\colon A \cup \{a\} \to \mathbb{R}$ durch die Vorschrift $f^{\#}(x) = f(x)$ für $x \neq a$ und $f^{\#}(a) = c$. Durch Umschreiben der Definitionen erhält man:

(2.2) Notiz. *Folgende Aussagen sind gleichwertig:*
(1) $\lim_{x\to a} f(x) = c$.
(2) $f^{\#}$ *ist bei* a *stetig.* □

(2.3) Folgerung. *Ist* $f\colon A \to \mathbb{R}$ *stetig im Häufungspunkt* a *von* A*, so gilt* $\lim_{x\to a} f(x) = f(a)$. □

(2.4) Folgerung. *Genau dann existiert* $\lim_{x\to a} f(x)$*, wenn die Funktion* $f|A \setminus \{a\}$ *sich zu einer bei* a *stetigen Funktion* $f^{\#}$ *auf* A *fortsetzen läßt. Es ist dann* $f^{\#}(a)$ *gleich* $\lim_{x\to a} f(x)$. □

Man kann die Aussage des nächsten Satz so formulieren: Eine stetige Funktion ist mit der Limesbildung vertauschbar.

(2.5) Vertauschungssatz. *Seien* $f\colon A \to B$ *und* $g\colon B \to \overline{\mathbb{R}}$ *Funktionen. Sei* a *Häufungspunkt von* A*, gelte* $\lim_{x\to a} f(x) = b \in B$ *und sei* g *bei* b *stetig. Dann existiert der Limes* $\lim_{x\to a}(g \circ f)(a)$ *und ist gleich* $g(\lim_{x\to a} f(x))$.

Beweis. Die Funktion $g \circ f^{\#}$ ist nach (2.3) und (1.3) bei a stetig. Sie stimmt auf $A \setminus \{a\}$ mit $g \circ f$ überein. Also hat letztere eine stetige Ergänzung $(g \circ f)^{\#}$ mit dem Wert $g(\lim_{x\to a} f(x))$ an der Stelle a. □

(2.6) Monotone Grenzwerte. *Sei* $f\colon A \to \overline{\mathbb{R}}$ *monoton steigend und sei* $b = \sup A$ *ein nicht in* A *enthaltener Häufungspunkt von* A*. Dann existiert* $\lim_{x\to b} f(x)$ *und ist gleich* $M = \sup f(A)$.

Beweis. Sei U eine Umgebung von M. Dann gibt es $N < M$, so daß $]N, M] \subset U$ ist. Nach Definition von M als Supremum gibt es $a \in A$, $a < b$ mit $N < f(a) < M$. Da f monoton wächst, ist $f([a, b]) \subset U$, das heißt, $f^{-1}(U)$ enthält eine Umgebung von b. □

Will man eine Funktion in einem Punkt stetig ergänzen, will man also die Existenz eines Limes beweisen, so steht man oft vor dem Problem, daß man den Grenzwert nicht kennt und deshalb die Definition (2.2) nicht anwenden kann. Das folgende, nach Cauchy benannte, *Grenzwertkriterium* ist deshalb von fundamentaler Bedeutung.

(2.7) Cauchy-Kriterium. *Sei $f \colon A \to \mathbb{R}$ gegeben. Sei a Häufungspunkt von $A \subset \overline{\mathbb{R}}$. Dann sind äquivalent:*
(1) *Es existiert $\lim_{x \to a} f(x)$.*
(2) *Zu jedem $\varepsilon > 0$ existiert eine Umgebung U von a gibt, so daß für $x, y \in U$ immer $|f(x) - f(y)| < \varepsilon$ gilt.*

Beweis. (2) $\Rightarrow$ (1). Sei U zunächst so gewählt, daß für $x, y \in U$ immer $|f(x) - f(y)| < 1$ gilt. Dann ist $|f(x)| < 1 + |f(y)|$. Indem wir y festhalten und x die Menge U durchlaufen lassen, sehen wir, daß f auf U beschränkt ist. Deshalb existiert
$$c(U) = \sup_{x \in U} f(x).$$
Aus $V \subset U$ folgt offenbar $c(V) \le c(U)$. Sei
$$c = \inf\{c(U) \mid U \text{ Umgebung von } a\}.$$
Wir behaupten: $\lim_{x \to a} f(x) = c$.

Zum Beweis sei ein $\varepsilon > 0$ fixiert. Wir wählen dann eine Umgebung U von a derart, daß für $x, y \in U$ immer $|f(x) - f(y)| < \varepsilon$ gilt. Nach Definition von c gibt es eine Umgebung V von a mit $f(x) < c + \varepsilon$ für alle $x \in V$. Wäre $f(x) \le c - \varepsilon$ für alle $x \in V$, so wäre $c(V) \le c - \varepsilon$, im Widerspruch zur Definition von c. Also gibt es $y \in V$ mit $c - \varepsilon < f(y)$ und folglich $|f(y) - c| < \varepsilon$. Wir fixieren ein derartiges y. Ist $x \in V$, so gilt demnach
$$|f(x) - c| \le |f(y) - c| + |f(x) - f(y)| < 2\varepsilon.$$
Es folgt die Behauptung.

(2) $\Rightarrow$ (1). Sei $\lim_{x \to a} f(x) = c$. Sei $\varepsilon > 0$ fixiert und die Umgebung U von a so gewählt, daß für $x \in U$ die Ungleichung $|f(x) - c| < \varepsilon/2$ gilt. Dann folgt aus der Dreiecksungleichung, daß für $x, y \in U$ immer $|f(x) - f(y)| < \varepsilon$ ist. □

3 Rechnen mit Funktionen

Seien $f \colon X \to \mathbb{R}$ und $g \colon X \to \mathbb{R}$ Funktionen und sei $c \in \mathbb{R}$. Dann erklären wir neue Funktionen $X \to \mathbb{R}$ durch: $f + g \colon x \mapsto f(x) + g(x)$, $fg = f \cdot g \colon x \mapsto f(x)g(x)$, $cf \colon x \mapsto cf(x)$ und $\frac{f}{g} \colon x \mapsto \frac{f(x)}{g(x)}$. Die Quotientenbildung im letzten Fall ist natürlich nur erlaubt, wenn immer $g(x) \neq 0$ ist. Sinngemäß nennen wir $f+g$ die *Summe* und fg das *Produkt* der Funktionen f und g. Ferner definieren wir $|f|: x \mapsto |f(x)|$, $\max(f,g): x \mapsto \max(f(x), g(x))$ und $\min(f,g) = \min(f(x), g(x))$. Für diese Definitionen kann X eine beliebige Menge sein.

(3.1) Beispiele. Die identische Funktion $f: \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, $x \mapsto x$ ist stetig. Eine konstante Funktion ist stetig. Mittels $||x| - |a|| \leq |x - a|$ sieht man, daß die Betragsfunktion $\mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$ stetig ist.

Die Funktion $\mathbb{R} \to \mathbb{R}, x \mapsto x^2$ ist stetig. Es ist nämlich $|(a+h)^2 - a^2| = |h||2a+h| \leq |h|(2|a|+|h|) \leq |h|(2|a|+1) < \varepsilon$ sofern $|h| < \delta < \min(1, (2|a|+1)^{-1})$ ist.

Die Funktion $f: \mathbb{R}^* \to \mathbb{R}, x \mapsto x^{-1}$ ist stetig. Ist nämlich $0 < x$ und $0 < \varepsilon < x^{-1}$, so ist $f^{-1}U_\varepsilon(x^{-1}) =]x(1+\varepsilon x)^{-1}, x(1-\varepsilon x)^{-1}[$ ein offenes Intervall um x. $\diamond$

(3.2) Satz. *Seien $f, g: A \to \mathbb{R}$ an der Stelle a stetig. Dann sind auch die folgenden Funktionen stetig: $f+g$, $f \cdot g$, f/g; im letzten Fall hat man, wie üblich, $g(x) \neq 0$ vorauszusetzen. Ferner sind $|f|$, $\max(f,g)$ und $\min(f,g)$ bei a stetig.* $\square$

Beweis. Sei $\varepsilon > 0$ gegeben. Für x aus dem Schnitt U der Umgebungen $f^{-1}U_\varepsilon(f(a))$ und $g^{-1}U_\varepsilon(f(a))$ ist

$$\begin{aligned} &|(f \pm g)(x) - (f \pm g)(a)| = |(f(x) - f(a) \pm (g(x) - g(a)| \\ &\leq |f(x) - f(a)| + |g(x) - g(a)| < 2\varepsilon. \end{aligned}$$

Also enthält das Urbild von $U_{2\varepsilon}(f(a))$ bei $f \pm g$ die Umgebung U.

Wegen (1.3) und (3.1) ist $f^2 = f \cdot f$ stetig. Aus $(f+g)^2 - f^2 - g^2 = 2fg$ und dem schon Bewiesenen folgt, daß $2fg$ stetig ist und dann (kleine Überlegung) auch fg.

Wegen (1.3) und (3.1) ist $x \mapsto 1/g(x)$ stetig und dann nach dem schon Gezeigten auch das Produkt $x \mapsto f(x)/g(x)$.

Die weiteren Behauptungen ergeben sich aus (1.3), (3.1) und den Relationen $\max(f,g) = \frac{1}{2}(f+g+|f-g|)$ und $\min(f,g) = \frac{1}{2}(f+g-|f-g|)$ zusammen mit dem schon Gezeigten. $\square$

Eine Funktion der Form $P: x \mapsto a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ heißt *Polynomfunktion* oder kurz *Polynom*, und zwar vom *Grad* n, wenn a_n ungleich Null ist ($a_j \in \mathbb{R}$). Eine Funktion der Form $x \mapsto P(x)/Q(x)$ mit Polynomen P und Q heißt *rationale Funktion*.

Aus (3.2) und dem Beispiel der konstanten und linearen Funktionen folgt durch wiederholte Anwendung (Induktion):

(3.3) Satz. *Jede rationale Funktion ist stetig.* $\square$

(3.4) Beispiel. Die n-te Potenz $f(x) = x^n$ ist auf $[0, \infty[$ streng monoton wachsend, also auch ihre Umkehrfunktion, die Funktion *n-te Wurzel*. Sie wird bekanntlich mit $x \mapsto x^{1/n} = \sqrt[n]{x}$ bezeichnet. Nach dem Umkehrsatz ist diese Funktion stetig. $\diamond$

(3.5) Satz. *Sei $f\colon \mathbb{R} \to \mathbb{R}$ eine stetige Funktion, die für alle x, y der Funktionalgleichung $f(x) + f(y) = f(x + y)$ genügt. Dann gibt es genau eine reelle Zahl λ, so daß $f(x) = \lambda x$ ist.*

Beweis. Wegen $f(0) = f(0+0) = f(0) + f(0)$ ist $f(0) = 0$. Durch vollständige Induktion schließt man auf $f(nx) = nf(x)$ für alle $x \in \mathbb{R}$ und $n \in \mathbb{N}$. Wegen $f(0) = f(x - x) = f(x) + f(-x)$ folgt $-f(x) = f(-x)$. Deshalb gilt $f(nx) = nf(x)$ sogar für alle $n \in \mathbb{Z}$. Ist $r = p/q$ eine rationale Zahl, so folgt $pf(x) = f(px) = qf((p/q)x)$, also $f(rx) = rf(x)$ für alle $r \in \mathbb{Q}$. Speziell gilt $f(r) = \lambda r$ mit $\lambda = f(1)$. Da f für rationale Zahlen mit der Funktion $x \mapsto \lambda x$ übereinstimmt und beide Funktionen stetig sind, so stimmen sie nach (1.5) überein. □

Wegen (2.3) liefern Aussagen und Regeln über stetige Funktionen Aussagen und Regeln über Limites.

(3.6) Satz. *Sei a Häufungspunkt von A. Seien $f, g\colon A \to \mathbb{R}$ gegeben und gelte $\lim_{x\to a} f(x) = L$ und $\lim_{x\to a} g(x) = M$. Dann gilt:*

(1) $\lim_{x\to a}(f + g)(x)$ *existiert und ist gleich $L + M$.*
(2) $\lim_{x\to a}(f \cdot g)(x)$ *existiert und ist gleich $L \cdot M$.*
(3) *Ist $g(x) \neq 0$ für $a \neq x \in A$ und ist $M \neq 0$, so existiert $\lim_{x\to a}(f/g)(x)$ und ist gleich L/M.*
(4) *Ist $f(x) \leq g(x)$ für alle $x \neq a$, so gilt $L \leq M$.*

Beweis. Wir verwenden die Hilfsfunktionen $f^{\#}$ und $g^{\#}$. Sie sind bei a stetig. Also ist auch $f^{\#} + g^{\#}$ bei a stetig. Somit gilt

$$L + M = (f^{\#} + g^{\#})(a) = \lim_{x\to a}(f^{\#} + g^{\#})(x) = \lim_{x\to a}(f + g)(x).$$

Die erste Gleichung folgt aus den Definitionen, die zweite aus (2.3) und die dritte, weil der Limes nur von den Funktionswerten abhängt, die verschieden von a sind. Entsprechend verfährt man mit den anderen Fällen. □

4 Aufgaben und Ergänzungen

1. Stetigkeit ist eine lokale Eigenschaft. Das bedeutet: $f\colon A \to \overline{\mathbb{R}}$ ist genau dann bei $a \in A$ stetig, wenn es eine Umgebung U von a gibt, so daß die Einschränkung $f|U$ von f auf U bei a stetig ist.

2. Sei A Vereinigung der offenen Teilmengen $(U_j \mid j \in J)$. Eine auf A definierte Funktion f ist genau dann stetig, wenn alle Einschränkungen $f|U_j$ stetig sind.

3. Einsperrsatz. Seien $f, g, h\colon A \to \overline{\mathbb{R}}$ Funktionen. Es gelte für alle $x \neq a$ die Ungleichung $f(x) \leq g(x) \leq h(x)$. Existieren die Limites $\lim_{x\to a} f(x)$ und $\lim_{x\to a} h(x)$ und sind sie beide gleich L, so existiert auch $\lim_{x\to a} g(x)$ und ist gleich L.

4. Sei $M = \sup S$ und gelte $M \notin S$. Dann ist M Häufungspunkt von S.

5. Wir geben einen zweiten Beweis für den Monotoniesatz. Sei $a < b$. Die Rechnung

$$a = (1 - t)a + ta \leq (1 - t)a + tb \leq (1 - t)b + tb = b$$

zeigt, daß wir eine stetige Funktion $p\colon [0,1] \to [a,b], t \mapsto (1-t)a+tb$ definieren können. Sie hat die stetige Umkehrfunktion $q\colon x \mapsto (x-a)/(b-a)$. Falls f nicht streng monoton wäre, so gäbe es zwei Paare (x,y) und (z,w) mit den Eigenschaften

$$f(x) - f(y) > 0, \quad x < y \quad \text{und} \quad f(z) - f(w) < 0, \quad z < w.$$

Die stetige Hilfsfunktion (definiert wegen (4.1))

$$a\colon [0,1] \to \mathbb{R}, \quad t \mapsto f((1-t)x + tz) - f((1-t)y + tw)$$

hat die Werte $a(0) > 0$ und $a(1) < 0$. Nach dem Zwischenwertsatz ist für ein $0 < u < 1$ dann $a(u) = 0$. Da f injektiv ist, bedeutet das $(1-u)x + uz = (1-t)y + uw$ oder

$$0 = (1-u)(y-x) + u(w-z),$$

was nicht sein kann, da die rechte Seite offenbar positiv ist. Also ist f streng monoton.

Im vorstehenden Beweis wurde mit der Hilfsfunktion a ein kleiner Trick verwendet. Dieser Trick hat folgenden geometrischen Hintergrund. Man betrachte die Teilmenge der Ebene $D = \{(x,y) \mid x,y \in [a,b], x < y\}$. Diese Menge ist ein Dreieck. Auf D betrachte man die Funktion von zwei Veränderlichen $\mu\colon (x,y) \mapsto f(x)-f(y)$. Falls f nicht streng monoton ist, so nimmt sie positive und negative Werte an. Man betrachtet nun die Verbindungsstrecke von einem Punkt mit positvem Wert zu einem mit negativem Wert und macht dadurch aus μ eine Funktion a einer Veränderlichen, auf die man den Zwischenwertsatz anwenden kann.

6. Die Funktionen $f\colon [0,\infty[\to \mathbb{R}$, $x \mapsto x^2$ und $f\colon]0,1] \to \mathbb{R}$, $x \mapsto x^{-1}$ sind nicht gleichmäßig stetig. Die Funktion $f\colon [0,\infty[\to \mathbb{R}$, $x \mapsto \sqrt[n]{x}$ ist gleichmäßig stetig.

7. Gilt $\lim_{x\to a} f(x) = c$ für $f\colon A \to \overline{\mathbb{R}}$ und hat $B \subset A$ auch a als HP, so gilt für die Einschränkung $g = f|B$ die Aussage $\lim_{x\to a} g(x) = c$.

8. Sei $x \mapsto f(x)$ eine rationale Funktion. Wie bestimmt man $\lim_{x\to\infty} f(x)$?

4 Differentialrechnung

1 Die Ableitung

Sei $f\colon J \to \mathbb{R}$ eine auf einem Intervall J definierte Funktion und seien b, a zwei verschiedene Zahlen aus dem Intervall J. Der Quotient

$$\textbf{(1.1)} \qquad \frac{f(b) - f(a)}{b - a}$$

heißt die *Änderungsrate der Funktion* im Intervall von a nach b. Ein Quotient der Form (1.1) wird *Differenzenquotient* von f genannt.

Der Graph der Funktion

(1.2) $$x \mapsto f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

ist eine Gerade durch die Punkte $(a, f(a))$ und $(b, f(b))$, wie man durch Einsetzen von $x = a, b$ sofort verifiziert. Der Differenzenquotient (1.1) ist die *Steigung* dieser Geraden. Die Gerade (1.2) wird *Sekante* von f durch die Punkte $(a, f(a))$ und $(b, f(b))$ genannt.

Eine Funktion $f \colon J \to \mathbb{R}$ heißt *im Punkt $a \in J$ differenzierbar*, wenn es eine bei a stetige Funktion $\Psi_a \colon J \to \mathbb{R}$ gibt, so daß für alle $x \in J$ die Gleichung

$$f(x) = f(a) + (x - a)\Psi_a(x)$$

gilt. Der Funktionswert $\Psi_a(a)$ heißt *Ableitung* oder *Differentialquotient* von f im Punkt a und wird mit

$$f'(a) \quad \text{oder} \quad \frac{df}{dx}(a)$$

bezeichnet. Wir nennen f *differenzierbar*, wenn f in jedem Punkt von J differenzierbar ist. Die dadurch gegebene Funktion

$$f' = \frac{df}{dx} \colon J \to \mathbb{R}, \quad x \mapsto f'(x)$$

heißt dann die *Ableitung* von f. Wir *differenzieren* f, indem wir die Ableitung f' bilden, und dieser Prozeß heißt *Differentiation.* ◇

Erläuterung. Für $x \neq a$ ist

$$\Psi_a(x) = \frac{f(x) - f(a)}{x - a}$$

der Differenzenquotient. Da Limites eindeutig bestimmt sind, ist der Wert $\Psi_a(a)$ eindeutig durch die Funktion $f|J \setminus \{a\}$ bestimmt. Wir haben die Funktion Ψ_a mit dem Index a versehen, weil man natürlich im allgemeinen für jedes a eine andere Funktion braucht. Obgleich die Funktion Ψ_a für alle $x \neq a$ mit dem Differenzenquotienten übereinstimmt, muß man sie generell als die „bessere Funktion“ ansehen. Übrigens sind ja auch, geometrisch betrachtet, Sekante und Tangente (siehe unten) gleichartige Objekte.

Etwas anders formuliert: f ist bei a differenzierbar, wenn es eine reelle Zahl $f'(a)$ und eine bei a stetige Funktion Δ_a so gibt, daß

$$f(x) = f(a) + (x - a)f'(a) + (x - a)\Delta_a(x)$$

ist und $\Delta_a(a) = 0$. In der Terminologie der Grenzwerte ist die Ableitung durch

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

definiert; der Differentialquotient ist also der Grenzwert des Differenzenquotienten. Die Gerade $x \mapsto f(a) + f'(a)(x - a)$ durch den Punkt $(a, f(a))$ heißt *Tangente* an den Graphen von f im Punkt a. Die Ableitung ist also deren Steigung. Aus den Rechenregeln über stetige Funktionen erhalten:

(1.3) Folgerung. *Eine im Punkt a differenzierbare Funktion ist dort auch stetig.* $\square$

(1.4) Beispiele. Sei $f(x) = \lambda x + \mu$ eine lineare Funktion. Dann ist $f(x) - f(a) = (x-a)\lambda$, also $\Psi_a(x) = \lambda$. Demnach ist f differenzierbar und $f'(a) = \lambda$.

Wegen $x^{-1} - a^{-1} = (x-a)(-(xa)^{-1})$ ist $\Psi_a(x) = -(xa)^{-1}$ und $f\colon x \mapsto x^{-1}$ differenzierbar mit der Ableitung $f'(a) = \Psi_a(a) = -a^{-2}$. Hier sieht man deutlich, wie Ψ_a von a abhängt.

Die Funktion $f(x) = |x|$ ist im Punkt $a = 0$ nicht differenzierbar, aber in allen anderen Punkten. Der Graph hat im Punkt $(0,0)$ eine „Ecke". Für $x > 0$ ist die Ableitung gleich 1, für $x < 0$ gleich -1. Eine solche Funktion mit einem „Sprung" an der Stelle Null kann durch keinen Funktionswert an der Stelle Null stetig gemacht werden. $\diamond$

Die Ableitung einer Funktion ist wieder eine Funktion, die wir ebenfalls untersuchen können. Die Ableitung der Ableitung (wenn es sie gibt) heißt zweite Ableitung f''. So fortfahrend können wir gegebenenfalls die n-te Ableitung $f^{(n)}$ von f bilden. Gibt es die n-te Ableitung von f und ist diese überdies stetig, so heißt f n-mal *stetig differenzierbar.* Für die n-te Ableitung gibt es auch die Bezeichnung

$$\frac{d^n f}{dx^n}.$$

Existiert $f^{(n)}$ für jede natürliche Zahl n, so heißt f *unendlich oft* differenzierbar. Es gibt stetige Funktionen, die an keiner Stelle differenzierbar sind. Es ist möglich, daß f' existiert, aber nicht stetig ist. Beispiele dafür werden wir noch kennenlernen.

2 Ableitungsregeln

Ableitungen werden fast niemals mit der Definition einer Ableitung berechnet. Ausgehend von bekannten Ableitungen verwendet man vielmehr vorzugsweise Rechenregeln. In den folgenden Beweisen unterdrücken wir in Ψ_a die Abhängigkeit von a und lassen diesen Index weg. Ein dennoch vorkommender Index hat dann eine ganz andere Bedeutung, die aber aus dem Kontext klar sein dürfte.

(2.1) Satz. *Seien $f, g\colon J \to \mathbb{R}$ in a differenzierbar und λ, μ reelle Zahlen. Dann sind auch die Funktionen $\lambda f + \mu g$, $f \cdot g$ und (falls $g(x) \neq 0$) f/g in a differenzierbar und es gilt:*

$$\begin{aligned}
(\lambda f + \mu g)'(a) &= \lambda f'(a) + \mu g'(a). \\
(f \cdot g)'(a) &= f'(a) \cdot g(a) + f(a) \cdot g'(a). \\
\left(\frac{f}{g}\right)'(a) &= \frac{f'(a) \cdot g(a) - f(a) \cdot g'(a)}{g(a)^2}.
\end{aligned}$$

Die erste Regel heißt die Linearität *der Ableitung, die zweite Regel nennen wir die* Produktregel *von Leibniz und die dritte die* Quotientenregel.

Beweis. Sei $f(x) = f(a) + (x-a)\Psi_f(x)$ und $g(x) = g(a) + (x-a)\Psi_g(x)$. Dann ist

$$f(x)g(x) = f(a)g(a) + (x-a)\Big(\Psi_f(x)g(a) + f(a)\Psi_g(x) + (x-a)\Psi_f(x)\Psi_g(x)\Big).$$

Die Funktion in der großen Klammer ist bei a stetig und hat dort den Funktionswert $f'(a)g(a)+f(a)g'(a)$; das zeigt die Produktregel. Ähnlich argumentiert man für die Linearität. Die Quotientenregel ergibt sich, wie weiter unten erläutert, aus anderen Regeln. □

Aus (2.1) folgt durch wiederholte Anwendung (Induktion), daß Polynome und rationales Funktionen differenzierbar sind, denn diese lassen sich aus den schon als differenzierbar erkannten linearen Funktionen durch Summen-, Produkt-, und Quotientenbildung herstellen. Außerdem sagen einem die Regeln, wie die Ableitungen zu berechnen sind. Es gilt nämlich:

(2.2) Beispiel. Sei n eine natürliche Zahl. Die Potenzfunktion $f(x) = x^n$ hat die Ableitung $f'(x) = nx^{n-1}$. Das wird mit den Regeln durch Induktion nach n bewiesen, denn die Aussage ist uns für $n = 1$ schon bekannt und für den Induktionsschritt wenden wir die Produktregel auf die Zerlegung $x^{n+1} = x^n \cdot x$ an. Mit der Quotientenregel erhalten wir daraus für die Funktion $f(x) = x^{-n}$ die Ableitung $f'(x) = -nx^{-n-1}$. Also gilt

$$f(x) = x^n, \quad f'(x) = nx^{n-1}$$

für alle ganzen Zahlen n. Mit den Potenzfunktionen können wir dann wegen der Linearität und der Quotientenregel sofort alle rationalen Funktionen differenzieren. Die Polynomfunktion $f(x) = \sum_{k=0}^n a_k x^k$ hat also die Ableitung $f'(x) = \sum_{k=1}^n k a_k x^{k-1}$. ◇

(2.3) Kettenregel. *Sei $f\colon A \to B$ bei a und $g\colon B \to C$ bei $b = f(a)$ differenzierbar. Dann ist $g \circ f$ bei a differenzierbar und es gilt*

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a).$$

Diese letzte Formel wird Kettenregel *genannt.*

Beweis. Sei $f(x) = f(a) + (x-a)\Psi_f(x)$ und $g(x) = g(b) + (x-b)\Psi_g(x)$. Es folgt

$$\begin{aligned} g(f(x)) &= g(f(a)) + (f(x) - f(a))\Psi_g(f(x)) \\ &= g(f(a)) + (x-a)\Psi_f(x)\Psi_g(f(x)). \end{aligned}$$

Mit den Rechenregeln für stetige Funktionen folgt die Behauptung. □

(2.4) Beispiel. Sei $g(x) \neq 0$. Die Funktion $h(x) = (g(x))^{-1}$ ist die Verkettung von $i(x) = x^{-1}$ mit g. Die Kettenregel zusammen mit der Kenntnis der Ableitung von i liefert in diesem Fall das Ergebnis der Quotientenregel $h'(x) = -g'(x)/g(x)^2$. Zusammen mit der Produktregel erhalten wir dann die allgemeine Quotientenregel. $\diamond$

(2.5) Beispiel. Zwei einfache Spezialfälle der Kettenregel: Die Funktion $x \mapsto f(x+c)$ hat an der Stelle x die Ableitung $f'(x+c)$; und die Funktion $x \mapsto f(\lambda x)$ hat an der Stelle x die Ableitung $\lambda f'(\lambda x)$. $\diamond$

(2.6) Satz. *Seien $f\colon A \to B$ und $g\colon B \to A$ Umkehrfunktionen voneinander. Sei f stetig, bei a differenzierbar und gelte $f'(a) \neq 0$. Dann ist g bei $b = f(a)$ differenzierbar und es gilt*

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

Beweis. Sei $f(x) = f(a) + (x-a)\Psi(x)$. Es ist $f'(a) = \Psi(a) \neq 0$. Wegen der Stetigkeit von Ψ bei a, ist $\Psi(x) \neq 0$ in einer Umgebung U von a. Dort gilt also

$$x = a + \frac{f(x) - f(a)}{\Psi(x)}, \qquad x \in U.$$

Wir setzen $x = g(y)$ ein, was für y in der Umgebung $V = g^{-1}(U)$ von b möglich ist. Es ergibt sich

$$g(y) = g(b) + \frac{x - b}{\Psi(g(y))}, \qquad y \in V.$$

Die Funktion $y \mapsto \Psi(g(y))^{-1}$ ist bei $y = b$ stetig und hat dort den behaupteten Funktionswert $f'(g(b))$. $\square$

(2.7) Beispiel. Die differenzierbare Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$ hat eine Umkehrfunktion. Diese ist aber im Nullpunkt nicht differenzierbar, weil $f'(0) = 0$ ist. Geometrisch gesprochen: Die Umkehrfunktion von f hat an der Stelle Null eine senkrechte Tangente. $\diamond$

(2.8) Beispiel. Sei $f(x) = x^n$ und $g(x) = \sqrt[n]{x}$. Dann ergibt sich nach der letzten Formel

$$g'(x) = \frac{1}{n(\sqrt[n]{x})^{n-1}}$$

als Ableitung der n-ten Wurzel. $\diamond$

3 Der Mittelwertsatz

Der Mittelwertsatz (3.3) ist der Hauptsatz der Differentialrechnung. Sei $f\colon J \to \mathbb{R}$ eine auf dem offenen Intervall J erklärte Funktion. Wir sagen, f habe an der Stelle $a \in J$ ein *lokales Maximum* (bzw. *lokales Minimum*), wenn es ein $t > 0$ gibt, so daß für alle x im Intervall $[a-t, a+t] \cap J$ die Ungleichung $f(x) \leq f(a)$ (bzw. $f(x) \geq f(a)$) gilt. Das lokale Maximum an der Stelle a heißt *isoliert* oder *strikt*, wenn für alle $c \neq a$, $|c-a| < t$ gilt $f(c) < f(a)$. Analog für das Minimum.

(3.1) Satz. *Hat die differenzierbare Funktion $f\colon J \to \mathbb{R}$ im Innern des Intervalles an der Stelle a ein lokales Maximum oder Minimum, so ist dort $f'(a) = 0$.*

Beweis. Liege bei a ein lokales Maximum, das heißt gelte für $a-t \le x \le a+t$ immer $f(x) \le f(a)$. Dann ist der Differenzenquotient

$$\frac{f(x)-f(a)}{x-a} = \Psi(x)$$

für $a < x \le a+t$ kleinergleich Null und für $a-t \le x < a$ größergleich Null. Die Behauptung folgt nun aus III(1.4). □

(3.2) Satz von Rolle. *Sei $f\colon [a,b] \to \mathbb{R}$ stetig und auf $]a,b[$ differenzierbar. Ferner gelte $f(a) = f(b)$. Dann gibt es eine Stelle $\xi \in]a,b[$ mit $f'(\xi) = 0$.*

Beweis. Der Satz ist sicherlich richtig, wenn f konstant ist. Andernfalls gibt es im Innern des Intervalls eine Stelle ξ, an der die stetige Funktion f ein Maximum oder ein Minimum hat. Der Satz von Rolle folgt deshalb aus dem vorigen Satz über lokale Extremwerte. □

(3.3) Mittelwertsatz. *Sei $f\colon [a,b] \to \mathbb{R}$ stetig und auf $]a,b[$ differenzierbar. Dann gibt es eine Stelle $\xi \in]a,b[$, an der die Gleichung*

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

gilt. □

Beweis. Der Satz von Rolle wird auf die Funktion g (Subtraktion der Geraden (1.2) von f)

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

angewendet. Dann ist nämlich $g(a) = g(b) = 0$, und $g'(\xi) = 0$ liefert genau die Formel des Mittelwertsatzes. □

Der Mittelwertsatz hat eine anschauliche Bedeutung. Wir betrachten die Sekante durch die Punkte $(a, f(a))$ und $(b, f(b))$. Wir verschieben sie sodann parallel solange nach oben oder unten, bis sie den Graphen nur noch „berührt“, das heißt zur Tangente degeneriert. Gibt es etwa unterhalb der Sekante Punkte des Graphen, so irgendwann bei Parallelverschiebung einen „letzten“ $(\xi, f(\xi))$, der von der verschobenen Sekante getroffen wird. Von der Zahl ξ im Mittelwertsatz wird lediglich die Existenz bewiesen; man weiß nicht, wo sie liegt. Wie die anschauliche Interpretation nahelegt, ist sie auch im allgemeinen nicht eindeutig bestimmt; es kann sogar unendlich viele geeignete ξ geben.

(3.4) Satz. *Die auf einem Intervall J definierte differenzierbare Funktion f habe die Ableitung Null. Dann ist f eine konstante Funktion. Haben $f, g\colon J \to \mathbb{R}$ dieselbe Ableitung, so gilt mit einer Konstanten c die Gleichung $f(x) = g(x) + c$ für alle $x \in J$.*

Beweis. Die zweite Aussage folgt aus der ersten, angewendet auf $f - g$. Für die erste Aussage fixieren wir $a \in J$ und erhalten aus dem Mittelwertsatz, daß $f(x) - f(a)$ immer Null ist. $\square$

(3.5) Satz. *Habe die auf einem Intervall erklärte Funktion im Innern des Intervalls eine überall positive (bzw. negative) Ableitung. Dann ist die Funktion streng monoton wachsend (bzw. fallend).*

Beweis. Ist $f'(\xi) > 0$, so folgt für $x > a$ aus dem Mittelwertsatz die Ungleichung $f(x) > f(a)$. $\square$

(3.6) Verallgemeinerter Mittelwertsatz. *Seien $f, g\colon [a, b] \to \mathbb{R}$ stetig und auf $]a, b[$ differenzierbar. Sei $g'(x) \neq 0$ für $x \in]a, b[$. Dann ist $g(a) \neq g(b)$, und es gibt ein $\xi \in]a, b[$ mit*

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

Beweis. Es ist $g(a) \neq g(b)$, weil andernfalls ein $\xi \in]a, b[$ mit $g'(\xi) = 0$ existierte. Auf die Hilfsfunktion $h\colon [a, b] \to \mathbb{R}$

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{g(b) - g(a)}(g(x) - g(a))$$

läßt sich der Satz von Rolle anwenden. Er liefert die Behauptung. $\square$

(3.7) Bemerkung. Satz (3.6) folgt nicht etwa, indem (3.1) auf f und g angewendet wird. Wir beschreiben einen anschaulichen Hintergrund zu (3.6). Durch $[a, b] \to \mathbb{R}^2, t \mapsto (f(t), g(t))$ wird eine Kurve in der Ebene gegeben. Der Vektor $v = (f(b) - f(a), g(b) - g(a))$ ist als Verbindungsvektor von Anfangs- und Endpunkt der Bahnkurve zu deuten. Der Satz besagt, daß es ein ξ gibt, so daß der Geschwindigkeitsvektor $w = (f'(\xi), g'(\xi))$ parallel zu v ist. (In dieser Formulierung muß man nicht $g'(x) \neq 0$ voraussetzen.) Daß w parallel zu v ist, ist äquivalent dazu, daß w senkrecht zum gedrehten Vektor $(g(b) - g(a), -(f(b) - f(a))$ ist, daß also das Skalarprodukt

$$f'(\xi)(g(b) - g(a)) - g'(\xi)(f(b) - f(a))$$

gleich Null ist. Das begründet den Ansatz für h. $\diamond$

(3.8) Satz. *Sei $f\colon J \to \mathbb{R}$ eine auf dem offenen Intervall J differenzierbare Funktion und sei $f'(a) = 0$.*

(1) *Ist $f'(a + h) \cdot h > 0$ für alle hinreichend kleinen $h \neq 0$, so ist a ein isoliertes lokales Minimum von f. Dies ist insbesondere der Fall, wenn f' um a streng monoton wächst, oder wenn $f''(a)$ existiert und positiv ist.*

(2) *Ist a ein isoliertes lokales Minimum von f', so wächst f lokal um a streng monoton.*

Beweis. (1) Für $t \neq 0$ ist nach dem Mittelwertsatz $f(a+t) = f(a)+f'(a+h)t$ für ein $|h| < t$. Aus der Voraussetzung folgt $f'(a+h)t > 0$ für alle genügend kleinen t. Die Voraussetzung besagt, daß $f'(a+h)$ dasselbe Vorzeichen wie $h \neq 0$ hat; wegen $f'(a) = 0$ folgt dies, wenn f' um a streng monoton wächst. Existiert $f''(a)$, so gibt es eine bei Null stetige Funktion Φ mit $f'(a+h) = f'(a) + \Phi(h)h = \Phi(h)h$ mit $\Phi(0) = f''(a) > 0$. Wegen der Stetigkeit von Φ bei Null ist dann auch $\Phi(h) > 0$ für alle hinreichend kleinen h.
(2) Wegen $f'(a) = 0$ ist $f'(a+h) > 0$ für kleine $h \neq 0$. Die Behauptung folgt jetzt aus (3.5). □

Der vorstehende Satz läßt sich auf höhere Ableitungen übertragen.

4 Die Regeln von de l'Hospital

Die *Regeln von de l'Hospital* sind im nächsten Satz formuliert. Sie erlauben in manchen Fällen die Bestimmung von Grenzwerten.

(4.1) Satz. *Seien $f, g\colon]a,b[\to \mathbb{R}$ differenzierbar. Sei $g'(x)$ immer von Null verschieden. Sei*

$$\lim_{x\to a} \frac{f'(x)}{g'(x)} = A \in \overline{\mathbb{R}}.$$

Dann gilt

$$\lim_{x\to a} \frac{f(x)}{g(x)} = A$$

unter einer der folgenden Voraussetzungen:
(1) *Es ist* $\lim_{x\to a} f(x) = 0 = \lim_{x\to a} g(x)$.
(2) *Es ist* $\lim_{x\to a} g(x) = \infty$.

Beweis. Sei $-\infty \le A < \infty$. Sei $q \in \mathbb{R}$ mit $A < q$ gewählt. Wir zeigen zunächst: Es gibt $c \in]a,b[$, so daß für $x \in]a,c[$ gilt:

$$g(x) \neq 0 \qquad \text{und} \qquad \frac{f(x)}{g(x)} < q.$$

Dazu wählen wir $A < r < q$. Wegen der Voraussetzung gibt es ein $c_1 \in]a,b[$, so daß für $x \in]a,c_1[$ gilt

$$\frac{f'(x)}{g'(x)} < r.$$

Nach dem verallgemeinerten Mittelwersatz haben wir für $a < x < y < c_1$ eine Gleichung der Form

$$\frac{f(x)-g(x)}{g(x)-g(y)} = \frac{f'(t)}{g'(t)} < r \tag{1}$$

mit einem t zwischen x und y. Es gibt höchstens eine Stelle $z \in]a,b[$, an der $g(z) = 0$ ist, da anderfalls nach dem Satz von Rolle nicht immer $g'(x) \neq 0$ wäre.

Sei c_1 so klein gewählt, daß $g(x) \neq 0$ ist für $x \in]a, c_1[$. Mit der Voraussetzung (1) des Satzes folgt nun

$$\frac{f(y)}{g(y)} = \lim_{x \to a} \frac{f(x) - f(y)}{g(x) - g(y)} \leq r < q.$$

In diesem Fall wählen wir $c = c_1$.

Nun zur Voraussetzung (2). Sei $y \in]a, c_1[$ fest gewählt. Wegen $\lim_{x \to a} g(x) = \infty$ gibt es ein $c_2 \in]a, c_1[$, so daß für $x \in]a, c_2[$ die Ungleichungen $g(x) > g(y)$ und $g(x) > 0$ gelten. Multiplikation der Zeile (1) mit $(g(x) - g(y))/g(x)$ führt auf

$$\frac{f(x)}{g(x)} < r - r\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)}.$$

Wegen Voraussetzung (2) gibt es ein $c_3 \in]a, c_2[$, so daß für $x \in]a, c_3[$ gilt

$$g(x) > -\frac{4rg(y)}{q-r} \qquad \text{und} \qquad g(x) > \frac{4f(y)}{q-r}$$

und folglich

$$\frac{q-r}{4} > -\frac{rg(y)}{g(x)} \qquad \text{und} \qquad \frac{q-r}{4} > \frac{4f(y)}{g(x)}$$

und deshalb schließlich

$$r - r\frac{g(y)}{g(x)} + \frac{f(y)}{g(x)} < r + \frac{q-r}{4} + \frac{q-r}{4} < q.$$

Wir wählen in diesem Fall $c = c_3$.

Sei $-\infty < A \leq \infty$. Sei $p \in \mathbb{R}$ gewählt mit $p < A$. Es gibt dann $d \in]a, b[$, so daß für $x \in]a, d[$ gilt:

$$g(x) \neq 0 \qquad \text{und} \qquad p < \frac{f(x)}{g(x)}.$$

Der Beweis ist analog zum voranstehenden Beweis.

Ist $A = -\infty$, so besagt die unter Schritt 1 bewiesene Aussage gerade

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

Ist $A = \infty$, so zeigt Schritt 2 ebenso

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \infty.$$

Ist $A \in \mathbb{R}$ und $\varepsilon > 0$, so wählen wir $q = A + \varepsilon$ und $p = A - \varepsilon$ und erhalten

$$A - \varepsilon < \frac{f(x)}{g(x)} < A + \varepsilon$$

für $a < x < \min(c, d)$. □

5 Aufgaben und Ergänzungen

1. Sei f n-mal (stetig) differenzierbar und habe eine differenzierbare Umkehrfunktion g. Dann ist auch g n-mal (stetig) differenzierbar.

2. Die Funktion $x \mapsto x \cdot |x|$ ist differenzierbar. Bestimme die Ableitung. Ist diese Funktion auch zweimal differenzierbar?

3. Gelegentlich hat man einseitige Ableitungen zu betrachten. Sei $f\colon J \to \mathbb{R}$ gegeben und liege a im Innern des Intervalls J. Sei f_+ die Einschränkung von f auf $J \cap [a, \infty[$ und f_- die Einschränkung auf $]-\infty, a] \cap J$. Ist f im Punkt a differenzierbar, so sind auch $f_\pm$ dort differenzierbar und haben dieselbe Ableitung. Existieren die *linksseite Ableitung* $f'_-(a)$ und die *rechtsseitige Ableitung* $f'_+(a)$ und sind diese gleich, so ist f in a differenzierbar.

4. Sei f auf dem offenen Intervall stetig und auf $J \setminus \{a\}$ differenzierbar. Existiert $\lim_{x \to a} f'(x) = c$, so ist f im Punkt a differenzierbar, und es gilt $f'(a) = c$. Zum Beweis wende man den Mittelwertsatz auf den Differenzenquotienten an der Stelle a an.

5. Sei $f\colon]a, b[\to \mathbb{R}$ differenzierbar. An einer Stelle x_0 gelte $f'(x_0) > 0$. Beweise oder widerlege: Es gibt eine Umgebung U von x_0 derart, daß f auf U monoton wächst.

6. Sei $f\colon [a, b] \to \mathbb{R}$ differenzierbar und $f'(a) < f'(b)$. Obgleich f' nicht notwendig stetig ist, nimmt f' jeden Wert zwischen $f'(a)$ und $f'(b)$ an.

7. Sei $f\colon J \to \mathbb{R}$ n-mal differenzierbar und gelte $f^{(n)} = 0$. Dann ist f ein Polynom höchstens vom Grad $n-1$.

8. Eine Funktion ist n-mal differenzierbar, wenn ihre Ableitung $(n-1)$-mal differenzierbar ist. Diese simple Bemerkung verwende man, um der Reihe nach induktiv zu beweisen: Seien f und g n-mal differenzierbar. Dann sind auch (sofern definiert) $f+g$, $f \cdot g$, $f \circ g$ n-mal differenzierbar. Damit zeigt man: Hat f eine differenzierbare Umkehrfunktion h, so ist auch h n-mal differenzierbar.

9. Sei $f\colon \mathbb{R} \to \mathbb{R}$ differenzierbar und gelte $f \circ f = f$. Dann ist entweder f konstant oder es gilt $f(x) = x$ für alle x. Die Betragsfunktion zeigt, daß die Folgerung für stetige Funktionen nicht immer gilt.

5 Integration

1 Integration stetiger Funktionen

Sei $a \leq b$ und bezeichne $C[a, b]$ die Menge der stetigen Funktionen $f\colon [a, b] \to \mathbb{R}$. Sei $B[a, b]$ die Menge der beschränkten Funktionen $g\colon [a, b] \to \mathbb{R}$, also der Funktionen g, deren Bildmenge $g[a, b]$ in $\mathbb{R}$ beschränkt ist. Diese Mengen bilden übrigens bezüglich der Addition und Skalarmultiplikation von Funktionen einen $\mathbb{R}$-Vektorraum.

(1.1) Definition. Eine *Integration für stetige Funktionen* ist eine Familie von Abbildungen $(I_a^b \mid a, b \in \mathbb{R}, a \le b)$

$$I_a^b \colon C[a,b] \to \mathbb{R}$$

mit den Eigenschaften:

(1) Für alle $x \in [a,b]$ gelte $m \le f(x) \le M$. Dann ist

$$m(b-a) \le I_a^b(f) \le M(b-a).$$

(2) Ist $f \in C[a,b]$ und $c \in [a,b]$, so gilt

$$I_a^c(f|[a,c]) + I_c^b(f|[c,b]) = I_a^b(f).$$

Ist $f \in C[a,b]$ und gilt $a \le c \le d \le b$, so bezeichnen wir $I_c^d(f|[c,d])$ der Einfachheit halber mit $I_c^d(f)$. Die Zahl $I_a^b(f)$ nennen wir das *Integral* von f über $[a,b]$ und benutzen die Bezeichnungen

$$I_a^b(f) = \int_a^b f(x)\,dx = \int_a^b f.$$

Man nennt b die obere und a die untere *Integrationsgrenze*. Die Vereinbarung $\int_b^a = -\int_a^b$ wird sich als sinnvoll und zweckmäßig erweisen. $\diamond$

(1.2) Satz. *Es gibt genau eine Integration für stetige Funktionen.*

Die Existenz wird durch (1.3) gezeigt; die Eindeutigkeit beweisen wir im nächsten Abschnitt.

Eine *Zerlegung* Z des Intervalles $[a,b]$ ist eine Familie $Z = (x_0, x_1, \ldots, x_n)$ von Punkten aus $[a,b]$ mit $a \le x_0 \le x_1 \le \ldots \le x_n = b$. Die *Untersumme* von $f \in B[a,b]$ zu Z sei

$$U(f,Z) = \sum_{i=0}^{n-1} (x_{i+1} - x_i) \inf(f[x_i, x_{i+1}]).$$

Wir betrachten die Menge aller Untersummenwerte

$$US(f) = \{\lambda \mid \text{es gibt eine Zerlegung } Z \text{ von } [a,b] \text{ mit } \lambda = U(f,Z)\}.$$

Für jedes $\lambda \in US(f)$ gilt $\lambda \le (b-a)\max(f)$. Also ist $US(f)$ nach oben beschränkt. Wir setzen

$$_*I_a^b(f) = \sup US(f)$$

und nennen diese Zahl das *Unterintegral* der beschränkten Funktion $f \colon [a,b] \to \mathbb{R}$.

(1.3) Notiz. *Das Unterintegral hat die Eigenschaften* (1) *und* (2) *aus Definition* (1.1), *wobei* I *durch* $_*I$ *zu ersetzen ist.*

Beweis. (1) Sei $m \le f(x) \le M$ für alle x. Dann gilt für jede Untersumme

$$m(b-a) \le U(f,Z) \le M(b-a).$$

Da also $M(b-a)$ obere Schranke für $US(f)$ ist, folgt $I_a^b(f) \leq M(b-a)$. Da $I_a^b(f)$ obere Schranke für alle $U(f,Z)$ ist, folgt $m(b-a) \leq I_a^b$.
(2) Sei $f \in B[a,b]$ und sei $c \in [a,b]$. Sei $Z_1 = (x_0,\ldots,x_n)$ eine Zerlegung von $[a,c]$ und $Z_2 = (y_0,\ldots,y_m)$ eine Zerlegung von $[c,b]$. Dann ist $Z = (x_0,\ldots,x_n,y_0,\ldots,x_m)$ eine Zerlegung von $[a,b]$. Es gilt

$$_*I_a^b(f) \geq U(f,Z) = U(f,Z_1) + U(f,Z_2).$$

Also ist, bei festem Z_2 durch $_*I_a^b(f) - U(f,Z_2)$ eine obere Schranke der $U(f,Z_1)$ gegeben; es folgt $_*I_a^b(f) \geq {_*I_a^c}(f) + U(f,Z_2)$. Nun argumentiert man ebenso für Z_2 und erhält insgesamt $_*I_a^b(f) \geq {_*I_a^c}(f) + {_*I_c^b}(f)$.

Sei umgekehrt $Z = (z_0,\ldots,z_t)$ eine Zerlegung von $[a,b]$. Sei $0 < i \leq t$ so gewählt, daß $z_{i-1} \in [a,c]$ ist und $z_i \in [c,b]$. Dann ist $Z^1 = (z_0,\ldots,z_{i-1},c)$ eine Zerlegung von $[a,c]$ und $Z^2 = (c,z_i,\ldots,z_t)$ eine Zerlegung von $[c,b]$. Mittels

$$(z_i - z_{i-1}) \inf f[z_i,z_{i-1}] \leq (c - z_{i-1}) \inf f[z_{i-1},c] + (z_i - c) \inf f[c,z_i]$$

erkennt man $U(f,Z) \leq U(f,Z^1) + U(f,Z^2) \leq {_*I_a^c}(f) + {_*I_c^b}(f)$. Durch Übergang zum Supremum folgt $_*I_a^b(f) \leq {_*I_a^c}(f) + {_*I_c^b}(f)$. Insgesamt haben wir die zweite der Eigenschaften (1.1) nachgewiesen. □

Ebenso kann man mit den *Obersummen*

$$O(f,Z) = \sum_{i=0}^{n-1} (x_{i+1} - x_i) \sup(f[x_i,x_{i+1}])$$

durch das Infimum über alle diese Werte das *Oberintegral* $^*I_a^b(f)$ von f definieren. Es gilt nach Konstruktion $_*I_a^b(f) \leq {^*I_a^b}(f)$, aber diese Werte sind nicht immer gleich. Wir nennen f *integrierbar im Sinne von Riemann*, wenn Ober- und Unterintegral übereinstimmen. Den gemeinsamen Wert bezeichnen wir dann als *Riemann–Integral* von f

$$\int_a^b f(t)\,dt.$$

2 Hauptsatz der Differential- und Integralrechnung

Sei eine beliebige Integration (1.1) stetiger Funktionen gegeben, die wir durch das Integralzeichen notieren, obgleich die Eindeutigkeit noch nicht bewiesen ist. Sei J ein nichtentartetes Intervall und $f\colon J \to \mathbb{R}$ eine Funktion. Wir nennen $F\colon J \to \mathbb{R}$ *Stammfunktion* von f, wenn F differenzierbar ist und die Ableitung f hat.

(2.1) Hauptsatz. *Sei $f\colon J \to \mathbb{R}$ stetig. Dann gilt:*
(1) *Für jedes $a \in J$ wird durch*

$$F\colon J \to \mathbb{R}, \quad x \mapsto \int_a^x f(t)\,dt = F(x)$$

eine Stammfunktion F von f gegeben.

(2) *Ist G eine Stammfunktion von f, so gilt für alle $a, b \in J$ die Gleichung* $\int_a^b f(t)\,dt = G(b) - G(a) = G\Big|_a^b$.

Beweis. (1) Wir zeigen etwas allgemeiner, daß F an der Stelle x die Ableitung $f(x)$ hat, wenn f bei x stetig ist; das heißt, wir zeigen

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

Es ist $\int_a^{x+h} f(t)\,dt = \int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt$. Das folgt für $a \le x \le x+h$ direkt aus der Eigenschaft (2) in (1.1) und sonst aus dieser Eigenschaft zusammen mit der Festsetzung $\int_a^b = -\int_b^a$. Folglich ist $F(x+h) - F(x) = \int_x^{x+h} f(t)\,dt$. Sei $\varepsilon > 0$ gegeben. Wegen der Stetigkeit von f an der Stelle $x \in J$ gibt es ein $\delta > 0$, so daß für $|t| < \delta$ und $x + t \in J$ die Ungleichung

$$-\varepsilon + f(x) < f(x+t) < f(x) + \varepsilon$$

gilt. Mittels (1.1.1) folgt für $0 < h < \delta$

$$(-\varepsilon + f(x))h < \int_x^{x+h} f(t)\,dt < (f(x) + \varepsilon)h$$

und für $-\delta < h < 0$ eine analoge Ungleichung, so daß sich deshalb insgesamt

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| < \varepsilon$$

ergibt, womit die gewünschte Grenzwertaussage gezeigt ist.
(2) Es sind $x \mapsto \int_a^x f(t)\,dt$ und $x \mapsto G(x) - G(a)$ beides differenzierbare Funktionen, deren Ableitung f ist und die an der Stelle $x = a$ den Wert Null haben. Also stimmen sie überein. □

Die Ableitbarkeit von $x \mapsto \int_a^x f$ an einzelnen Stellen benutzt nur die Stetigkeit an diesen Stellen und gilt deshalb sowohl für das Ober- als auch das Unterintegral.

(2.2) Notiz. *Ist F eine Stammfunktion von f, so hat jede andere Stammfunktion die Form $x \mapsto F(x) + c$ mit einer Konstanten c, und alle diese Funktionen sind auch Stammfunktionen.* □

Wir zeigen nun die Eindeutigkeit einer Integration nach (1.1). Sind nämlich I und $\tilde{I}$ zwei Integrationen, so sind $x \mapsto I_a^x(f)$ und $x \mapsto \tilde{I}_a^x(f)$ beides Stammfunktionen von f. Da sie an der Stelle a übereinstimmen, sind sie gleich. Wegen der Eindeutigkeit stimmen für stetige Funktionen Ober- und Unterintegral überein, sie sind also Riemann–integrierbar.

Ein Motiv für die Integralrechnung ist die Definition und Berechnung von Flächeninhalten. Ist $f\colon [a, b] \to \mathbb{R}$ eine Riemann–integrierbare Funktion mit nichtnegativen Funktionswerten, so wird der *Flächeninhalt* der ebenen Punktmenge

$$\{(x, y) \in \mathbb{R}^2 \mid a \le x \le b,\ 0 \le y \le f(x)\}$$

als das Integral $\int_a^b f(t)\,dt$ definiert. Die Axiome (1) und (2) in (1.1) haben dann eine unmittelbar einsichtige geometrische Bedeutung.

Sei $Z = (x_0, \dots, x_n)$ eine Zerlegung des Intervalls $[a,b]$. Die Zahl $\max\{x_{i+1} - x_i \mid i = 0, \dots, n-1\}$ heiße die *Feinheit* der Zerlegung. Sei $f\colon [a,b] \to \mathbb{R}$ gegeben. Eine beliebige Summe der Form

$$\sum_{i=0}^{n-1} f(\xi_i)(x_{i+1} - x_i), \qquad x_i \le \xi_i \le x_{i+1}$$

heiße *Rechtecksumme* von f zu Z.

(2.3) Satz. *Sei $f\colon [a,b] \to \mathbb{R}$ stetig. Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß für jede Zerlegung Z, deren Feinheit kleiner als δ ist und jede Rechtecksumme R von f zu Z die Abschätzung $\left|\int_a^b f(t)\,dt - R\right| < \varepsilon$ gilt.*

Beweis. Sei $\varepsilon > 0$ gegeben. Sei $\delta > 0$ so gewählt, daß $|f(x) - f(y)| < \varepsilon/(2(b-a))$ ist, sofern $|x - y| < \delta$ ist; wegen der gleichmäßigen Stetigkeit von f ist das möglich. Sei $Z = (x_0, \dots, x_n)$ eine Zerlegung von $[a,b]$ mit einer Feinheit kleiner als δ. Dann gilt

$$\left|\int_a^b f(t)\,dt - R\right| = \left|\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} (f(t) - f(\xi_i))\,dt\right| \le \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} |f(t) - f(\xi_i)|\,dt.$$

Unter den Integralen schätzen wir durch $\varepsilon/(2(b-a))$ ab und erhalten dadurch das gewünschte Resultat. □

3 Eigenschaften des Integrals

Aus dem Hauptsatz erhalten wir leicht durch „Integration von Differentiationsregeln“ Regeln für die Integration.

(3.1) Satz. *Sei $a \le b$, seien $f, g\colon [a,b] \to \mathbb{R}$ stetig und seien $\lambda, \mu \in \mathbb{R}$. Dann gilt:*

(1) $\int_a^b (\lambda f + \mu g) = \lambda \int_a^b f + \mu \int_a^b g$.

(2) *Ist $f \le g$, so ist $\int_a^b f \le \int_a^b g$.*

(3) $|\int_a^b f| \le \int_a^b |f|$.

Die Eigenschaften nennen wir der Reihe nach Linearität, Monotonie *und* Dreiecksungleichung.

Beweis. (1) Wir betrachten beide Seiten als Funkionen in der Veränderlichen b. Nach dem Hauptsatz haben sie dieselbe Ableitung. Da sie an der Stelle $b = a$ übereinstimmen, sind sie gleich.

(2) Wegen $0 \le g(x) - f(x)$ folgt aus (1.1.1) die Ungleichung $0 \le \int_a^b (g - f)$. Nun wende man Teil (1) an.

(3) Wir setzen in (2) $g = |f|$ und bedenken, daß auch g stetig ist. Es folgt aus (2) $\int_a^b f \le \int_a^b |f|$. Dieselbe Ungleichung gilt auch für $-f$ statt f. □

Wir machen noch einen Zusatz zur Monotonie. Ist $f \le g$ und gibt es x mit $f(x) < g(x)$, so ist $\int f < \int g$. Zum Beweis beachte man, daß $h(x) = g(x) - f(x) > 0$ impliziert, daß diese Ungleichung auch noch auf einem kleinen Intervall $[c,d] \subset [a,b]$ um x gilt. Dann hat h auf $[c,d]$ ein positives Minimum, und es folgt $0 < \int_c^d h \le \int_a^b h$.

(3.2) Mittelwertsatz der Integralrechnung. *Seien $f, g \in C([a,b])$ und gelte immer $g(x) \ge 0$. Dann gibt es ein $\xi \in]a,b[$ für das die Gleichung $\int_a^b f(t)g(t)\,dt = f(\xi)\int_a^b g(t)\,dt$ gilt. Im Fall $g = 1$ ergibt sich $\int_a^b f(t)\,dt = f(\xi)(b-a)$.*

Beweis. Mit dem Maximum M und dem Minimum m von f gilt $mg(x) \le f(x)g(x) \le Mg(x)$. Es folgt $m\int_a^b g \le \int_a^b fg \le M\int_a^b g$ und mit dem Zwischenwertsatz folgt die Behauptung, wobei man mit dem Zusatz zur Monotonie erkennt, daß ξ wirklich im Innern des Intervalls gewählt werden kann. □

(3.3) Transformationsformel. *Seien J_1 und J_2 nichtentartete Intervalle. Seien $f\colon J_1 \to J_2$ und $g\colon J_2 \to \mathbb{R}$ stetig. Sei f differenzierbar und f' stetig. Dann gilt für alle $a, b \in J_1$*

$$\int_{f(a)}^{f(b)} g(y)\,dy = \int_a^b g(f(x)) \cdot f'(x)\,dx.$$

Beweis. Wir betrachten beide Seiten der behaupteten Gleichheit als Funktionen von b. Die linke Seite hat nach dem Hauptsatz an der Stelle x die Ableitung $g(f(x))f'(x)$ und die rechte Seite nach dem Hauptsatz und der Kettenregel ebenfalls. Beide Seiten stimmen für $b = a$ überein. □

Die Transformationsformel merkt man sich oft so: Man ersetzt (substituiert) die Veränderliche y vermöge $y = f(x)$ und ersetzt dy vermöge $dy = f'(x)\,dx$. Mit den Integrationsgrenzen muß man dann noch aufpassen. Deshalb spricht man auch von der *Substitutionsregel*.

(3.4) Beispiele. Hier sind einfache Anwendungen. Mit $f(x) = x+c$ und folglich $f'(x) = 1$ erhält man die Formel

$$\int_{a+c}^{b+c} g(t)\,dt = \int_a^b g(t+c)\,dt.$$

Mit $f\colon [a,b] \to [ac, bc], x \mapsto cx$ und $g\colon [ac, bc] \to \mathbb{R}$ erhält man wegen $f'(x) = c$

$$\int_a^b g(ct)c\,dt = \int_{ac}^{bc} g(t)\,dt.$$

So gilt zum Beispiel für $b > 0$ und $c > 0$

$$\int_c^{bc} \frac{1}{x}\,dx = \int_1^b \frac{1}{cx}c\,dx = \int_1^b \frac{1}{x}\,dx,$$

eine Formel, die alsbald interessant wird.

Gelte $g(x) = g(-x)$ für $g\colon [-a,a] \to \mathbb{R}$. Dann ist $\int_0^a g(x)\,dt = \int_{-a}^0 g(x)\,dx$. Zum Beweis verwende man die Substitution $f(x) = -x$. □

(3.5) Partielle Integration. *Seien $f, g\colon J \to \mathbb{R}$ stetig differenzierbar. Dann gilt für alle $a, b \in J$*

$$\int_a^b f(x)g'(x)\,dx + \int_a^b g(x)f'(x)\,dx = f(b)g(b) - f(a)g(a).$$

Beweis. Man betrachte beide Seiten als Funktion von b und differenziere. □

In den Anwendungen versucht man eine Situation zu erreichen, in der eine der Funktionen $f'(x)g(x)$ oder $f(x)g'(x)$ einfacher als die andere ist.

4 Die Taylorsche Formel

Wir erinnern daran, daß mit $n!$ (*n-Fakultät*) das Produkt der natürlichen Zahlen von 1 bis n bezeichnet wird, $n! = 1 \cdot 2 \cdot \ldots \cdot n$.

(4.1) Taylorsche Formel. *Sei $f\colon J \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion. Sei $a \in J$. Dann gilt*

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i + R_n f(x,a), \quad R_n f(x,a) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt.$$

Diese Identität heißt Taylorsche Formel.

Beweis. Induktion nach n. Sei $n = 0$. Nach dem Hauptsatz ist $f(x) = f(a) + \int_a^x f'(t)\,dt$; das ist der Induktionsanfang. Wir nehmen an, daß die Aussage für k-mal stetig differenzierbare Funktionen gilt $(0 \le k \le n)$. Durch partielle Integration erhalten wir

$$\int_a^x (x-t)^{n-1} f^{(n)}(t)\,dt = -\frac{(x-t)^n}{n} f^{(n)}(t)\Big|_a^x + \int_a^x \frac{(x-t)^n}{n} f^{(n+1)}(t)\,dt.$$

Das setzen wir in die Induktionsvoraussetzung ein und erhalten den Induktionsschritt. □

Wir nennen $\sum_{i=0}^n \frac{f^{(i)}(a)}{i!}(x-a)^i$ das n-te *Taylor–Polynom* von f an der Stelle a und $R_n f(x,a)$ das zugehörige *Restglied*. Das Restglied können wir mit dem Mittelwertsatz der Integralrechnung (3.2) umformen

$$R_n f(x,a) = \frac{1}{n!} f^{(n+1)}(\xi) \int_a^x (x-t)^n\,dt = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

mit einem ξ zwischen a und x. Damit erhält die Taylorsche Formel eine Form, die analog zur Formel des Mittelwertsatzes ist. Wenden wir den Mittelwertsatz in der zweiten Form von (3.2) an, so erhalten wir eine Darstellung

$$R_n f(x,a) = \frac{1}{n!}(1-\theta)^n (x-a)^{n+1} f^{(n+1)}(a + \Theta(x-a))$$

mit einem $\Theta \in]0,1[$.

(4.2) Beispiel. Wir wenden die Taylorsche Formel auf die Kurvendiskussion an. Sei $f\colon [a,b] \to \mathbb{R}$ n-mal stetig differenzierbar ($n \geq 2$). Sei $x_0 \in]a,b[$ und gelte $f^{(j)}(x_0) = 0$ für $1 \leq j \leq n-1$ und $f^{(n)}(x_0) \neq 0$. Dann gilt also mit einem $\Theta \in]0,1[$

$$f(x) = f(x_0) + \frac{f^{(n)}(x_0 + \Theta(x-x_0))}{n!}(x-x_0)^n.$$

Aus dieser Formel sieht man:
Sei n eine gerade Zahl.

(1) Ist $f^{(n)}(x_0) > 0$, so hat f an der Stelle x_0 ein lokales Minimum.
(2) Ist $f^{(n)}(x_0) < 0$, so hat f an der Stelle x_0 ein lokales Maximum.

Sei n ungerade. Dann ist x_0 keine Extremstelle.
Zum Beweis benutzt man, daß die Ungleichungen in (1) und (2) wegen Stetigkeit auch noch in einer Umgebung von x_0 gelten. ◇

5 Aufgaben und Ergänzungen

1. Sei $f\colon [c,d] \to \mathbb{R}$ stetig differenzierbar. Sei $f(x) - f(a) = (x-a)\Phi_a(x)$ mit einer bei a stetigen Funkion Φ_a. Dann gilt

$$\Phi_a(x) = \int_0^1 f'(tx + (1-t)a)\, dt.$$

Damit zeigt man, daß $\Phi_a(x)$ von den beiden Veränderlichen a und x in folgendem Sinne stetig abhängt: Seien a und x fixiert. Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für alle b und y mit $|b - | < \delta$ und $|y - x| < \delta$ die Abschätzung $|\Phi_b(y) - \Phi_a(x)| < \varepsilon$ gilt.

6 Exponentialfunktion. Logarithmus

1 Logarithmus

Der natürliche *Logarithmus* ist die Funktion

$$\log\colon \mathbb{R}_+ \to \mathbb{R}, \quad x \mapsto \int_1^x \frac{1}{t}\, dt.$$

Man schreibt auch $\ln(x)$ an Stelle von $\log(x)$. Nach dem Hauptsatz IV(2.1) ist log differenzierbar, und es gilt $\log'(x) = x^{-1}$. Demnach ist der Logarithmus streng monoton wachsend. Mit der Substitutionsregel erhalten wir

$$\int_1^{xy} \frac{1}{t}\, dt = \int_1^x \frac{1}{t}\, dt + \int_x^{xy} \frac{1}{t}\, dt = \int_1^x \frac{1}{t}\, dt + \int_1^y \frac{1}{t}\, dt.$$

Also genügt der Logarithmus der Funktionalgleichung

(1.1) $$\log(xy) = \log(x) + \log(y).$$

Definitionsgemäß ist $\log(1) = 0$. Wegen der strengen Monotonie ist also $\log(x) \neq 0$, wenn $x \neq 1$ ist. Aus der Funktionalgleichung folgt induktiv $\log(x^n) = n\log(x)$ für natürliche Zahlen n, und diese Regel gilt dann wegen $0 = \log(1) = \log(x^n x^{-n}) = \log(x^n) + \log(x^{-n})$ auch für alle $n \in \mathbb{Z}$. Da wegen der Monotonie für $x > 1$ der Logarithmus positiv ist und für $x < 1$ negativ, erhalten wir daraus

$$\lim_{x\to\infty} \log(x) = \infty, \qquad \lim_{x\to 0} \log(x) = -\infty.$$

Demnach ist $\log\colon \mathbb{R}_+ \to \mathbb{R}$ nach dem Zwischenwertsatz eine bijektive Abbildung.

2 Exponentialfunktion

Der Logarithmus hat eine differenzierbare Umkehrfunktion. Sie heißt *Exponentialfunktion* $\exp\colon \mathbb{R} \to \mathbb{R}_+$. Aus den Regeln über die Differentiation der Umkehrfunktion folgt die fundamentale Differentialgleichung

(2.1) $$\exp'(x) = \exp(x).$$

Aus der Funktionalgleichung des Logarithmus erhalten wir die Funktionalgleichung

(2.2) $$\exp(x + y) = \exp(x) \cdot \exp(y).$$

Die Exponentialfunktion ist streng monoton wachsend. Es gilt

$$\lim_{x\to\infty} \exp(x) = \infty, \qquad \lim_{x\to-\infty} \exp(x) = 0.$$

Die Exponentialfunktion ist die wichtigste Funktion der Analysis.

3 Die Exponentialfunktion zur Basis a

Sei $a > 0$. Wir definieren für $x \in \mathbb{R}$ die Symbole $\exp_a(x)$ und a^x (gelesen: a hoch x) durch

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

Mit der durch $\exp(1) = e$ definierten Zahl e ist dann insbesondere $\exp(x) = \exp_e(x) = e^x$. Die *Exponentialfunktion zur Basis* a $\exp_a\colon \mathbb{R} \to \mathbb{R}_+$ ist dann eine differenzierbare Funktion mit der Ableitung $\exp_a'(x) = \log(a)\exp_a(x)$. Für eine ganze Zahl n hat a^n die elementar-mathematische Bedeutung. Es ist $1^x = 1$. Aus den schon bekannten Eigenschaften von log und exp folgen durch Umrechnung fast unmittelbar:

(3.1) Eigenschaften der Potenz a^x.

(1) $a^{x+y} = a^x a^y, \quad (a^x)^y = a^{xy}, \quad a^x b^x = (ab)^x, \quad \log(a^x) = x\log(a).$

(2) Sei $a > 1$. Dann gilt $x < y \Leftrightarrow a^x < a^y$.

(3) Sei $a < 1$. Dann gilt $a < b \Leftrightarrow a^x > a^y$.
(4) Sei $x > 0$. Dann gilt $a < b \Leftrightarrow a^x < b^x$.
(5) Sei $x < 0$. Dann gilt $a < b \Leftrightarrow a^x > b^x$. □

Wir runden die Betrachtung ab und studieren stetige Funktionen $f\colon \mathbb{R} \to \mathbb{R}$, die der Funktionalgleichung $f(x+y) = f(x)f(y)$ des Potenzgesetzes genügen. Hat eine solche Funktion eine Nullstelle, $f(x_0) = 0$, so ist sie wegen $f(x) = f(x-x_0)f(x_0) = 0$ überhaupt gleich Null. Diesen Fall schließen wir deshalb aus. Wegen $f(x) = f(x/2)^2$ nimmt dann f nur positive Werte an.

(3.2) Satz. *Zu jedem $a > 0$ gibt es genau eine stetige Funktion $f\colon \mathbb{R} \to \mathbb{R}_+$ mit den Eigenschaften:*
(1) $f(x+y) = f(x)f(y)$.
(2) $f(1) = a$.
Diese Funktion ist differenzierbar, und es gilt $f'(x) = \log(a)f(x)$.

Beweis. Sei f mit den Eigenschaften (1) und (2) gegeben. Die Funktion $\ell\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \log(f(x))$ ist definiert, stetig und erfüllt $\ell(x+y) = \ell(x) + \ell(y)$. Nach II(3.5) gibt es deshalb eine reelle Zahl λ, so daß $\ell(x) = \lambda x$ ist. Folglich ist f die Funktion $f(x) = \exp(\lambda x)$. Eine derartige Funktion ist differenzierbar und genügt wegen (2.1) und der Kettenregel der Gleichung $f'(x) = \lambda f(x)$. Wegen $f(1) = \exp(\lambda)$ ist die Zahl λ gleich $\log f(1) = \log(a)$. Umgekehrt hat $x \mapsto \exp(\log(a)\cdot x)$ immer die im Satz genannten Eigenschaften. □

4 Der Logarithmus zur Basis a

Die Umkehrfunktion von $\exp_a$ für $a \neq 1$ wird *Logarithmus zur Basis a* genannt $\log_a\colon \mathbb{R}_+ \to \mathbb{R}$. Es gilt $\log_a'(x) = (\log a)^{-1}x^{-1}$. Der Logarithmus zur Basis a erfüllt die Funktionalgleichung $\log_a(xy) = \log_a(x) + \log_a(y)$. Ferner gilt die Umrechnungsformel $\log(a)\cdot\log_a(x) = \log(b)\cdot\log_b(x)$.

(4.1) Satz. *Sei $a > 0$ und $a \neq 1$. Es gibt genau eine stetige Funktion $g\colon \mathbb{R}_+ \to \mathbb{R}$ mit den Eigenschaften*
(1) $g(xy) = g(x) + g(y)$
(2) $g(a) = 1$.
Die Funktion ist differenzierbar, und es gilt $g'(x) = \log(a)^{-1}x^{-1}$.

Beweis. Sei $g\colon \mathbb{R}_+ \to \mathbb{R}$ mit den Eigenschaften (1) und (2) gegeben. Dann erfült $\ell\colon x \mapsto g(\exp(x))$ die Funktionalgleichung $\ell(x+y) = \ell(x) + \ell(y)$ und hat deshalb die Form $g(\exp(x)) = \lambda x$. Demnach ist $g(x) = \lambda\log(x)$. Der Wert λ bestimmt sich aus $1 = g(a) = \lambda\log(a)$. Die weiteren Behauptungen sind klar. □

5 Potenzen mit reellen Exponenten

(5.1) Satz. *Sei $\alpha \in \mathbb{R}$. Die Funktion $p_\alpha\colon \mathbb{R}_+ \to \mathbb{R}_+$, $x \mapsto x^\alpha$ ist differenzierbar und hat die Ableitung $p_\alpha'(x) = \alpha x^{\alpha-1}$.*

Beweis. Definitionsgemäß ist $x^\alpha = \exp(\alpha \log(x))$. Die Ableitung nach der Kettenregel ist $\alpha \cdot x^{-1} \cdot \exp(\alpha \log(x) = \alpha x^{\alpha-1}$. □

Wegen $(x^\alpha)^\beta = x^{\alpha\beta}$ ist $x \mapsto x^{1/\alpha}$ die Umkehrfunktion von $x \mapsto x^\alpha$. Insbesondere ist die n-te Wurzel von $x > 0$ durch $x^{1/n}$ im Sinne dieser Funktion gegeben.

6 Die Zahl e

Die Definition der Ableitung liefert die folgende Gleichungskette

$$1 = \log'(1) = \lim_{h\to 0} \frac{\log(1+h) - \log(1)}{h} = \lim_{h \to 0} \log(1+h)^{\frac{1}{h}} = \lim_{u\to\infty} \log\left(1 + \frac{1}{u}\right)^u .$$

Wenden wir auf diese Gleichung die stetige Funktion exp an, so erhalten wir nach II(2.6)

$$e = \exp(1) = \lim_{u\to\infty} \left(1 + \frac{1}{u}\right)^u .$$

Der rechts stehende Grenzwert tritt bei der sogenannten „stetigen Verzinsung“ auf.

Die Taylorsche Formel liefert für die Exponentialfunktion

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} e^\xi$$

mit einer Zahl ξ zwischen 0 und x. Für $x = 1$ und $n = 2$ entnimmt man daraus zunächst $e < 3$ und dann

$$\left| e - \sum_{k=0}^{n} \frac{1}{k!} \right| < \frac{3}{(n+1)!}.$$

Die Reihensumme bis $n = 7$ liefert die ersten drei richtigen Dezimalstellen nach dem Komma $e = 2.718\ldots$.

Die Fehlerabschätzung kann man auch dazu benutzen, ein interessantes theoretisches Resultat herzuleiten:

(6.1) Notiz. *e ist irrational.*

Beweis. Angenommen $e = p/q$, $p, q \in \mathbb{N}$. Dann liefert die Taylorsche Formel bis zur Stelle $q \geq 2$

$$0 < \left| q! \left(e - \sum_{k=0}^{q} \frac{1}{k!} \right) \right| < \frac{3}{q+1} \leq 1.$$

Der Ausdruck zwischen den Betragszeichen ist aber eine ganze Zahl. Widerspruch. □

Aus der Taylorschen Formel sehen wir, daß für $x \geq 0$ und $n \in \mathbb{N}$ die Ungleichung $e^x > \frac{x^{n+1}}{(n+1)!}$ gilt. Daraus entnehmen wir leicht $\lim_{x\to\infty} x^n e^{-x} = 0$. Man sagt deshalb: Die Exponentialfunktion wächst schneller als jede Potenz.

7 Die Differentialgleichung $y' = \alpha y$

Sei α eine reelle Zahl. Wir fragen nach differenzierbaren Funktionen f, für die immer

$$f'(x) = \alpha f(x)$$

gilt. Wir sagen dann, sie sind *Lösungen* der *Differentialgleichung* $y' = \alpha y$. Der folgenden Satz ist der Hauptsatz über diese Differentialgleichung.

(7.1) Satz. *Seien α und β reelle Zahlen. Die Funktion*

$$u\colon \mathbb{R} \to \mathbb{R}, \quad u(x) = \beta \exp(\alpha x)$$

ist die einzige differenzierbare Funktion $u\colon \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften

(1) $u'(x) = \alpha u(x)$
(2) $u(0) = \beta$.

Beweis. Nach der Kettenregel hat die Funktion $x \mapsto \beta \exp(\alpha x)$ jedenfalls diese Eigenschaften. Sei u eine weitere derartige Funktion. Wir haben im letzten Abschnitt gesehen, daß $\exp(\alpha x) \neq 0$ ist. Wir können deshalb die Funktion $x \mapsto v(x) = u(x)\exp(-\alpha x)$ bilden. Deren Ableitung ergibt sich nach den Ableitungsregeln und $u'(x) = \alpha u(x)$ als Null. Also ist v eine konstante Funktion. Ihr Wert $v(0)$ ergibt sich aus der Voraussetzung als 1. □

Die Menge der differenzierbaren Funktionen $u\colon \mathbb{R} \to \mathbb{R}$, die der Differentialgleichung $y' = \alpha y$ genügen, ist bezüglich punktweiser Addition und Skalarmultiplikation ein $\mathbb{R}$-Vektorraum L (*Lösungsraum*). Der letzte Satz besagt, daß dieser Raum eindimensional ist. Jedes feste x_0 liefert einen Isomorphismus $L \to \mathbb{R}$, $u \mapsto u(x_0)$. Man sagt dazu, daß die Lösung durch die *Anfangsbedingung* an der Stelle x_0 festgelegt ist. Ist $u(x_0) = \lambda$, so ist $u(x) = \lambda e^{\alpha(x-x_0)}$ die zugehörige Lösung.

(7.2) Satz. *Seien $a \neq 0$, b und λ reelle Zahlen. Die Funktion*

$$f\colon \mathbb{R} \to \mathbb{R}, \quad f(x) = \left(\lambda + \frac{b}{a}\right) e^{a(x-x_0)} - \frac{b}{a}$$

ist die einzige differenzierbare Funktion $\mathbb{R} \to \mathbb{R}$ mit den Eigenschaften

(1) $f'(x) = af(x) + b$
(2) $f(x_0) = \lambda$.

Beweis. Zunächst rechnet man leicht nach, daß die angegebene Funktion die Eigenschaften (1) und (2) hat. Sei g eine zweite Funktion mit diesen Eigenschaften. Dann hat $h = f - g$ die Eigenschaften $h' = ah$ und $h(x_0) = 0$ und ist deshalb nach dem vorigen Satz gleich Null □

8 Hyperbolische Funktionen

Aus der Exponentialfunktion werden einige andere gebräuchliche Funktionen definiert. Diese sind

$$\cosh(x) = \frac{1}{2}(e^x + e^{-x}), \quad \sinh(x) = \frac{1}{2}(e^x - e^{-x})$$

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad \coth(x) = \frac{\cosh(x)}{\sinh(x)} = \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

Sie heißen der Reihe nach *Cosinus hyperbolicus, Sinus hyperbolicus, Tangens hyperbolicus, Cotangens hyperbolicus.* Die ersten drei sind auf $\mathbb{R}$ definiert, während coth auf $\mathbb{R}^*$ erklärt ist.

9 Die Areafunktionen

Die (geeignet definierte) Umkehrfunktion der hyperbolischen Funktion $y = f(x)$ wird mit $x = \mathrm{Ar}f(y)$ bezeichnet (Ar sprich: area). $\mathrm{Ar}\sinh(x)$ ist für $x \in \mathbb{R}$ definiert und wächst streng monoton von $-\infty$ nach ∞. Der hyperbolische Cosinus besitzt zwei Umkehrfunktionen, eine im Intervall $]\infty]$ und eine im Intervall $[0, \infty[$. Wir verwenden letztere. Der Definitionsbereich von $\mathrm{Ar}\cosh$ ist $[1, \infty[$. Es ist $\mathrm{Ar}\tan$ in $]-1, 1[$ definiert, der Definitionsbereich von $\mathrm{Ar}\ \mathrm{coth}$ besteht aus den Stücken $]-\infty, -1[$ und $]1, \infty[$. Die Ableitungen errechnen sich zu

$$\mathrm{Ar}\sinh'(y) = \frac{1}{\sqrt{1+y^2}}, \qquad \mathrm{Ar}\cosh'(y) = \frac{1}{\sqrt{y^2-1}}, \quad y > 1$$

$$\mathrm{Ar}\tanh'(y) = \frac{1}{1-y^2}, \quad |y| < 1, \qquad \mathrm{Ar}\ \coth'(y) = \frac{1}{y^2-1}, \quad |y| > 1.$$

10 Aufgaben und Ergänzungen

1. Aus den Eigenschaften der Exponentialfunktion errechnet man die Additionstheoreme $\cosh(x+y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y)$ und $\sinh(x+y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y)$, sowie die Ableitungen $1 = \cosh^2(x) - \sinh^2(x)$, $\cosh'(x) = \sinh(x)$, $\sinh'(x) = \cosh(x)$, $\tanh'(x) = 1 - \tanh^2(x)$, $\coth'(x) = 1 - \coth^2(x)$. Die ebene Punktmenge $H = \{(x, y) \in \mathbb{R}^2 \mid x^2 - y^2 = 1\}$ ist eine Hyperbel. Deswegen ist $(\cosh(x), \sinh(x)$ für jedes x ein Punkt dieser Hyperbel.

2. Man kann die Umkehrfunktionen durch den Logarithmus ausrechnen:
$\mathrm{Ar}\sinh(y) = \log(y + \sqrt{1+y^2})$, $\mathrm{Ar}\cosh(y) = \log(y + \sqrt{y^2-1})$, $\mathrm{Ar}\tanh(y) = \frac{1}{2}\log\frac{1+y}{1-y}$, $\mathrm{Ar}\ \coth(y) = \frac{1}{2}\log\frac{y+1}{y-1}$.

3. Die Funktionen $\exp_a$ haben das folgende Grenzverhalten:

$$\lim_{x\to\infty} a^x = \begin{cases} \infty & a > 1 \\ 0 & a > 1 \end{cases} \qquad \lim_{x\to-\infty} a^x = \begin{cases} 0 & a > 1 \\ \infty & a < 1. \end{cases}$$

4. Die Funktionen $\log_a$ haben das folgende Verhalten bei Annäherung an die Grenzen des Definitionsbereiches:

$$\lim_{x\to\infty} = \begin{cases} \infty & a > 1 \\ -\infty & 0 < a < 1 \end{cases} \qquad \lim_{x\to 0} = \begin{cases} -\infty & a > 1 \\ \infty & 0 < a < 1. \end{cases}$$

5. Die Funktionen $x \mapsto x^\alpha$ für $\alpha \neq 1$ haben das folgende Grenzverhalten:

$$\lim_{x\to\infty} x^\alpha = \begin{cases} \infty & \alpha > 0 \\ 0 & \alpha < 0 \end{cases} \qquad \lim_{x\to 0} x^\alpha = \begin{cases} 0 & \alpha > 0 \\ \infty & \alpha < 0. \end{cases}$$

6. Sei $r \in \mathbb{N}$. Die Funktion $f\colon \mathbb{R} \to \mathbb{R}$, definiert durch

$$f(x) = \begin{cases} x^{-r}\exp(-x^{-2}) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

ist stetig.

7. Zeige durch Induktion nach n, daß die n-te Ableitung von $x \mapsto \exp(-x^{-2})$ die Form $x \mapsto P_n(x)x^{-r(n)}\exp(-x^{-2})$ hat, worin P_n ein geeignetes Polynom ist und $r(n) \in \mathbb{N}$.

8. Zeige, daß die Funktion $f\colon \mathbb{R} \to \mathbb{R}$

$$f(x) = \begin{cases} \exp(-x^{-2}) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

unendlich oft differenzierbar ist. Für alle $n \in \mathbb{N}$ gilt $f^{(n)}(0) = 0$.

9. Sei $\alpha > 0$. Dann gilt

$$\lim_{x\to\infty} \frac{\log x}{x^\alpha} = 0 \quad \text{und} \quad \lim_{x\to 0} x^\alpha \log x = 0.$$

10. $\lim_{x\to\infty} x^{\frac{1}{x}} = 1 = \lim \sqrt[n]{n}$. Zum Beweis zeige zunächst $\lim_{x\to\infty} \log(x^{\frac{1}{x}}) = \lim_{x\to\infty} \frac{1}{x}\log x = 0$. Wende darauf exp und den Vertauschungssatz an.

7 Trigonometrische Funktionen

1 Sinus und Cosinus

Die Funktionen Sinus und Cosinus werden, ähnlich wie die Exponentialfunktion, zunächst durch Eigenschaften definiert. Wir nehmen in diesem Abschnitt an, daß es Funktionen mit den Eigenschaften (1.1) gibt und leiten weitere Eigenschaften her. Es ist zweckmäßig, den Existenzbeweis zu verschieben.

(1.1) Definition. Die Funktionen *Sinus* sin und *Cosinus* cos sind differenzierbare Funktionen $\sin\colon \mathbb{R} \to \mathbb{R}$ und $\cos\colon \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften $\sin'(x) = \cos(x)$, $\cos'(x) = -\sin(x)$ sowie $\cos(0) = 1$, $\sin(0) = 0$. Aus dieser Definition folgt unmittelbar, daß Sinus und Cosinus beliebig oft differenzierbar sind. $\diamond$

Es gilt ferner $\cos^2(x) + \sin^2(x) = 1$, denn die Ableitung der linken Seite ist $-2\cos(x)\sin(x) + 2\sin(x)\cos(x) = 0$; also handelt es sich um eine konstante Funktion, die wegen (1.1) den Wert 1 hat. Es folgt, daß die Funktionswerte beider Funktionen im Intervall $[-1, 1]$ liegen und daß $(\cos(x), \sin(x))$ ein Punkt der Ebene auf dem Kreis vom Radius 1 um den Nullpunkt ist.

Geometrisch wird x als Winkel, gemessen im Bogenmaß gegen den Uhrzeigersinn, zwischen der positiven x-Achse und dem Halbstrahl vom Nullpunkt durch $(\cos(x), \sin(x))$ interpretiert. Die mathematische Begründung dafür werden wir nachliefern.

2 Die Differentialgleichung $y'' = -y$

Die zweite Ableitung der Funktionen sin und cos ist gleich dem Negativen der Ausgangsfunktion ist. Man sagt, sie genügen der Differentialgleichung $y'' = -y$.

(2.1) Satz. *Sei $f\colon \mathbb{R} \to \mathbb{R}$ eine zweimal differenzierbare Funktion, die der Gleichung $f''(x) = -f(x)$ für alle $x \in \mathbb{R}$ genügt. Dann gibt es reelle Zahlen α und β, so daß für alle $x \in \mathbb{R}$ die Gleichung $f(x) = \alpha \sin(x) + \beta \cos(x)$ gilt. Darin ist $\beta = f(0)$ und $\alpha = f'(0)$.*

Beweis. Seien α und β wie im Satz angegeben durch f festgelegt. Wir betrachten die Funktion $g(x) = f(x)\sin(x) + f'(x)\cos(x)$. Sie hat die Ableitung

$$f'(x)\sin(x) + f(x)\cos(x) + f''(x)\cos(x) - f'(x)\sin(x),$$

die wegen $f'' = -f$ gleich Null ist. Also handelt es sich um eine Konstante, die durch Einsetzen von $x = 0$ zu α bestimmt wird. Ebenso ergibt sich, daß $h(x) = f(x)\cos(x) - f'(x)\sin(x) = \beta$ ist. Wir lösen die Gleichungen nach $f(x)$ auf, $g(x)\sin(x) + h(x)\cos(x) = f(x)$, was nach dem schon Gezeigten die Behauptung des Satzes liefert. $\square$

Dieser Satz besagt insbesondere, daß durch die Eigenschaften (1.1) die Funktionen sin und cos eindeutig festgelegt sind.

3 Additionstheoreme

Die *Additionstheoreme* für $\sin(x)$ und $\cos(x)$ sind die im nächsten Satz niedergelegten Relationen.

(3.1) Satz. *Für alle a, x gilt*

$$\cos(a + x) = \cos(a)\cos(x) - \sin(a)\sin(x)$$

$$\sin(a + x) = \sin(a)\cos(x) + \cos(a)\sin(x).$$

Daraus ergeben sich speziell die Relationen

$$\sin(2x) = 2\sin(x)\cos(x), \quad \cos(2x) = \cos^2(x) - \sin^2(x), \quad 1 + \cos(2x) = 2\cos^2(x).$$

Beweis. Beide Seiten der behaupteten Gleichung erfüllen bei festem a als Funktionen von x die Voraussetzungen des Satzes (2.1). Dieser liefert die Formeln (3.1). □

Ebenfalls aus Satz (2.1) ergeben sich die folgenden Relationen $\cos(-x) = \cos(x)$ und $\sin(-x) = -\sin(x)$, denn die Funktionen $x \mapsto \cos(-x)$ und $x \mapsto \sin(-x)$ erfüllen auch die Differentialgleichung $y'' = -y$. Eine Funktion f mit $f(x) = f(-x)$ nennt man *gerade*, eine mit $f(-x) = -f(x)$ *ungerade*.

4 Nullstellen und Perioden

(4.1) Lemma. *Die Funktion* cos *hat eine kleinste positive Nullstelle.*

Beweis. Angenommen, $\cos(x) \neq 0$ für alle $x > 0$. Da $\cos(0) = 1$ ist, folgt nach dem Zwischenwertsatz, daß $\cos(x) > 0$ ist für $x > 0$. Demnach wäre sin für $x \geq 0$ streng monoton steigend und folglich $\sin x > 0$ für $x > 0$. Da cos und sin dem Betrag nach höchstens gleich 1 sind, wäre für $x \geq 0$ immer $0 \leq \cos^2 x \leq \cos x$ und $0 \leq \sin^2 x \leq \sin x$. Nach dem Hauptsatz und Rechenregeln würde folgen

$$x = \int_0^x dt = \int_0^x (\cos^2 t + \sin^2 t)\, dt \leq \int_0^x (\cos t + \sin t)\, dt = 1 - \cos x + \sin x \leq 3,$$

was für $x > 3$ unmöglich ist.

Sei S die Menge der positiven Nullstellen von cos und s ihr Infimum. Wegen der Stetigkeit von cos ist $\cos s = 0$ und wegen $\cos 0 = 1$ ist $s > 0$. □

Wir definieren die Kreiszahl π nun durch $2s = \pi$. Wegen $\sin' = \cos$ ist sin im Intervall $[0, \pi/2]$ streng monoton wachsend, also $\sin(\pi/2) > 0$. Wegen (1.4) ist dann $\sin(\pi/2) = 1$. Aus (3.1) folgt $\cos(\pi) = -1$. Indem man diese Ergebnisse in die Additionstheoreme einsetzt, erhält man nacheinander die folgenden Relationen.

$$\begin{array}{rclcrcl}
\cos(x \pm \frac{\pi}{2}) & = & \mp \sin(x) & \qquad & \sin(x \pm \frac{\pi}{2}) & = & \pm \cos(x) \\
\cos(x + \pi) & = & -\cos(x) & & \sin(x + \pi) & = & -\sin(x) \\
\cos(x + 2\pi) & = & \cos(x) & & \sin(x + 2\pi) & = & \sin(x)
\end{array}$$

(4.2) Notiz. *Die Nullstellenmenge von* cos *ist* $\{\frac{\pi}{2} + \pi n \mid n \in \mathbb{Z}\}$, *diejenige von* sin *ist* $\{\pi n \mid n \in \mathbb{Z}\}$.

Beweis. Nach Definition von π und wegen $\cos(x) = \cos(-x)$ hat cos im Intervall $]-\pi/2, \pi/2[$ keine Nullstellen. Sei x_0 irgendeine Nullstelle des Cosinus. Dann gibt es eine ganze Zahl k, so daß $-\pi/2 < x_1 = x_0 - k\pi \leq \pi/2$ ist. Wegen $\cos(x \pm \pi) = -\cos(x)$ sieht man, daß auch x_1 eine Nullstelle ist. Also ist $x_1 = \pi/2$. Ähnlich schließt man beim Sinus. □

Eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ heißt *periodisch* mit der Periode a, wenn für alle x die Gleichung $f(x + a) = f(x)$ gilt. Die Funktionen sin und cos sind also periodisch mit der Periode 2π. Periodische Funktionen sind das mathematische Werkzeug für die Beschreibung von Schwingungsvorgängen.

5 Tangens und Cotangens

Die Funktionen *Tangens* $\tan(x)$ und *Cotangens* $\cot(x)$ werden durch

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \qquad \cot(x) = \frac{\cos(x)}{\sin(x)}$$

definiert. Ihr Definitionsbereich sind die reellen Zahlen ohne die Nullstellen des jeweiligen Nenners. Aus den Eigenschaften von sin und cos leitet man die folgenden Aussagen über diese Funktionen her.

Beide Funktionen haben die Periode π und sind ungerade. Ihre Ableitungen sind

(5.1) $$\tan'(x) = 1 + \tan^2(x), \qquad \cot'(x) = -1 - \cot^2(x).$$

Deshalb ist der Tangens im Intervall $]-\frac{\pi}{2}, \frac{\pi}{2}[$ streng monoton wachsend und hat als Wertebereich die Menge der reellen Zahlen. Der Cotangens ist im Intervall $]0, \pi[$ streng monoton fallend und der Wertebereich ist ebenfalls die Menge der reellen Zahlen. Es gelten die *Additionstheoreme*

(5.2) $$\tan(x+y) = \frac{\tan(x) + \tan(y)}{1 - \tan(x)\tan(y)}, \qquad \cot(x+y) = \frac{\cot(x)\cot(y) - 1}{\cot(x) - \cot(y)}.$$

Im ersten Fall dürfen x, y und $x+y$ keine Nullstellen von cos sein. Sie folgen aus den Additionstheoremen für sin und cos durch Umrechnung.

6 Umkehrfunktionen

Die Umkehrfunktionen der trigonometrischen Funktionen werden *Arcus-Funktionen* genannt. Da die trigonometrischen Funktionen wegen ihrer Periodizität keine globale Umkehrfunktionen haben, muß man zur Definition der Umkehrfunktionen den Definitionsbereich geeignet einschränken. Wir treffen die folgenden Vereinbarungen.

Die Umkehrfunktion von $\tan\colon]-\frac{\pi}{2}, \frac{\pi}{2}[\to \mathbb{R}$ heißt *arcus tangens*

$$\arctan\colon \mathbb{R} \to]-\frac{\pi}{2}, \frac{\pi}{2}[\,.$$

Die Umkehrfunktion von $\cot\colon]0, \pi[\to \mathbb{R}$ heißt *arcus cotangens*

$$\operatorname{arccot}\colon \mathbb{R} \to]0, \pi[\,.$$

Der Sinus ist im Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$ streng monoton wachsend und hat die Umkehrfunktion *arcus sinus*

$$\arcsin\colon [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}],$$

und der Cosinus ist im Intervall $[0, \pi]$ streng monoton fallend und hat die Umkehrfunktion *arcus cosinus*

$$\arccos\colon [-1, 1] \to [0, \pi].$$

Für die Ableitungen dieser Funktionen erhält man aus der Umkehrregel

$$\begin{aligned} \arcsin'(x) &= \frac{1}{\sqrt{1-x^2}} & \arccos'(x) &= -\frac{1}{\sqrt{1-x^2}} \\ \arctan'(x) &= \frac{1}{1+x^2} & \operatorname{arccot}'(x) &= -\frac{1}{1+x^2}. \end{aligned}$$

In der ersten Zeile muß natürlich $x \neq \pm 1$ sein. Durch diese Formeln erhält man auch neue Stammfunktionen. Aus dem Zusammenhang zwischen Sinus und Cosinus und den Periodizitäten gewinnt man Umkehrfunktionen zu anderen Definitionsintervallen.

Wir haben bisher noch nicht nachgewiesen, daß die trigonometrischen Funktionen existieren. An dieser Stelle könnte man folgenden Weg einschlagen: Man definiert den arcus tangens durch

$$\arctan(x) = \int_0^x \frac{1}{1+t^2}\, dt,$$

sodann den Tangens (im Intervall $]-\pi/2, \pi/2[$) als Umkehrfunktion und entwickelt von da aus die Eigenschaften der trigonometrischen Funktionen.

7 Die komplexe Exponentialfunktion

Die Additionstheoreme lassen ihre eigentliche Bedeutung erkennen, wenn man komplexe Zahlen oder lineare Algebra verwendet.

Wir benutzen die folgende Bezeichnung (*Eulersche Formel*)

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

Ein Grund für die Exponentialschreibweise: Die Additionstheoreme sind äquivalent zu der Formel

$$e^{i(t+u)} = e^{it} e^{iu},$$

indem man sie in Real- und Imaginärteil zerlegt. (Warum man gerade e als Basis für diese Exponentialfunktion verwendet, wird bei Gelegenheit der Exponentialreihe deutlich werden.) Wir setzen für eine beliebige komplexe Zahl $z = a + bi$ $e^z = e^a e^{ib}$ und erhalten dann das für alle komplexen Zahlenpaare (z, w) gültige Additionstheorem

(7.2) $$e^{z+w} = e^z e^w.$$

Wir schreiben auch $\exp(z) = e^z$ und nennen $z \mapsto \exp(z)$ die *komplexe Exponentialfunktion.* Wegen der Periodizität von sin und cos gilt $e^{z+2\pi i n} = e^z$ für $z \in \mathbb{C}$ und $n \in \mathbb{Z}$. Deshalb hat exp im Komplexen keine (globale) Umkehrfunktion, und man kann nicht ohne Vorsicht den Logarithmus im Komplexen definieren. Die zu e^{it}, $t \in \mathbb{R}$ konjugiert-komplexe Zahl ist e^{-it}. Die Zahlen e^{it} liegen alle auf dem Einheitskreis S^1 der komplexen Zahlen vom Betrag 1. Jede Zahl auf dem Einheitskreis hat die Form e^{it}, wie der nächste Satz zeigt.

(7.3) Satz. *Die Abbildung* $\kappa\colon\,]-\pi, \pi] \to S^1$, $t \mapsto (\cos t, \sin t)$ *ist bijektiv.*

Beweis. Wir nehmen aus dem Kreis $S^1 = \{(x,y) \mid x^2+y^2=1\}$ den Punkt $(-1,0)$ heraus, $S^1_- = S^1 \backslash \{(-1,0)\}$. Eine kleine Rechnung zeigt: Die Abbildungen

$$\alpha\colon \mathbb{R} \to S^1_-,\ t \mapsto \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right), \qquad \beta\colon S^1_- \to \mathbb{R},\ (x,y) \mapsto \frac{y}{1+x}$$

sind (wohldefiniert und) Umkehrfunktionen voneinander. Unsere Kenntnis der Nullstellen von sin zeigt, daß für $t \in]-\pi,\pi[$ der Wert $\kappa(t)$ in S^1_- liegt. Der Satz (7.3) folgt jetzt offenbar, wenn wir zeigen: Die Einschränkung $\kappa_0\colon]-\pi,\pi[\to S^1_-$ von κ ist bijektiv. Da β bijektiv ist, genügt es zu zeigen, daß $\beta\kappa_0$ bijektiv ist. Es ist

$$\beta\kappa_0(t) = \frac{\sin(t)}{1+\cos(t)} = \tan(\frac{t}{2}).$$

Die letzte Gleichung benutzt die Additionstheoreme (3.1). Diese Funktion ist uns schon als bijektiv bekannt. □

Die Abbildung β hat die folgende geometrische Bedeutung. Die Gerade durch die Punkte $(-1,0)$ und $(x,y) \in S^1_-$ schneidet die Parallele zur y-Achse durch $(1,0)$ an der Stelle $(1, 2\beta(x,y))$. Die Zuordnung $(x,y) \mapsto 2\beta(x,y)$ wird als *stereographische Projektion* bezeichnet. Anschaulich ist klar, daß die stereographische Projektion eine Bijektion ist.

Aus dem Vorangehenden erkennt man, daß jede komplexe Zahl die Form

$$z = re^{i\varphi}, \qquad r \geq 0,\ 0 \leq \varphi < 2\pi$$

hat. Darin heißen (r,φ) die *Polarkoordinaten* von z. Ist $t \neq 0$, so sind (r,φ) im angegebenen Bereich durch z eindeutig bestimmt.

8 Drehungen und Spiegelungen

Die Multiplikation $z \mapsto e^{i\varphi}z$ ist, geometrisch gesprochen, die Drehung der komplexen Zahlenebene $\mathbb{C}$ um den Winkel φ gegen den Uhrzeigersinn. Diese Drehung der Ebene $\mathbb{R}^2$, aufgefaßt als $\mathbb{R}$-lineare Abbildung, hat bezüglich der Standardbasis $e_1 = (1,0), e_2 = (0,1)$ die Matrix

$$D(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

Die Additionstheoreme sind äquivalent zur Matrizengleichung der Drehungsmatrizen

$$D(\varphi)D(\psi) = D(\varphi+\psi),$$

die ja geometrisch einen guten Sinn hat. Die *Spiegelung* an dem Unterraum durch $(\cos\varphi, \sin\varphi)$ wird durch die Matrix

$$S(\varphi) = \begin{pmatrix} \cos 2\varphi & \sin 2\varphi \\ \sin 2\varphi & -\cos 2\varphi \end{pmatrix}$$

vermittelt. Drehungen und Spiegelungen bilden die orthogonale Gruppe $O(2)$.

9 Bogenmaß. Kurvenlänge

Die *euklidische Norm* $\|x\|$ eines Vektors $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ ist durch

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

definiert. Die Zahl $d(x,y) = \|x - y\|$ heißt *euklidischer Abstand* von x und y. Eine Abbildung $f\colon J \to \mathbb{R}^n$ hat die Form $f(x) = (f_1(x), \dots, f_n(x))$ und ist deshalb dasselbe wie ein System von n Abbildungen $f_1, \dots, f_n\colon J \to \mathbb{R}$. Wir nennen f_k die k-te Komponente von f. Ist $J \subset \mathbb{R}$, so heißt f stetig (differenzierbar) an der Stelle $x \in J$, wenn alle f_k an dieser Stelle stetig (differenzierbar) sind. Wir setzen

$$f'(t) = (f_1'(t), \dots, f_n'(x)).$$

Eine Abbildung $w\colon J \to \mathbb{R}^3$ wird als *Kurve* oder *Weg* im Raum $\mathbb{R}^3$ vorgestellt, oder als Bewegungslinie eines Punkts, wenn $t \in J$ als Zeitparameter angesehen wird. In diesem Fall ist $w'(t)$ der *Geschwindigkeitsvektor* von w zur Zeit t.

Sei $w\colon [a,b] \to \mathbb{R}^n$ stetig differenzierbar. Wir definieren als die *Länge* $L(w)$ *der Kurve* w das Integral

$$\textbf{(9.1)} \qquad L(w) = \int_a^b \|w'(t)\|\, dt.$$

Es ist $w\colon [0,t] \to S^1$, $t \mapsto (\cos(t), \sin(t))$ für $0 \le t < 2\pi$ eine bei $(1,0)$ startende Kurve auf dem Einheitskreis. Ihre Länge ist nach der eben gegebenen Definition wegen $w'(t) = (-\sin(t), \cos(t))$ gleich

$$\int_0^t \|w'(t)\|\, dt = \int_0^t (\cos^2(t) + \sin^2(t))^{1/2}\, dt = \int_0^t dt = t.$$

Also ist die Zahl t in $\cos(t)$ als Bogenmaß des orientierten Winkels zu interpretieren. Wir sehen auch, daß ein Kreis vom Radius R den Umfang $2\pi R$ hat.

Die Definition der Kurvenlänge wird durch einen Grenzprozeß aus Streckenlängen motiviert. Sei $Z = (x_0, \dots, x_n)$ eine Zerlegung des Intervalls $[a,b]$. Die euklidische Länge der Strecke von $f(x_{i-1})$ nach $f(x_i)$ ist $\|f(x_i) - f(x_{i-1})\|$. Da die Gerade die kürzeste Verbindung zwischen zwei Punkten ist, so ist die „wirkliche Länge“ der Kurve mindestens gleich

$$L(f,Z) = \sum_{i=0}^{n-1} \|f(x_i) - f(x_{i-1})\|.$$

Je feiner die Zerlegung ist, desto besser wird die Kurve durch den Streckenzug approximiert. Deshalb liegt es nahe, das Supremum der Werte $L(f,Z)$ zu betrachten.

Wir definieren: Ein Kurve $f\colon [a,b] \to \mathbb{R}^n$ heißt *rektifizierbar*, wenn das Supremum $L(f) = \sup L(f,Z)$, gebildet über alle Zerlegungen z von $[a,b]$, als reelle Zahl existiert.

(9.2) Satz. *Ist f stetig differenzierbar, so ist $L(f)$ durch das Integral* (9.1) $\int_a^b \|f'(t)\|\, dt$ *gegeben.* □

10 Integrale

Ist J ein Intervall und $f\colon J \to \mathbb{C}$ stetig, so definieren wir

$$\int_a^b f(t)\,dt = \int_a^b u(t)\,dt + i\int_a^b v(t)\,dt,$$

wenn $f(t) = u(t) + iv(t)$ die Zerlegung in Real- und Imaginärteil ist. Ebenso definieren wir $f'(t) = u'(t) + iv'(t)$, wenn u und v differenzierbar sind.

(10.1) Satz. *Seien $f, g\colon [a,b] \to \mathbb{R}$ stetig und seien λ, μ komplexe Zahlen. Dann gilt:*

(1) $\int_a^b(\lambda f + \mu g) = \lambda\int_a^b f + \mu\int_a^b g$

(2) $\left|\int_a^b f\right| \le \int_a^b |f|$

(3) *Ist f stetig differenzierbar, so ist $\int_a^b f'(t)\,dt = f(b) - f(a)$.*

Beweis. (1) folgt aus der analogen Aussage für reelle Funktionen durch Zerlegung in Real- und Imaginärteil. Ebenso (3) aus dem Hauptsatz der Differential- und Integralrechnung.
(2) Wir schreiben $\int_a^b f = r\alpha$ mit $r \ge 0$ und $|\alpha| = 1$. Damit rechnen wir

$$\begin{aligned} & |\int f| = |\alpha^{-1}\int f| = \alpha^{-1}\int f = \int \alpha^{-1} f \\ = & \int \mathrm{Re}(\alpha^{-1}f) + i\mathrm{Im}(\alpha^{-1}f) = \int \mathrm{Re}(\alpha^{-1}f) \le \int |\mathrm{Re}(\alpha^{-1}f| \\ \le & \int |\alpha^{-1}f| = \int |f|. \end{aligned}$$

(Hier ist natürlich Re der Real- und Im der Imaginärteil.) □

Die Funktion $t \mapsto e^{\lambda t}$ hat für alle $\lambda \in \mathbb{C}$ die Ableitung $t \mapsto \lambda e^{\lambda t}$. Das sieht aus wie eine Kettenregel, muß aber mit unseren Definitionen mittels Zerlegung in Real- und Imaginärteil nachgewiesen werden. Für $\lambda \ne 0$ ist also $t \mapsto \lambda^{-1}e^{\lambda t}$ eine Stammfunktion von $t \mapsto e^{\lambda t}$, und deshalb gilt in diesem Fall

(10.2)
$$\int_a^b e^{\lambda t}\,dt = \lambda^{-1}(e^{\lambda b} - e^{\lambda a}).$$

Speziell ist $\int_a^{a+2\pi} e^{int}\,dt = 0$, sofern $0 \ne n \in \mathbb{Z}$. Daraus erhält man durch Umrechnung mit den Formeln

(10.3)
$$\cos t = \frac{1}{2}(e^{it} + e^{-it}), \qquad \sin t = \frac{1}{2i}(e^{it} - e^{-it})$$

die Integrale

$$\int_a^{a+2\pi} \cos mt \cdot \cos nt\,dt \quad = \quad \pi\delta_{m,n}$$

$$\int_a^{a+2\pi} \cos mt \cdot \sin nt \, dt = 0$$
$$\int_a^{a+2\pi} \sin mt \cdot \sin nt \, dt = \pi\delta_{m,n},$$

worin $\delta_{m,n}$ das *Kronecker–Symbol* (Einheitsmatrix) ist: $\delta_{m,n} = 1$ für $m = n$ und $= 0$ sonst.

(10.4) Beispiel. Wir berechnen die Fläche des Kreises. Die Fläche des Halbkreises vom Radius 1 wird durch $\int_{-1}^{1} \sqrt{1-x^2}\, dx$ gegeben. Wir machen die Substitution $x = \sin t$ und erhalten nach (10.2) das Integral

$$\int_{-\pi/2}^{\pi/2} \cos^2 t \, dt = \frac{\pi}{2}.$$

Die Fläche des Halbkreises vom Radius R ist $\int_{-R}^{R} \sqrt{R^2 - x^2}\, dx$. Das Integral wird durch die Substitution $x = Rt$ in $R^2 \int_{-1}^{1} \sqrt{1-t^2}\, dt$ verwandelt. Also ist πR^2 die Fläche des Kreises vom Radius R. $\diamond$

11 Existenz von Sinus und Cosinus

Die Funktion

$$F\colon [-1,1] \to \mathbb{R}, \quad x \mapsto \frac{1}{2}x\sqrt{1-x^2} + \int_x^1 \sqrt{1-t^2}\, dt$$

hat auf $]-1,1[$ die Ableitung

$$F'(x) = -\frac{1}{2\sqrt{1-x^2}}$$

und ist deshalb streng monoton fallend. Es ist $F(-1) = \int_{-1}^{1} \sqrt{1-t^2}\, dt$ und $F(1) = 0$. Wir bezeichnen $F(-1)$ mit $\pi/2$. Sei $G\colon [0,\pi/2] \to [-1,1]$ die Umkehrfunktion von F. Für $x \in [0,\pi]$ definieren wir $\cos(x) = G(\frac{x}{2})$ und $\sin(x) = \sqrt{1-x^2}$. Nach der Umkehrformel berechnen wir die Ableitungen dieser Funktionen zu $\cos'(x) = -\sin(x)$ und $\sin'(x) = \cos(x)$. Es folgt außerdem $\cos(\pi/2) = 0$ und $\sin(\pi/2) = 1$. Durch die Festsetzungen $\cos(x) = \cos(-x)$ und $\sin(x) = -\sin(-x)$ für $x \in [-\pi, 0]$ sowie $\cos(x) = \cos(x + 2\pi k)$ und $\sin(x) = \sin(x + 2\pi k)$ für $k \in \mathbb{Z}$ und $x \in [-\pi,\pi]$ erhalten wir stetige Funktionen $\cos\colon \mathbb{R} \to [-1,1]$ und $\sin\colon \mathbb{R} \to [-1,1]$. Man weist sie ohne Mühe als differenzierbar nach. Sie erfüllen die Bedingungen (1.1). Die hier festgelegte Zahl $\pi/2$ ist dann auch nach Konstruktion die kleinste positive Nullstelle des Cosinus. Zum Nachweis der Differenzierbarkeit an den „Nahtstellen“ benutzt man das

(11.1) Lemma. *Sei $f\colon J \to \mathbb{R}$ eine auf dem nichtentarteten Intervall J stetige Funktion. Sei $p \in J$ und f auf $J \setminus \{p\}$ differenzierbar. Gelte $\lim_{x\to p} f'(x) = a$. Dann ist f bei p differenzierbar, und es gilt $f'(p) = a$.*

Beweis. Nach Voraussetzung über den Limes gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß $|f'(p+t) - a| < \varepsilon$ für $0 < |t| < \delta$. Es folgt mit dem Mittelwertsatz

$$\left| \frac{f(p+h) - f(x)}{h} - a \right| = |f'(\xi) - a| < \varepsilon,$$

wenn $|p| < \delta$ ist. Also ist der Limes des Differenzenquotienten wie behauptet. □

Einen weiteren Existenzbeweis geben wir nach der Behandlung der unendlichen Reihen. Dadurch wird dann auch die Beziehung zur Exponentialfunktion präzisiert.

12 Aufgaben und Ergänzungen

1. Verifiziere die folgenden Werte: $\cos\frac{\pi}{3} = \frac{1}{2}$, $\cos\frac{\pi}{4} = \frac{1}{2}\sqrt{2}$, $\cos\frac{\pi}{6} = \frac{1}{2}\sqrt{3}$, $\cos\frac{2\pi}{5} = \frac{1}{4}(\sqrt{5}-1)$, $\cos\frac{\pi}{8} = \frac{1}{2}\sqrt{2+\sqrt{2}}$.
2. Seien a und c reelle Zahlen. Sei $c > 0$ Sei $f\colon[-1,1] \to \mathbb{R}$ definiert durch

$$f(x) = \begin{cases} x^a \sin x^{-c} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Zeige die folgenden Aussagen:
(1) f stetig $\iff a > 0$.
(2) $f'(0)$ existiert $\iff a > 1$.
(3) f' beschränkt $\iff a \geq 1 + c$.
(4) f' stetig $\iff a > 1 + c$.

3. Jede stetige Funktion $f\colon\mathbb{R} \to \mathbb{C} \setminus \{0\}$ mit $f(x+y) = f(x)f(y)$ hat die Form $f(t) = \exp(\lambda t)$ mit einem $\lambda \in \mathbb{C}$.

8 Folgen und Reihen

1 Folgen

Ein *Folge* komplexer Zahlen ist eine Funktion $a\colon\mathbb{N} \to \mathbb{C}$. Für den Funktionswert $a(n)$ schreiben wir oft a_n und nennen ihn das n-te Glied der Folge. Wir notieren die Folge dann in der Form $(a_n \mid n \in \mathbb{N})$ oder kurz (a_n); oder wir schreiben einfach $a_1, a_2, a_3, \ldots$, wenn wir wissen, wie es weitergeht, wenn also das „Bildungsgesetz" der a_n bekannt oder ersichtlich ist. Auch Abbildungen $\{n \in \mathbb{Z} \mid n \geq k\} \to \mathbb{C}$ bezeichnen wir als Folge $a_k, a_{k+1}, \ldots$ Ist $\mu\colon\mathbb{N} \to \mathbb{N}$ eine injektive monoton wachsende Abbildung, so heißt $n \mapsto a_{\mu(n)}$ eine *Teilfolge* der Folge (a_n).

Liegt für jedes $n \in \mathbb{N}$ eine Aussage $A(n)$ vor und gilt $A(n)$ für alle $n \in \mathbb{N}$ bis auf endlich viele, so sagen wir, $A(n)$ gelte für *fast alle* oder *schließlich alle* n.

Da eine Folge eine Funktion ist, haben wir im zweiten Kapitel schon definiert, was $\lim_{n\to\infty} a_n = a$ bedeutet. Wir wiederholen die relevanten Definitionen und Sätze.

(1.1) Definition. Ein Folge $(a_n \mid n \in \mathbb{N})$ *konvergiert* zum *Grenzwert* a, wenn zu jeder Zahl $\varepsilon > 0$ eine Zahl N existiert, so daß für alle $n \geq N$ die Ungleichung $|a - a_n| < \varepsilon$ gilt. Wir sagen dann auch, a ist der *Limes* der Folge. Ist der Limes Null, so sprechen wir von einer *Nullfolge*. Eine Folge heißt *konvergent*, wenn sie einen Grenzwert hat, andernfalls *divergent*. Die symbolische Gleichung

$$\lim_{n\to\infty} a_n = a$$

(gelesen: Limes a_n für n gegen Unendlich ist gleich a) ist eine Abkürzung für den Satz: Die Folge (a_n) konvergiert zum Grenzwert a. Ist das der Fall, so verwenden wir das Symbol $\lim_{n\to\infty} a_n$ auch als Bezeichnung für die Zahl a. Wir lassen das Symbol $n \to \infty$ am lim-Zeichen auch manchmal weg.

Wir verwenden ferner noch die folgenden Bezeichnungen: $\lim_{n\to\infty} a_n = \infty$ (bzw. $\lim_{n\to\infty} a_n = -\infty$) bedeutet: Zu jedem $A \in \mathbb{R}$ gibt es ein N, so daß für alle $n \geq N$ die Ungleichung $a_n > A$ (bzw. $a_n < A$) gilt. Wir sprechen in diesem Falle aber nicht von Konvergenz. $\diamond$

(1.2) Bemerkungen. Der Limes einer Folge ist durch die Folge eindeutig bestimmt. Endlich viele Glieder am Anfang der Folge spielen für die Konvergenzuntersuchung keine Rolle, das heißt $(a_n \mid n \in \mathbb{N})$ konvergiert genau dann, wenn $a_k, a_{k+1}, a_{k+2}, \ldots$ konvergiert, und der Grenzwert ist in beiden Fällen derselbe. Diese Bemerkung wird stillschweigend verwendet. Konvergiert eine Folge, so hat jede Teilfolge denselben Limes. $\diamond$

(1.3) Beispiele. Die Folge $a_n = \frac{1}{n}$ ist eine Nullfolge I(2.4). Die Folge (q^n) ist für $|q| < 1$ eine Nullfolge I(2.5). Die Folge $a_n = (-1)^n$ divergiert. Die konstante Folge $a_n = a$ hat den Limes a.

Es gilt $\lim_{n\to\infty} \sqrt[n]{n} = 1$. Die Ungleichung $\sqrt[n]{n} < 1+\varepsilon$ ist nämlich äquivalent zu $n < (1+\varepsilon)^n$. Aus der binomischen Formel entnimmt man $(1+\varepsilon)^n > \frac{1}{2}n(n-1)\varepsilon^2$. Also gilt $\sqrt[n]{n} < 1 + \varepsilon$, sofern $n \geq N = 2\varepsilon^{-2} + 1$ ist. $\diamond$

(1.4) Definition. Ein Folge (a_n) heißt *Cauchy-Folge*, wenn es zu jedem $\varepsilon > 0$ ein N gibt, so daß für alle $m, n > N$ die Ungleichung $|a_m - a_n| < \varepsilon$ gilt. $\diamond$

(1.5) Notiz. *Sei (z_n) eine Folge komplexer Zahlen und sei $z_n = x_n + iy_n$ die Zerlegung von z_n in Realteil x_n und Imaginärteil y_n. Genau dann konvergiert (z_n), wenn die beiden Folgen (x_n) und (y_n) konvergieren, und es gilt dann $\lim z_n = \lim x_n + i \lim y_n$. Genau dann ist (z_n) eine Cauchy-Folge, wenn (x_n) und (y_n) Cauchy-Folgen sind.*

Beweis. Die Aussagen folgen unmittelbar aus den Betragsrelationen

$$|x - u| \leq |z - w| \leq |x - u| + |y - v|,$$

worin $z = x + iy$ und $w = u + iv$ die Zerlegungen in Real- und Imaginärteil sind. □

Der folgende Satz wird *Konvergenzkriterium von Cauchy* genannt. Er ist von fundamentaler Bedeutung, da er gestattet, die Konvergenz ohne Kenntnis des Grenzwertes nachzuweisen. Der Satz ist ein Spezialfal von III(2.8). Wir beweisen ihn noch einmal nach derselben Methode.

(1.6) Satz. *Eine Folge konvergiert genau dann, wenn sie eine Cauchy-Folge ist.* □

Beweis. Wegen (1.5) genügt es, Folgen reeller Zahlen zu betrachten. Konvergiert (x_n) gegen x, so zeigt $|x_m - x_n| \le |x_m - x| + |x_n - x|$ unmittelbar, daß (x_n) eine Cauchy-Folge ist

Sei umgekehrt (x_n) eine Cauchy-Folge. Dann gibt es ein N, so daß $|x_m - x_n| < 1$ ist für $m, n > N$. Wegen $|x_m| < |x_n| + 1$ bei festem n und beliebigem $m > N$ sehen wir, daß die Wertemenge der Folge beschränkt ist.

Eine Zahl K heiße *obere Fastschranke* von (x_n), wenn $x_n > K$ nur für endlich viele n gilt. Eine obere Schranke der Folgenwerte ist auch eine obere Fastschranke, und jede obere Fastschranke ist größergleich dem Infimum der Folgenwerte. Sei x das Infimum der oberen Fastschranken. Wir behaupten: Die Cauchy-Folge (x_n) konvergiert gegen x.

Sei $\varepsilon > 0$ gegeben und sei $|x_m - x_n| < \varepsilon$ für $m, n > N$. Es ist $x + \varepsilon$ obere Fastschranke, also gibt es $M \ge N$, so daß $x_n \le x + \varepsilon$ für $n > M$. Es ist $x - \varepsilon$ keine obere Fastschranke. Wir können deshalb ein $k > M$ wählen, so daß $x_k > x - \varepsilon$ ist. Für dieses k ist also $|x - x_k| < \varepsilon$. Es folgt

$$|x - x_n| \le |x_k - x| + |x_k - x_n| < \varepsilon + \varepsilon$$

für $n > M$. □

Das im Beweis verwendete Infimum x der oberen Fastschranken einer nach oben beschränkten Folge (x_n) reeller Zahlen heißt *Limes superior* $\limsup(x_n) = x$ der Folge. Ebenso hat man das Supremum der *untereren Fastschranken*, den *Limes inferior* $\liminf$ der Folge. Wir setzen außerdem $\limsup(x_n) = \infty$, wenn die Folge nach oben nicht beschränkt ist. Die Folge $((-1)^n \mid n \in \mathbb{N})$ hat den Limes superior 1 und den Limes inferior -1. Ein beschränkte Folge konvergiert genau dann, wenn Limes superior und Limes inferior übereinstimmen.

Die folgenden Rechenregeln sind Spezialfälle über das Rechnen mit Grenzwerten von Funktionen III(4.5). Es sei an die Beziehung zwischen Konvergenz und Stetigkeit erinnert. Sei $\mathbb{N}^{-1} = \{n^{-1} \mid n \in \mathbb{N}\}$ und $S = \{0\} \cup \mathbb{N}^{-1}$. Jeder Folge $a\colon \mathbb{N} \to \mathbb{C}$ ordnen wir die Funktion $\tilde{a}\colon \mathbb{N}^{-1} \to \mathbb{C}, n^{-1} \mapsto a(n)$ zu. Wir erweitern diese Funktion durch $\tilde{a}(0) = a$ zu einer Funktion auf S. Genau dann gilt $\lim_{n\to\infty} a_n = a$, wenn $\tilde{a}\colon S \to \mathbb{C}$ bei 0 stetig ist.

(1.7) Rechenregeln für Grenzwerte. *Seien (a_n) bzw. (b_n) konvergente Folgen mit den Grenzwerten a bzw. b. Dann gilt:*

(1) $\lim_{n\to\infty}(a_n + b_n) = a + b$.
(2) $\lim_{n\to\infty}(a_n b_n) = ab$.
(3) *Gilt* $a_n \neq 0$ *und* $a \neq 0$, *so ist* $\lim_{n\to\infty}(b_n/a_n) = b/a$.
(4) *Gilt für fast alle* n *die Ungleichung* $a_n \leq b_n$, *so ist* $a \leq b$. *Ist speziell* $m \leq a_n \leq M$ *für fast alle* n, *so gilt* $m \leq a \leq M$.
(5) *Ist* (c_n) *eine weitere Folge, gilt* $a_n \leq c_n \leq b_n$ *für fast alle* n *und ist* $a = b$, *so konvergiert* (c_n) *zum Limes* a.

In Teil (1) ist natürlich gemeint, daß die Folge mit dem n*-ten Glied* $a_n + b_n$ *betrachtet wird. Analog in (2) und (3). Speziell gilt für eine komplexe Zahl* λ *die Gleichheit* $\lim_{n\to\infty}(\lambda a_n) = \lambda a$. □

(1.8) Beispiel. Sei (a_n) eine monoton wachsende Folge reeller Zahlen. Dann gilt $\lim_{n\to\infty} a_n = \sup\{a_n \mid n \in \mathbb{N}\}$. Ist (a_n) eine monoton fallende Folge reeller Zahlen, so gilt $\lim_{n\to\infty} a_n = \inf\{a_n \mid n \in \mathbb{N}\}$. Diese Aussagen folgen unmittelbar aus der Definition eines Supremums oder Infimums. ◇

Die Stetigkeit läßt sich mittels Folgenkonvergenz formulieren.

(1.9) Satz. *Sei* $A \subset \overline{\mathbb{R}}$ *und sei* $f\colon A \to \overline{\mathbb{R}}$ *eine Funktion. Genau dann ist* f *bei* $a \in A$ *stetig, wenn für jede Folge* $x\colon \mathbb{N} \to A$ *mit* $\lim_{n\to\infty}(x_n) = a$ *die Gleichung* $\lim_{n\to\infty} f(x_n) = f(a)$ *gilt.*

Beweis. Ist f bei a stetig, so ist die behauptete Gleichung ein Spezialfall des Vertauschungssatzes III(4.7).

Ist, umgekehrt, f bei a nicht stetig, so gibt es eine Umgebung U von $f(a)$, so daß in jeder Umgebung V von a ein Element x_V mit $f(x_V) \notin U$ existiert. Ist $a \in \mathbb{R}$, so wählen wir ein solches Element $x_n \in U_{1/n}(a)$. Dann konvergiert (x_n) gegen a. Wegen $\lim f(x_n) = f(a)$, gibt es ein N, so daß $f(x_n)$ für $n \geq N$ in U liegt. Widerspruch. Ebenso im Fall $a = \pm\infty$. □

2 Reihen

Sei $(a_n \mid n \in \mathbb{N})$ eine Folge komplexer Zahlen. Die Summe

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{k=1}^{n} a_k$$

heißt n-te *Partialsumme* der Folge (a_n). Die Folge $(s_n \mid n \in \mathbb{N})$ der Partialsummen wird als *unendliche Reihe* mit den Gliedern a_n bezeichnet und durch das Symbol

$$\sum_{j=1}^{\infty} a_j = a_1 + a_2 + a_3 + \cdots$$

notiert. Natürlich kann man nicht unendlich viele Zahlen addieren, weshalb dieses Symbol zunächst keinen Zahlenwert repräsentiert. Die unendliche Reihe heißt *konvergent* zum Grenzwert s, wenn $s = \lim_{n\to\infty} s_n$ ist. In diesem Fall schreiben wir

$$\sum_{j=1}^{\infty} a_n = s,$$

und die linke Seite des Gleichheitszeichens steht dann auch für den Zahlenwert s. Eine entsprechnende Symbolik wird verwendet, wenn die Folge von einer anderen Stelle an durchnumeriert wird. Der Beginn bei Null tritt häufig auf. Eine Reihe, die nicht konvergiert, heißt *divergent*.

(2.1) Bemerkung. Endlich viele Glieder sind für die Konvergenz einer Reihe belanglos. Ist $\sum_{j=1}^{\infty} a_j$ konvergent, so auch $\sum_{j=k}^{\infty} a_j$ für jedes $k \geq 2$ (und umgekehrt), und es gilt

$$\sum_{j=1}^{\infty} a_j = \sum_{j=1}^{k-1} a_j + \sum_{j=k}^{\infty} a_j,$$

wie eine leichte Überlegung zeigt. Diese Bemerkung wird meist stillschweigend verwendet. $\diamond$

(2.2) Beispiel. (1.8) nimmt die folgende Form an: Sind die a_n nichtnegativ und ist die Folge der Partialsummen beschränkt, so konvergiert $\sum a_n$. $\diamond$

Das Cauchy-Kriterium lautet wegen $s_m - s_n = \sum_{j=n+1}^{m} a_j$, $m > n$ für Reihen:

(2.3) Konvergenzkriterium von Cauchy für Reihen. *Eine unendliche Reihe konvergiert genau dann, wenn zu jedem $\varepsilon > 0$ ein N existiert, so daß für $m > n \geq N$ immer die Ungleichung $|a_{n+1} + \cdots + a_m| < \varepsilon$ besteht.* $\square$

Eine unmittelbare Konsequenz dieses Kriteriums ist, daß die Glieder einer konvergenten Reihe eine Nullfolge bilden. Eine Reihe $\sum_{j=1}^{\infty} a_j$ heißt *absolut konvergent*, wenn $\sum_{j=1}^{\infty} |a_j|$ konvergiert. Die Dreiecksungleichung $|\sum_{j=n+1}^{m} a_j| \leq \sum_{j=n+1}^{m} |a_j|$ und das Cauchy-Kriterium zeigen, daß eine absolut konvergente Reihe konvergiert.

(2.4) Majorantenkriterium *Seien (a_n) und (b_n) Folgen und gelte immer $|b_n| \leq a_n$. Dann heißt die Reihe $\sum a_j$ eine* Majorante *der Reihe $\sum b_j$. Konvergiert $\sum a_j$, so konvergiert $\sum b_j$ absolut.*

Beweis. Wegen der Dreiecksungleichung und der Voraussetzung gilt

$$\left|\sum_{j=n+1}^{m} b_j\right| \leq \sum_{j=n+1}^{m} |b_j| \leq \sum_{j=n+1}^{m} a_j.$$

Jetzt wende man das Cauchy-Kriterium an. $\square$

(2.5) Notiz. *Die Reihe $\sum_{j=0}^{\infty} q^j$ heißt* geometrische Reihe. *Für $|q| < 1$ konvergiert die geometrische Reihe zum Grenzwert $(1-q)^{-1}$, für $|q| \geq 1$ divergiert sie.*

Beweis. Für die Partialsumme $s_n = 1 + q + q^2 + \cdots + q^{n-1}$ gilt die Formel

$$s_n = \frac{1-q^n}{1-q}.$$

Es ist

$$\left| \frac{1}{1-q} - \frac{1-q^n}{1-q} \right| = |q|^n \frac{1}{|1-q|}$$

und die rechte Seite ist Glied einer Nullfolge. Die Divergenz folgt, weil q^n keine Nullfolge ist, wenn $|q| \geq 1$ ist. □

Die geometrische Reihe ist eine wichtige Vergleichsreihe, wie die folgenden Konvergenzkriterien zeigen.

(2.6) Quotientenkriterium. *Sei $0 \leq q < 1$. Für fast alle n sei $a_n \neq 0$ und gelte*

$$\left| \frac{a_{n+1}}{a_n} \right| \leq q.$$

Dann konvergiert $\sum a_n$ absolut. Gilt für fast alle n $|a_{n+1}| \geq |a_n| > 0$, so divergiert die Reihe.

Beweis. Sei die Bedingung für alle $n \geq 0$ erfüllt. Aus der Voraussetzung folgt induktiv

$$|a_n| \leq q|a_{n-1}| \leq \ldots \leq q^n |a_0|.$$

Also ist die geometrische Reihe, bis auf den konstanten Faktor $|a_0|$, eine Majorante. Im zweiten Fall bilden die a_n keine Nullfolge. □

(2.7) Wurzelkriterium. *Die Reihe $\sum_{i=1}^{\infty} z_i$ konvergiert absolut, wenn es ein $q \in]0,1[$ gibt, so daß für fast alle n die Ungleichung $\sqrt[n]{|z_n|} \leq q$ gilt. Gilt für unendlich viele n die Ungleichung $\sqrt[n]{|z_n|} \geq 1$, so divergiert die Reihe.*

Beweis. Wegen $|z_n| \leq q^n$ ist die geometrische Reihe eine Majorante. □

(2.8) Integralkriterium. *Sei $f\colon [1,\infty[\to \mathbb{R}_+$ stetig und monoton fallend. Genau dann konvergiert $\sum_{n=1}^{\infty} f(n)$, wenn die Folge $(\int_1^n f(t)\,dt \mid n \in \mathbb{N})$ konvergiert.*

Beweis. Wegen $f(n) \leq \int_{n-1}^n f(t)\,dt \leq f(n-1)$ folgt das aus dem Majorantenkriterium. □

(2.9) Beispiel. Die Reihe $\sum_{j=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}$ ist nach dem dem Quotientenkriterium für $|x| < 1$ absolut konvergent. Nach dem Leibniz–Kriterium (2.15) konvergiert sie auch für $x = 1$. ◇

(2.10) Beispiel. Die Reihe

$$\sum_{k=0}^{\infty} \frac{z^k}{k!}$$

heißt *Exponentialreihe*. Es ist

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{z}{n+1}\right| < \frac{1}{2},$$

wenn $n > 2|z| - 1$ ist, also für fast alle n. Nach dem Quotientenkriterium konvergiert deshalb die Exponentialreihe für jede komplexe Zahl z absolut. Wir zeigen alsbald, daß der Grenzwert mit der Zahl $\exp(z)$ übereinstimmt, die wir im letzten Kapitel definiert haben. Zur Unterscheidung schreiben wir vorläufig $\exp^{\#}(z)$ für den Grenzwert. ◇

(2.11) Beispiel. Die Reihe $\sum n^{-\alpha}$ konvergiert für reelles $\alpha > 1$. Das läßt sich nicht mit dem Quotientenkriterium und dem Wurzelkriterium feststellen. Zwar ist $|a_{n+1}/a_n|$ immer kleiner als 1, es gibt aber kein $q < 1$, das die Bedingung aus (2.6) erfüllt. Das Integralkriterium führt aber zum Ziel, denn es ist

$$\int_1^n \frac{1}{t^\alpha}\,dt = \frac{1}{\alpha - 1}(1 - n^{1-\alpha}) \le \frac{1}{\alpha - 1},$$

und die Konvergenz folgt damit aus (2.8). ◇

(2.12) Beispiel. Die sogenannte *harmonische Reihe* $\sum_{j=1}^{\infty} \frac{1}{j}$ divergiert. Das folgt zum Beispiel aus dem Integralkriterium. ◇

(2.13) Lemma. *Seien $(a_n \mid n \in \mathbb{N})$ und $(b_n \mid n \in \mathbb{N})$ Folgen. Sei $A_n = \sum_{k=1}^n a_k$ die n-te Partialsumme und sei $A_0 = 0$ gesetzt. Für $1 \le p \le q$ gilt dann die Formel der* partiellen Summation

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q-1} A_n(b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p.$$

Beweis. Die Umformungen

$$\begin{aligned}
\sum_{n=p}^{q} a_n b_n &= \sum_{n=p}^{q} (A_n - A_{n-1}) b_n \\
&= \sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1} \\
&= A_q b_q + \sum_{n=p}^{q-1} A_n b_n - \sum_{n=p}^{q-1} A_n b_{n+1} - A_{p-1} b_p
\end{aligned}$$

konstituieren einen Beweis. □

(2.14) Satz. *Sei $(b_n \mid n \in \mathbb{N})$ eine monoton fallende Nullfolge reeller Zahlen. Sei $(a_n \mid n \in \mathbb{N})$ eine Folge komplexer Zahlen. Es gebe ein $C > 0$ derart, daß für alle $n \in \mathbb{N}$ gilt $|\sum_{i=1}^n a_i| \le C$. Dann konvergiert $\sum a_i b_i$, und es gilt $|\sum_{i=1}^n a_i b_i| \le C b_1$.*

Beweis. Wir verwenden das vorige Lemma und erhalten mit der Dreiecksungleichung

$$\left|\sum_{n=p}^{q} a_n b_n\right| \le C \left(\sum_{n=p}^{q-1}(b_n - b_{n-1}) + b_q\right) + |A_{p-1}|b_p = Cb_q + |A_{p-1}|b_p.$$

Für $p = 1$ folgt zunächst die behauptete Ungleichung.

Sei $\varepsilon > 0$ gegeben. Sei N so gewählt, daß für $p > N$ die Ungleichung $b_p < \varepsilon/(2C)$ erfüllt ist. Dann folgt für $q > p > N$ aus der soeben hergeleiteten Ungleichung

$$\left|\sum_{n=p}^{q} a_n b_n\right| \le 2Cb_p < \varepsilon.$$

Aus dem Cauchy-Kriterium folgt die Behauptung. □

(2.15) Leibniz–Kriterium. *Sei $(b_n \mid n \in \mathbb{N})$ eine monoton fallende Nullfolge. Nach (2.14) konvergiert dann $\sum_{n\ge1}(-1)^{n-1}b_n$ (alternierende Reihe), und der Rest $\sum_{n\ge p+1}(-1)^{n-1}b_n$ hat einen Betrag kleinergleich b_{p+1}.* □

(2.16) Umordnungssatz. *Sei $\sum_{n=1}^{\infty} a_n$ absolut konvergent und sei $\tau\colon \mathbb{N} \to \mathbb{N}$ eine Bijektion. Dann konvergiert auch die umgeordnete Reihe $\sum_{n=1}^{\infty} a_{\tau(n)}$ absolut gegen denselben Grenzwert.*

Beweis. Sei s_n die n-te Partialsumme der Ursprungsreihe und t_n die der mittels τ umgeordneten. Sei $\varepsilon > 0$ gegeben und sei m so gewählt, daß $\sum_{k=m}^{m+l} |a_k| < \varepsilon$ für alle $l \in \mathbb{N}$. Wir wählen N, so daß für $r > N$ sowohl $r > m$ als auch $\tau(r) > m$ gilt. Die Differenz $s_r - t_r$ ist dann für solche r eine Summe von Gliedern $\pm a_j$ mit einem Index $j > m$. Daher ist für $r > N$

$$|s_r - t_r| \le \sum_{j>m} |a_j| \le \varepsilon.$$

Daraus folgt schon, daß mit (s_n) auch (t_n) konvergiert, und zwar gegen denselben Grenzwert (1.7.5). Wenden wir dieselbe Überlegung auf die Reihe $\sum |a_n|$ an, so erkennen wir die absolute Konvergenz der umgeordneten Reihe. □

(2.17) Beispiel. Sei $\sum c_k$ eine konvergente, aber nicht absolut konvergente Reihe reeller Zahlen. Zu jedem $x \in \mathbb{R}$ existiert eine Umordnung $\rho\colon \mathbb{N} \to \mathbb{N}$, so daß $\sum_{k=1}^{\infty} a_{\rho(k)} = x$ ist.

Zum Beweis sei (a_k) die Folge der positiven und (b_k) die Folge der negativen Glieder der Reihe c_k. Weil $\sum c_k$ konvergiert, aber nicht absolut, gilt

$$\sum a_k = \infty, \qquad \sum b_k = -\infty.$$

Die Umordnung ρ bestimmen wir jetzt rekursiv wie folgt: Ist $\sum_{k=1}^{\ell-1} c_{\rho(k)} < x$, so sei $c_{\rho(\ell)}$ das nächste bislang noch nicht verbrauchte Glied der Folge (a_k), andernfalls das nächste noch nicht verbrauchte Glied der Folge (b_k). Da sowohl (a_k) als auch (b_k) eine Nullfolge ist, schließt man auf die Konvergenz der umgeordneten Reihe gegen x. $\diamond$

Aus (1.7.1) folgt durch Umschreiben in Reihenform:

(2.18) Notiz. *Seien* $\sum a_j$ *und* $\sum b_j$ *konvergente Reihen. Dann ist* $\sum(a_j + b_j)$ *konvergent, und es gilt* $\sum a_j + \sum b_j = \sum(a_j + b_j)$*. Ebenso gilt immer* $\sum \lambda a_j = \lambda \sum a_j$*.* $\square$

Die Multiplikation von Reihen ist komplizierter. Seien $\sum_{m\ge 0} a_m$ und $\sum_{n\ge 0} b_n$ zwei konvergente Reihen. Wenn wir das Produkt $(\sum a_m)(\sum b_n)$ nach dem Distributivgesetz formal ausmultiplizieren, bekommen wir eine „Reihe“ mit den Summanden $a_m b_n$. Um aus dieser „Reihe“ eine Reihe zu machen, müssen wir die Summanden in bestimmter Weise anordnen. Wenn man an die Multiplikationsformel für Polynome denkt, so bietet sich folgendes an: Man setzt $c_k = \sum_{i=0}^k a_i b_{k-i}$ und betrachtet die Reihe $\sum_{k\ge 0} c_k$. Sie wird das *Cauchy-Produkt* der gegebenen Reihen genannt.

(2.19) Satz. *Das Cauchy-Produkt zweier absolut konvergenter Reihen konvergiert absolut, und der Grenzwert ist das Produkt der Grenzwerte der gegebenen Reihen.*

Beweis. Sei $(a_n \mid n \ge 0)$ eine Folge mit n-ter Partialsumme A_n und Grenzwert A; und sei ebenso (b_n) eine zweite mit Partialsumme B_n und Grenzwert B. Wir setzen $c_n = \sum_{i=0}^n a_i b_{n-i}$ und bezeichnen mit C_n die n-te Partialsumme der Folge (c_n). Dann gilt zunächst nach (1.7.2) $\lim(A_n B_n) = \lim A_n \lim B_n = AB$. Ferner ist

$$|A_n B_n - C_n| \le \sum_{i+j>n} |a_i||b_j| \le \left(\sum_{n/2<i} |a_i|\right)\left(\sum_{n/2<j} |b_j|\right).$$

Ist nun N so gewählt, daß für $n > N$ sowohl $|AB - A_n B_n| < \varepsilon$ ist als auch die beiden Summen in den Klammern kleiner als ε, so ist insgesamt $|AB - C_n| \le |AB - A_n B_n| + |A_n B_n - C_n| < \varepsilon + \varepsilon\varepsilon$. Damit folgt $\lim C_n = AB$. Die absolute Konvergenz folgt genauso durch Betrachtung der Folgen $(|a_n|)$ und $(|b_n|)$. $\square$

(2.20) Beispiel. Wir verwenden die binomische Formel

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}, \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

und sehen, daß das Cauchy-Produkt zweier Exponentialreihen die Relation $\exp^{\#}(x+y) = \exp^{\#}(x)\exp^{\#}(y)$ erfüllt. $\diamond$

(2.21) Beispiel. Das Cauchy-Produkt der Reihe $\sum_{n\ge 0}(-1)^n(n+1)^{-1/2}$ mit sich selbst divergiert. Es ist nämlich

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

Wegen $(n-k+1)(k+1) = (\frac{n}{2}+1)^2 - (\frac{n}{2}-k)^2 \leq (\frac{n}{2}+1)^2$ ist $|c_n| \geq \frac{2(n+1)}{n+2}$, also ist nicht einmal (c_n) eine Nullfolge. ◇

3 Unendliche Produkte

Sei $(z_n \mid n \in \mathbb{N})$ eine Folge komplexer Zahlen. Wir betrachten die Folge der *Partialprodukte* $p_n = \prod_{i=0}^n z_n$. Einen eventuell vorhandenen Grenzwert bezeichnen wir mit $\prod_{i=0}^\infty z_n$. Sind die z_n sämtlich positive Zahlen, so kann man die Betrachtung wegen $\log p_n = \sum_{i=0}^n \log z_n$ auf unendliche Reihen zurückführen, das heißt, (p_n) konvergiert genau dann gegen eine von Null verschiedene Zahl p, wenn $(\log p_n)$ gegen $\log p$ konvergiert. Die durch die für $s > 1$ konvergierende Reihe gegebene Funktion

$$\zeta(s) = \sum_{n=1}^\infty \frac{1}{n^s}, \qquad s > 1$$

heißt *Riemannsche Zetafunktion.* Sie hat die *Eulersche Produktdarstellung*

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}},$$

worin das Produkt über alle Primzahlen p ersteckt wird. Das beweisen wir durch eine Grenzwertüberlegung. Dabei benutzen wir ein Resultat aus der Zahlentheorie: Eine natürliche Zahl besitzt eine (bis auf die Reihenfolge) eindeutige Darstellung als Produkt von Primzahlen. Damit erhalten wir

$$\sum_{n=1}^N \frac{1}{n^s} \leq \prod_{p\leq N} \frac{1}{1-p^{-s}} = \prod_{p\leq N}\left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\dots\right) = \sum_{n\in P(N)} \frac{1}{n^s} \leq \sum_{n=1}^\infty \frac{1}{n^s},$$

wobei $P(N)$ die Menge der natürlichen Zahlen ist, die nur Primteiler kleinergleich N haben. Für $0 < x \leq 1/2$ ist $\log((1-x)^{-1}) \leq 2x$ (denn das ist für $x = 1/2$ richtig und folgt allgemein daraus durch Monotoniebetrachtung), und deshalb ist für $s > 1$

$$\log\zeta(s) = \sum_p \log\frac{1}{1-p^{-s}} \leq 2\sum_p \frac{1}{p^s}.$$

Wegen $\lim_{s\to 1}\zeta(s) = \infty$ folgt die Divergenz der Reihe $\sum_p p^{-1}$, das heißt, die Primzahlen sind so dicht gesät, daß die Summe ihrer Reziproken divergiert.

4 Aufgaben und Ergänzungen

1. Zu einem Paar natürlicher Zahlen (r, s) betrachte die folgende Umordnung der alternierenden Reihe

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots :$$

Es werden der Reihe nach abwechselnd immer r Glieder des positiven Teils und s Glieder des negativen Teils verwendet. Zeige, daß die Reihe gegen den Grenzwert

$$\log 2 + \frac{1}{2}\log\frac{r}{s}$$

konvergiert.

2. Für eine Folge $(a_n \mid n \in \mathbb{N})$ sei

$$s_n = \frac{1}{n}\sum_{j=1}^{n}.$$

Konvergiert (a_n), so konvergiert (s_n) gegen denselben Grenzwert. Es gibt divergente Folgen (a_n), für die (s_n) konvergiert. Konvergiert $(a_n + s_n)$, so konvergiert auch (a_n).

3. Eine Reihe ist genau dann absolut konvergent, wenn ihr Cauchy-Produkt mit jeder konvergenten Reihe konvergiert.

4. Für welche positiven Zahlen a konvergiert die induktiv durch

$$a_1 = a, \qquad a_{n+1} = a^{a_n}$$

definierte Folge?

5. Zeige, daß die Folge $(C_n \mid n \in \mathbb{N})$ mit

$$C_n = \left(\sum_{j=1}^{n}\frac{1}{j}\right) - \log n$$

konvergiert. Der Grenzwert heißt *Eulersche Konstante*.

6. Sei $f\colon \mathbb{R}_+ \to \mathbb{R}$ stetig. Für jedes $\varepsilon > 0$ konvergiere die Folge $(f(n\varepsilon) \mid n \in \mathbb{N})$. Existiert dann auch der Limes $\lim_{x\to\infty} f(x)$?

9 Folgen und Reihen von Funktionen

1 Gleichmäßige Konvergenz

Sei E zunächst eine beliebige Menge und $(f_n \mid n \in \mathbb{N})$ eine Folge von Funktionen $f_n\colon E \to \mathbb{C}$. Falls für jedes $x \in E$ die Folge $(f_n(x))$ konvergiert, so definieren wir $f\colon E \to \mathbb{C}$ durch $f(x) = \lim_{n\to\infty} f_n(x)$. Wir sagen dann, (f_n) konvergiert (auf E) *punktweise* gegen f und schreiben $f = \lim_{n\to\infty} f_n$. Ebenso sagen wir, $\sum f_n$ konvergiert punktweise gegen f, wenn für jedes $x \in E$ der Limes $\sum f_n(x)$ existiert und gleich $f(x)$ ist.

(1.1) Definition. Sei $(f_n \mid n \in \mathbb{N})$ eine Folge von Funktionen $f_n\colon E \to \mathbb{C}$. Wir sagen, (f_n) *konvergiert gleichmäßig* auf E gegen f, wenn zu jedem $\varepsilon > 0$ ein N existiert, so daß für alle $n \geq N$ und alle $x \in E$ stets $|f_n(x) - f(x)| < \varepsilon$ gilt. Wir sagen, $\sum_{n=1}^{\infty} f_n$ konvergiert gleichmäßig, wenn die zugehörige Folge der Partialsummen gleichmäßig konvergiert. $\diamond$

Zur gleichmäßigen Konvergenz gibt es ein Cauchy–Kriterium (1.2) und ein Majorantenkriterium (1.3).

(1.2) Satz. *Die Folge ($f_n\colon E \to \mathbb{C}$) konvergiert genau dann gleichmäßig, wenn es zu jedem $\varepsilon > 0$ ein N gibt, so daß für alle $m, n \geq N$ und alle $x \in E$ stets $|f_n(x) - f_m(x)| < \varepsilon$ ist.*

Beweis. Sei (f_n) gleichmäßig konvergent mit f als Grenzfunktion. Nach Definition der gleichmäßigen Konvergenz gibt es zu $\varepsilon > 0$ ein N, so daß für $n \geq N$ und $x \in E$ immer $|f_n(x) - f(x)| < \varepsilon/2$ ist. Die Dreiecksungleichung $|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f_m(x) - f(x)|$ liefert dann die genannte Cauchy–Bedingung.

Sei, umgekehrt, die Bedingung des Satzes erfüllt. Dann ist insbesondere für jedes $x \in E$ die Folge $(f_n(x))$ eine Cauchy–Folge. Ihren Grenzwert bezeichnen wir mit $f(x)$. Ist nun $|f_m(x) - f_n)x)| < \varepsilon$ für alle $m, n \geq N$, $x \in E$, so folgt mittels VII(1.9) $|f_m(x) - f(x)| = \lim_{n\to\infty} |f_m(x) - f_n(x)| \leq \varepsilon$ für $m \geq N$ und $x \in E$. □

(1.3) Satz. *Sei ($f_n\colon E \to \mathbb{C}$) eine Folge von Funktionen und (M_n) eine Folge positiver Zahlen. Es gelte $|f_n(x)| \leq M_n$ für alle $x \in E$. Ferner sei $\sum M_n$ konvergent. Dann konvergiert $\sum f_n$ gleichmäßig.*

Beweis. Das folgt vermöge $|\sum_{i=m}^n f_i(x)| \leq \sum_{i=m}^n M_i$ aus dem gewöhnlichen Cauchy–Kriterium und dem vorigen Satz. □

2 Stetigkeit der Grenzfunktion

Die Folge der Funktionen $f_n\colon [0,1] \to [0,1], x \mapsto \sqrt[n]{x}$ konvergiert punktweise gegen die Funktion f mit $f(0) = 0$ und $f(x) = 1$, $x > 0$. Die Grenzfunktion von stetigen Funktionen ist also nicht notwendig stetig.

(2.1) Satz. *Konvergiert eine Folge (f_n) an der Stelle x_0 stetiger Funktionen gleichmäßig gegen f, so ist f bei x_0 stetig.*

Beweis. Seien $x, y \in E$. Die Dreiecksungleichung liefert

$$|f(y) - f(x_0)| \leq |f(y) - f_n(y)| + |f_n(y) - f_n(x_0)| + |f_n(x_0) - f(x_0)|.$$

Wir schätzen die drei Summanden gesondert ab. Sei $\varepsilon > 0$ gegeben. Wegen der gleichmäßigen Konvergenz gibt es ein n, so daß für alle $z \in E$ die Ungleichung $|f(z) - f_n(z)| < \varepsilon/3$ erfüllt ist. Wir wählen ein derartiges n. Wegen der Stetigkeit von f_n in x_0 gibt es dann ein $\delta > 0$, so daß für alle $y \in E$ mit $|y - x_0| < \delta$ die Ungleichung $|f_n(y) - f_n(x_0)| < \varepsilon/3$ gilt. Insgesamt ist dann für $|y - x_0| < \delta$ auch $|f(y) - f(x_0)| < \varepsilon$, womit die Stetigkeit von f bei x_0 gezeigt ist. □

(2.2) Bemerkung. Die Stetigkeit ist eine lokale Eigenschaft. Deshalb kann man die Voraussetzung in (2.1) abschwächen. Gibt es zu jedem $x \in E$ eine Umgebung U, so daß die Einschränkungen $f_n|U$ gleichmäßig gegen $f|U$ konvergieren, so ist f stetig. ◇

Wenn man die Stetigkeit mittels Folgenkonvergenz formuliert, so läuft die Stetigkeit der Grenzfunktion auf eine Vertauschung von Limesbildungen hinaus.

(2.3) Satz. *Die Folge stetiger Funktionen $f_n: E \to \mathbb{C}$ konvergiere punktweise gegen eine Funktion $f: E \to \mathbb{C}$. Die Funktion f ist genau dann stetig, wenn für jede Folge $(x_i \in E)$, die in E konvergiert, die Vertauschungsrelation*

$$\lim_{n\to\infty} \lim_{i\to\infty} f_n(x_i) = \lim_{i\to\infty} \lim_{n\to\infty} f_n(x_i)$$

gilt.

Beweis. Gelte $x = \lim_{i\to\infty} x_i = x$. Da f_n stetig ist, so ist $\lim_{i\to\infty} f_n(x_i) = f_n(x)$. Die Vertauschungsrelation lautet also

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{i\to\infty} f(x_i).$$

Die Behauptung folgt nun aus VII(1.9). □

3 Integration der Grenzfunktion

(3.1) Satz. *Sei $(f_n: [a,b] \to \mathbb{C})$ eine Folge stetiger Funktionen, die gleichmäßig gegen f konvergiert $(a < b)$. Dann existiert $\lim_{n\to\infty} \int_a^b f_n(x)\,dx$ und ist gleich $\int_a^b f(x)\,dx$.*

Beweis. Wir wissen schon nach (2.1), daß f stetig ist und damit integrierbar. Sei $\varepsilon > 0$ gegeben und N so gewählt, daß für $n \geq N$ und $x \in [a,b]$ die Ungleichung $|f_n(x) - f(x)| < \varepsilon/2(b-a)$ gilt. Dann liefert

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b (f_n - f) \right| \leq \int_a^b |f_n - f| \leq (b-a) \cdot \frac{\varepsilon}{2(b-a)} < \varepsilon$$

die behauptete Grenzwertaussage. □

(3.2) Folgerung. *Seien $f_n: [a,b] \to \mathbb{C}$ stetig und konvergiere $\sum f_n$ gleichmäßig. Dann ist $\int_a^b \sum f_n = \sum \int_a^b f_n$. Man sagt: Die Reihe läßt sich* gliedweise *integrieren.* □

(3.3) Beispiel. Es kann vorkommen, daß (f_n) nicht konvergiert, aber $\lim \int_a^b f_n$ existiert. Als Beleg definieren wir $f_n: [0,1] \to \mathbb{R}$ durch

$$f_n(x) = \begin{cases} 0 & 2^{-n} \leq x \leq 1 \\ 2^{n+1}(2^{-n} - x) & 0 \leq x \leq 2^{-n}. \end{cases}$$

Es ist $\int_0^1 f_n = 1$. ◇

(3.4) Beispiel. Es kann vorkommen, daß (f_n) gegen eine stetige Funktion f konvergiert und daß $\lim \int f_n$ existiert, aber von $\int f$ verschieden ist. Als Beleg definieren wir $f_n: [0,1] \to \mathbb{R}$ wie folgt: $f_n(0) = 0$, $f_n(2^{-n-1}) = 2^n$, $f_n(2^{-n}) = 0$, $f_n(1)$; zwischen diesen Werten verlaufe der Graph jeweils gradlinig. Dann entsteht eine stetige Funktion. Es ist $\lim f_n$ die Nullfunktion. Es ist $\int_0^1 f_n = 2^{-1}$. ◇

4 Differentiation der Grenzfunktion

(4.1) Satz. *Sei $a < b$. Sei $(f_n\colon [a,b] \to \mathbb{C})$ eine Folge differenzierbarer Funktionen. Alle Ableitungen f_n' seien stetig. Die Folge (f_n') konvergiere gleichmäßig gegen g. Es gebe ein $x_0 \in [a,b]$, so daß $(f_n(x_0))$ konvergiert. Dann konvergiert (f_n) gleichmäßig gegen eine differenzierbare Funktion f, und es gilt $f' = g$.*

Beweis. Nach dem Hauptsatz gilt

$$f_n(x) = \int_{x_0}^{x} f_n'(t)\,dt + f_n(x_0).$$

Wegen (3.1) konvergieren beide Summanden auf der rechten Seite. Wir setzen $c = \lim f_n(x_0)$ und $f(x) = \lim(f_n(x))$. Dann gilt nach (3.1)

$$f(x) = \lim\left(\int_{x_0}^{x} f_n'(t)\,dt + c_n\right) = \int_{x_0}^{x} g(t)\,dt + c.$$

Also ist f differenzierbar und hat die Ableitung g.

Sei nun $\varepsilon > 0$ gegeben. Sei N so gewählt, daß für $n > N$ und $x \in [a,b]$ gilt

$$|f_n'(x) - g(x)| < \frac{\varepsilon}{2(b-a)}, \qquad |f_n(x_0) - f(x_0)| < \frac{\varepsilon}{2}.$$

Dann ist

$$|f_n(x) - f(x)| \le |f_n(x_0) - f(x_0)| + \int_a^x |f_n' - g| \le (x-a)\frac{\varepsilon}{2(b-a)} + \frac{\varepsilon}{2} < \varepsilon.$$

Also konvergiert (f_n) gleichmäßig gegen f. □

(4.2) Folgerung. *Sei $\sum f_n$ eine Reihe von Funktionen $f_n\colon [a,b] \to \mathbb{C}$ mit stetiger Ableitung. Sei $\sum f_n'$ gleichmäßig konvergent und sei $\sum f_n$ konvergent zur Summe f. Dann ist f differenzierbar, und es gilt $f' = \sum f_n'$. Man sagt, die Reihe kann gliedweise differenziert werden.* □

(4.3) Beispiel. Sei $f_n(x) = n^{-1/2}\sin nx$. Dann konvergiert (f_n) gleichmäßig gegen die Nullfunktion. Es ist f_n stetig differenzierbar, aber (f_n') konvergiert nicht. ◇

5 Potenzreihen

Eine Reihe der Form $P(z) = \sum_{n=0}^{\infty} a_n z^n$ heißt *Potenzreihe P*.

(5.1) Hilfssatz. *Die Potenzreihe $P(z_0)$ konvergiere. Sei $r < |z_0|$. Dann konvergiert die Reihe auf $D(r) = \{z \in \mathbb{C} \mid r \ge |z|\}$ absolut und gleichmäßig.*

Beweis. Da die Gleider einer konvergenten Reihe eine Nullfolge bilden, gibt es ein M, so daß $|a_n z_0^n| < M$ für alle $n \in \mathbb{N}_0$. Dann ist $|a_n z^n| \le Mq^n$ mit $q = |r/z_0|$. Somit ist $\sum Mq^n$ eine Majorante. □

Der *Konvergenzradius* $R = R(P) \in \overline{\mathbb{R}}$ einer Potenzreihe P werde definiert durch $R = \sup\{r \in \mathbb{R} \mid P(r) \text{ konvergiert}\}$.

(5.2) Satz. *Sei P eine Potenzreihe mit Konvergenzradius R. Dann gilt:*

(1) *Für $|z| < R$ konvergiert die Reihe.*

(2) *Für $|z| > R$ divergiert die Reihe.*

Beweis. Sei $|z| < R$. Dann gibt es ein r mit $|z| < r < R$, und $P(r)$ konvergiert. Nach (5.1) konvergiert dann $P(z)$ absolut. Sei $|z| > R$. Wäre $P(z)$ konvergent, so wäre nach (5.1) auch $P(r)$ für r mit $|z| > r > R$ konvergent, im Widerspruch zur Supremumseigenschaft von R. □

(5.3) Notiz. *Der Konvergenzradius kann durch die* Formel von Hadamard *bestimmt werden:*

$$R = \frac{1}{L}, \quad L = \limsup \sqrt[n]{|a_n|}.$$

Hier ist $L = \infty$ zugelassen, und in der Formel interpretieren wir die undefinierten Ausdrücke durch $1/0 = \infty$ und $1/\infty = 0$.

Beweis. Sei $z \neq 0$, $0 < L < |z|^{-1}$, $\varepsilon > 0$ und $L + \varepsilon < |z|^{-1}$. Dann gilt nach Definition von L für fast alle n die Ungleichung $\sqrt[n]{|a_n|} \leq L + \varepsilon < |z|^{-1}$, also $\sqrt[n]{|a_n z^n|} < q$ mit $q = (L+\varepsilon)|z| < 1$. Nach dem Wurzelkriterium konvergiert $P(z)$ absolut. Ist $L > |z|^{-1}$, so gibt es für $L - \varepsilon > |z|^{-1}$ unendlich viele n mit $\sqrt[n]{|a_n|} > L - \varepsilon$, also mit $\sqrt[n]{|a_n z^n|} > 1$, woraus die Divergenz von $P(z)$ folgt. Ähnlich behandelt man die Grenzfälle $L = 0$ und $L = \infty$. □

(5.4) Satz. *Die durch formale Differentiation entstehende Reihe $\sum n a_n x^{n-1}$ hat denselben Konvergenzradius wie $\sum a_n x^n$.*

Beweis. Sei R der Konvergenzradius der Ausgangsreihe und R' derjenige der abgeleiteten Reihe. Ist $|z| < R'$, so konvergiert $\sum n a_n z^n$ absolut. Weil $|a_n z^n| \leq |n a_n z^n|$ ist, so konvergiert auch $\sum a_n z^n$ absolut. Also ist $R' \leq R$.

Sei $0 < |z| < R$. Wir wählen ein r zwischen $|z|$ und R. Weil $\lim \sqrt[n]{n} = 1$ ist und $r/|z| > 1$, so ist $\sqrt[n]{n} \leq r/|z|$ für fast alle n. Mithin ist $|a_n||\sqrt[n]{n} z|^n \leq |a_n| r^n$ für fast alle n. Weil $\sum |a_n| r^n$ konvergiert, so konvergiert auch $\sum |n a_n z^n|$. Also ist $|z| \leq R'$, und wir schließen auf $R \leq R'$. □

(5.5) Folgerung. *Seien a_n komplexe Zahlen und habe $\sum a_n z^n$ den Konvergenzradius $R > 0$. Dann wird durch $f\colon]-R, R[\to \mathbb{C}$, $x \mapsto \sum a_n x^n$ eine differenzierbare Funktion dargestellt, und es gilt $f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$. Wegen* (5.4) *ist f sogar beliebig oft differenzierbar.* □

(5.6) Lemma. *Sei $P(z) = a_0 + a_1 z + a_2 z^2 + \cdots$ eine Potenzreihe mit einem Konvergenzradius $R > 0$. Sei $a_0 \neq 0$. Es gibt $0 < \varepsilon \leq R$, so daß für $|z| < \varepsilon$ immer $P(z) \neq 0$ ist.*

Beweis. Wir schreiben $P(z) = a_0 + zQ(z)$ mit $Q(z) = a_1 + a_2 z + \ldots$. Dann hat auch Q den Konvergenzradius R. Für $|z| \leq r < R$ ist $|Q(z)|$ beschränkt, etwa $|Q(z)| \leq M$. Dann ist $|P(z)| \geq |a_0| - |z||Q(z)| > 0$, sofern $|z| < |a_0| M^{-1}$. □

(5.7) Identitätssatz. *Seien $P(z) = \sum a_n z^n$ und $Q(z) = \sum b_n z^n$ Potenzreihen mit einem Konvergenzradius $r > 0$. Es gebe eine Nullfolge (z_i), $0 \neq z_i \in \mathbb{C}$ mit $P(z_i) = Q(z_i)$. Dann gilt $a_i = b_i$ für alle i.*

Beweis. Wir betrachten die Reihe $R(z) = P(z) - Q(z)$. Falls nicht immer $a_j = b_j$ gilt, so hat $R(z)$ die Form $z^m(c_0 + c_1 z + \cdots)$, $c_0 \neq 0$, $m \in \mathbb{N}_0$. Es ist $R(z_i) \neq 0$ für alle i. Nach dem vorigen Lemma gibt es ein $\varepsilon > 0$, so daß $R(z) \neq 0$, sofern $|z| < \varepsilon$. Da aber die z_i eine Nullfolge bilden, ist das nicht möglich. □

(5.8) Folgerung. *Ist $\sum a_n x^n$ eine gerade Funktion, so ist $a_k = 0$ für ungerade k. Ist diese Funktion ungerade, so sind die Koeffizienten zu geradem Index Null.* □

Sei $P(z) = a_0 + a_1 z + a_2 z^2 + \cdots$ eine Potenzreihe, die einen Konvergenzradius $R > 0$ hat. Ferner sei $P(0) = a_0 \neq 0$. Wir bilden mit einer Reihe $Q(z) = b_0 + b_1 z + \cdots$ formal das Cauchy–Produkt $P(z)Q(z) = c_0 + c_1 z + \cdots$. Es gilt also

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0.$$

Da $P(z)$ in einem Umgebung der Null von Null verschieden ist, kann man dort die Funktion $Q\colon z \mapsto P(z)^{-1}$ betrachten. Falls Q um Null in eine Potenzreihe entwickelbar ist, so muß gelten $c_0 = 1$ und $c_n = 0$ für $n > 0$. Wir können in diesem Fall die b_n rekursiv bestimmen:

$$b_0 = a_0^{-1}, \qquad b_n = -a_0^{-1}(a_1 b_{n-1} + \cdots + a_n b_0).$$

Falls die mit den so bestimmten b_n gebildete Reihe einen Konvergenzradius $S > 0$ hat, so stellt sie Q im Kreis $|z| < S$ dar. Zur Untersuchung der Konvergenz ist es keine wesentliche Einschränkung, $a_0 = 1$ anzunehmen. Wegen der Konvergenz von P gibt es ein $r > 0$, so daß für alle n die Abschätzung $|a_n z^n| \leq 1$ gilt. Damit folgt induktiv aus der Rekursionformel für b_n, daß $|b_n| \leq 2^{n-1} r^{-n}$ ist. Also konvergiert $\sum b_n z^n$ für $|z| < \frac{r}{2}$. Damit haben wir gezeigt:

(5.9) Satz. *Ist $P(z) = \sum a_n z^n$ um Null konvergent und gilt $a_0 \neq 0$, so besitzt auch $z \mapsto P(z)^{-1}$ in einer geeigneten Umgebung von Null eine Potenzreihenentwicklung.* □

(5.10) Beispiel. Die Reihen $\sum z^n$, $\sum n^{-1} z^n$ und $\sum n^{-2} z^n$ haben alle den Konvergenzradius 1. Die erste Reihe divergiert für alle z mit $|z| = 1$; die zweite divergiert für $z = 1$, aber nicht für $z \neq 1$ vom Betrag eins; die dritte konvergiert für alle z vom Betrag eins. Auf dem Rand des Konvergenzkreises sind also unterschiedliche Konvergenzverhältnisse möglich. ◇

6 Taylor–Reihen

Sei $f\colon J \to \mathbb{R}$ auf dem Intervall J beliebig oft differenzierbar. Wir haben die Taylorsche Formel

$$f(x) = \sum_{j=0}^{n} \frac{f^{(j)}(a)}{j!}(x-a)^j + R_n f(x,a).$$

Falls $\lim_{n\to\infty} R_n f(x,a) = 0$ ist, so folgt

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

Die Reihe auf der rechten Seite heißt *Taylor–Reihe* von f um die Stelle a. Sie läßt sich immer dann bilden, wenn f unendlich oft differenzierbar ist. Zweierlei kann jedoch vorkommen:

(1) Die Taylor–Reihe konvergiert nicht.

(2) Die Taylor–Reihe konvergiert zwar, aber ihr Grenzwert ist nicht $f(x)$.

Die Funktion aus Aufgabe V(11.6) hat eine konvergente Taylor–Reihe um Null, nämlcih die Nullreihe, aber sie offenbar nicht durch diese Taylor–Reihe dargestellt. Die Taylor–Reihe ist eine Potenzreihe in $x-a$. Sie hat einen Konvergenzradius R, das heißt, für $|x-a| < R$ konvergiert die Reihe absolut und für $|x-a| > R$ divergiert sie. Falls die Reihe für alle $x \in J \cap \{y \mid |y-a| < R\}$ den Wert $f(x)$ hat, so sagen wir, f werde durch die *Taylor–Reihe um a dargestellt.*

Aus unserer Kenntnis des Restes $R_n f(x,a)$ leiten wir her:

(6.1) Satz. *Es gebe ein $K > 0$, so daß für fast alle $n \in \mathbb{N}$ und alle $x \in J$ gilt $|f^{(n)}(x)| \le K^n$. Dann ist für $x \in J$ die Funktion f um a durch ihre Taylor–Reihe darstellbar.*

Beweis. Wir schätzen das Restglied der Taylor–Formel ab

$$|R_n f(x,a)| = \left| \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi) \right| \le \frac{|x-a|^{n+1}}{(n+1)!} K^{n+1}.$$

Die rechte Seite bildet aber eine Nullfolge, wie wir von der konvergenten Exponentialreihe her wissen. □

Ist umgekehrt $\sum_{j=0}^{\infty} c_j (x-a)^j$ eine für $|x-a| < R$ konvergente Reihe mit dem Wert $f(x)$, so wissen wir nach (5.5), daß f dort beliebig oft differenzierbar ist. Die Ableitungen können durch gliedweises Differenzieren gewonnen werden. Damit folgt induktiv $n! \cdot c_j = f^{(j)}(a)$. Das heißt, die Reihe ist die Taylor–Reihe der durch sie dargestellten Funktion.

7 Der Abelsche Grenzwertsatz

(7.1) Satz. *Seien $b_n \colon E \to \mathbb{R}$ und $a_n \colon E \to \mathbb{C}$ Funktionen, die den folgenden Bedingungen genügen:*

(1) *Für jedes $x \in E$ ist die Folge $(b_n(x))$ monoton fallend.*

(2) *(b_n) konvergiert auf E gleichmäßig gegen die Nullfunktion.*

(3) *Es gibt ein $C > 0$, so daß für alle $n \in \mathbb{N}$ und alle $x \in E$ die Abschätzung $|\sum_1^n a_j(x)| \le C$ gilt.*

Dann konvergiert $\sum_1^\infty a_n b_n$ *gleichmäßig.*

Beweis. Wie im Beweis von VII(2.14) erhalten wir durch partielle Summation die Abschätzung

$$\left|\sum_{n=p}^{q} a_n(x)b_n(x)\right| \leq 2Cb_p(x).$$

Die gleichmäßige Konvergenz der b_p gegen die Nullfunktion besagt, daß zu $\varepsilon > 0$ ein N existiert, so daß für alle $x \in E$ und alle $p > N$ die Abschätzung $|b_p(x)| < \varepsilon/(2C)$ gilt. Die Behauptung folgt nun mittels (1.3), angewendet auf die Folge der Partialsummen. □

(7.2) Satz. *Seien* $b_n\colon E \to \mathbb{R}$ *und* $a_n\colon E \to \mathbb{R}$ *Funktionen, die den folgenden Bedingungen genügen:*

(1) *Für jedes* $x \in E$ *ist* $(b_n(x))$ *monoton fallend.*

(2) *Es gibt ein* $M > 0$*, so daß für alle* $n \in \mathbb{N}$ *und alle* $x \in E$ *gilt* $|b_n(x)| \leq M$.

(3) $\sum a_n$ *konvergiert gleichmäßig.*

Dann konvergiert $\sum a_n b_n$ *gleichmäßig.*

Beweis. Sei $A(x) = \sum_1^\infty a_n(x)$. Partielle Summation liefert wieder

$$\begin{aligned}\sum_{j=p}^{q} a_j b_j &= \sum_{j=p}^{q-1} A_j(b_j - b_{j+1}) + A_q b_q - A_{p-1}b_p \\ &= \sum_{j=p}^{q-1}(A_j - A)(b_j - b_{j+1}) + (A_q - A)b_q - (A_{p-1} - A)b_p.\end{aligned}$$

Zu $\varepsilon > 0$ sei N so gewählt, daß $|A_k(x) - A(x)| \leq \varepsilon$ für $x \in E$ und $k \geq N$, was nach der Voraussetzung (3) möglich ist. Sei $q > p > N$. Dann folgt

$$\left|\sum_{j=p}^{q} a_j(x)b_j(x)\right| \leq \varepsilon \sum_{j=p}^{q-1}(b_j(x) - b_{j+1}(x)) + 2\varepsilon M = \varepsilon(b_p(x) - b_q(x)) + 2\varepsilon M \leq 4\varepsilon M.$$

Die Behauptung folgt nun wiederum aus (1.3) □

(7.3) Abelscher Grenzwertsatz. *Die Reihe* $\sum_0^\infty c_n R^n$ *sei für* $0 < R \in \mathbb{R}$ *konvergent. Dann konvergiert* $\sum_0^\infty c_n x^n$ *auf dem Intervall* $[0, R]$ *gleichmäßig.*

Beweis. Wir wenden den voranstehenden Satz in der folgenden Weise an. Sei $b_n(x) = (x/R)^n$ und $a_n(x) = c_n R^n$. Dann fällt $(b_n(x))$ monoton, wenn $x \in [0, R]$ ist, und es ist $|b_n(x)| \leq 1$. Ferner konvergiert $\sum a_n$ gleichmäßig, da die Summanden konstant sind. □

8 Aufgaben und Ergänzungen

1. Durch wiederholte gliedweise Differentiation der geometrischen Reihe zeige für $n \in \mathbb{N}$

$$\frac{1}{(1-z)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} z^k.$$

Zeige, daß diese Reihe den Konvergenzradius 1 hat. Benutze das Ergebnis, um die Formel

$$\sum_{i+j=r} (-1)^j \binom{m+i}{i} \binom{k}{j} = \binom{m-k}{r}$$

zu beweisen (Cauchy-Produkt).

2. Bestimme für jedes $x \in \mathbb{R}$ den Grenzwert

$$\lim_{n\to\infty} \left(\lim_{m\to\infty} [\cos(n!\pi x)]^{2m} \right).$$

10 Die elementaren Funktionen

1 Sinus und Cosinus

Wir definieren Funktionen Sinus $\sin\colon \mathbb{C} \to \mathbb{C}$ und Cosinus $\cos\colon \mathbb{C} \to \mathbb{C}$ durch die für alle $z \in \mathbb{C}$ absolut konvergenten Reihen

$$\sin(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k+1}}{(2k+1)!}, \qquad \cos(z) = \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{(2k)!}.$$

Die besagte Konvergenz ergibt sich mühelos aus dem Quotientenkriterium. Mittels VIII(4.1) erkennt man für reelle z sofort die Relationen $\sin'(z) = \cos(z)$ und $\cos'(z) = -\sin(z)$. Wir können deshalb VI(2.1) anwenden und sehen, daß es sich um die dort betrachteten Funktionen handelt. Wir können durch diese Reihen also die Funktionen Sinus und Cosinus definieren.

2 Die Exponentialfunktion

Wir haben in VII(2.10) die absolut konvergente Exponentialreihe kennengelernt. Wir haben im sechsten Kapitel die komplexe Exponentialfunktion $\exp(z)$ definiert.

(2.1) Satz. *Für alle komplexen z gilt*

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

.

Beweis. Wir bezeichnen den Wert der Exponentialreihe vorübergehend zur Unterscheidung mit $\exp_{\#}(z)$. Die für reelle z gliedweise differenzierte Reihe ist wiederum $\exp_{\#}(x)$. Also hat $\exp_{\#}$ nach VII(4.1) die Eigenschaften V(7.1), und deshalb gilt $\exp(x) = \exp_{\#}(x)$, $x \in \mathbb{R}$.

Die Exponentialreihe genügt nach der Überlegung zum Cauchy–Produkt der Funktionalgleichung $\exp_{\#}(z+w) = \exp_{\#}(z)\exp_{\#}(w)$ für alle komplexen z und w. Wir müssen deshalb noch $\exp_{\#}(iy)$ für reelles y bestimmen.

Durch Einsetzen in die Exponentialreihe erhalten wir die Relation $\exp_{\#}(iy) = \cos(y) + i\sin(y)$. Wir sehen also, daß sich bei reellen y unsere frühere Festsetzung ergibt. □

Definieren wir jetzt $\exp(z)$, $\sin(z)$ und $\cos(z)$ durch die genannten Reihen, so sehen wir unmittelbar, daß die Eulersche Formel VI(7.1) für alle komplexen Zahlen gilt. Ferner haben wir damit die Bezeichnung e^{it} gerechfertigt, weil alles durch die Exponentialreihe definiert ist. Ebenso gelten die Formeln VI(10.3) für alle komplexen t.

3 Der Logarithmus

Durch gliedweise Integration der geometrischen Reihe $(1+x)^{-1} = \sum_0^\infty (-1)^n x^n$ erhält man für $|x| < 1$ die Logarithmenreihe

$$\log(1+x) = \sum_1^\infty (-1)^{n-1}\frac{x^n}{n}.$$

Die Reihe konvergiert auch noch für $x = 1$ und liefert nach dem Abelschen Grenzwertsatz

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots.$$

Wegen der langsamen Konvergenz kann man damit aber $\log(2)$ nicht berechnen. Etwas besser wird die Situation, wenn man die Reihe

$$\log\frac{1+x}{1-x} = 2\sum_{n=0}^\infty \frac{x^{2n+1}}{2n+1}$$

verwendet, die für $x = \frac{1}{3}$ die Entwicklung

$$\log 2 = 2\sum_{n=0}^\infty \frac{1}{(2n+1)3^{2n+1}}$$

liefert. Die ersten vier Glieder ergeben schon 4 richtige Dezimalstellen $\log 2 = 0.6931\ldots$.

4 Arcus tangens

Die geometrische Reihe $(1+x^2)^{-1} = \sum_{n=0}^{\infty}(-1)^n x^{2n}$ konvergiert, wenn $|x| < 1$ ist. Durch gliedweise Intergration über $[0,x]$ erhält man eine Reihenentwicklung

$$\arctan(x) = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{2n+1},$$

die den Konvergenzradius 1 hat. Der Arcus tangens ist auf $\mathbb{R}$ definiert; der Konvergenzradius kann aber nur unter Benutzung der komplexen Zahlen gut verstanden werden. Die Nullstellen $\pm i$ des Nenners $1+x^2$ verhindern einen größeren Konvergenzradius.

Nach dem Abelschen Grenzwertsatz erhalten wir die Reihendarstellung

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots.$$

Sie eignet sich nicht zur Berechnung von π. Mit dem Additionstheorem des Arcus tangens

$$\arctan x + \arctan y = \arctan \frac{x+y}{1-xy}, \qquad xy < 1$$

erhält man die Gleichung

$$\frac{\pi}{4} = \arctan\frac{1}{2} + \arctan\frac{1}{3}.$$

Mit der obigen Reihe liefert das

$$\frac{\pi}{4} = \left(\frac{1}{2} - \frac{1}{3\cdot 2^3} + \frac{1}{5\cdot 2^5} - \cdots\right) + \left(\frac{1}{3} - \frac{1}{3\cdot 3^3} + \frac{1}{5\cdot 3^5} - \cdots\right),$$

und damit kann man schon besser rechnen. Man kann das Additionstheorem aber noch geschickter anwenden und damit die 1706 von John Machin gefundene Formel

$$\frac{\pi}{4} = 4\arctan\frac{1}{5} - \arctan\frac{1}{239}$$

herleiten, mit der man aus der arctan-Reihe eine schneller konvergente Reihenentwicklung bekommt. Zur Herleitung wendet man das Additionstheorem für $x = y = \frac{1}{5}$ und $x = y = \frac{5}{12}$ an, erhält damit

$$2\arctan\frac{1}{5} = \arctan\frac{5}{12}, \quad 2\arctan\frac{5}{12} = \arctan\frac{120}{119}$$

und benutzt dann noch

$$\arctan 1 + \arctan\frac{1}{239} = \arctan\frac{120}{119}.$$

5 Binomialreihe

Sei $\alpha \in \mathbb{R}$. Wir definieren für $n \in \mathbb{N}$ verallgemeinerte Binomialkoeffizienten

$$\binom{\alpha}{n} = \frac{\alpha \cdot (\alpha - 1) \cdot (\alpha - 2) \cdot \ldots \cdot (\alpha - n + 1)}{n!}, \qquad \binom{\alpha}{0} = 1.$$

Damit gelten die Relationen

$$n\binom{\alpha}{n} = \alpha\binom{\alpha - 1}{n - 1}, \quad n \geq 1; \qquad \binom{\alpha - 1}{n + 1} + \binom{\alpha - 1}{n} = \binom{\alpha}{n + 1}, \quad n \geq 0.$$

Nach dem Quotientenkriterium hat die Reihe

$$\sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

für $\alpha \notin \mathbb{N}$ den Konvergenzradius 1.

(5.1) Satz. *Für $|x| < 1$ gilt die Reihenentwicklung*

$$(1 + x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n.$$

Beweis. Wir bezeichnen die linke Seite der zu beweisenden Gleichung mit $f_\alpha(x)$ und die rechte Seite mit $g_\alpha(x)$. Mit den angegebenen Relationen der Binomialkoeffizienten rechnet man die Gleichungen

$$g'_\alpha(x) = \alpha g_\alpha(x), \qquad g_\alpha(x) = (1 + x)g_{\alpha-1}(x)$$

nach. Damit liefert die Quotientenregel, daß der Quotient g_α/f_α die Ableitung Null hat. Wegen $f_\alpha(0) = 1 = g_\alpha(0)$ folgt nun die Behauptung. □

6 Arcus sinus

Die Funktion arcsin: $]-1, 1[\to]-\pi/2, \pi/2[$ hat an der Stelle x die Ableitung $(1 - x^2)^{-1/2}$. Wir entwickeln diese Funktion in die Binomialreihe und integrieren gliedweise. Wir erhalten die für $|x| < 1$ gültige Entwicklung

$$\arcsin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n + 1} \binom{\frac{-1}{2}}{n} x^{2n+1}.$$

Wegen $\sin(\pi/6) = 1/2$ erhalten wir daraus eine Reihe für $\pi/6$, wenn wir $x = 1/2$ einsetzen.

7 Tangens und Cotangens

Wir haben eine Reihenentwicklung, die nach VIII(5.9) in einer Umgebung von Null existiert,

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n.$$

Die dabei auftretenden Koeffizienten B_n heißen *Bernoulli-Zahlen*. Aus der Kenntnis der Exponentialreihe erhalten wir über das Cauchy–Produkt die Rekursionsformel

$$\sum_{i=0}^{n} \binom{n+1}{i} b_i = 0.$$

Mit der Darstellung

$$\frac{x}{e^x - 1} + \frac{x}{2} = \frac{x}{2} \cdot \frac{e^{x/2} + e^{-x/2}}{e^{x/2} - e^{-x/2}} = \frac{x}{2} \coth \frac{x}{2}$$

sehen wir, weil diese Funktion gerade ist, daß B_{2n+1} für $n > 0$ gleich Null ist.

Aus dieser Reihenentwicklung erhält man durch Umformung die folgenden Reihen

$$\begin{aligned}
\coth(x) &= \frac{1}{x} \sum_{n=0}^{\infty} \frac{B_{2n}}{(2n)!} 4^n x^{2n} \\
\tanh(x) &= \sum_{n=1}^{\infty} \frac{B_{2n}}{(2n)!} 4^n (4^n - 1) x^{2n-1} \\
\tan(x) &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{B_{2n}}{(2n)!} 4^n (4^n - 1) x^{2n-1} \\
\cot(x) &= \frac{1}{x} \sum_{n=0}^{\infty} (-1)^n \frac{B_{2n}}{(2n)!} 4^n x^{2n}
\end{aligned}$$

Die erste haben wir im wesentlichen oben schon hergestellt. Die zweite folgt daraus durch die Formel $\tanh(x) = 2\coth(2x) - \coth(x)$. Für die weiteren verwenden wir die Relationen $\tanh(ix) = -i \tan(x)$ und $\coth(ix) = i \cot(x)$.

11 Verschiedene Beispiele und Anwendungen

1 Differentialgleichungen zweiter Ordnung

Wir betrachten die Differentialgleichung $y'' + 2ay' + y = 0$ mit konstanten Koeffizienten $a, b \in \mathbb{C}$. Der *Lösungsraum* $L = L_{\mathbb{C}}$ besteht aus allen differenzierbaren

Funktionen $f: \mathbb{R} \to \mathbb{C}$ mit der Eigenschaft

$$f''(t) + 2af'(t) + bf(t) = 0, \qquad t \in \mathbb{R}.$$

Mit $f, g \in L$ und $\alpha, \beta \in \mathbb{C}$ ist auch $\alpha f + \beta g \in L$. Wir sagen deshalb, L ist ein Vektorraum über $\mathbb{C}$. Wir nennen $P(x) = x^2 + 2ax + b$ das *charakteristische Polynom* der Differentialgleichung. Wir unterscheiden zwei Fälle:

(1) $a^2 \neq b$. Dann hat $P(x)$ zwei verschiedene Nullstellen λ_1 und λ_2.

(2) $a^2 = b$. Dann hat $P(x)$ nur eine (doppelte) Nullstelle $\lambda = -a$.

(1.1) Satz. *Der Lösungsraum L ist zweidimensional. Im Fall $a^2 \neq b$ hat er die Basis $f_1(t) = \exp(\lambda_1 t)$, $f_2(t) = \exp(\lambda_2 t)$. Im Fall $a^2 = b$ hat er die Basis $f_1(t) = \exp(\lambda t)$, $f_2(t) = t\exp(\lambda t)$.*

Beweis. Setzen wir die Funktion $g(t) = \exp(\mu t)$ in die linke Seite der Differentialgleichung ein, so erhalten wir $(\mu^2+2a\mu+b)\exp(\mu t)$. Da die Exponentialfunktion immer von Null verschieden ist, sehen wir, daß g genau dann in L liegt, wenn μ eine Nullstelle des charakteristischen Polynoms ist. Ebenso rechnet man nach, daß $t\exp(-at)$ im Fall $a^2 = b$ eine Lösung ist. Also sind die im Satz genannten Funktionen Lösungen.

Sei $\lambda_1 \neq \lambda_2$ und $f \in L$. Wir bilden die Hilfsfunktionen

$$g_1(t) = (\lambda_2 f(t) - f'(t))\exp(-\lambda_1 t), \quad g_2(t) = (\lambda_1 f(t) - f'(t))\exp(-\lambda_2 t).$$

Man rechnet nach, daß g_1 und g_2 die Ableitung Null haben, wobei man benutzt, daß f der Differentialgleichung genügt. Also sind diese Funktionen konstant, etwa $g_j(t) = c_j$. Die dadurch gegebenen Gleichungen kann man dann nach den f_j auflösen und erhält

$$(\lambda_2 - \lambda_1)f(t) = c_1\exp(\lambda_1 t) + c_2\exp(\lambda_2 t).$$

Die beiden Exponentialfunktionen erzeugen also den Raum L. Sie sind aber auch linear unabhängig und bilden somit eine Basis.

Im zweiten Fall $a^2 = b$ bilden wir aus $f \in L$ die Hilfsfunktion $u(t) = f(t)\exp(-\lambda t)$ und rechnen mittels der Differentialgleichung für f nach, daß u die zweite Ableitung Null hat. Folglich hat u die Form $u(t) = c_1 + c_2 t$. Die genannten Funktionen bilden also wiederum ein Erzeugendensystem von L. Sie sind linear unabhängig und somit eine Basis. □

Als Folgerung aus dem Vorstehenden erhalten wir mit etwas linearer Algebra:

(1.2) Satz. *Sei $t_0 \in \mathbb{R}$ fixiert. Dann ist $L \to \mathbb{C}^2$, $t \mapsto (f(t_0), f'(t_0))$ ein Isomorphismus von Vektorräumen. Wir sagen, eine Lösung f sei durch die Anfangwerte $f(t_0)$ und $f'(t_0)$* bestimmt. □

Sind a und b reell, so ist man an reellen Lösungen interessiert. Wir betrachten dann den $\mathbb{R}$-Vektorraum $L = L_{\mathbb{R}}$ der reellen Lösungen $f: \mathbb{R} \to \mathbb{R}$ der Differentialgleichung. Jetzt unterscheiden wir drei Fälle:

(1) Sind λ_1, λ_2 verschiedene reelle Nullstellen der charakteristischen Gleichung, so ist wiederum $\exp(\lambda_1 t)$, $\exp(\lambda_2 t)$ eine Basis.

(2) Hat die charakteristische Gleichung eine nichtreelle Nullstelle, so ist auch die zweite Nullstelle nichtreell und beide konjugiert-komplex $\lambda_1 = u+iv$, $\lambda_2 = u-iv$. Wir rechnen dann die komplexe Exponentialfunktion in Sinus und Cosinus um und erhalten die Basis $e^{ut}\sin vt$, $e^{ut}\cos vt$.

(3) Es gibt nur eine reelle Nullstelle λ. Dann haben wir wieder die Basis $e^{\lambda t}$, $te^{\lambda t}$.

2 Konvexe und konkave Funktionen

Sei $f\colon]a,b[\to \mathbb{R}$ eine differenzierbare Funktion. Ist die Ableitung monoton wachsend, so gilt für $x_1 < x < x_2$ nach dem Mittelwertsatz

$$f'(\xi_1) = \frac{f(x)-f(x_1)}{x-x_1} \le \frac{f(x_2)-f(x)}{x_2-x} = f'(\xi_2).$$

Erfüllt eine Funktion $f\colon]a,b[\to \mathbb{R}$ für alle $x_1 < x < x_2$ die Ungleichung

(2.1)
$$\frac{f(x)-f(x_1)}{x-x_1} \le \frac{f(x_2)-f(x)}{x_2-x},$$

so nennen wir sie *konvex*. Gilt stattdessen immer $<, \ge, >$, so heißt sie *streng konvex, konkav, streng konkav*. Durch Umrechnung folgt aus dieser Ungleichung

(2.2)
$$\frac{f(x)-f(x_1)}{x-x_1} \le \frac{f(x_2)-f(x_1)}{x_2-x_1} \le \frac{f(x_2)-f(x)}{x_2-x}.$$

Das bedeutet: Für eine konvexe Funktion ist der Differenzenquotient

$$\frac{f(x)-f(y)}{x-y}$$

in jeder der Variablen x und y monoton wachsend. Setzen wir

$$t_1 = \frac{x_1-x}{x_2-x_1}, \qquad t_2 = \frac{x-x_1}{x_2-x_1},$$

so sehen wir, daß die Ungleichungen (2.1) äquivalent sind zu der Bedingung: Es gilt

(2.3)
$$f(t_1x_1 + t_2x_2) \le t_1 f(x_1) + t_2 f(x_2)$$

für alle $x_1 < x_2$, $t_1 > 0$, $t_2 > 0$, $t_1 + t_2 = 1$.

(2.4) Notiz. *Sind $f_1, \dots, f_n$ konvexe Funktionen, so ist auch deren Summe $\sum_j f_j$ konvex. Ist f konvex und $c \in \mathbb{R}$, so ist auch $f + c$ konvex.* □

Sei f konvex und differenzierbar. Aus (2.1) folgt durch die Grenzübergänge $x_1 \to x$ und $x \to x_2$, daß $f'(x_1) \le f'(x_2)$ ist. Insgesamt liefern diese Überlegungen:

(2.5) Satz. *Eine differenzierbare Funktion $f\colon]a,b[\to \mathbb{R}$ ist genau dann (streng) konvex, wenn ihre Ableitung (streng) monoton ist.* □

Die geometrische Bedeutung der strengen Konvexität ist: Der Graph von f zwischen x_1 und x_2 verläuft unterhalb der Verbindungsstrecke durch $(x_1, f(x_1))$ und $(x_2, f(x_2))$.

(2.6) Satz. *Sei $f\colon]a,b[\to \mathbb{R}$ konvex. Seien $t_1,\dots,t_n$ nichtnegative Zahlen mit der Summe 1. Dann gilt für beliebige $x_j \in]a,b[$ die Ungleichung*

$$f(t_1x_1 + \dots + t_nx_n) \le t_1f(x_1) + \dots + t_nf(x_n).$$

Beweis. Induktion nach n. Für $n = 1$ ist nichts zu beweisen und für $n = 2$ handelt es sich um die Definition der Konvexität. Ist $t_1 = 1$, so ist nichts zu beweisen. Ist $t_1 \ne 1$, so schreiben wir

$$\sum_{i=1}^{n+1} t_ix_1 = t_1x_1 + (1-t_1)\sum_{i=2}^{n+1}\frac{t_i}{1-t_1}x_i, \quad \tau_j = \frac{t_j}{1-t_1}, \quad \sum_j \tau_j = 1.$$

Jetzt wende man erst die Definition der Konvexität und dann die Induktionsvoraussetzung an. □

(2.7) Beispiel. Seien $z_1,\dots,z_n$ positive Zahlen. Seien $t_1,\dots,t_n$ positive reelle Zahlen mit der Summe 1. Dann gilt

$$z_1^{t_1}z_2^{t_2}\cdots z_n^{t_n} \le t_1z_1 + \dots + t_nz_n.$$

Insbesondere erhalten wir

$$(z_1z_2\cdots z_n)^{1/n} \le \frac{1}{n}(z_1 + \dots + z_n).$$

Beweis. Wir setzen $z_j = \exp(x_j)$ und benutzen die Konvexität der Exponentialfunktion, die uns nach (2.3) bekannt ist,

$$\prod z_j^{t_j} = \prod(\exp(t_jx_j) = \exp(\sum t_jx_j) \le \sum t_j\exp(x_j) = \sum t_jz_j.$$

Man nennt $\frac{1}{n}\sum z_j$ das *arithmetische Mittel* und $(\prod z_j)^{1/n}$ das *geometrische Mittel* der Zahlen $z_1,\dots,z_n$. ◇

3 Die Ungleichungen von Hölder und Minkowski

Ist $z = (z_1,\dots,z_n) \in \mathbb{C}$ und $p > 0$, so nennt man die Zahl

$$\|z\|^p = \left(\sum_{j=1}^n |zj|^p\right)^{1/p}$$

die p-*Norm* von z. Im Fall $p = 2$ ist das die euklidische Norm $\|z\| = \|z\|_2$.

(3.1) Höldersche Ungleichung. *Sei $p > 1$, $q > 1$ und $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt*

$$\sum_{k=1}^{n} |z_k w_k| \le \|z\|_p \cdot \|w\|_q$$

für beliebige Vektoren $z = (z_1, \ldots, z_n)$ und $w = (w_1, \ldots, w_n)$ des $\mathbb{C}^n$.

Beweis. Es genügt, denn Fall $z \neq 0$ und $w \neq 0$ zu behandeln. Nach (2.5) gilt für positive Zahlen a und b

$$ab \le \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Das liefert in unserem Fall

$$\frac{|z_k w_k|}{\|z\|_p \|w\|_q} \le \frac{1}{p} \cdot \frac{|z_k|^k}{\|z\|_p^p} + \frac{1}{q} \cdot \frac{|w|^q}{\|w\|_q^q}.$$

Durch Summation dieser Ungleichungen über k erhalten wir die behauptete Ungleichung. □

(3.2) Bemerkung. In der linearen Algebra nennt man

$$\langle\, z, w \,\rangle = \sum_{k=1}^{n} z_k \overline{w}_k$$

das *hermitesche Produkt* der Vektoren z und w (*Skalarprodukt* im Falle reeller Vektoren). Ist $p = 2$, so erhalten wir aus (3.1) die *Cauchy–Schwarz'sche Ungleichung*

$$|\langle\, z, w \,\rangle| \le \|z\|_2 \|w\|_2,$$

die man allerdings auch anders mit den Hilfsmitteln der linearen Algebra beweisen kann. ◇

(3.3) Minkowskische Ungleichung. *Für $p \ge 1$ gilt*

$$\|z + w\|_p \le \|z\|_p + \|w\|_p.$$

Diese Ungleichung heiß die Dreiecksungleichung *der p-Norm.*

Beweis. Für $p = 1$ handelt es sich um die Dreiecksungleichung für den Betrag. Sei $p > 1$ und $\frac{1}{p} + \frac{1}{q} = 1$. Wir kürzen ab:

$$S = \left(\sum (|z_j| + |w_j|)^p\right)^{1/p}, \; A = \left(\sum |z_j|^p\right)^{1/p}, \; B = \left(\sum |w_j|^p\right)^{1/p}, \; s_j = |z_j| + |w_j|.$$

Wir haben dann $S \le A + B$ zu zeigen. Es ist nach der Hölderschen Ungleichung

$$\begin{aligned} S^p &= \sum s_j^p = \sum |z_j| s_j^{p-1} + \sum |w_j| s_j^{p-1} \\ &\le (\left(\sum |z_j|^p\right)^{1/p} \left(\sum s_j^{(p-1)q}\right)^{1/q} + \left(\sum |w_j|^p)\right)^{1/p} \left(\sum s_j^{(p-1)q}\right)^{1/q}. \end{aligned}$$

Wegen $(p-1)q = p$ lautet diese Ungleichung $S^p \leq (A+B)S^{p/q}$. Wegen $p/q = p-1$ folgt daraus durch Division mit S^{p-1} die Behauptung. □

Durch Anwendung auf Rechtecksummen und anschließenden Grenzübergang erhalten wir die Höldersche und Minkowskische Ungleichung für Integrale. Sind $f, g\colon [a, b] \to \mathbb{C}$ stetige Funktionen, so setzen wir

$$\langle f, g \rangle = \int_a^b f(t)\overline{g(t)}\, dt, \qquad \|f\|_p = \left(\int_a^b |f(t)|^p\, dt\right)^{1/p}.$$

(3.4) Satz. *Seien $f, g\colon [a, b] \to \mathbb{C}$ stetige Funktionen, sei $p > 1, q > 1$ und $\frac{1}{p} + \frac{1}{q} = 1$. Dann gelten die Ungleichungen*

$$|\langle f, g \rangle| \leq \|f\|_p \|g\|_q \qquad \|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Beweis. Aus der Hölderschen Ungleichung folgt

$$\sum_{j=0}^{n-1} |f(\xi_j)g(\xi_j)| \frac{b-a}{n} \leq \left(\sum |f(\xi_j)|^p \frac{b-a}{n}\right)^{1/p} \left(\sum |g(\xi_j)|^q \frac{b-a}{n}\right)^{1/q}.$$

Ist $\xi_j \in [a + \frac{b-a}{n}j, a + \frac{b-a}{n}(i+1)]$, so stehen hier Rechtecksummen zu einer Zerlegung von $[a, b]$. Der Grenzübergang $n \to \infty$ gemäß IV(2.3) liefert die erste der behaupteten Ungleichungen (Benutzung der Stetigkeit von $x \mapsto x^{1/p}$, Erhalt von $\leq$ bei Grenzübergängen, und dergleichen). Analog für die zweite. □

(3.5) Beispiel. Sei $0 \leq a < b$. Sei $f(t) \geq 0$ und $f\colon [a, b] \to \mathbb{R}$ stetig. Wir setzen $F(x) = \int_a^b t^x f(t)\, dt$. Dann ist $x \mapsto \log F(x)$ konvex. Zum Beweis bemerkt man, daß aus der Hölderschen Ungleichung $F(\frac{x}{p} + \frac{y}{q}) \leq F(x)^{1/p} F(y)^{1/q}$ folgt. ◇

4 Uneigentliche Integrale

Wir wiederholen das Cauchy–Kriterium:

(4.1) Satz. *Sei $f\colon A \to \mathbb{R}$ gegeben. Sei a Häufungspunkt von $A \subset \overline{\mathbb{R}}$. Dann sind äquivalent:*

(1) *Es existiert* $\lim_{x \to a} f(x)$.

(2) *Zu jedem $\varepsilon > 0$ existiert eine Umgebung U von a gibt, so daß für $x, y \in U$ immer $|f(x) - f(y)| < \varepsilon$ gilt.*

□

(4.2) Folgerung. *Sei $f\colon [a, b[\to \mathbb{R}$ stetig. Es existiert $\lim_{x \to b} \int_a^x f(t)\, dt$ genau dann, wenn es zu jedem $\varepsilon > 0$ eine Umgebung U von b so gibt, daß für $x, y \in U$, $x < y$ immer $|\int_x^y f(t)\, dt| < \varepsilon$ gilt.* □

Sei $f\colon]a,b[\to \mathbb{R}$ stetig. Wir sagen, $\int_a^b f(t)\,dt$ *konvergiert*, wenn für ein $c \in]a,b[$ (und damit für alle diese c) die Grenzwerte

$$\lim_{x\to b} \int_c^x f(t)\,dt, \qquad \lim_{x\to a} \int_x^c f(t)\,dt$$

existieren. Wir setzen dann $\int_a^b f(t)\,dt$ gleich der Summe dieser Limites (das ist unabhängig von der Wahl von c wegen der Intervalladditivität der Integrale und Rechenregeln über Grenzwerte). Integrale dieser Form werden manchmal *uneigentliche Integrale* genannt. Aus dem Cauchy–Kriterium (4.1) folgt unmittelbar das

(4.3) Majorantenkriterium. *Gelte $|f(x)| \le g(x)$ für alle $x \in]a,b[$. Konvergiert $\int_a^b g(t)\,dt$, so konvergiert $\int_a^b f(t)\,dt$ absolut, das heißt $\int_a^b |f(t)|\,dt$ konvergiert.* □

(4.4) Beispiel. Sei $f\colon [a,\infty[\to \mathbb{R}$ stetig, und gelte $|f(t)| \le Mt^{-1-\alpha}$ mit einem $\alpha > 0$ für alle $t \ge N \ge a$. Dann konvergiert $\int_a^\infty f(t)\,dt$ absolut. ◇

(4.5) Beispiel. Sei $f\colon]0,a] \to \mathbb{R}$ stetig. Sei $\alpha > -1$. Es gebe $\delta > 0$ und $M > 0$, so daß für $0 < x < \delta$ die Ungleichung $|f(x)| < Mx^\alpha$ gilt. Dann konvergiert $\int_0^a f(t)\,dt$ absolut. ◇

(4.6) Beispiel. Das Integral

$$\int_0^\infty \frac{\sin x}{x}\,dx$$

konvergiert, aber nicht absolut. Wegen $\lim_{x\to 0} \frac{\sin x}{x} = 1$ haben wir nur $\lim_{t\to\infty} \int_0^\infty \frac{\sin x}{x}\,dx$ zu untersuchen. Sei $0 < x < y$. Partielle Integration liefert

$$\int_x^y \frac{\sin t}{t}\,dt = \left(-\frac{\cos y}{y} + \frac{\cos x}{x}\right) - \int_x^y \frac{\cos t}{t^2}\,dt.$$

Mit dem Majorantenkriterium folgt jetzt leicht die Behauptung durch Abschätzung der Summanden auf der rechten Seite.

Um zu zeigen, daß $\int_0^\infty |\sin t| t^{-1}\,dt$ divergiert, vergleiche man das Integral in geeigneter Weise mit der harmanischen Reihe. ◇

(4.7) Beispiel. Das sogenannte *Fresnelsche Integral*

$$\int_0^\infty \sin(t^2)\,dt = \frac{1}{2}\int_0^\infty \frac{\sin u}{\sqrt{u}}\,du$$

konvergiert ebenfalls (Substitution $u = t^2$). Bemerkenswert ist die Konvergenz, obgleich die Funktion $\sin(t^2)$ bei $t \to \infty$ nicht gegen Null strebt. ◇

(4.8) Beispiel. Das uneigentliche Integral $\int_0^1 x^\beta \log x\,dx$ konvergiert für $\beta > -1$. Sei $\varepsilon > 0$ und $\beta - \varepsilon > -1$. Es ist $\lim_{x\to 0} x^\varepsilon \log x = 0$. Es gibt deshalb $M > 0$, so daß $|x^\varepsilon \log x| < M$ für $0 < x \le 1$. Wegen

$$|x^\beta \log x| = |x^\varepsilon \log x \cdot x^{\beta-\varepsilon}| \le Mx^{\beta-\varepsilon}$$

folgt jetzt die Behauptung mittels (4.5). ◇

5 Die Gamma-Funktion

Die sogenannte Γ-Funktion erweitert die Funktion „Fakultä“ zu einer auf $\mathbb{R}_+$ definierten Funktion.

(5.1) Satz. *Es gibt genau eine Funktion* $\Gamma\colon]0,\infty[\to]0,\infty[$*, die die folgenden Eigenschaften hat:*

(1) *Es gilt die Funktionalgleichung* $\Gamma(x+1) = x\Gamma(x)$.
(2) $\Gamma(1) = 1$.
(3) $\log \circ \Gamma$ *ist konvex.*

Die durch diese Eigenschaften bestimmte Funktion besitzt die Darstellung

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

als uneigentliches Integral. Sie läßt sich außerdem durch den Gaußschen Grenzwert

$$\Gamma(x) = \lim_{n\to\infty} \frac{n! \cdot n^x}{x(x+1)\cdots(x+n)}$$

erhalten.

Beweis. Wir zeigen, daß das im Satz genannte uneigentliche Integral konvergiert und die dadurch gegebene Funktion Γ die drei Eigenschaften hat. Die Konvergenz an der unteren Grenze folgt durch die Abschätzung $t^{x-1}e^{-t} \le t^{x-1}$. Die Konvergenz an der oberen Grenze folgt so: Wegen $\lim_{t\to\infty} t^{x+1}e^{-t} = 0$ gibt es ein $M > 0$, so daß für $t > M$ die Abschätzung $t^{x-1}e^{-t} \le t^{-2}$ gilt.

Partielle Integration liefert

$$\int_r^R t^x e^{-t}\, dt = -t^x e^{-t}|_{t=r}^{t=R} + x\int_r^R t^{x-1}e^{-t}\, dt.$$

Die Grenzübergänge $r \to 0$ und $R \to \infty$ liefern $x\Gamma(x) = \Gamma(x+1)$. Mittels $\Gamma(1) = 1$ folgt dann $\Gamma(n+1) = n!$. Nach (3.5), zusammen mit Grenzübergängen $r \to 0$ und $R \to \infty$ wie eben, sehen wir, daß $\log\Gamma$ konvex ist.

Sei nun eine Funktion $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ mit den drei genannten Eigenschaften gegeben. Wir setzen $g(x) = \log f(x)$. Es gilt dann $g(x+1) = g(x) + \log x$, $g(1) = 0$, und g ist konvex. Durch Induktion folgt $f(n+1) = n!$ und $g(x) = \log n!$. Wir betrachten nun für $0 < x \le 1$ die Differenzenquotienten von g auf den Intervallen

$$[n, n+1], \quad [n+1, n+1+x], \quad [n+1, n+2].$$

Wegen der Konvexität sind diese monoton wachsend; das besagt in unserem Falle

$$\log n \le \frac{g(n+1+x) - g(n+1)}{x} \le \log(n+1).$$

Aus $g(x+1) = g(x) + \log(x)$ folgt induktiv

$$g(n+1+x) = g(x) + \log(x(x+1)\cdots(x+n)).$$

Daraus erhält man durch Umrechnung

$$0 \le g(x) - \log \frac{n^x \cdot n!}{x(x+1)\cdots(x+n))} \le x \log\left(1 + \frac{1}{n}\right).$$

Der Grenzübergang $n \to \infty$ zeigt, daß der Limes

$$\lim_{n\to\infty} \log \frac{n^x \cdot n!}{x(x+1)\cdots(x+n)}$$

existiert und gleich $g(x)$ ist. Damit haben wir gezeigt, daß $f(x)$ auf dem Intervall $]0,1]$ durch die drei Eigenschaften eindeutig bestimmt ist. Wegen der Funktionalgleichung ist $f(x)$ überhaupt eindeutig bestimmt. □

Die *Beta-Funktion* wird definiert durch

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

Darin sind x und y positive Zahlen. Nach (4.??) konvergiert das Integral.

(5.2) Satz. *Es gilt*

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Beweis. Wir zeigen, daß die Funktion

$$f(x) = \frac{\Gamma(x+y)}{\Gamma(y)} B(x,y)$$

für jedes feste y die Eigenschaften (5.1) hat. Zunächst einmal gilt

$$B(x+1,y) = \frac{x}{x+y} b(x,y).$$

Um das zu zeigen, wendet man partielle Integration an auf

$$B(x+1,y) = \int_0^1 \left(\frac{t}{1-t}\right)^x \cdot (1-x)^{x+y-1}\,dt.$$

Aus der Definition entnimmt man leicht $B(1,y) = y^{-1}$. Wir haben

$$\log f(x) = \log \Gamma(x+y) + \log B(x,y) - \log \Gamma(y).$$

Wegen (3.5) sind die beiden ersten Summanden rechts konvex. Daraus folgt, daß die gesamte Linearkombination konvex ist (3.2). □

(5.3) Beispiel. Aus (5.2) erhalten wir $\Gamma(\frac{1}{2})^2 = \int_0^1 t^{-1/2}(1-t)^{-1/2}\,dt$. Wir machen die Substitution $t = \sin^2\theta$ und bestimmen damit den Wert des Integrals als π. Also ist $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. ◇

(5.4) Beispiel. In dem Γ-Integral machen wir die Substitution $t = s^2$ und erhalten $\Gamma(x) = 2\int_0^\infty s^{2s-1}e^{-s^2}\,ds$. Wir setzen $x = 1/2$ und gewinnen damit aus (5.3) die Formel

$$\int_{-\infty}^{\infty} e^{-x^2}\,ds = \sqrt{\pi}.$$

Für die Funktion unter dem Integralzeichen kann man keine Stammfunktion aus elementaren Funktionen finden. Es ist deshalb bemerkenswert, daß man trotzdem das bestimmte Integral durch bekannte Größen ausrechnen kann.

(5.5) Beispiel. Es gilt die Funktionalgleichung

$$\Gamma(x) = \tfrac{2^{x-1}}{\sqrt{\pi}}\Gamma(\tfrac{x}{2})\Gamma(\tfrac{x+1}{2}).$$

Zum Beweis zeigt man aus den schon hergeleiteten Eigenschaften der Γ-Funktion, daß die rechte Seite der behaupteten Gleichung die drei Eigenschaften aus (5.1) hat. ◇

6 Aufgaben und Ergänzungen

1. Sei $f\colon\,]a,b[\,\to \mathbb{R}$ gegeben. Bezeichne $f' + (x)$ und $f'_-(x)$ die rechts- und linksseite Ableitung einer Funktion. Folgende Aussagen sind äquivalent:

(1) f ist konvex.

(2) In jedem Punkt $x \in]a,b[$ existiert $f'_-(x)$ und $f'_+(x)$, und für $x_1 < x_2$ gilt $f'_-(x_1) \le f'_+(x_1) \le f'_-(x_2)$.

Eine Funktion, die in jedem Punkt eine rechts- und linksseitige Abbleitung hat ist stetig.

Eine konvexe Funktion ist genau dann streng konvex, wenn für $x_1 < x_2$ immer $f'_+(x_1) < f'_-(x_2)$ ist. Gibt es für eine konvexe Funktion f ein Tripel $x_1 < x < x_2$, so daß in (6.1) die Gleichheit besteht, so ist im Intervall $[x_1,x_2]$ die Funktion die Gerade durch $(x_1, f(x_1))$ und $(x_2, f(x_2))$. Ist f konvex, so verläuft f im Intervall $]a,x_1[$ nicht unterhalb der Geraden $f'_-(x)(x-x_1)+f(x_1)$ und im Intervall $[x_1,b[$ nicht unterhalb der Geraden $f'_+(x_1)(x-x_1)+f(x_1)$. Eine streng konvexe differenzierbare Funktion verläuft also oberhalb jeder ihrer Tangenten (bis auf den Berührpunkt). Eine zweimal differenzierbare Funktion f ist genau dann streng konvex, wenn die zweite Ableitung überall positiv ist.

Da für eine konvexe Funktion f'_- und f'_+ monoton wachsend sind, haben sie nur abzählbar viele Unstetigkeitsstellen. Ist f'_- an der Stelle x stetig, so folgt aus $f'_+(x) \le f'_-(y)$ für $y > x$ durch den Grenzübergang $y \to x$, daß $f'_-(x) = f'_+(x)$ ist. An dieser Stelle ist also f differenzierbar. Eine konvexe Funktion ist also bis auf abzählbar viele Stellen differenzierbar.

2. Die stetige Funktion $f\colon\,]a,b[\,\to \mathbb{R}$ erfülle

$$f(\tfrac{1}{2}(x_1+x_2)) \le \tfrac{1}{2}(f(x_1)+f(x_2))$$

für alle x_1, x_2 aus dem Definitionsbereich. Dann ist f konvex. Zum Beweis zeige man (6.3) durch Induktion nach n zunächst für alle t_1 der Form $a2^{-n}$, $a \in \mathbb{N}$ und benutze sodann, daß jede Zahl in $[0,1]$ Häufungspunkt derartiger t_1 ist.

3. $x \mapsto a^x$ ist für $a \neq 1$, $a > 0$ streng konvex. $x \mapsto x^\alpha$ ist für $\alpha > 1$ und $\alpha < 0$ streng konvex und für $0 < \alpha < 1$ streng konkav. $\log_a$ ist für $a > 1$ streng konkav und für $a < 1$ streng konvex.

4. *Bernoullische Ungleichung.* Sei $a > -1$ und $a \neq 1$. Dann gilt: Für $0 < x < 1$ ist $(1+a)^x < 1 + ax$; für $x > 1$ ist $(1+a)^x > 1 + ax$. Als Anwendung dieser Ungleichungen zeige man: Die Funktion $\mathbb{R}_+ \to \mathbb{R}, x \mapsto (1+\frac{1}{x})^x$ ist streng monoton wachsend; $x \mapsto (1+\frac{1}{x})^{x+1}$ ist streng monoton fallend. Für jedes x liegt zwischen diesen beiden Werten die Zahl e.

5. Gib eine unbeschränkte Funktion $f\colon [0,\infty[\to \mathbb{R}$ an, deren Integral $\int_0^\infty f(x)\,dt$ (absolut) konvergiert.

6. Zeige, daß der Gaußsche Grenzwert für alle reellen Zahlen $x \neq -n$, $n \in \mathbb{N}_0$ existiert.

7. Sei $f\colon [a,b] \to \mathbb{R}$ stetig. Zeige, daß $f\colon x \mapsto \int_a^b t^x f(t)\,dt$ n-mal differenzierbar ist und die n-te Ableitung $f^{(n)}(x) = \int_a^b t^x (\log t)^n f(t)\,dt$ ist. Wende das auf die Γ-Funktion an und zeige, daß sie n-mal differenzierbar ist ($n \in \mathbb{N}$).

12 Rationale Funktionen

Die einfachsten Funktionen sind die Polynome und die rationalen Funktionen. Wir stellen einige algebraische Eigenschaften solcher Funktionen zusammen. Die Zerlegung von Polynomen in Linearfaktoren benutzt allerdings analytische Eigenschaften der Zahlen. Wir wenden die algebraischen Resultate auf die Integrationstheorie an.

1 Division mit Rest

Im Rechenbereich der Polynome gibt es ähnlich wie bei den ganzen Zahlen eine Division mit Rest. Sie besagt:

(1.1) Satz. *Seien A und $B \neq 0$ Polynome. Dann gibt es genau eine Darstellung der Form $A = QB + R$ mit Polynomen Q und R, worin R einen Grad kleiner als B hat.* □

Ist im letzten Satz $R = 0$, so sagen wir, B sei ein *Teiler* von A. Ist insbesondere $B(z) = z - a$, so haben wir eine Darstellung der Form $A(z) = Q(z)(z-a) + A(a)$. Ist $A(a) = 0$, also a eine Nullstelle von A, so wird A von $z - a$ geteilt.

Der Satz über die Division mit Rest gilt für Polynome über einem beliebigen Körper. Er ist die Grundlage für die Teilbarkeitstheorie von Polynomen.

2 Nullstellen von Polynomen

Der Körper der komplexen Zahlen ist *algebraisch abgeschlossen*. Das heißt:

(2.1) Satz. *Seien $a_0, \ldots, a_n$ komplexe Zahlen. Sei $n \geq 1$, $a_n \neq 0$ und $P(z) = \sum_{k=0}^{n} a_k z^k$. Dann gilt es eine komplexe Zahl z_0, die die Gleichung $P(z_0) = 0$ erfüllt.*

Beweis. Ohne wesentliche Einschränkung sei $a_n = 1$ (sonst betrachte man $a_n^{-1}P(z)$.) Sei $m = \inf\{|P(z)| \mid z \in \mathbb{C}\}$. Ist $|z| = R > 1$, so gilt

$$|P(z)| \geq R^n(1 - |a_{n-1}|R^{-1} - \ldots - |a_0|R^{-n}) \geq R^n(1 - R^{1-n}(|a_0| + \ldots + |a_{n-1}|)).$$

Für $R^{n-1} > 2(|a_0| + \ldots + |a_{n-1}|)$ ist also $|P(z)| \geq \frac{1}{2}R^n$. Es gibt demnach ein M, so daß für $|z| > M$ die Abschätzung $|P(z)| > m + 1$ gilt. Auf der kompakten Teilmenge $\{z \in \mathbb{C} \mid |z| \leq M\}$ wird ein Minimum angenommen und dieses ist nach der eben durchgeführten Abschätzung gleich $m = P(z_0)$. Wir zeigen $m = 0$.

Sei $m > 0$. Wir setzen $Q(z) = P(z + z_0)/P(z_0)$. Dann ist $Q(0) = 1$ und $|Q(z)| \geq 1$ für alle $z \in \mathbb{C}$. Außerdem ist $z \mapsto Q(z)$ nicht konstant. Es gibt dann k, $1 \leq k \leq n$, so daß sich die Funktion Q in der Form

$$Q(z) = 1 + b_k z^k + \cdots + b_n z^n$$

schreiben läßt. Es gibt $\varphi \in \mathbb{R}$, so daß $e^{ik\varphi} b_k = -|b_k|$ ist. Es folgt für eine reelle Zahl $r > 0$ mit $|b_k|r^k < 1$ die Abschätzung

$$\begin{aligned} |Q(re^{i\varphi}| &= |1 + b_k e^{ik\varphi} r^k + \cdots + b_n e^{in\varphi} r^n| \\ &\leq |1 - |b_k| r^k| + |b_{k+1}| r^{k+1} + \cdots + |b_n| r^r \\ &= 1 - r^k(|b_k| - r|b_{k+1} r - \cdots - |b_n| r^{n-k}) \end{aligned}$$

Für genügend kleine r wäre dann aber $|Q(re^{i\varphi})| < 1$. Widerspruch. □

Die geometrische Idee hinter dieser Rechnung ist folgende: Nachdem man einmal gezeigt hat, daß $|P(z)|$ ein Minimum an einer Stelle z_0 annimmt, betrachtet man in der komplexen Zahlenebene z_0 und den Bildpunkt $P(z_0)$. Man hat die Vorstellung, daß beim Durchlaufen eines kleinen Kreises um z_0 die Bildpunkte um $P(z_0)$ herumlaufen sollten. Wenn dem aber so ist, so muß es Bildpunkte mit kleinerem Betrag geben, sofern $P(z_0) \neq 0$ ist.

(2.2) Notiz. *Sei $P(z) = \sum_{i=0}^{n} a_j z^j$ ein Polynum mit $a_n = 1$. Es gibt komplexe Zahlen $z_1, \ldots, z_n$, so daß für alle $z \in \mathbb{C}$ gilt $P(z) = \prod_{i=1}^{n}(z - z_i)$.*

Beweis. (Induktion nach dem Grad n des Polynoms.) Für $n = 1$ ist das richtig mit $z_1 = -a_0$. Nach (2.1) gibt es z_n mit $P(z_n) = 0$. Dann ist P durch $z - z_n$ teilbar und wir können auf den Quotienten die Induktionsvoraussetzung anwenden. □

Die Produktzerlegung des letzten Satzes nennt man eine *Zerlegung des Polynoms in Linearfaktoren.* Die z_i heißen die *Nullstellen* des Polynoms. Ein Polynom $P(z)$ sei in der Form

$$P(z) = \prod_{j=1}^{m} (z - z_j)^{n_j}, \qquad n_j \in \mathbb{N}$$

mit paarweise verschiedenen z_j geschrieben. Dann nennen wir n_j die *Vielfachheit* der Nullstelle z_j. Ein Polynom vom Grad n hat also höchstens n verschiedene Nullstellen. Die Nullstellen und ihre Vielfachheiten sind eindeutig durch die Funktion $z \mapsto P(z)$ bestimmt. Hat das Polynom reelle Koeffizienten, so kommt mit der Nullstelle w auch die konjugiert-komplexe $\overline{w}$ als Nullstelle vor, und beide Vielfachheiten sind gleich. Wegen $(z-(a+bi))(z-(a-bi)) = z^2 - 2az + a^2 + b^2$ läßt sich also ein reelles Polynom in lineare und quadratische Faktoren zerlegen.

3 Interpolation

Ein Polynom P vom Grad n ist eindeutig durch die Werte $P(x_j)$ an $n+1$ paarweise verschiedenen Stellen x_j bestimmt, denn gäbe es ein weiteres Polynum Q mit diesen Werten, so wäre $P-Q$ ein Polynom vom Grad höchstens n mit $n+1$ verschiedenen Nullstellen.

Umgekehrt gibt es zu jeder Vorgabe von Stellen $x_0, \ldots, x_n$ und Werten $y_0, \ldots, y_n$ ein Polynom P vom Grad höchstens n mit $P(x_j) = y_j$. Die Herstellung eines solchen Polynoms bezeichnet man als *Interpolation.* Zum Beweis betrachten wir das durch

$$L_k(x) = \frac{\prod_{j,j\neq k}(x - x_j)}{\prod_{j,j\neq k}(x_k - x_j)}$$

definierten, nach *Lagrange* benannte, Polynom. Es hat den Grad n und die Werte $L_k(x_j) = \delta_{ik}$ (geschrieben mit dem Kronecker–Symbol). Das gesuchte Interpolationspolynom ist dann einfach die Linearkombination

$$P(x) = \sum_{j=0}^{n} y_j L_j(x).$$

4 Partialbruchzerlegung

Ein Polynom $P(z)$ sei in der Form

$$P(z) = \prod_{j=1}^{m} (z - z_j)^{n_j}, \qquad n_j \in \mathbb{N}$$

mit paarweise versiedenen z_j geschrieben. Sei $Q(z) = b_0 + b_1 z + \cdots + b_s z^s$ ein weiteres Polynom. Die Darstellung der ratinalen Funktion $Q(z)/P(z)$ in der Form des nächsten Satzes wird *Partialbruchzerlegung* genannt.

(4.1) Satz. *Es gibt ein Polynum $R(z)$ und komplexe Zahlen a_{kl}, $1 \leq k \leq m$, $1 \leq l \leq n_k$, so daß für alle komplexen Zahlen $z \notin \{z_1, \ldots, z_m\}$ gilt*

$$\frac{Q(z)}{P(z)} = R(z) + \sum_{k=1}^{m} \sum_{l=1}^{n_k} \frac{a_{kl}}{(z - z_k)^l}.$$

Beweis. P hat den Grad $\sum_{k=1}^m n_k$. Wir führen den Beweis durch Induktion nach n. Sei $n = 1$. Es ist $Q(z) - Q(z_1) = Q^*(z)$ ein Polynom mit $Q^*(z_1) = 0$. Also gilt $Q^*(z) = (z - z_1)R(z)$ mit einem geeigneten Polynom R. Demnach ist für $z \neq z_1$

$$\frac{Q(z)}{P(z)} = \frac{Q(z_1) + (z - z_1)R(z)}{z - z_1} = R(z) + \frac{Q(z_1)}{z - z_1}.$$

Mit $a_{11} = Q(z_1)$ folgt in diesem Fall die Behauptung.

Sei die Behauptung für alle Q/P mit einem Grad von P kleinergleich n richtig. Sei

$$A = \frac{Q(z_1)}{\prod_{k=2}^m (z - z_k)^{n_k}},$$

wobei unter einem leeren Produkt die Zahl 1 verstanden werde. Dann ist

$$\frac{Q(z)}{P(z)} - \frac{A}{(z - z_1)^{n_1}} = \frac{Q(z) - A\prod_{k=2}^m (z - z_k)^{n_k}}{P(z)}$$

und der Zähler $Q^*(z)$ rechts ein Polynom mit $Q^*(z_1) = 0$. Dieser Zähler hat deshalb die Form $(z - z_1)Q^{**}(z)$. Es folgt

$$\frac{Q(z)}{P(z)} - \frac{A}{(z - z_1)^{n_1}} = \frac{Q^{**}(z)}{(z - z_1)^{n_1 - 1}\prod_{k=2}^m (z - z_k)^{n_k}}.$$

Die rechte Seite hat nach Induktionsvoraussetzung die gewünschte Gestalt. □

Sei der Grad von Q kleiner als der Grad n von P, und P habe paarweise verschiedene Nullstellen $x_1, \dots, x_n$. Dann gilt

$$\frac{Q(x)}{P(x)} = \sum_{j=1}^n \frac{A_j}{x - x_j}.$$

Die Konstante A_j ist durch

$$A_j = \frac{Q(x_j)}{\prod_{k \neq j}(x_j - x_k)}$$

gegeben. Das erhält man aus $(x - x_j)Q(x)/P(x)$ durch den Grenzübergang $x \mapsto x_j$.

Will man bei reellen Polynomen die komplexen Zahlen vermeiden, so muß man auch Potenzen quadratischer Polynome im Nenner zulassen. Eine Partialbruchzerlegung hat dann die folgende Form.

$$\frac{Q(x)}{P(x)} = R(x) + \sum_{i=1}^m \sum_{j=1}^{n_i} \frac{a_{ij}}{(x - x_i)^j} + \sum_{k=1}^r \sum_{l=1}^{s_k} \frac{c_{kl}x + d_{kl}}{(x^2 + 2a_l x + b_l)^l}.$$

Die quadratische Polynome im Nenner haben keine reellen Nullstellen, das heißt es gilt $b_l - a_l^2 > 0$.

Um diese Form aus der komplexen zu beweisen, fasse man zwei Summanden zu konjugiert-komplexen Nullstellen zusammen

$$\frac{\alpha}{(x-w)^n} + \frac{\beta}{(x-\overline{w})^n},$$

bringe auf den Hauptnenner und führe im Zähler Division mit Rest bezüglich $(x-w)(x-\overline{w})$ durch.

5 Integration rationaler Funktionen

Um Stammfunktionen rationaler Funktionen zu bestimmen, hat man wegen der Partialbruchzerlegung im wesentlichen nur Polynome zu integrieren, sowie Funktionen der Form $x \mapsto (x-z)^{-n}$, $z \in \mathbb{C}$, $n \in \mathbb{N}$. Eine Stammfunktion von $x \mapsto (x-w)^{-n-1}$ für $n \neq 0$ ist $x \mapsto -n^{-1}(x-w)^{-n}$; und für $w = a+bi$ und $n = 0$

$$x \mapsto \frac{1}{2}\log((x-a)^2+b^2) - i \arctan\frac{b}{x-a}.$$

Im reellen Fall der quadratischen Polynome hat man als Stammfunktion von $x(1+x^2)^{-n}$ die Funktion $\frac{1}{2}\log(1+x^2)$, falls $n=1$ ist, und $\frac{1}{2n-2}(1+x^2)^{1-n}$, falls $n>1$ ist. Eine Stammfunktion von $(1+x^2)^{-1}$ ist $\arctan x$. Die Stammfunktion F_n von $(1+x^2)^n$ berechnet man induktiv aus der Rekursionsformel

$$F_{n+1}(x) = \frac{x}{2n(1+x^2)^n} + \frac{2n-1}{2n}F_n(x)$$

für $n \in \mathbb{N}$. Durch quadratische Ergänzung und Substitution führt man allgemeinere quadratische Polynome darauf zurück. Sei $P(x) = x^2+2ax+b$ mit $D = b-a^2 > 0$. Dann ist

$$\int_\alpha^\beta \frac{dx}{P(x)^n} = \frac{\sqrt{D}}{D^n}\int_{g(\alpha)}^{g(\beta)} \frac{du}{(u^2+1)^n}$$

mit $g(t) = D^{-1/2}(x+a)$. Ähnlich ist

$$\int_\alpha^\beta \frac{x}{P(x)^n}\,dx = \frac{\sqrt{D}}{D^n}\int_{g(\alpha)}^{g(\beta)} \frac{\sqrt{D}u-a}{(u^2+1)^n}\,du.$$

Diese Integrale werden dann nach dem vorgenannten Verfahren weiter behandelt.

13 Metrische Räume

Die methodischen Grundbegriffe der Analysis sind Stetigkeit, Grenzwert und Konvergenz. Für die allgemeine Theorie ist der Begriff der Stetigkeit am besten geeignet. Wir stellen eine Begriffswelt bereit, die es erlaubt, den Stetigkeitsbegriff in großer Allgemeinheit zu entwickeln. Die Terminologie soll möglichst geometrisch sein.

1 Metrik

Eine *Metrik* auf einer Menge E ist eine Abbildung $d\colon E \times E \to \mathbb{R}$ mit den Eigenschaften:

(1) $d(x,y) = 0$ genau dann, wenn $x = y$ ist.

(2) $d(x,y) = d(y,x)$ für alle $x, y \in E$.

(3) $d(x,z) \le d(x,y) + d(y,z)$ für alle $x, y, z \in E$.

Ein *metrischer Raum* ist ein Paar (E,d), das aus einer Menge E und einer Metrik d auf E besteht. Die Elemente von E heißen die *Punkte* des Raumes. Wir üblich bezeichnen wir einen metrischen Raum (E,d) meist nur durch die zugrundeliegende Menge E. Die Eigenschaft (3) heißt *Dreiecksungleichung* der Metrik. Wegen

$$0 = d(x,x) \le d(x,y) + d(y,x) = 2d(x,y)$$

sind die Werte einer Metrik nichtnegativ. Wir nennen $d(x,y)$ den *Abstand* der Punkte x und y.

(1.1) Beispiel. Durch die Festsetzung $d(x,y) = |x-y|$ wird die *Standardmetrik* auf $\mathbb{R}$ oder $\mathbb{C}$ gegeben. Wenn nichts anderes gesagt wird, soll $\mathbb{R}$ oder $\mathbb{C}$ immer in dieser Weise als metrischer Raum aufgefaßt werden. ◇

(1.2) Beispiel. Die *euklidische Metrik* auf $\mathbb{R}^n$ wird durch $d(x,y) = \|x-y\|$ definiert. Darin ist $\|x\| = (\sum_i x_i^2)^{1/2}$ die *euklische Norm* des Vektors $x = (x_1, \ldots, x_n)$. Die euklidische Metrik wird als die Standardmetrik angesehen. ◇

(1.3) Beispiel. Der $\mathbb{R}^n$ trägt zahlreiche andere Metriken. Zwei weitere sind

$$\begin{aligned} d_1(x,y) &= \sum_{i=1}^{n} |x_i - y_i| \\ d_\infty(x,y) &= \max(|x_i - y_i| \mid 1 \le i \le n). \end{aligned}$$

Darin sind wieder $x = (x_i)$ und $y = (y_i)$ beliebige Vektoren des $\mathbb{R}^n$. ◇

(1.4) Beispiel. Sei E eine beliebige Menge. Durch die Vorschrift $d(x,x) = 0$ und $d(x,y) = 1$ für $x \ne y$ wird auf E eine Metrik definiert. Wir nennen sie die *diskrete Metrik*. ◇

Die Zahl $\inf\{d(a,b) \mid a \in A,\, b \in B\}$ heißt *Abstand der Mengen* $A, B \subset E$. Dieser Abstand ist Null, wenn $A \cap B \ne \emptyset$, aber die Umkehrung gilt im allgemeinen nicht. Wir nennen $\sup\{d(a_1,a_2) \mid a_j \in A\} \in \overline{\mathbb{R}}$ den *Durchmesser* von $A \subset E$.

2 Offene Mengen. Abgeschlossene Mengen. Umgebungen

Sei (E, d) ein metrischer Raum. Eine Menge der Form

$$U_\varepsilon(x) = \{y \in E \mid d(y, x) < \varepsilon\}$$

heißt *ε-Umgebung* von x. Eine Menge $U \subset E$ heißt *offene Menge* des metrischen Raumes, wenn zu jedem $x \in U$ ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subset U$ existiert. Die leere Menge gilt als offen. Aus der Definition bestätigt man unmittelbar: Beliebige Vereinigungen von offenen Mengen sind offen; ein Durchschnitt endlich vieler offener Mengen ist offen.

Eine Teilmenge von E heißt *abgeschlossen*, wenn ihr Komplement offen ist. Aus den Eigenschaften offener Mengen erhält man durch Komplementbildung: Beliebige Durchschnitte abgeschlossener Mengen sind abgeschlossen; ein Durchschnitt endlich vieler abgeschlossener Mengen ist abgeschlossen. Der ganze Raum und die leere Menge sind abgeschlossen. Es gibt also durchaus Mengen, die sowohl offen als auch abgeschlossen sind.

(2.1) Beispiel. Eine Menge der Form $U_\varepsilon(x)$ ist offen. Sei nämlich $y \in U_\varepsilon(x)$ und sei $0 < \delta < \varepsilon - d(x, y)$. Ist $d(y, z) < \delta$, so folgt

$$d(x, z) \leq d(y, z) + d(x, y) < \delta + d(x, y) < \varepsilon.$$

Also ist $U_\delta(y) \subset U_\varepsilon(x)$. $\diamond$

(2.2) Beispiel. Ein offenes Intervall $]a, b[\subset \mathbb{R}$ ist eine offene Teilmenge des metrischen Raumes $\mathbb{R}$. Ein abgeschlossenes Intervall $[a, b]$ ist eine abgeschlossene Teilmenge von $\mathbb{R}$, denn das Komplement ist die Vereinigung der offenen Intervalle $]-\infty, a[$ und $]b, \infty[$. Die Mengen $[a, \infty[$ und $]-\infty, a]$ sind ebenfalls abgeschlossen in $\mathbb{R}$. Die Menge $M = [a, b[$, $a < b$ ist weder offen noch abgeschlossen; zum Punkt a gibt es nämlich keine Menge $U_\varepsilon(a)$, die in M enthalten ist, weshalb M nicht offen ist. Und ähnlich sieht man, daß das Komplement von M nicht offen ist. $\diamond$

Sei $A \subset E$. Eine offene Menge U von E, die A enthält, heißt *offene Umgebung* von A. Eine Menge $B \subset X$ heißt *Umgebung* von A, wenn sie eine offene Umgebung enthält. Besteht A nur aus einem Punkt x, so sprechen wir auch von Umgebungen von x. Es ist also U genau dann eine Umgebung von x, wenn es ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subset U$ gibt.

(2.3) Notiz. *Eine Menge ist genau dann offen, wenn sie Umgebung jedes ihrer Punkte ist.*

Beweis. Ist U offen, so ist nach Definition von „Umgebung“ U Umgebung jedes Elementes aus U. Sei umgekehrt die Menge U Umgebung jedes ihrer Elemente. Dann gibt es also zu $u \in U$ eine offene Menge U_u mit $u \in U_u \subset U$. Demnach ist $U = \bigcup_{u \in U} U_u$ als Vereinigung offener Mengen offen. $\square$

3 Abgeschlossene Hülle. Inneres

Sei X ein metrischer Raum. Der Durchschnitt aller abgeschlossenen Mengen von X, die A enthalten, wird mit $\overline{A}$ bezeichnet und *abgeschlossene Hülle* von A in X genannt. Sie ist als Schnitt abgeschlossener Mengen abgeschlossen, und A ist genau dann abgeschlossen, wenn $A = \overline{A}$ ist. Eine Teilmenge A heißt *dicht in* X, wenn $\overline{A} = X$ ist.

Sei $A \subset X$. Das *Innere* von A oder der *offene Kern* von A ist die Vereinigung aller in A enthaltenen offenen Mengen und wird mit A° bezeichnet. Der Kern ist offen, und A ist genau dann offen, wenn $A = A^\circ$ ist. Ein Punkt $a \in A^\circ$ wird *innerer* Punkt von A genannt. Eine Menge heißt *nirgends dicht*, wenn das Innere ihrer abgeschlossenen Hülle leer ist.

Ein Punkt $x \in X$ heißt *Berührpunkt* von $A \subset X$, wenn jede Umgebung von x einen nichtleeren Durchschnitt mit A hat, und *Häufungspunkt* von A, wenn jede Umgebung von x einen nichtleeren Durchschnitt mit $A \setminus \{x\}$ hat. Hat $x \in A$ eine Umgebung U, deren Schnitt mit A nur x enthält, so ist x ein *isolierter Punkt* von A.

(3.1) Satz. *Die Menge der Berührpunkte von A ist gleich der abgeschlossenen Hülle von A.*

Beweis. Sei $x \in \overline{A}$ und U Umgebung von x. Wir haben $U \cap A \neq \emptyset$ zu zeigen. Angenommen, das sei nicht der Fall. Sei $V \subset U$ eine offene Umgebung von x. Ist $U \cap A$ leer, so auch $V \cap A$, und deshalb gilt $A \subset X \setminus V$. Da $X \setminus V$ abgeschlossen ist, folgt $\overline{A} \subset X \setminus V$ und also $V \cap \overline{A} = \emptyset$. Letzteres ist ein Widerspruch zu $x \in V$ und $x \in \overline{A}$.

Sei umgekehrt x Berührpunkt von A. Ist x nicht in $\overline{A}$ enthalten, so liegt x in der offenen Menge $X \setminus \overline{A}$. Nach der Definition eines Berührpunktes müßte dann $(X \setminus \overline{A}) \cap A \neq \emptyset$ sein, was aber nicht der Fall ist. □

(3.2) Beispiel. Es gilt $\overline{[a,b[} = [a,b]$, falls $a < b$ ist. Der Punkt b ist nämlich Berührpunkt von $[a,b[$, und deshalb ist $[a,b[\cup \{b\}$ in der abgeschlossenen Hülle enthalten. Die umgekehrte Inklusion ist klar, da $[a,b]$ abgeschlossen ist. ◇

(3.3) Beispiel. Die Menge $\mathbb{Q}$ ist dicht in $\mathbb{R}$, das heißt es gilt $\overline{\mathbb{Q}} = \mathbb{R}$. Da jede Umgebung einer reellen Zahl x eine rationale Zahl enthält, ist x nämlich Berührpunkt von $\mathbb{Q}$. ◇

(3.4) Beispiel. Sei $C \subset \mathbb{R}$ und $s = \sup C \in \mathbb{R}$. Dann ist s ein Berührpunkt von C und ein Häufungspunkt, wenn s nicht in C liegt. Eine abgeschlossene beschränkte Teilmenge der reellen Zahlen enthält also ihr Supremum und Infimum, das heißt, sie hat ein Maximum und ein Minimum. ◇

4 Stetigkeit

Seien (E_1, d_1) und (E_2, d_2) metrische Räume. Eine Abbildung $f \colon E_1 \to E_2$ heißt *stetig an der Stelle* $a \in E_1$, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß für alle $x \in E_1$ mit $d_1(x, a) < \delta$ die Ungleichung $d_2(f(x), f(a)) < \varepsilon$ gilt. Wir nennen f *stetig*, wenn f an jeder Stelle $a \in E_1$ stetig ist. Ferner heißt f *gleichmäßig stetig*, wenn zu jedem $\varepsilon > 0$ ein $\delta > 0$ existiert, so daß für alle $x, y \in E_1$ mit $d_1(x, y) < \delta$ stets $d_2(f(x), f(y)) < \varepsilon$ gilt.

Die Stetigkeit läßt sich auch mit dem Umgebungsbegriff umformulieren:

(4.1) Notiz. *Folgende Aussagen sind äquivalent zur Stetigkeit von $f\colon E_1 \to E_2$ an der Stelle a:*

(1) *Das Urbild jeder Umgebung von $f(a)$ ist eine Umgebung von a.*
(2) *Das Urbild jeder Umgebung von $f(a)$ enthält eine Umgebung von a.*
(3) *Zu jeder Umgebung V von $f(a)$ gibt es eine Umgebung U von a, die durch f nach V abgebildet wird, $f(U) \subset V$.*
(4) *Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ mit $f(U_\delta(a)) \subset U_\varepsilon(f(a))$ (oder, äquivalent dazu, $f^{-1}U_\varepsilon(f(a)) \supset U_\delta(a)$).*

In (2) *und* (3) *kann „Umgebung“ auch durch „offene Umgebung“ ersetzt werden.* □

Ist $E_1 = E_2 = \mathbb{R}$ und $d_1 = d_2$ die Standardmetrik (1.2), so haben wir den üblichen Begriff der elementaren Analysis für die (gleichmäßige) Stetigkeit einer Funktion $f\colon \mathbb{R} \to \mathbb{R}$.

Unmittelbar aus der Definition der Stetigkeit folgt:

(4.2) Notiz. *Seien $f\colon E_1 \to E_2$ und $g\colon E_2 \to E_3$ Abbildungen zwischen metrischen Räumen. Ist f bei a und g bei $f(a)$ stetig, so ist $g \circ f$ bei a stetig.* □

Sei (E, d) ein metrischer Raum und $F \subset E$. Durch die Einschränkung von $d \times d$ auf $F \times F$ erhalten wir eine (induzierte) Metrik auf F, die wir wieder mit demselben Buchstaben bezeichnen. Wenn nichts anderes gesagt wird, versehen wir Teilmengen immer mit dieser Metrik und sprechen dann von einem *Teilraum* oder *Unterraum* von E. Insbesondere betrachten wir Teilmengen $X \subset \mathbb{R}^n$ als Teilraum bezüglich der euklidischen Metrik.

(4.3) Notiz. *Sei $f\colon E_1 \to E_2$ eine Abbildung zwischen metrischen Räumen. Seien $F_j \subset E_j$ Teilmengen. Ist f bei $a \in F_1$ stetig, so ist dort auch die Einschränkung $f|F_1$ stetig. Die Inklusion $\iota\colon F_1 \to E_1, x \mapsto x$ ist immer stetig. Ist $f(E_1) \subset F_2$, so ist f genau dann bei a stetig, wenn die zugehörige (und genauso bezeichnete) Abbildung $f\colon E_1 \to F_2$ bei a stetig ist. (Für die Feststellung der Stetigkeit spielt also die Bildmenge keine Rolle.)* □

(4.4) Satz. *Eine Abbildung $f\colon E \to F$ zwischen metrischen Räumen ist genau dann stetig, wenn das Urbild jeder offenen (bzw. jeder abgeschlossenen) Menge wieder offen (bzw. abgeschlossen) ist.*

Beweis. Der Beweis als leichte Umformung aus den Definitionen mag als Aufgabe dienen. Siehe aber das dritte Kapitel, wo analoge Aussagen in einem allgemeineren Kontext hergeleitet werden. □

(4.5) Beispiel. Sei (E, d) ein metrischer Raum und $A \subset E$ eine nichtleere Teilmenge. Wir setzen $d_A(x) = \inf\{d(x, a) \mid a \in A\}$ und bezeichnen diese Zahl als *Abstand* von x und A. Eine kleine Überlegung zeigt $|d_A(x) - d_A(y)| \le d(x,y)$. Also ist $d_A\colon E \to \mathbb{R}$ gleichmäßig stetig. $\diamond$

(4.6) Satz. *Seien $f, g\colon E \to F$ an der Stelle a stetig. Sei $f(a) \ne g(a)$. Dann gibt es eine Umgebung U von a, so daß für alle $x \in U$ immer noch $f(x) \ne g(x)$ gilt.*

Beweis. Sei $\varepsilon = d(f(a), g(a))$. Wegen der vorausgesetzten Stetigkeit gibt es ein $\delta > 0$, so daß für $d(x,a) < \delta$ gilt $d(f(a), f(x)) < \varepsilon/2$ und $d(g(a), g(x)) < \varepsilon/2$. Bestünde für ein $u \in U_\delta(a)$ die Gleichheit $f(u) = g(u)$, so würde die Dreiecksungleichung den Widerspruch

$$d(f(a), g(a)) \le d(f(a), f(u)) + d(g(u), g(a)) < \varepsilon$$

liefern. $\square$

Genauso wie für Funktionen einer reellen Veränderlichen beweist man:

(4.7) Satz. *Seien $f, g\colon X \to \mathbb{C}$ bei $a \in X$ stetig. Dann sind auch $f+g$, $f \cdot g$, f/g (sofern definiert), $|f|$, $\max(f,g)$ und $\min(f,g)$ bei a stetig.* $\square$

(4.8) Beispiel. Die Projektionen $\mathrm{pr}_j\colon \mathbb{R}^n \to \mathbb{R}$, $(x_1, \ldots, x_n) \mapsto x_j$ auf den j-ten Faktor sind stetig. Mit dem vorigen Satz zeigt man induktiv, daß alle Polynome in n Veränderlichen

$$(x_1, \ldots, x_n) \mapsto \sum a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n}$$

stetig sind ($a_{i_1 \ldots i_n} \in \mathbb{R}$, Summe mit endlich vielen Termen). Analog für komplexe Polynome. $\diamond$

(4.9) Satz. *Seien $f, g\colon X \to \mathbb{R}$ stetig. Dann ist $\{x \in X \mid f(x) \le g(x)\}$ abgeschlossen und $\{x \in X \mid f(x) < g(x)\}$ offen.*

Beweis. Die Aussage über die Abgeschlossenheit folgt durch „Komplementbildung“ aus der Aussage über die Offenheit. Letztere ist aber eine Folgerung aus (4.6) und (2.3). $\square$

(4.10) Folgerung. *Seien $f, g\colon E \to F$ stetig. Dann ist die Koinzidenzmenge $\{e \in E \mid f(e) = g(e)\}$ abgeschlossen. Zwei stetige Funktionen sind gleich, wenn sie auf einer dichten Menge ihres Definitionsbereichs übereinstimmen.* $\square$

(4.11) Beispiel. Sei $M_n(\mathbb{R})$ der Vektorraum aller (n,n)-Matrizen mit reellen Einträgen. Diesen Vektorraum können wir auch als $\mathbb{R}^{n^2}$ auffassen. Die Determinante einer solchen Matrix ist ein Polynom in den Einträgen. Also ist $\det\colon M_n(\mathbb{R}) \to \mathbb{R}$ stetig. Die Menge $GL(n, \mathbb{R}) = \det^{-1}(\mathbb{R} \setminus \{0\})$ ist die Menge der invertierbaren Matrizen darin. Als Urbild einer offenen Menge ist diese Menge also offen. $\diamond$

(4.12) Beispiel. Die euklidische Norm als Abbildung $\mathbb{R}^n \to \mathbb{R}, x \mapsto \|x\|$ ist stetig. Also sind die Mengen

$$D^n = \{x \in \mathbb{R} \mid \|x\| \le 1\}, \quad S^{n-1} = \{x \in \mathbb{R} \mid \|x\| = 1\}$$

als Urbilder abgeschlossener Mengen abgeschlossen. Das Innere von D^n ist $U_1(0)$, und die abgeschlossene Hülle von $U_1(0)$ ist D^n. Das Innere von S^{n-1} ist leer. Wir nennen S^{n-1} die $(n-1)$-*dimensionale Einheitssphäre.* $\diamond$

5 Grenzwerte

Seien E und F metrische Räume, sei $A \subset E$. Sei $z \in E$ *Häufungspunkt* von A. Sei $f\colon A \to F$ eine Abbildung. Das Symbol

$$\lim_{a \to z} f(a) = c$$

bedeutet: Zu jedem $\varepsilon > 0$ existiert ein $\delta > 0$, so daß für $a \in U_\delta(z) \cap A \setminus \{z\}$ immer $d(f(a), c) < \varepsilon$ gilt. Wir verwenden dieselbe Limes-Terminologie wie in der elementaren Analysis. Wir können auch hier den Grenzwertbegriff auf die Stetigkeit zurückführen. Wir definieren $f^{\#}\colon A \cup \{z\} \to F$ durch $f^{\#}(a) = f(a)$, sofern $a \neq z$ ist, und $f^{\#}(z) = c$. Dann gilt:

(5.1) Notiz. *Folgende Aussagen sind äquivalent:*
(1) $\lim_{a\to z} f(a) = c$.
(2) $f^{\#}$ *ist bei* z *stetig.* $\square$

(5.2) Folgerung. *Ist* $z \in A$ *Häufungspunkt von* A *und ist* $f\colon A \to F$ *bei* z *stetig, so gilt* $\lim_{a\to z} f(a) = f(z)$. $\square$

Ein Limes $\lim_{a\to z} f(a)$ ist, wenn er existiert, eindeutig durch $f|A \setminus \{z\}$ bestimmt (4.3). Ferner gilt auch jetzt mit gleichem Beweis wie früher der

(5.3) Vertauschungssatz. *Seien* E, F *und* G *metrische Räume. Sei* $A \subset E$, *sei* $z \in E$ *Häufungspunkt von* A *und seien* $f\colon A \to F$ *und* $g\colon F \to G$ *Abbildungen. Sei* $\lim_{a\to z} f(a) = b$ *und sei* g *bei* b *stetig. Dann existiert* $\lim_{a\to z}(g \circ f)(a)$ *und ist gleich* $g(\lim_{a\to z} f(a))$. $\square$

6 Folgen

Sei $(x_n \mid n \in \mathbb{N})$ eine Folge von Elementen x_n aus dem metrischen Raum E. Wir sagen, (x_n) *konvergiert gegen* $x \in E$ und schreiben dann

$$\lim_{n\to\infty} x_n = x,$$

wenn zu jedem $\varepsilon > 0$ ein $N \in \mathbb{R}$ existiert, so daß für $n > N$ stets $d(x_n, x) < \varepsilon$ gilt. Gibt es ein derartiges x, so heißt (x_n) *konvergent*, andernfalls *divergent*. Die Folge (x_n) heißt *Cauchy–Folge*, wenn zu jedem $\varepsilon > 0$ ein $N \in \mathbb{R}$ existiert, so daß für $m, n > N$ stets $d(x_m, x_n) < \varepsilon$ gilt. Wir nennen E *vollständig*, wenn jede Cauchy–Folge (x_n) aus Elementen $x_n \in E$ konvergiert.

Die reellen und komplexen Zahlen mit der Standardmetrik sind vollständig. Zur letzten Definition bemerken wir zweierlei.

(6.1) Notiz. *Konvergiert* (x_n) *gegen* x, *so ist* (x_n) *eine Cauchy–Folge.*

Beweis. Das folgt sofort mit Hilfe der Dreiecksungleichung $d(x_m, x_n) \leq d(x_m, x) + d(x, x_n)$. □

(6.2) Notiz. *Der Grenzwert einer Folge ist, wenn er existiert, durch die Folge eindeutig bestimmt.*

Beweis. Gelte $x = \lim_{n\to\infty} x_n = y$. Sei $\varepsilon > 0$ gegeben und N so gewählt, daß $d(x_n, x) < \varepsilon$ und $d(x_n, y) < \varepsilon$ für $n > N$. Dann ist $d(x, y) \leq d(x, x_n) + d(x_n, y) < 2\varepsilon$. Da dieses für jedes $\varepsilon > 0$ gilt, folgt $d(x, y) = 0$ und damit nach einer Grundeigenschaft einer Metrik $x = y$. □

(6.3) Satz. *Sei A Teilmenge des metrischen Raumes E. Genau dann liegt x in $\overline{A}$, wenn es eine Folge (x_n) in A gibt, die gegen x konvergiert.*

Beweis. Ist $x = \lim x_n$, so liegt x nach (6.2) in $\overline{A}$. Sei $X \in \overline{A}$. Dann ist $U_{1/n} \cap A \neq \emptyset$. Wir wählen x_n darin. Die resultierende Folge konvergiert gegen x. □

(6.4) Satz. *Sei $f\colon E_1 \to E_2$ eine Abbildung zwischen metrischen Räumen. Dann sind äquivalent:*

(1) *f ist bei $a \in E_1$ stetig.*
(2) *Für jede Folge (a_n) in E_1 mit $\lim_{n\to\infty} a_n = a$ gilt $\lim_{n\to\infty} f(a_n) = f(a)$.*

Beweis. (1) ⇒ (2). Sei $\varepsilon > 0$ gegeben. Sei $\delta > 0$ so gewählt, daß für $d_1(x, a) < \delta$ immer $d_2(f(x), f(a)) < \varepsilon$ gilt. Sei $N > 0$ so gewählt, daß für $n > N$ immer $d_1(a_n, a) < \delta$ ist. Für diese n gilt dann $d_2(f(a_n), f(a)) < \varepsilon$. Damit ist $\lim_{n\to\infty} f(a_n) = f(a)$ gezeigt.
(2) ⇒ (1). Angenommen, f sei bei a nicht stetig. Dann gibt es ein $\varepsilon > 0$, so daß für jedes $\delta > 0$ ein $x \in E_1$ existiert mit $d_1(x, a) < \delta$ aber $d_2(f(x), f(a)) \geq \varepsilon$. Wir wählen zu jedem $\delta = n^{-1}$, $n \in \mathbb{N}$, ein $x = a_n$ mit dieser Eigenschaft. Nach Konstruktion gilt dann $\lim_{n\to\infty} a_n = a$. Wegen $d_2(f(a_n), f(a)) \geq \varepsilon$ gilt dagegen nicht $\lim_{n\to\infty} f(a_n) = f(a)$. Widerspruch. □

Sei X ein metrischer Raum und $(x_n \mid n \in \mathbb{N})$ eine Folge in X. Ein Element $z \in X$ heißt *Häufungswert* der Folge, wenn jede Umgebung von z unendlich viele Folgenglieder enthält. Sei $HW(x_n)$ die Menge der Häufungswerte.

Ist $\mu\colon \mathbb{N} \to \mathbb{N}$ eine injektive, monoton wachsende Abbildung, so heißt $(x_{\mu(n)} \mid n \in \mathbb{N})$ eine *Teilfolge* von (x_n). Sei $T(x_n)$ die Menge der Elemente $z \in X$, die Grenzwerte konvergenter Teilfolgen sind.

Sei $X(n) = \{x_n, x_{n+1}, x_{n+2}, \ldots\}$ und $H(x_n) = \bigcap_{n=1}^{\infty} \overline{H(n)}$.

(6.5) Satz. *Für jede Folge (x_n) in X sind die Mengen $HW(x_n)$, $H(x_n)$ und $T(x_n)$ gleich.*

Beweis. (1) Sei $z \in HW(x_n)$. Da jede Umgebung U von z unendlich viele Folgenglieder enthält, ist $U \cap X(n) \neq \emptyset$. Folglich ist z Berührpunkt von $X(n)$ und demnach in der abgeschlossenen Hülle enthalten. Das gilt für jedes n; also liegt z in $H(x_n)$.
(2) Sei $z \in T(x_n)$ und sei $(x_{\mu(n)})$ eine Teilfolge, die gegen z konvergiert. Ist U eine Umgebung von z, so gibt es wegen der Konvergenz ein N, so daß für $n > N$ gilt

$x_{\mu(n)} \in U$. Also enthält U unendlich viele Folgenglieder, das heißt $z \in HW(x_n)$.
(3) Sei $z \in H(x_n)$. Wir konstruieren induktiv eine Teilfolge, die gegen z konvergiert. Seien $x_{\mu(j)}$, $1 \leq j \leq n-1$ so gegeben, daß $d(z, x_{\mu(j)}) < j^{-1}$ ist. Da der Durchschnitt $U_{1/n}(z) \cap X(\mu(n-1)+1)$ nicht leer ist, können wir ein $\mu(n) > \mu(n-1)$ mit $d(z, x_{\mu(n)}) < n^{-1}$ wählen. Die resultierende Folge konvergiert wie gewünscht. □

(6.6) Folgerung. *Sei (x_n) eine Folge in der abgeschlossenen Teilmenge A des metrischen Raumes X. Dann liegen alle $\overline{X(n)}$ in A, also auch alle Häufungswerte.* □

(6.7) Folgerung. *Sei A eine abgeschlossene Teilmenge des vollständigen metrischen Raumes E. Dann ist A vollständig. Eine Cauchy–Folge in A ist eine Cauchy–Folge in E, hat dort einen Limes, der nach* (6.6) *in A liegt.* □

7 Das Cauchy–Kriterium

(7.1) Cauchy–Kriterium. *Sei $f\colon A \to F$ eine Abbildung aus einer Teilmenge A des metrischen Raumes E. Sei z Häufungspunkt von A. Sei F vollständig. Zu jedem $\varepsilon > 0$ gebe es ein $\delta > 0$, so daß für alle $a, b \in U_\delta(z) \cap A \setminus \{z\}$ die Ungleichung $d(f(a), f(b)) < \varepsilon$ gilt. Dann existiert $\lim_{a\to z} f(a)$.*

Beweis. Sei $\varepsilon > 0$ fixiert und $\delta > 0$ mit der Eigenschaft des Satzes gewählt. Sei $z = \lim a_n$, $a_n \in A$. Das ist möglich, weil z Häufungspunkt von A ist. Sei N so, daß $d(a_n, z) < \delta/2$ für $n > N$. Dann ist $d(a_m, a_n) < \delta$ und folglich $d(f(a_m), f(a_n)) < \varepsilon$ für $m, n > N$, das heißt, $(f(a_m))$ ist eine Cauchy–Folge in F. Da F vollständig ist, hat sie einen Grenzwert c. Ist nun $d(z, a) < \delta/2$ und wählen wir a_m so, daß $d(z, a_m) < \delta/2$ und $d(c, f(a_m)) < \varepsilon/2$, so ist $d(a, a_m) < \delta$ und folglich $d(c, f(a)) \leq d(a, f(a_m)) + d(f(a_m), f(a)) < \varepsilon$. Damit ist $\lim_{a\to z} f(a) = c$ gezeigt. □

8 Fortsetzung stetiger Abbildungen

(8.1) Satz. *Seien E und F metrische Räume. Sei F vollständig. Sei $A \subset E$ und sei $\varphi\colon A \to F$ gleichmäßig stetig. Dann gibt es genau eine gleichmäßig stetige Abbildung $\Phi\colon \overline{A} \to F$, die f erweitert.*

Beweis. Nach (4.4) ist eine Erweiterung eindeutig bestimmt. Sei $z \in \overline{A} \setminus A$. Da φ gleichmäßig stetig ist, so existiert $\lim_{a\to z} \varphi(a)$, weil das Cauchy–Kriterium (7.1) erfüllt ist. Wir nennen diesen Limes $\Phi(z)$. Es bleibt zu zeigen, daß Φ gleichmäßig stetig ist. Zu $\varepsilon > 0$ wählen wir $\delta > 0$ so, daß für $a, b \in A$ mit $d(a, b) < 3\delta$ immer $d(\varphi(a), \varphi(b) < \varepsilon/3$ gilt. Seien $x, y \in \overline{A}$ mit einem Abstand $d(x, y)\delta$. Es gibt $a, b \in A$ mit $d(x, a) < \delta$, $d(y, b)\delta$ und $d(\Phi(x), \varphi(a)) < \varepsilon/3$, $d(\Phi(y), \varphi(b)) < \varepsilon/3$. Aus der Dreiecksungleichung folgt, daß $d(a, b) < 3\delta$ ist. Insgesamt bekommt man $d(\Phi(x), \Phi(y)) \leq d(\Phi(x), \varphi(a)) + d(\varphi(a), \varphi(b)) + d(\varphi(b), \Phi(y)) < \varepsilon$. □

9 Produkte

Seien (E_j, d_j), $1 \leq j \leq n$ metrische Räume. Auf dem Produkt $E_1 \times E_2 \times \cdots \times E_n = \prod_{j=1}^n E_j$ definieren wir eine Metrik d durch die Festsetzung

$$d((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max_j d_j(x_j, y_j).$$

Die Axiome einer Metrik sind leicht nachzuweisen. Diese Metrik ist der Produktstruktur gut angepaßt, denn es gilt

$$U_\varepsilon(x_1, \ldots, x_n) = U_\varepsilon(x_1) \times \cdots \times U_\varepsilon(x_n).$$

Wir nennen deshalb diese Metrik die *Produktmetrik*. Diese Metrik hat außerdem die Eigenschaft, daß wir sie schrittweise aufbauen können; zum Beispiel stimmen die so auf $(E_1 \times E_2) \times E_3$ und $E_1 \times (E_2 \times E_3)$ und $E_1 \times E_2 \times E_3$ gegebenen Metriken überein.

Die offenen Mengen der Produktmetrik hängen nur von den offenen Mengen der beteiligten Räume ab. Es gilt nämlich:

(9.1) Notiz. *Eine Menge $U \subset \prod E_j$ ist genau dann offen, wenn sie Vereinigung von Mengen der Form $\prod U_j$, $U_j \subset E_j$ offen, ist.*

Beweis. Es genügt, den Fall $n = 2$ zu behandeln. Wir zeigen, daß $U \subset E_1 \times E_2$ genau dann offen ist, wenn U Vereinigung von Mengen der Form $U_1 \times U_2$, $U_j \subset E_j$ offen ist. Sind $U_j \subset E_j$ offen und ist $u_j \in U_j$, so gibt es $\varepsilon > 0$ mit $U_\varepsilon(u_j) \subset U_j$. Also ist $U_1 \times U_2$ Vereinigung von offenen Mengen der Form $U_\varepsilon(u_1, u_2)$ und damit offen. Ist $U \subset E_1 \times E_2$ offen, so ist nach Definition der Produktmetrik U Vereinigung von Mengen der Form $U_\varepsilon(u_1) \times U_\varepsilon(u_2)$. □

(9.2) Satz. *Seien E_1, E_2, F metrische Räume. Sei $f = (f_1, f_2)\colon F \to E_1 \times E_2$, $x \mapsto f(x) = (f_1(x), f_2(x))$ eine Abbildung. Sie ist genau dann bei a stetig, wenn die $f_j(x)$ beide bei a stetig sind.* □

(9.3) Bemerkung. Für die Untersuchung der Stetigkeit von Abbildungen zwischen Teilmengen euklidischer Räume spielt es keine Rolle, welche der Metriken d_1, d_2, d_∞ wir verwenden. Der Grund dafür liegt darin, daß alle diese Metriken dieselben offenen Mengen liefern und die Stetigkeit allein mit Hilfe der offenen Mengen formuliert werden kann. Dieser Standpunkt wird im dritten Kapitel durch Betrachtung topologischer Räume präzisiert und ausgebaut. ◇

10 Aufgaben und Ergänzungen

1. Eine Cauchy–Folge, die eine konvergente Teilfolge hat, konvergiert.

2. Sei A eine nichtleere Teilmenge des metrischen Raumes E. Dann ist $\overline{A} = d_A^{-1}(0)$, also die Menge der Punkte, die von A den Abstand Null haben.

3. Eine endliche Menge in einem metrischen Raum ist abgeschlossen.

4. Beweise oder widerlege: In einem metrischen Raum ist die abgeschlossene Hülle von

$U_\varepsilon(z)$ immer gleich $\{e \in E \mid d(e,z) \leq \varepsilon\}$.

5. Sei (E,d) ein metrischer Raum. Trage $E \times E$ die Produktmetrik. Dann ist $d\colon E \times E \to \mathbb{R}$ gleichmäßig stetig.

14 Normierte Vektorräume

Viele metrische Räume der Analysis entstehen aus normierten Vektorräumen. Insbesondere sind Funktionenräume wichtig, also Vektorräume, deren Elemente Funktionen sind. Wir verwenden Vektorräume über $\mathbb{R}$ oder $\mathbb{C}$ und setzen $\mathbb{K} = \mathbb{R}, \mathbb{C}$, wenn beide Fälle erlaubt sind.

1 Norm

Sei V ein $\mathbb{K}$-Vektorraum. Eine *Norm* auf V ist eine Abbildung $N\colon V \to \mathbb{R}$ mit den Eigenschaften:

(1) $N(v) = 0$ genau dann, wenn $v = 0$ ist.

(2) $N(\lambda v) = |\lambda| N(v)$ für $\lambda \in \mathbb{K}$ und $v \in V$.

(3) $N(v + w) \leq N(v) + N(w)$.

Ein *normierter Vektorraum* (V, N) ist ein Paar, das aus einem Vektorraum V und einer Norm N auf V besteht. Die Eigenschaft (3) heißt *Dreiecksungleichung* der Norm. Eine Norm wird oft durch $N(v) = \|v\|$ bezeichnet, eventuell noch mit einem Index an $\| - \|$ zur Unterscheidung mehrerer Normen. Wir bezeichnen normierte Räume meist nur durch die zugrundeliegende Menge.

(1.1) Notiz. *Sei (V, N) ein normierter Vektorraum. Dann wird durch die Festsetzung $d(x,y) = N(x - y)$ eine Metrik auf V gegeben.* □

Wir fassen einen normierten Vektorraum in der vorgenannten Weise als metrischen Raum auf. Ist ein normierter Vektorraum als metrischer Raum vollständig, so nennt man ihn *Banach-Raum.* Wegen $d(x, 0) = N(v)$ nimmt eine Norm übrigens nur nichtnegative Werte an.

(1.2) Beispiele. Auf $\mathbb{R}^n$ haben wir für $1 \leq p$ die p-Norm

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

Im Fall $p = 2$ ist das die euklidische Norm. Die Dreiecksungleichung für die p-Norm ist die früher hergeleitete Minkowskische Ungleichung. Im Fall $p = 1$ ist $\|x\|_p = \sum_{i=1}^n |x_i|$, und die zugehörige Metrik haben wir schon in (1.3) betrachtet. Wir haben außerdem die *Maximumnorm* $\|x\|_\infty = \max(|x_i|)$ mit der zugehörigen Metrik d_∞ aus (1.3). Entsprechende Normen gibt es auch auf dem $\mathbb{C}^n$. ◇

(1.3) Bemerkung. Ist (V, N) ein komplexer normierter Vekorraum und fassen wir V durch Strukturvergessen als reellen Vektorraum $V_{\mathbb{R}}$ auf, so ist dieselbe Abbildung N auch eine Norm auf $V_{\mathbb{R}}$. ◇

(1.4) Notiz. *Sind* (v_n) *und* (w_n) *konvergente Folgen in einem normierten Vektorraum, so gilt* $\lim(v_n + w_n) = \lim v_n + \lim w_n$ *und* $\lim \lambda v_n = \lambda \lim v_n$. □

2 Funktionenräume

Wir geben jetzt einige für die Analysis typische normierte Vektorräume aus Funktionen an.

(2.1) Beispiel. Sei X eine beliebige Menge und sei $B(X, \mathbb{R})$ der Vektorraum aller beschränkten Funktionen $f \colon X \to \mathbb{R}$. Dabei heißt eine Funktion f *beschränkt*, wenn

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in X\} \in \mathbb{R}$$

existiert. Durch $f \mapsto \|f\|_\infty$ wird die *Supremumsnorm* auf $B(X, \mathbb{R})$ definiert.

Sei $(f_n \mid n \in \mathbb{N})$ eine Folge in $B(X, \mathbb{R})$. Konvergiert (f_n) bezüglich der Supremumsnorm gegen f, so ist das äquivalent dazu, daß (f_n) im üblichen Sinn gleichmäßig gegen f konvergiert. Der Raum $(B(X, \mathbb{R}), \|-\|_\infty)$ ist vollständig, also ein Banach–Raum. Ist nämlich (f_n) eine Cauchy–Folge, so ist für jedes $x \in X$ auch $(f_n(x))$ eine Cauchy–Folge reeller Zahlen, konvergiert also gegen eine $f(x)$ genannte Zahl. Man bestätigt, daß $x \mapsto f(x)$ eine beschränkte Funktion ist und daß (f_n) in der Sup-Norm gegen f konvergiert. Im Fall $X = \{1, \ldots, n\}$ ist übrigens $B(X, \mathbb{R}) \cong \mathbb{R}^n$, und dieser Raum ist vollständig bezüglich der Maximumnorm. ◇

(2.2) Beispiel. Das vorstehende Beispiel läßt sich leicht verallgemeinern. Sei V ein normierter Vektorraum und X eine beliebige Menge. Eine Funktion $f \colon X \to V$ heißt *beschränkt*, wenn $x \mapsto \|f(x)\|$ als Funktion $X \to \mathbb{R}$ beschränkt ist. Aus den Grundeigenschaften einer Norm folgt, daß die Menge $B(X, V)$ der beschränkten Funktionen $X \to V$ bezüglich punktweiser Addition und Skalarmultiplikation ein Vektorraum ist. Für $f \in B(X, V)$ setzen wir

$$\|f\|_\infty = \sup\{\|f(x)\| \mid x \in X\}.$$

Die Zuordnung $f \mapsto \|f\|_\infty$ ist dann eine Norm auf $B(X, V)$.

Ist V ein Banach–Raum, so ist $B(X, V)$ ebenfalls ein Banach–Raum. Den Nachweis lassen wir als Aufgabe. ◇

(2.3) Beispiel. Sei $C^0[a, b] = C([a, b], \mathbb{R})$ der Vektorraum der stetigen Funktionen $[a, b] \to \mathbb{R}$. Da eine solche Funktion beschränkt ist, so handelt es sich um einen Untervektorraum von $B([a, b], \mathbb{R})$. Der Raum $C^0[a, b], \|-\|_\infty$ ist ein Banach–Raum, da Konvergenz in diesem Raum die übliche gleichmäßige Konvergenz ist. Wir wisen aber schon, das ein gleichmäßiger Limes stetiger Funktionen wieder stetig ist.

Auf dem Vektorraum $C^0[a,b]$ haben wir für $1 \le p < \infty$ auch die p-Norm

$$\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{1/p}.$$

Die Dreiecksungleichung dafür ist die früher hergeleitete Minkowskische Ungleichung. ◇

3 Lineare Abbildungen

Wir untersuchen die Stetigkeit von linearen Abbildungen zwischen normierten Vektorräumen.

(3.1) Satz. *Sei $f\colon V \to W$ eine lineare Abbildung zwischen normierten Vektorräumen. Dann sind äquivalent:*
(1) *f ist stetig.*
(2) *f ist am Nullpunkt stetig.*
(3) *Es gibt eine Zahl $C > 0$, so daß für alle $v \in V$ die Ungleichung*

$$\|f(v)\| \le C\|v\|$$

gilt.
Eine stetige lineare Abbildung ist gleichmäßig stetig. (Wir bezeichnen die Normen in V und W mit demselben Symbol.)

Beweis. (1) $\Rightarrow$ (2) ist klar.
(2) $\Rightarrow$ (3). Wegen der Stetigkeit am Nullpunkt gibt es ein $\delta > 0$, so daß aus $\|u\| \le \delta$ folgt $\|f(u)\| \le 1$. Sei $v \ne 0$ ein Vektor aus V. Dann hat $u = \delta\|v\|^{-1}v$ die Norm kleinergleich δ. Es ergibt sich $\|f(v)\| \le \delta^{-1}\|v\|$.
(3) $\Rightarrow$ (1). Wegen $\|f(u)-f(v)\| \le C\|u-v\|$ erkennen wir, daß f sogar gleichmäßig stetig ist. □

Im Kontext der normierten Vektorräume wird eine lineare Abbildung auch *linearer Operator* genannt. Wegen (3) in (3.1) heißt eine stetige lineare Abbildung auch *beschränker Operator*. Das Infimum der C, für die $\|f(v)\| \le C\|v\|$ gilt, wird mit $\|f\|$ bezeichnet und *Norm des Operators f* genannt. Diese Norm hängt natürlich von den in V und W vorgegebenen Normen ab. Der nächste Satz zeigt, daß diese Verwendung des Wortes Norm gerechfertigt ist.

(3.2) Satz. *Sei $L(V,W)$ die Menge der beschränkten Operatoren $V \to W$ zwischen normierten Vektorräumen. Dann ist $L(V,W)$ bezüglich der vorstehend definierten Normen ein normierter Vektorraum. Ist W ein Banach–Raum so auch $L(V,W)$.*

Beweis. Zunächst einmal handelt es sich bei $L(V,W)$ um einen Vektorraum (mit punktweise definierter Addition und Skalarmultiplikation; Untervektorraum von $\mathrm{Hom}(V,W)$). Seien nämlich $f_1, f_2 \in L(V,W)$. Dann gilt

$$\|(f_1+f_2)(v)\| \le \|f_1(v)\| + \|f_2(v)\| \le (\|f_1\| + \|f_2\|) \cdot v.$$

Also ist $f_1 + f_2$ ein beschränkter Operator, und es gilt

$$\|f_1 + f_2\| \le \|f_1\| + \|f_2\|.$$

Analog folgt $\|\lambda f\| = |\lambda \cdot \|f\|$. Ist $\|f\| = 0$, so ist für alle $v \in V$ auch $\|f(v)\| = 0$ und deshalb $f(v) = 0$; also ist f das Nullelement in $L(V, W)$.

Sei nun W ein Banach–Raum und (f_n) eine Cauchy–Folge in $L(V, W)$. Sei $\varepsilon > 0$ fixiert und N so gewählt, daß für $m, n > N$ die Abschätzung $\|f_m - f_n\| < \varepsilon$ gilt. Für $v \in V$ gilt dann $\|f_m(v) - f_n(v)\| \le \varepsilon \|v\|$ und deshalb ist $(f_n(v))$ eine Cauchy–Folge in W. Sei $f(v)$ ihr Limes. Aus (1.4) und der Linearität der f_m zeigt man, daß $v \mapsto f(v)$ eine lineare Abbildung ist. Da die Norm eine stetige Abbildung ist, folgt aus $\|f_m(v) - f_n(v)\| \le \varepsilon \|v\|$ die Ungleichung $\|f_m(v) - f(v)\| \le \varepsilon \|v\|$, und damit schließen wir auf $\|f_m - f\| \le \varepsilon$. Also konvergiert (f_n) in $L(V, W)$ gegen f. □

(3.3) Notiz. *Seien $f : U \to V$ und $g : V \to W$ beschränkte Operatoren. Dann ist auch $g \circ f$ beschränkt, und es gilt $\|g \circ f\| \le \|g\| \cdot \|f\|$.* □

(3.4) Satz. *Sei $A : V \to W$ ein beschränkter Operator. Dann gilt*

$$\|A\| = \sup_{x, \|x\|=1} \|Ax\|.$$

Beweis. Aus $\|Ax\| \le \|A\| \|x\|$ folgt $\|Ax\| \le \|A\|$ für alle x mit $\|x\| = 1$. Also ist das fragliche Supremum $L \le \|A\|$. Sei umgekehrt $v \neq 0$. Dann hat $\|v\|^{-1} v$ die Norm 1 und aus

$$\left\| A\left(\frac{v}{\|v\|}\right) \right\| \le L, \qquad \|Av\| \le L\|v\|$$

folgt $\|A\| \le L$. □

(3.5) Beispiel. Jede lineare Abbildung $f : \mathbb{R}^m \to \mathbb{R}^n$ ist stetig, wenn man etwa die euklidische Norm zugrundelegt. Unstetige lineare Abbildungen treten also erst bei unendlich-dimensionalen Vektorräumen auf. ◇

(3.6) Beispiel. Sei $\mathbb{R}^\infty$ der Vektorraum aller Folgen $(a_n \mid n \in \mathbb{N})$, bei denen nur endlich viele $a_n \neq 0$ sind. Die Formel für die euklidische Norm definiert auch auf $\mathbb{R}^\infty$ eine Norm. Die Vektoren e_k, die die k-te Komponente gleich Eins haben und alle anderen gleich Null, haben die Norm 1 und bilden eine Basis im Sinne der linearen Algebra. Eine lineare Abbildung $f : \mathbb{R}^\infty \to \mathbb{R}$ wird durch die Werte $f(e_k)$ bestimmt. Setzen wir $f(e_k) = k$, so ist der Operator f nicht beschränkt. ◇

(3.7) Beispiel. Das Integral ist eine lineare Abbildung

$$\int : C^0[a, b] \to \mathbb{R}, \quad f \mapsto \int_a^b f(t)\, dt.$$

Es gilt $|\int_a^b f(t)\, dt| \le (b - a)\|f\|_\infty$. Demnach ist $\int$ ein beschränkter linearer Operator, also eine stetige lineare Abbildung

$$\int : (C^0[a, b], \| - \|_\infty) \to (\mathbb{R}, | - |).$$

Als Spezialfall von I(6.4) erhalten wir die früher schon bewiesenen Aussage

$$\lim \int_a^b f_n = \int_a^b \lim f_n.$$

Hierin ist $\lim f_n$ bezüglich $\| - \|_\infty$ zu lesen. ◇

(3.8) Beispiel. Sei $C^1[a,b]$ der Vektorraum der stetig differenzierbaren Funktionen $[a,b] \to \mathbb{R}$. Die Ableitung ist eine lineare Abbildung

$$D\colon C^1[a,b] \to C^0[a,b].$$

Wir definieren auf $C^1[a,b]$ die C^1-Norm als die Summe zweier Sup-Normen durch

$$\|f\|_{C^1} = \sup |f| + \sup |f'|$$

und setzen, der Einheitlichkeit halber,

$$\|f\|_{C^0} = \sup |f|.$$

Dann ist $\|Df\|_{C^0} = \sup |f'| \leq \|f\|_{C^1}$ und demnach D ein beschränkter Operator bezüglich dieser Normen. Wir können $C^1[a,b]$ auch als Unterraum von $C^0[a,b]$ mit der C^0-Norm betrachten. Dann ist D nicht beschränkt, da im Fall $[a,b] = [0,2\pi]$ gilt: $f_n\colon x \mapsto \sin x$ hat die C^0-Norm 1 und Df_n hat die C^0-Norm n. ◇

(3.9) Beispiel. $C^1[a,b]$ mit der C^1-Norm ist vollständig. Eine Folge (f_n) in diesem Raum konvergiert nämlich , wenn sowohl (f_n) als auch (f'_n) im gewöhnlichen Sinne gleichmäßig konvergiert. Wir haben gesehen, daß dann die Limesfunktion wieder stetig differenzierbar ist. ◇

(3.10) Satz. *Seien E und F normierte Vektorräume. Sei F vollständig, sei $A \subset E$ ein Untervektorraum und sei $\varphi\colon A \to F$ eine stetige lineare Abbildung mit Norm L. Dann gibt es genau eine stetige lineare Abbildung $\Phi\colon \overline{A} \to F$, die φ erweitert; sie hat dieselbe Norm L.*

Beweis. Nach I(8.1) gibt es eine stetige Erweiterung Φ. Sie ist linear: Die Abbildung $(x,y) \mapsto \Phi(x+y) - \Phi(x) - \Phi(y)$ ist auf $A \times A$ gleich Null, also auch auf der Hülle $\overline{A} \times \overline{A}$ von $A \times A$; ebenso für die Skalarmultiplikation. Also ist Φ linear. Eine Relation $\|\varphi(z)\| \leq (L+\varepsilon)\|z\|$ bleibt bei Grenzübergängen erhalten; also gilt sie auch für Φ; das zeigt die Behauptung über die Norm.

4 Aufgaben und Ergänzungen

1. Seien d_1 und d_2 Metriken auf E. Sie heißen *äquivalent*, wenn es positive Zahlen a und b gibt, so daß für alle $x,y \in E$ die Ungleichung $ad_1(x,y) \leq d_2(x,y) \leq bd_1(x,y)$ gilt. Äquivalenz von Metriken auf E ist eine Äquivalenzrelation. Äquivalente Metriken führen zu demselben Begriff von Cauchy–Folge. Auch der Begriff der gleichmäßigen Stetigkeit ändert sich nicht, wenn die Metrik eines beteiligten Raumes durch eine äquivalente ersetzt wird.

2. Seien N_1 und N_2 Normen auf dem Vektorraum V. Sie heißen *äquivalent*, wenn es positive Zahlen a und b so gibt, daß für alle $v \in V$ die Ungleichung $aN_1(v) \leq N_2(v) \leq bN_1(v)$ gilt. Äquivalente Normen führen zu äquivalenten Metriken. Äquivalenz von Normen ist eine Äquivalenzrelation auf der Menge der Normen auf V.

Im vierten Abschnitt werden wir zeigen, daß je zwei Normen auf dem $\mathbb{R}^n$ äquivalent sind. Für die Normen $\| - \|_1, \| - \|_2, \| - \|_\infty$ weise man diese Tatsache durch eine direkte Rechnung nach.

15 Topologische Grundbegriffe

Der Grundbegriff der Stetigkeit hängt nur mittelbar von einer Metrik ab. Er läßt sich mit dem Grundbegriff der offenen Menge formulieren. Wird letzterer zugrundegelegt, so erhält man den Begriff eines topologischen Raumes.

1 Topologische Räume

Die grundlegenden mengentheoretischen Eigenschaften der offenen Mengen werden im Begriff einer Topologie zusammengefaßt.

Sei X eine Menge und $P(X)$ die Menge aller Teilmengen von X. Eine Teilmenge $\mathcal{O} \subset P(X)$ heißt *Topologie* auf X, wenn sie die folgenden Eigenschaften hat:

(1) Jede Vereinigung von Mengen aus $\mathcal{O}$ ist eine Menge aus $\mathcal{O}$.
(2) Durchschnitte von endlich vielen Mengen aus $\mathcal{O}$ liegen in $\mathcal{O}$.
(3) Die leere Menge und X liegen in $\mathcal{O}$.

Ein *topologischer Raum* ist ein Paar $(X, \mathcal{O})$, das aus einer Menge X und einer Topologie $\mathcal{O}$ auf X besteht. Die Mengen in $\mathcal{O}$ heißen die *offenen Mengen* des topologischen Raumes $(X, \mathcal{O})$. Wir sagen auch, X *trägt* die Topologie $\mathcal{O}$. Wie üblich bezeichnen wir einen topologischen Raum oft nur durch die zugrundeliegende Menge X.

Eine Menge $A \subset X$ heißt *abgeschlossen* in $(X, \mathcal{O})$, wenn ihr Komplement $X \setminus A$ offen in $(X, \mathcal{O})$ ist. Das System aller abgeschlossenen Mengen hat die zu (1) – (3) dualen Eigenschaften, die sich daraus durch Komplementbildung ergeben:

(1) Der Schnitt beliebig vieler abgeschlossener Mengen ist abgeschlossen.
(2) Eine Vereinigung endlich vieler abgeschlossener Mengen ist abgeschlossen.
(3) Die leere Menge und X sind abgeschlossen.

(1.1) Notiz. *Sei $\mathcal{O}(E)$ die Menge aller offenen Mengen eines metrischen Raumes E. Dann ist $\mathcal{O}(E)$ eine Topologie auf E.* □

Ist $\mathcal{O}$ eine Topologie auf X und gibt es eine Metrik d auf X, so daß $\mathcal{O} = \mathcal{O}(X, d)$ ist, so heißt die Topologie *metrisch* und der topologische Raum $(X, \mathcal{O})$ *metrisierbar*.

Die Begriffe Umgebung, abgeschlossene Hülle, Inneres, Berührpunkt werden wie für metrische Räume definiert.

(1.2) Der Raum $\overline{\mathbb{R}}$. Um die offenen Mengen von $\overline{\mathbb{R}}$ zu definieren, erklären wir zunächst die Umgebungen von Punkten. Ist $x \in \mathbb{R}$, so heiße U Umgebung von x, wenn es ein $\varepsilon > 0$ mit $U_\varepsilon(x) \subset U$ gibt. Ist $x = \infty$, so heiße U Umgebung von x, wenn es ein $M \in \mathbb{R}$ gibt mit $]M, \infty] \subset U$. Ist $x = -\infty$, so heiße U Umgebung von x, wenn es ein $m \in \mathbb{R}$ gibt mit $[-\infty, m[\subset U$. Eine Menge heiße schließlich offen, wenn sie Umgebung jedes ihrer Elemente ist. Man bestätigt leicht, daß die so definierten offenen Mengen eine Topologie bilden. ◇

Ist $(X, \mathcal{O})$ ein topologischer Raum und $A \subset X$, so ist

$$\mathcal{O}|A = \{U \subset A \mid \text{es gibt } V \in \mathcal{O} \text{ mit } U = A \cap V\}$$

eine Topologie auf A. Sie heißt die *induzierte Topologie* oder die *Teilraumtopologie* oder die *Relativtopologie* von A bezüglich des Raumes $(X, \mathcal{O})$. Der Raum $(A, \mathcal{O}|A)$ heißt *Unterraum* oder *Teilraum* von $(X, \mathcal{O})$, oder kurz: A Unterraum von X.

(1.3) Notiz. *Sei (E, d) ein metrischer Raum und $F \subset E$. Die Teilraumtopologie $\mathcal{O}(E, d)|F$ ist die Topologie, die von der auf F eingeschränkten Metrik d induziert wird.* □

(1.4) Beispiel. Wenn wir die in (1.4) definierte Topologie von $\overline{\mathbb{R}}$ auf die Teilmenge $\mathbb{R}$ einschränken, so erhalten wir die Standardtopologie von $\mathbb{R}$. ◇

Man hat zu beachten, daß durch die Relativtopologie jede Teilmenge ihre eigene Topologie bekommt. Es kann zum Beispiel $U \subset A$ in der Relativtopologie von A offen sein, ohne in X offen zu sein. Zum Beispiel ist ja der gesamte Raum immer offen, also ist $C = [a, b[$ offen im Raum C, aber nicht offen im Raum $\mathbb{R}$. In dem Teilraum $[0, 1[\cup]1, 2]$ von $\mathbb{R}$ sind die beiden Teilmengen $[0, 1[$ und $]1, 2]$ sowohl offen als auch abgeschlossen.

(1.5) Satz. *Über die Menge $\mathcal{U}(x)$ der Umgebungen eines Punktes $x \in X$ gelten die folgenden Aussagen:*

(1) *Jede Obermenge einer Umgebung von x ist eine Umgebung von x.*
(2) *Der Durchschnitt endlich vieler Umgebungen von x ist eine Umgebung von x.*
(3) *Jede Umgebung von x enthält x.*
(4) *Ist U eine Umgebung von x, so gibt es eine weitere Umgebung V von x, so daß U Umgebung jedes Punktes von V ist.* □

Den leichten Nachweis aus den Definitionen übergehen wir. Es ist jedoch gelegentlich nützlich, zu wissen, daß der Umgebungsbegriff auch zur axiomatischen Definition eines topologischen Raumes benutzt werden kann. Felix Hausdorff, der in seinem 1914 erschienenen Buch „Grundzüge der Mengenlehre“ den Grundbegriff geprägt hat, ist so vorgegangen [**?**]. Es gilt:

(1.6) Satz. *Sei X eine Menge. Jedem Element x von X sei ein System $\mathcal{U}(x)$ von Teilmengen, genannt Umgebungen von x, so zugeordnet, daß die im voranstehenden Satz genannten Aussagen gelten. Dann gibt es genau eine Topologie auf X, so daß $\mathcal{U}(x)$ das Umgebungssystem von x bezüglich dieser Topologie ist.*

Beweis. Wir definieren eine Menge $\mathcal{T} \subset P(X)$ durch: $U \in \mathcal{T} \Leftrightarrow$ zu jedem $x \in U$ gibt es ein $V \in \mathcal{U}(x)$ mit $V \subset U$. Dann zeigt man, daß $\mathcal{T}$ eine Topologie ist. Gibt es eine Topologie $\mathcal{S}$ mit den im Satz genannten Eigenschaften, so muß offenbar $\mathcal{S} \supset \mathcal{T}$ sein; und da eine offene Menge Umgebung jedes ihrer Punkte ist, so ist auch $\mathcal{S} \subset \mathcal{T}$.

Es muß noch gezeigt werden, daß $\mathcal{U}(x)$ aus den Umgebungen von x bezüglich $\mathcal{T}$ besteht. Sei $U \in \mathcal{U}(x)$, und sei $W \subset U$ die Teilmenge der $u \in U$, zu denen es ein $W_u \in \mathcal{U}(u)$ mit $W_u \subset U$ gibt. Wir zeigen, daß W offen in der Topologie $\mathcal{T}$ ist und x enthält. Die Relation $x \in W$ ist klar, weil $U = W_x$ gewählt werden kann. Sei $z \in W$, $W_z \subset U$, $W_z \in \mathcal{U}(z)$. Es gibt nach (1.2.4) zu W_z eine Menge $V_z \in \mathcal{U}(z)$, so daß $W_z \in \mathcal{U}(v)$ für alle $v \in V_z$. Aus $v \in V_z$ folgt also $W_z \in \mathcal{U}(v)$, $W_z \subset U$ und damit $v \in W$ nach Definition von W. Da $z \in W$ beliebig war, ist $W \in \mathcal{T}$ und deshalb wegen $W \subset U$ die Menge U eine Umgebung von x in $\mathcal{T}$. □

Ein System von Umgebungen eines Punktes x heißt *Umgebungsbasis* des Punktes x, wenn in jeder Umgebung von x eine Menge dieses Systems enthalten ist. Die Umgebungen sind dann einfach die Obermengen von Mengen dieses Systems. Ein Raum erfüllt das *erste Abzählbarkeitsaxiom*, wenn jeder Punkt eine abzählbare Umgebungsbasis hat. Abzählbare Umgebungsbasen sind der Grund dafür, daß man Grenzwerttheorie mit Folgen betreiben kann. In allgemeinen Räumen muß man Folgen zu Netzen verallgemeinern. Eine Folge $(x_i \mid i \in \mathbb{N})$ in einem topologischen Raum X heißt *konvergent mit dem Grenzwert* x, wenn zu jeder Umgebung U von x ein N existiert, so daß für $m > N$ das Element x_m in U enthalten ist.

Eine Teilmenge $\mathcal{B}$ aller offenen Mengen $\mathcal{O}$ eine Raumes $(X, \mathcal{O})$ heißt *Basis der Topologie* $\mathcal{O}$, wenn jede Menge $U \in \mathcal{O}$ Vereinigung von Elementen aus $\mathcal{B}$ ist. Eine Teilmenge $\mathcal{S}$ von $\mathcal{O}$ heißt *Subbasis*, wenn die Menge aller endlichen Durchschnitte aus $\mathcal{S}$ eine Basis von $\mathcal{O}$ ist. Sind $\mathcal{O}_1$ und $\mathcal{O}_2$ Topologien auf X und gilt $\mathcal{O}_1 \subset \mathcal{O}_2$, so heißt $\mathcal{O}_1$ *gröber* als $\mathcal{O}_2$ und $\mathcal{O}_2$ *feiner* als $\mathcal{O}_1$. Die feinste Topologie hat alle Teilmengen als offene Mengen; sie heißt die *diskrete Topolgie*. Die gröbste Topologie auf X hat nur X und $\emptyset$ als offene Teilmengen; sie heißt die *Klumpentopologie*.

Ein Raum erfüllt das *zweite Abzählbarkeitsaxiom*, wenn er eine abzählbare Basis hat. Das zweite Abzählbarkeitsaxiom impliziert das erste. Räume mit abzählbarer Basis werden manchmal *separabel* genannt.

Die Topologie eines metrischen Raumes X hat die folgende weitere Eigenschaft:

(T_2) Zu je zwei verschiedenen Punkten x und y gibt es disjunkte Umgebungen U von x und V von y.

Hat ein topologischer Raum diese Eigenschaft T_2, so sagen wir, er sei ein *Hausdorff–Raum*, *hausdorffsch* oder *separiert*.

2 Abgeschlossene Hülle und Inneres

Der nächste Satz sammelt einige Eigenschaften der abgescglossenen Hülle und des Inneren. Der Satz I(5.1) gilt mit gleichem Beweis auch in topologischen Räumen.

(2.1) Satz. *Seien A und B Teilmengen des Raumes X. Dann gilt:*
(1) *Aus $A \subset B$ folgt $\overline{A} \subset \overline{B}$.*
(2) $\overline{A \cup B} = \overline{A} \cup \overline{B}$.
(3) $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.
(4) *Aus $A \subset B$ folgt $A^\circ \subset B^\circ$.*
(5) $(A \cap B)^\circ = A^\circ \cap B^\circ$.
(6) $(A \cup B)^\circ \supset A^\circ \cup B^\circ$.
(7) $X \setminus \overline{A} = (X \setminus A)^\circ$.
(8) $\overline{X \setminus A} = X \setminus A^\circ$.

Beweis. (1) Es gilt $A \subset B \subset \overline{B}$. Da $\overline{B}$ abgeschlossen ist und A enthält, gilt $\overline{A} \subset \overline{B}$.
(2) Wegen $A \subset A \cup B$ ist nach (1) $\overline{A} \subset \overline{A \cup B}$ und folglich $\overline{A} \cup \overline{B} \subset \overline{A \cup B}$. Wegen $A \subset \overline{A}$ und $B \subset \overline{B}$ ist $A \cup B \subset \overline{A} \cup \overline{B}$. Da $\overline{A} \cup \overline{B}$ abgeschlossen ist, folgt $\overline{A \cup B} \subset \overline{A} \cup \overline{B}$.
(3) Aus $A \subset \overline{A}$ und $B \subset \overline{B}$ folgt $A \cap B \subset \overline{A} \cap \overline{B}$. Da $\overline{A} \cap \overline{B}$ abgeschlossen ist, folgt $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$.
Der Beweis für (4), (5) und (6) ist „dual“ zum Beweis für (1), (2) und (3).
(7) $X \setminus \overline{A}$ ist als Komplement einer abgeschlossenen Menge offen und in $X \setminus A$ enthalten. Also gilt $X \setminus \overline{A} \subset (X \setminus A)^\circ$.

Wegen $X \setminus A \supset (X \setminus A)^\circ$ liegt (Komplementbildung) A in der abgeschlossenen Menge $X \setminus (X \setminus A)^\circ$ und folgt liegt auch $\overline{A}$ darin. Durch abermalige Komplementbildung erhält man $X \setminus \overline{A} \subset (X \setminus A)^\circ$.
Der Beweis von (8) ist „dual“ zum Beweis von (7). □

Der *Rand* von A in X wird durch $\overline{A} \cap (\overline{X \setminus A})$ definiert. Wir bezeichnen den Rand mit $\mathrm{Rd}(A)$.

3 Stetigkeit

Wir kommen nun zur allgemeinen Formulierung der Stetigkeit mit Hilfe von Umgebungen und offenen Mengen.

Sei $f\colon X \to Y$ eine Abbildung zwischen topologischen Räumen. Sie heißt *stetig an der Stelle* x, wenn das Urbild jeder Umgebung von $f(x)$ eine Umgebung von x ist. Äquivalent dazu sind: Das Urbild jeder Umgebung von $f(x)$ enthält eine Umgebung von x. Zu jeder Umgebung V von $f(x)$ gibt es eine Umgebung U

von $f(x)$ mit $f(U) \subset V$. Im Falle der metrischen Räume genügt es, mit ε-Umgebungen zu arbeiten, das heißt, f ist stetig bei x, wenn zu jedem $\varepsilon > 0$ ein $\delta >$ existiert mit $f(U_\delta(x)) \subset U_\varepsilon(f(x))$. Weiterhin heißt f *stetig*, wenn f an allen Stellen stetig ist. Verkettungen stetiger Abbildunge sind offenbar stetig. Wir erläutern nun, wie der Begriff der Stetigkeit mit den anderen topologischen Grundbegriffen zusammenhängt.

(3.1) Satz. *Sei $f\colon X \to Y$ eine Abbbildung zwischen topologischen Räumen. Folgende Aussagen sind äquivalent:*
(1) *f ist stetig.*
(2) *Das Urbild jeder offenen Menge ist offen.*
(3) *Für jede Teilmenge B von Y gilt $f^{-1}(B^\circ) \subset (f^{-1}(B))^\circ$.*
(4) *Für jede Teilmenge B von Y gilt $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$.*
(5) *Für jede Teilmenge A von X gilt $f(\overline{A}) \subset \overline{f(A)}$.*
(6) *Das Urbild jeder abgeschlossenen Menge ist abgeschlossen.*

Beweis. (1) $\Leftrightarrow$ (2). Sei $V \subset Y$ offen. Dann ist V eine Umgebung jedes Punktes $v \in V$. Also ist nach Definition der Stetigkeit $U = f^{-1}(V)$ eine Umgebung jedes Punktes dieser Menge. Eine Menge ist aber genau dann offen, wenn sie Umgebung jedes ihrer Elemente ist.

Sei umgekehrt V eine Umgebung von $f(x)$. Nach Definition von Umgebung und (2) ist $f^{-1}(V)$ eine Umgebung von x. Also ist f bei x stetig.
(2) $\Rightarrow$ (3). $f^{-1}(B^\circ)$ ist als Urbild einer offenen Menge offen und in $f^{-1}(B)$ enthalten. Nach Definition des Inneren folgt die behauptete Inklusion.
(3) $\Rightarrow$ (4). Wir benutzen (2.2.7) und mengentheoretische Dualität

$$\begin{aligned} X \setminus \overline{f^{-1}(B)} = (X \setminus f^{-1}(B))^\circ = f^{-1}(X \setminus B)^\circ \quad \supset \\ f^{-1}((X \setminus B)^\circ) = f^{-1}(X \setminus \overline{B}) = X \setminus f^{-1}(\overline{B}). \end{aligned}$$

Die Inklusion gilt wegen (3). Durch Komplementbildung ergibt sich die Behauptung.
(4) $\Rightarrow$ (5). Wir haben

$$f^{-1}(\overline{f(A)}) \supset \overline{f^{-1}f(A)} \supset \overline{A},$$

wobei die erste Ungleichung wegen (4) gilt und die zweite wegen $f^{-1}f(A) \supset A$. Die Inklusion zwischen den äußeren Termen ist aber äquivalent zur Behauptung.
(5) $\Rightarrow$ (6). Sei $B \subset Y$ abgeschlossen. Mittels (5) erhalten wir

$$f(\overline{f^{-1}(B)}) \subset \overline{f(f^{-1}(B))} \subset \overline{B} = B.$$

Also gilt $\overline{f^{-1}(B)} \subset f^{-1}(B)$; die umgekehrte Inklusion ist aber klar; also besteht Gleichheit, und deshalb ist $f^{-1}(B)$ abgeschlossen.
(6) $\Rightarrow$ (2) gilt wegen mengentheoretischer Dualität. □

(3.2) Satz. *Sei X ein topologischer Raum und V ein normierter Vektorraum über $\mathbb{K}$. Sind $f, g\colon X \to V$ bei $a \in X$ stetig und ist $\lambda \in \mathbb{K}$, so sind auch $f + g$ und λf bei a stetig.*

Beweis. Sei $\varepsilon > 0$ gegeben. Da f und g bei a stetig sind, gibt es eine Umgebung U von a, so daß für $x \in U$ die Ungleichungen

$$\|f(x) - f(a)\| < \varepsilon/2, \quad \|g(x) - g(a)\| < \varepsilon/2$$

gelten. Für $x \in U$ gilt dann

$$\|(f+g)(x) - (f+g)(a)\| \le \|f(x) - f(a)\| + \|g(x) - g(a)\| < \varepsilon,$$

was die Stetigkeit von $f + g$ bei a belegt. Analog für λf. □

Eine stetige Abbildung $f\colon X \to Y$ heißt *Homöomorphismus*, wenn sie eine stetige Umkehrabbildung $g\colon Y \to X$ besitzt. Gibt es einen Homöomorphismus $f\colon X \to Y$, so heißen die Räume X und Y *homöomorph*.

Eine stetige Abbildung $f\colon (Y, \mathcal{S}) \to (X, \mathcal{O})$ heißt *Einbettung*, wenn f injektiv ist und die durch f vermittelte Abbildung $(Y, \mathcal{S}) \to (f(Y), \mathcal{O}|f(Y))$, $y \mapsto f(y)$ ein Homöomorphismus ist.

(3.3) Satz. *Ist A Teilraum von X, so ist die Inklusion $i\colon A \to X$ stetig. Ist Y ein weiterer Raum und $f\colon Y \to X$ eine Abbildung, deren Bild in A enthalten ist, so ist $f\colon Y \to X$ genau dann stetig, wenn die Abbildung $\varphi\colon Y \to A$, $y \mapsto f(y)$ stetig ist.*

Beweis. Ist U offen in X, so ist $i^{-1}(U) = A \cap U$, und das ist nach der Definition offen in A. Ist φ stetig, so auch die Zusammensetzung $f = i \circ \varphi$. Ist f stetig und V offen in A, so gibt es eine offene Menge U in X, für die $V = A \cap U$ ist. Also ist $\varphi^{-1}(V) = f^{-1}(U)$ offen. □

(3.4) Satz. *Sei $f\colon X \to Y$ eine Abbildung zwischen topologischen Räumen und sei X Vereinigung der Teilmengen $(X_j \mid j \in J)$. Sind die X_j offen, so ist f genau dann stetig, wenn alle Einschränkungen $f_j = f|X_j$ stetig sind. Sind die X_j abgeschlossen, so gilt die analoge Aussage, falls J endlich.*

Beweis. Wir behandeln den Fall abgeschlossener X_j. Ist f stetig, so auch die Zusammensetzung mit der Inklusion $i_j\colon X_j \subset X$; das ist aber gerade die Einschränkung. Seien die Einschränkungen stetig und sei $C \subset Y$ abgeschlossen. Dann ist $f^{-1}(C) = \bigcup f_j^{-1}(C)$ eine endliche Vereinigung abgeschlossener Mengen und folglich abgeschlossen. □

4 Produkte

Sei $((X_j, \mathcal{O}_j) \mid j \in J)$ eine Familie topologischer Räume. Die Produktmenge $X = \prod_{j \in J} X_j$ ist die Menge aller Familien $(x_j \mid j \in J)$ mit $x_j \in X_j$. Wir haben die Projektionen auf die Faktoren $\mathrm{pr}_i\colon X \to X_i$, $(x_j) \mapsto x_i$. Wir nennen eine Teilmenge $B \subset X$ Basismenge, wenn es endlich viele $j_1, \ldots, j_n \in J$ und offene Mengen $U_k \in \mathcal{O}_{j_k}$ gibt, mit denen

$$B = \bigcap_{i=1}^{n} \mathrm{pr}_{j_i}^{-1}(U_i)$$

ist. Diese Mengen bilden die Basis einer Topologie $\mathcal{O}$ auf X, der sogenannten *Produkttopologie.* Wir nennen $(X, \mathcal{O})$ das *topologische Produkt* der $(X_j, \mathcal{O}_j)$.

(4.1) Satz. *Die Produkttopologie ist die gröbste Topologie, für die alle Projektionen* pr_j *stetig sind. Eine Abbildung* $f\colon Y \to X$ *eines topologischen Raumes* Y *nach* X *ist genau dann stetig, wenn die Komponentenabbildungen* $\mathrm{pr}_j \circ f$ *sämtlich stetig sind.*

Beweis. Soll pr_j stetig sein, so müssen zumindest die Mengen $\mathrm{pr}_j(U)$ für in X_j offenes U offen sein. Die Produkttopologie ist so definiert, daß diese Mengen eine Subbasis bilden. Das zeigt die erste Behauptung.

Ist $f\colon Y \to X$ stetig, so auch die Zusammensetzung $\mathrm{pr}_j \circ f$ von stetigen Abbildungen. Für die Umkehrung verwenden wir, daß eine Abbildung stetig ist, wenn die Urbilder von Subbasismengen offen sind. □

(4.2) Satz. *Die Projektionen* $\mathrm{pr}_j\colon X \to X_j$ *sind offene Abbildungen.*

Beweis. Sei U in X offen. Ist U ein endlicher Durchschnitt von in der Definition der Produkttopologie auftretenden Subbasismengen, so ist $\mathrm{pr}_j(U)$ offen. Wegen der Regel $\mathrm{pr}_j(\bigcup_\lambda A_\lambda) = \bigcup_\lambda \mathrm{pr}_j(A_\lambda)$ ist also das Bild jeder offenen Menge offen. □

Sind X_j, Y_j topologische Räume und $f_j\colon X_j \to Y_j$ Abbildungen, so ist

$$\prod f_j\colon \prod X_j \to \prod Y_j, \quad (x_j \mid j \in J) \mapsto (f_j(x_j) \mid j \in J)$$

das Produkt der f_j.

(4.3) Satz. *Ist kein* X_j *leer, so ist* $f = \prod f_j$ *genau dann stetig, wenn alle* f_j *stetig sind.*

Beweis. Seien die f_j stetig. Es ist $\mathrm{pr}_j \circ f = f_j \circ \mathrm{pr}_j$, also ist nach dem vorletzten Satz f stetig.

Ist umgekehrt f stetig, so ist jedenfalls die Zusammensetzung $\mathrm{pr}_j \circ f = f_j \circ \mathrm{pr}_j$ stetig. Da kein X_j leer ist, gilt für jede Menge $U \subset X_j$ die Gleichheit $\mathrm{pr}_j(\mathrm{pr}_j^{-1}(U)) = U$. Ist $U = f_j^{-1}(V)$, mit einer in Y_j offenen Menge V, so ist wegen der Stetigkeit von $f_j \circ \mathrm{pr}_j$ die Menge $\mathrm{pr}_j^{-1}(U)$ offen und deshalb nach dem letzten Satz auch U. □

Für zwei Faktoren schreiben wir natürlich $X_1 \times X_2$ für das Produkt der Räume und analog $f_1 \times f_2$ für das Produkt zweier Abbildungen. Die „identische Abbildung“ $X_1 \times (X_2 \times X_3) \to (X_1 \times X_2) \times X_3$ ist ein Homöomorphismus. Das folgt unmittelbar mit dem letzten Satz. Ganz allgemein ist das topologische Produkt assoziativ, das heißt mit beliebigen Klammerungen verträglich.

Die folgende Konstruktion verallgemeinert die Definition der Produkttopologie. Sei $(Y_j \mid j \in J)$ eine Familie topologischer Räume. Seien $f_j\colon X \to Y_j$ Abbildungen einer Menge X nach Y_j. Es gibt eine gröbste Topologie auf X, für die alle f_j stetig sind. Eine Abbildung eines topologischen Raumes Z nach X mit dieser Topologie ist genau dann stetig, wenn die Zusammensetzungen mit allen f_j stetig sind. Die definierte Topologie auf X heißt *Initialtopologie* bezüglich der Familie (f_j). Als Subbasismengen für die Initialtopologie werden die Urbilder offener Mengen bei den f_j verwendet.

(4.4) Satz. *Die folgenden Aussagen über einen topologischen Raum X sind gleichwertig:*

(1) *X ist ein Hausdorff–Raum.*
(2) *Jeder Punkt $x \in X$ ist Durchschnitt seiner abgeschlossenen Umgebungen.*
(3) *Die Diagonale $D = D_X = \{(x,x) \mid x \in X\}$ ist abgeschlosssen in $X \times X$.*

Beweis. (1) $\Rightarrow$ (2). Sei $x \neq y$. Wir wählen disjunkte offene Umgebungen U von x und V von y. Dann ist $X \setminus V$ eine abgeschlossene Umgebung von x, die y nicht enthält.

(1) $\Rightarrow$ (3). Sei $x \neq y$. Wir wählen disjunkte offene Umgebungen U von x und V von y. Dann ist $U \times V$ offen in $X \times Y$ und $D \cap (U \times V) = \emptyset$. Also ist $X \times X \setminus D$ als Vereinigung von Mengen des Typs $U \times V$ offen.

(3) $\Rightarrow$ (1). Sei $X \times X \setminus D$ offen. Ist $x \neq y$, so ist $(x,y) \in X \times X \setminus D$. Nach Definition der Produkttopologie gibt es eine Basismenge $U \times V$ dieser Topologie mit $(x,y) \in U \times V \subset X \times X \setminus D$. Das bedeutet aber, daß U und V disjunkte offene Umgebungen von x und y sind.

(2) $\Rightarrow$ (1). Ist $x \neq y$, so gibt es eine abgeschlossene Umgebung V von x, die y nicht enthält. Es ist dann $X \setminus V$ eine Umgebung von y, die zur Umgebung V von x disjunkt ist. □

(4.5) Satz. *Seien $f, g\colon X \to Y$ stetige Abbildungen in einen Hausdorff–Raum. Dann ist die Koinzidenzmenge $A = \{x \mid f(x) = g(x)\}$ abgeschlossen in X. Stimmen also zwei stetige Abbildungen in einen Hausdorff–Raum auf einer dichten Teilmenge überein, so sind sie gleich.*

Beweis. Die Diagonalabbildung $d\colon X \to X \times X$, $x \mapsto (x,x)$ ist stetig, wie unmittelbar aus Satz (4.2) folgt. Es ist $A = ((f \times g)d)^{-1}(D_Y)$. □

5 Aufgaben und Ergänzungen

1. Gibt alle Topologien auf einer Menge mit 2 und 3 Elementen an und klassifiziere sie bis auf Homöomorphie (siehe den nächsten Abschnitt).

2. Seien d_1 und d_2 äquivalente Metriken auf E. Dann sind die zugenörigen Topologien gleich.

3. Seien A und B Teilmengen eines Raumes X. Es gilt $\mathrm{Rd}(A) = \overline{A} \setminus A^\circ$. Eine Menge A ist genau dann abgeschlossen, wenn $\mathrm{Rd}(A) \subset A$ ist. Es gilt $\mathrm{Rd}(A \cup B) \subset \mathrm{Rd}(A) \cup \mathrm{Rd}(B)$.

4. Sei A Teilmenge eines topologischen Raumes X. Wieviele verschiedene Teilmengen von X kann man höchstens aus A durch die Prozesse „abgeschlossene Hülle“ und „Komplement“ erhalten? Beispiel für die Maximalzahl im Fall $X = \mathbb{R}$?

5. Die Abbildung $\varphi \colon \overline{\mathbb{R}} \to [-1,1]$

$$\varphi(x) = \begin{cases} \frac{x}{1+|x|} & x \in \mathbb{R} \\ \pm 1 & x = \pm\infty \end{cases}$$

ist ein Homöomorphismus. Damit zeigt man, daß die Topologie auf $\overline{\mathbb{R}}$ metrisch ist.

6. Seien X und Y topologische Räume. Sei X Vereinigung von A_1 und A_2. Sei $f \colon X \to Y$ eine Abbildung und f_j die Einschränkung von f auf A_j. Die f_j seien stetig. Unter der folgenden Voraussetzung ist f stetig:

$$(\overline{X \setminus A_1}) \cap (X \setminus A_2) = \emptyset, \quad (\overline{X \setminus A_2}) \cap (X \setminus A_1) = \emptyset.$$

7. Für Teilräume $A \subset X$ und $B \subset Y$ gilt $\mathrm{Rd}(A \times B) = \mathrm{Rd}(A) \times \overline{B} \cup \overline{A} \times \mathrm{Rd}(B)$.

8. Die Räume $]0,1[$, $]0,1]$ und $[0,1]$ sind paarweise nicht homöomorph. Dagegen sind $]0,1] \times [0,1]$ und $]0,1] \times]0,1[$ homöomorph. Man kann also in einem Homöomorphismus $X \times Z \cong Y \times Z$ im allgemeinen nicht Z wegkürzen.

9. Seien X_j, $j \in J$, topologische Räume und $A_j \subset X_j$ Teilmengen. Dann gilt:

(1) $\overline{\prod A_j} = \prod \overline{A_j}$.

(2) Ist kein A_j leer, so ist $\prod A_j$ genau dann in $\prod X_j$ abgeschlossen, wenn alle A_j in X_j abgeschlossen sind.

10. Seien (X_i, d_i) metrische Räume. Durch

$$d((x_1,y_1),(x_2,y_2)) = d_1(x_1,x_2) + d_2(y_1,y_2)$$

wird eine Metrik auf $X_1 \times X_2$ definiert. Sie induziert ebenfalls die Produkttopologie.

11. Eine Metrik $\delta(x,y)$ heißt *beschränkt* durch M, wenn für alle x,y immer $\delta(x,y) \le M$ ist. Sei (X,d) ein metrischer Raum. Dann wird durch

$$\delta(x,y) = \frac{d(x,y)}{1 + d(x,y)}$$

eine durch 1 beschränkte Metrik δ auf X definiert, die dieselbe Topologie wie d induziert.

12. Sei $((X_j, d_j) \mid j \in \mathbb{N})$ eine abzählbare Familie metrischer Räume mit durch 1 beschränkten Metriken d_j. Dann wird auf dem Produkt $\prod_{j=1}^\infty X_j$ durch

$$d((x_j),(y_j)) \mid j \in \mathbb{N}) = \sum_{n=1}^\infty 2^{-n} d_n(x_n,y_n)$$

eine Metrik definiert, die die Produkttopologie induziert.

13. Sei $\mathbb{R}$ homöomorph zu einem Produkt $X \times Y$. Dann besteht X oder Y nur aus einem Punkt.

14. Die Menge $Q_t(r,k) = \{(x,y) \in \mathbb{R}^r \times \mathbb{R}^k \mid \sum_{i=1}^r x_i^2 = t + \sum_{j=1}^k y_j^2\}$ ist für positive (r,k,t) homöomorph zu $S^{r-1} \times \mathbb{R}^k$.

15. Sei W ein Untervektorraum des normierten Raumes V. Dann ist auch $\overline{W}$ ein Untervektorraum. Zum Beweis betrachte man die stetige Abbildungen $q\colon W \times W \to W$, $(x,y) \mapsto x - y$ und $\mathbb{R} \times W \to W$, $(\lambda, x) \mapsto \lambda x$ und wende die Verträglichkeit mit der abgeschlossnen Hülle an.

16 Kompakte Räume

Die Vollständigkeit der reellen Zahlen hat den Überdeckungsatz von Heine–Borel als beweistechnisch wichtige Konsequenz. In der Tat könnte man diesen Satz auch als Vollständigkeitsaxiom postulieren. In einer axiomatischen Theorie wird die Aussage des Überdeckungssatzes als Eigenschaft der Kompaktheit formuliert. Die Kompaktheit ist das Äußerste an „Endlichkeit“, das man der Unendlichkeit der Zahlen und ähnlicher Objekte abgewinnen kann.

1 Kompaktheit

Sei A eine Teilmenge eines topologischen Raumes X. Eine *offene Überdeckung* von A ist eine Familie $(U_j \mid j \in J)$ von offenen Mengen $U_j \subset X$, deren Vereinigung A enthält. Die Menge A heißt *kompakt*, wenn es zu jeder offenen Überdeckung $(U_j \mid j \in J)$ von A eine endliche Teilmenge $K \subset J$ gibt, so daß $(U_j \mid j \in K)$ eine Überdeckung von A ist. Wir haben früher den *Überdeckungssatz von Heine–Borel* bewiesen:

(1.1) Satz. *Ein abgeschlossenes Intervall in $\overline{\mathbb{R}}$ ist kompakt.* □

(1.2) Satz. *Sei X kompakt, $A \subset X$ abgeschlossen und $f\colon X \to Y$ stetig. Dann sind auch A und $f(X)$ kompakt.*

Beweis. Sei $(U_j \mid j \in J)$ eine offene Überdeckung von A. Nehmen wir zu den U_j noch $X \setminus A$ hinzu, so erhalten wir eine offene Überdeckung von X, von denen endlich viele zur Überdeckung von X ausreichen.

Sei $(B_k \mid k \in K)$ eine offene Überdeckung von $f(X)$. Dann bilden die $(f^{-1}(B_k)$ eine offene Überdeckung von X. Wir wählen eine endliche Teilüberdeckung $(f^{-1}(B_j) \mid j \in K)$ aus, und dann liegt $f(X)$ in $\bigcup_{j \in K} B_j$. □

(1.3) Satz. *Sei A eine kompakte Teilmenge eines metrischen Raumes X. Dann ist A abgeschlossen.*

Beweis. Wir zeigen, daß $X \setminus A$ offen ist. Sei $z \in X \setminus A$. Zu jedem $a \in A$ wählen wir disjunkte offene Mengen $U(a)$ und $V(a)$ mit $z \in U(a)$ und $a \in V(a)$. Die $V(a)$ bilden eine offene Überdeckung von A. Da A kompakt ist, gibt es endlich viele, etwa $V(a_1), \dots, V(a_n)$, die A überdecken. Dann ist $U = \bigcap_{j=1}^n U(a_j)$ eine offene Umgebung, die disjunkt zu $V = \bigcup_{j=1}^n V(a_j)$ ist. Es ist $U \subset X \setminus A$. □

Allgemeiner genügt es im letzten Satz, X als Hausdorff–Raum vorauszusetzen. Wir haben genauer bewiesen, daß es dann zu jeder abgeschlossenen Menge A und jedem nicht darin gelegenen Punkt disjunkte Umgebungen gibt. Ist diese Aussage für einen topologischen Raum richtig, so sagt man, er erfülle das *Trennungsaxiom* T_3. Ein Raum heißt *regulär*, wenn er hausdorffsch ist und T_3 erfüllt.

Durch mengentheoretische Dualität erhält man aus der Definition der Kompaktheit:

(1.4) Satz. *Sei $A \subset X$ kompakt. Sei $(a_j \mid j \in J)$ eine Familie abgeschlossener Mengen und sei $A \cap (\bigcap_{j\in J} A_j) = \emptyset$. Dann gibt es eine endliche Teilmenge $K \subset J$, so daß $A \cap (\bigcap_{j\in K} A_j) = \emptyset$.*

Beweis. Sei $U_j = X \setminus A_j$. Dann ist $A \subset (X \setminus \bigcap_{j\in J} A_j = \bigcup_{j\in J} U_j$. Die U_j bilden also eine offene Überdeckung von A. Nun wende man die Definition der Kompaktheit an. □

(1.5) Folgerung. *Sei $A_1 \supset A_2 \subset A_3 \supset \dots$ eine geschachtelte Folge abgeschlossener nichtleerer Mengen. Liege A_1 in einer kompakten Teilmenge A. Dann ist der Schnitt $\bigcap_{j=1}^\infty A_j$ nicht leer.* □

(1.6) Beispiel. Sind in (1.5) die $A_j = [a_j, b_j]$ abgeschlossene Intervall von $\mathbb{R}$, so spricht man von einer *Intervallschachtelung*. Der Schnitt einer Intervallschachtelung ist also nicht leer. Bildet die Länge $b_j - a_j$ der Intervalle eine Nullfolge, so besteht dieser Schnitt also aus genau einem Punkt. Als Vollständigkeitsaxiom der reellen Zahlen könnten man auch postulieren, daß der Schnitt einer Intervallschachtelung nichtleer ist. ◇

(1.7) Satz. *Sei (x_n) eine Folge in einem kompakten metrischen Raum. Dann ist die Menge der Häufungswerte eine nichtleere kompakte Menge.*

Beweis. Das folgt aus I(6.5) und (1.5). □

Im Falle der reellen Zahlen ist der letzte Satz nach Bolzano und Weierstraß benannt. Eine beschränkte Folge reeller Zahlen hat eine nichtleere kompakte Menge von Häufungswerten; sie besitzt immer konvergente Teilfolgen.

2 Kompaktheit und Produkte

(2.1) Satz. *Sei B (bzw. C) kompakter Teilraum von X (bzw. Y). Sei ferner $(A_j \mid j \in J)$ eine Familie offener Teilmengen von $X \times Y$ mit $B \times C \subset \bigcup_{j\in J} A_j$.*

Dann gibt es offene Umgebungen U von B in X (und V von C in Y) und endliches $E \subset J$, so daß

$$U \times V \subset \bigcup_{j \in E} A_j.$$

Beweis. Sei zunächst $B = \{b\}$. Für jedes $c \in C$ gibt es offene Umgebungen M_c von b in X und N_c von c in Y derart, daß $M_c \times N_c$ in einer Menge $A_{j(c)}$, $j(c) \in J$ enthalten ist; das entnimmt man der Definition der Produkttopologie. Es ist $(N_c \mid c \in C)$ eine offene Überdeckung von C. Sei $C \subset (N_{c_1} \cup \ldots \cup N_{c_n})$, was wegen der Kompaktheit von C möglich ist. Sei

$$U = \bigcap_{i=1}^n M_{c_i}, \qquad V = \bigcup_{i=1}^n N_{c_i}.$$

Dann ist

$$\{b\} \times C \subset U \times V \subset \bigcup_{i=1}^n A_{j(c_i)}.$$

Sei nun B eine beliebige kompakte Teilmenge. Nach dem eben Gezeigten gibt es zu jedem $b \in B$ eine offene Umgebung U_b von b in X und eine offene Umgebung V_b von C in Y, so daß $U_b \times V_b$ in einer endlichen Vereinigung

$$\bigcup_{j \in J(b)} A_j, \qquad J(b) \subset J \text{ endlich,}$$

liegt. Es ist $(U_b \mid b \in B)$ eine offene Überdeckung von B. Sei $B \subset (U_{b_1} \cup \ldots \cup U_{b_m})$, was wegen der Kompaktheit von B möglich ist. Sei

$$U = \bigcup_{i=1}^m U_{b_i)}, \qquad V = \bigcap_{i=1}^m V_{b_i}.$$

Dann ist

$$B \times C \subset U \times V \subset \bigcup_{i=1}^m \bigcup_{j \in J(b_i)} A_j$$

und damit die Behauptung gezeigt. □

(2.2) Satz. *Seien X und Y kompakt. Dann ist $X \times Y$ kompakt.*

Beweis. Man wähle im vorigen Satz $B = X = U$ und $C = Y = V$. □

(2.3) Satz. *Eine Menge $A \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.*

Beweis. Sei A abgeschlossen und beschränkt. Beschränktheit heißt, es gibt ein $r > 0$, so daß $A \subset [-r, r]^n$. Nach (1.1) und (2.3) ist $[-r, r]^n$ kompakt und mithin auch A als abgeschlossene Menge (1.2).

Sei A kompakt. Die Mengen $U_n(0)$ überdecken A. Also gibt es ein n mit $A \subset U_n(0)$, das heißt A ist beschränkt. Als kompakte Teilmenge eines metrischen Raumes ist A auch abgeschlossen (1.3). □

3 Kompaktheit und Stetigkeit

(3.1) Satz. *Sei $f\colon X \to \mathbb{R}$ stetig und sei X kompakt. Dann hat f ein Minimum und ein Maximum.*

Beweis. Nach (1.2) ist $f(X)$ kompakt und nach (2.3) abgeschlossen und beschränkt. Also gibt es $M = \sup f(X) \in \mathbb{R}$. Da das Supremum einer Menge ein Berührpunkt dieser Menge ist und $f(X)$ abgeschlossen, ist $M \in f(X)$. □

(3.2) Satz. *Sei $f\colon X \to Y$ eine stetige Abbildung zwischen metrischen Räumen. Sei X kompakt. Dann ist f gleichmäßig stetig.*

Beweis. Der Beweis wird genauso wie für reelle Funktionen geführt. Statt des Betrages muß man nur die Metrik verwenden. □

Die gleichmäßige Stetigkeit ist keine rein topologische Eigenschaft, da man Umgebungen verschiedener Punkte der Größe nach vergleichen muß.

Eine Abbildung $f\colon X \to Y$ heißt *Identifizierung* oder *Quotientabbildung*, wenn sie surjektiv ist und eine Menge $B \subset Y$ genau dann offen (abgeschlossen) ist, wenn $f^{-1}(B)$ offen (abgeschlossen) ist. Eine Identifizierung ist stetig.

(3.3) Satz. *Sei $f\colon X \to Y$ eine bijekive stetige Abbildung. Sei X kompakt und Y ein Hausdorff–Raum. Dann ist f ein Homöomorphismus.*

Beweis. Da f bijektiv ist gibt es die Umkehrabbildung g. Die Stetigkeit von g folgt, wenn wir zeigen: f bildet eine abgeschlossene Menge auf eine abgeschlossene Menge ab. Sei $A \subset X$ abgeschlossen. Dann ist nach (1.2) A kompakt und $f(A)$ ebenfalls. Nach (1.3) ist dann $f(A)$ abgeschlossen. □

Eine Abbildung heißt *abgeschlossen* bzw. *offen*, wenn das Bild einer abgeschlossenen (offenen) Menge wieder abgeschlossen (offen) ist. Im Beweis von (3.3) haben eigentlich gezeigt:

(3.4) Notiz. *Eine stetige Abbildung eines kompakten Raumes in einen hausdorffschen ist abgeschlossen.* □

(3.5) Notiz. *Eine stetige surjektive offene (abgeschlossene) Abbildung $f\colon X \to Y$ ist eine Identifizierung.*

Beweis. Sei $B \subset Y$ gegeben und $f^{-1}(B)$ offen. Da f surjektiv ist, so ist $ff^{-1}(B) = B$. Ist f offen, so ist also B offen. Ist B offen, so auch $f^{-1}(B)$, da f stetig ist. □

(3.6) Notiz. *Sei $f\colon X \to Y$ eine Identifizierung und $g\colon Y \to Z$ eine Abbildung in einen topologischen Raum Z. Genau dann ist g stetig, wenn $g \circ f$ stetig ist.* □

Beweis. Mit f und g ist auch $g \circ f$ stetig. Sei umgekehrt $g \circ f$ stetig und $V \subset Z$ offen. Dann ist $f^{-1}(g^{-1}(V))$ offen und deshalb nach Definition einer Identifizierung auch $g^{-1}(V)$. Also sind Urbilder offener Mengen bei g offen und deshalb g stetig. □

(3.7) Satz von Dini. *Sei* $(f_n\colon X \to \mathbb{R})$ *eine Folge stetiger Funnktionen auf dem kompakten Raum* X. *Für* $t \in X$ *und* $n \in \mathbb{N}$ *gelte* $f_n(t) \le f_{n+1}(t)$. *Die Folge konvergiere punktweise gegen eine stetige Funktion* f. *Dann ist die Konvergenz gleichmäßig.*

Beweis. Sei $\varepsilon > 0$ gegeben. Zu jedem $x \in X$ wählen wir ein $n(x) \in \mathbb{N}$, so daß für $n \ge n(x)$ die Ungleichung $f(x) - f_n(x) < \varepsilon/3$ gilt. Da f und $f_{n(x)}$ stetig sind, gibt es eine Umgebung $V(x)$ von x, so daß für $y \in V(x)$ gilt

$$|f(y) - f(x)| < \frac{\varepsilon}{3}, \qquad |f_{n(x)}(y) - f_{n(x)}(x)| < \frac{\varepsilon}{3}.$$

Für jedes $y \in V(x)$ gilt dann nach der Dreiecksungleichung $f(y) - f_{n(x)}(y) < \varepsilon$. Der Raum X wird von endlich vielen $V(x)$ überdeckt, etwa von $V(x_1), \dots, V(x_m)$. Sei $N = \max(n(x_1), \dots, n(x_k))$. Für $n \ge N$ und alle $y \in X$ gilt dann $f(y) - f_n(y) < \varepsilon$. □

(3.8) Satz. *Je zwei Normen auf dem* $\mathbb{R}^n$ *sind äquivalent.*

Beweis. Sei N eine beliebige Norm und $\| - \|$ die Sup-Norm. Da Äquivalenz von Normen eine Äquivalenzrelation ist, wie eine kleine Überlegung zeigt, genügt es zu zeigen, daß diese beiden Normen äquivalent sind. Wir schreiben einen Vektor $x = (x_i) = \sum_i x_i e_i$ als Linearkombination der Standardbasis. Die Dreiecksungleichung für N liefert

$$N(x) = N(\sum x_i e_i) \le \sum |x_i| N(e_i) \le a\|x\|$$

mit $a = \sum_i N(e_i)$. Die Ungleichung $|N(x) - N(y)| \le N(x-y) \le a\|x-y\|$ zeigt, daß $N\colon (\mathbb{R}^n, \| - \|) \to \mathbb{R}$ stetig ist. Die Menge $S = \{x \mid \|x\| = 1\}$ ist kompakt, denn die durch $\| - \|$ induzierte Topologie auf $\mathbb{R}^n$ ist die Produkttopologie und S ist abgeschlossen und beschränkt. Da N stetig ist, hat $N(S)$ ein Minimum b. Es ist $b \ne 0$, da S nicht den Nullvektor enthält. Also gilt für $x \in S$ die Ungleichung $N(x) \ge b\|x\|$, und wegen der Homogenität einer Norm gilt diese Ungleichung dann auch für alle $x \in \mathbb{R}^n$. Also gilt immer $b\|x\| \le N(x) \le a\|x\|$, was definitionsgemäß die Äquivalenz der beiden Normen besagt. □

4 Der Approximationssatz von Stone–Weierstraß

Sei X ein kompakter Hausdorff–Raum und $C(X)$ der Vektorraum der stetigen Funktionen $f\colon X \to \mathbb{R}$. Auf $C(X)$ haben wir die Supremumsnorm $\|f\| = \sup\{|f(x)| \mid x \in X\}$. Dadurch wird $C(X)$ zu einem Banach–Raum. Wir sagen, eine Teilmenge A von $C(X)$ *trennt die Punkte*, wenn zu jedem Paar x, y verschiedener Punkte in X eine Funktion in A existiert, die auf x und y verschiedene Werte annimmt. Ein Untervektorraum A heißt *Unteralgebra*, wenn mit $f, g \in A$ auch das Produkt $f \cdot g \in A$ ist. Der Satz von *Stone–Weierstraß* lautet:

(4.1) Satz. *Sei* $A \subset C(X)$ *eine Unteralgebra, die die Punkte trennt und die konstanten Funktionen enthält. Dann ist die abgeschlossene Hülle von* A *gleich* $C(X)$.

Beweis. (1) Wir nehmen zunächst zusätzlich an, daß mit f und g auch $\max(f,g)$ und $\min(f,g)$ zu A gehören. Seien x_1 und x_2 verschiedene Punkte von X und seien a_i reelle Zahlen. Dann gibt es $h \in A$ mit $h(x_i) = a_i$. Nach Voraussetzung gibt es nämlich $g \in A$ mit $g(x_1) \neq g(x_2)$, und dann hat

$$h(x) = a_1 + (a_2 - a_1)\frac{g(x) - g(x_1)}{g(x_2) - g(x_1)}$$

die gewünschte Eigenschaft.

(2) Sei $f \in C(X)$ und $\varepsilon > 0$. Wir suchen $g \in A$, so daß die Ungleichungen $f - \varepsilon < g < f + \varepsilon$ gelten. Nach (1) wählen wir zu jedem Paar $x, y \in X$ eine Funktion $h_{x,y} \in A$ mit $h_{x,y}(x) = f(x)$ und $h_{x,y}(y) = f(y)$. Zu jedem $y \in X$ gibt es dann eine offene Umgebung U_y, so daß für $z \in U_y$ die Ungleichung $h_{x,y}(z) < f(z) + \varepsilon$ gilt. Werde X durch $U_{y(1)}, \ldots, U_{y(n)}$ überdeckt. Nach unserer Zusatzvoraussetzung liegt dann das Mininum h_x der $h_{x,y(j)}$ in A und erfüllt $h_x < f + \varepsilon$ sowie $h_x(x) = f(x)$. Es gibt sodann zu jedem $x \in X$ eine offene Umgebung V_x von x, so daß für $z \in V_x$ die Ungleichung $f(z) - \varepsilon < h_x(z)$ gilt. Wir überdecken X durch $V_{x(1)}, \ldots, V_{x(m)}$ und bilden das Maximum g der $h_{x(j)}$. Diese Funktion hat die gewünschten Eigenschaften.

(3) Es bleibt zu zeigen, daß die Zusatzvoraussetzung immer erfüllt ist. Wegen $2\max(f,g) = |f+g| + |f-g|$ und $2\min(f,g) = |f+g| - |f-g|$ genügt es zu zeigen, daß mit f auch $|f|$ in A liegt. Sei P ein Polynom in $t \in \mathbb{R}$, so daß für $t \in [-a, a]$ immer $|P(t) - |t|| < \varepsilon$ gilt, und sei $a = \|f\|$. Dann folgt $|P(f(x)) - |f(x)|| < \varepsilon$, und $x \mapsto P(f(x))$ liegt in A.

(4) Um geeignete Polynome P zu finden, reduzieren wir durch die Substitution $t \mapsto at$ auf den Fall $a = 1$. Wir definieren induktiv $P_1 = 0$ sowie $P_{n+1}(s) = P_n(s) + \frac{1}{2}[s^2 - P_n(s)^2]$. Diese Polynome konvergieren monoton wachsend und gleichmäßig gegen $|s|$. Um das einzusehen, verwendet man die Identität

$$|s| - P_{n+1}(s) = (|t| - P_n(s))(1 - \frac{1}{2}(|s| + P_n(s)),$$

die durch Umformung der Rekursionsformel erhalten wird. Sie liefert induktiv

$$0 \leq P_n(s) \leq P_{n+1}(s) \leq |s|.$$

Daraus folgt zunächst, daß $(P_n(s))$ immer konvergiert. Indem wir in der Rekursionsformel zum Limes übergehen, sehen wir, daß der Grenzwert $|s|$ ist. Aus dem Satz von Dini folgt die Gleichmäßigkeit der Konvergenz. Man kann aber auch direkt induktiv die Abschätzung

$$|s| - P_n(s) \leq |s|(1 - \frac{1}{2}|s|)^n < \frac{2}{n+1}$$

zeigen und braucht dann nicht den Satz von Dini. □

(4.2) Folgerung. *Sei $X \subset \mathbb{R}^n$ kompakt. Jede stetige Funktion $X \to \mathbb{R}$ ist gleichmäßiger Limes von Polynomen in n Variablen. Die Unteralgebra A der*

Polynome, eingeschränkt auf X, erfüllt nämlich die Voraussetzungen des letzten Satzes. □

(4.3) Beispiel. Sei $f\colon [a,b] \to \mathbb{R}$ stetig. Für jedes $n \in \mathbb{N}_0$ gelte $\int_a^b f(t)t^n\,dt = 0$. Dann ist $f = 0$. Zunächst gilt dann nämlich für jedes Polynom P auch $\int_a^b f(t)P(t)\,dt = 0$. Nach Weierstraß wählen wir dann P so, daß $\|f - P\| < \varepsilon$ ist. Es folgt $|\int_a^b f^2(t)\,dt| = |\int_a^b (f-P)f| \le \varepsilon\|f\|(b-a)$. Aus $\int f^2 = 0$ folgt aber $f = 0$. ◇

(4.4) Satz. *Sei $C(X,\mathbb{C})$ der Raum der stetigen Funkionen $X \to \mathbb{C}$ mit Supremumsnorm. Sei der komplexe Untervektorraum A eine Unteralgebra, die die konstanten Funktionen enthält, die Punkte trennt und mit f auch die konjugiert-komplexe Funktion $\overline{f}$ enthält. Dann ist die abgeschlossene Hülle von A gleich $C(X,\mathbb{C})$.*

Beweis. Ist f eine beliebige stetige Funktion, so genügt es zu zeigen, daß ihr Real- und Imaginärteil in der Hülle liegen. Das folgt aus dem Satz von Stone–Weierstraß, wenn gezeigt ist, daß die Algebra A_0 der reellwertigen Funktionen in A, die Voraussetzungen dieses Satzes erfüllt. Mit g ist aber der Realteil $\frac{1}{2}(f+\overline{f})$ und der Imaginärteil $\frac{1}{2i}(f-\overline{f})$ in A, also in A_0. Trennt g die Punkte x und y, trennt entweder der Realteil oder der Imaginärteil diese Punkte. Die anderen Voraussetzungen sind offenbar erfüllt. □

(4.5) Beispiel. Die Unteralgebra von $C(S^1,\mathbb{C})$, die von den Funktionen $z \mapsto z^n$ erzeugt wird, erfüllt die Voraussetzungen des letzten Satzes. ◇

Wir wollen die Situation des Beispiels auf periodische Funktionen $\mathbb{R} \to \mathbb{C}$ anwenden. Sei $f\colon S^1 \to \mathbb{C}$ stetig. Dann ist $g\colon \mathbb{R} \to \mathbb{C}$, $t \mapsto \exp(it))$ eine stetige Funktion der Periode 2π. Letzteres heißt $g(t+2\pi) = g(t)$ für alle $t \in \mathbb{R}$. Umgekehrt: Zu jeder stetigen Funktion $g\colon \mathbb{R} \to \mathbb{C}$ der Periode 2π gibt es genau eine stetige Funktion $f\colon S^1 \to \mathbb{C}$, so daß $f(\exp(it)) = g(t)$ ist. Es gibt nämlich genau eine Mengenabbildungen f mit dieser Eigenschaft, und die Stetigkeit folgt daraus, daß $\mathbb{R} \to S^1$, $t \mapsto \exp(it)$ eine Identifizierung ist. Damit erhalten wir aus (4.5):

(4.6) Folgerung. *Zu jeder stetigen Funktion $f\colon \mathbb{R} \to \mathbb{C}$ der Periode 2π und zu jedem $\varepsilon > 0$ gibt es $T(x) = \sum_{k=-n}^{n} c_k e^{ikx}$, so daß $\|f - T\| < \varepsilon$ ist.* □

Eine Funktion der Form $x \mapsto \sum_{k=-n}^{n} c_k e^{ikx}$ heißt *trigonometrisches Polynom.* In Analogie zu (4.3) erhält man nun:

(4.7) Folgerung. *Ist $f\colon \mathbb{R} \to \mathbb{C}$ stetig und gilt für jedes trigonometrische Polynom T die Gleichung $\int_0^{2\pi} \overline{T(x)}f(x)\,dx = 0$, so ist $f = 0$.* □

5 Kompakte metrische Räume

Eine Teilmenge $A \subset X$ heißt *relativ kompakt*, wenn ihre abgeschlossene Hülle kompakt ist. Ein metrischer Raum X heißt *präkompakt*, wenn gilt: Zu jedem ε

gibt es eine endliche Überdeckung von X durch Mengen mit einem Durchmesser kleiner als ε. Diese Bedingung ist äquivalent zu der folgenden: Zu jedem $\varepsilon > 0$ gibt es eine endliche Menge $F \subset X$, so daß für jedes $x \in X$ die Ungleichung $d(x, F) < \varepsilon$ besteht. Ein Raum heißt *folgenkompakt*, wenn jede Folge in diesem Raum einen Häufungswert hat. Nach (1.7) ist ein kompakter metrischer Raum folgenkompakt.

(5.1) Satz. *Sei X ein metrischer Raum. Folgende Aussagen über $A \subset X$ sind äquivalent:*
(1) *$\overline{A}$ ist kompakt.*
(2) *$\overline{A}$ ist folgenkompakt.*
(3) *Jede Folge in A hat einen Häufungspunkt in X.*
(4) *Jede Folge $(a_n \mid n \in \mathbb{N})$, $a_n \in A$, hat eine in X konvergente Teilfolge.*
(5) *$\overline{A}$ ist präkompakt und vollständig.*

Beweis. Die Implikationen (1) $\Rightarrow$ (2) $\Rightarrow$ (3) haben wir schon gezeigt.
(3) $\Rightarrow$ (4). Sei $(a_n \mid n \in \mathbb{N})$ eine Folge in $\overline{A}$. Es gibt $b_n \in A$ mit $d(b_n, a_n) < 2^{-n}$. Es gibt eine konvergente Teilfolge der (b_n) mit einem Limes b; und dann hat die entsprechende Teilfolge der (a_n) denselben Limes. Folglich gilt (3) auch für Folgen aus $\overline{A}$. Ist (a_n) eine Cauchy–Folge mit einer gegen a konvergenten Teilfolge, so konvergiert sie selbst schon gegen a. Also ist $\overline{A}$ vollständig.

Falls für $\alpha > 0$ die Menge $\overline{A}$ nicht durch endlich viele Mengen $B(x, \alpha) = \{y \mid d(x, y) \le \alpha\}$ überdeckt würde, so könnten wir induktiv eine Folge (x_n) mit $d(x_i, x_j) > \alpha$ für alle $i \ne j$ konstruieren, die dann keine konvergente Teilfolge hätte.
(4) $\Rightarrow$ (1). Sei $(U_j \mid j \in J)$ eine offene Überdeckung von $\overline{A}$. Wir definieren induktiv eine Folge von Mengen $B(x_n, 2^{-n}) = B_n$ wie folgt: Angenommen, $\overline{A}$ wird von keiner endlichen Teilfamilie der (U_j) überdeckt. Dann gibt es eine endliche Überdeckung durch Mengen $B(x, 1)$ von $\overline{A}$ und folglich ein $B(x_1, 1)$, das ebenfalls nicht durch endlich viele U_j überdeckt wird. Angenommen $B(x_{n-1}, 2^{-(n-1)})$ werde nicht durch endlich viele U_j überdeckt. Wir wählen eine endliche Überdeckung von $\overline{A}$ durch Mengen $B(y_k, 2^{-n})$. Unter den $B(y_k, 2^{-n})$, die $B(x_{n-1}, 2^{-(n-1)})$ schneiden, gibt es eine, etwa $B(x_n, 2^{-n})$, die nicht durch endlich viele U_j überdeckt wird. Für $n \le p < q$ gilt $d(x_p, x_q) \le d(x_p, x_{p+1}) + \ldots + d(x_{q-1}, x_q)$. Da $B_{n-1} \cap B_n \ne \emptyset$ ist, gilt

$$d(x_{n-1}, x_n) \le \frac{1}{2^{n-1}} + \frac{1}{2^n} \le \frac{1}{2^{n-2}}$$

und folglich

$$d(x_p, x_q) \le \frac{1}{2^{p-1}} + \ldots + \frac{1}{2^{q-2}} \le \frac{1}{2^{n-2}}.$$

Also ist (x_p) eine Cauchy–Folge. Diese hat eine konvergente Teilfolge, ist also konvergent, etwa gegen x. Es gibt ein λ, so daß $x \in U_\lambda$. Für ein $\alpha > 0$ ist $B(x, \alpha) \subset U_\lambda$. Es gibt ein n mit

$$d(x, x_n) < \frac{\alpha}{2}, \qquad \frac{1}{2^n} < \frac{\alpha}{2}.$$

Also ist $B(x_n, 2^{-n}) \subset B(x, \alpha) \subset U_\lambda$, im Widerspruch zur Nichtüberdeckbarkeit von $B(x_n, 2^{-n})$ durch endlich viele U_j. □

(5.2) Satz. *Sei X ein kompakter metrischer Raum. Sei $(A_j \mid j \in J)$ eine offene Überdeckung von X. Es gibt ein $\varepsilon > 0$ derart, daß für jedes $x \in X$ die Kugel $U_\varepsilon(x)$ in einer Menge A_j enthalten ist.*

Beweis. Da X kompakt ist, können wir ohne wesentliche Einschränkung annehmen, daß J endlich ist. Seien die $A_j \neq X$, da sonst die Aussage trivial ist. Auf X betrachten wir die mit der Metrik d gebildeten stetigen Funktionen $f_j(x) = d(x, X \setminus A_j)$ und $f(x) = \max(f_j(x) \mid j \in J)$. Die Menge $f(X)$ ist kompakt, also abgeschlossen in $\mathbb{R}$. Das Infimum $r = \inf f(X)$ ist größer als Null, denn $r = 0$ würde $f_j(x) = 0$, also $x \in X \setminus A_j$, also $x \in \bigcap(X \setminus A_j)$ implizieren; der letzte Durchschnitt ist aber leer. Sei $0 < \varepsilon < r$ und $x \in X$. Es ist $f(x) \geq r$. Nach Definition von r und f gibt es ein j mit $f_j(x) \geq r$, also $U_\varepsilon(x) \subset A_j$. □

Eine Zahl ε mit der im letzten Satz genannten Eigenschaft heißt *Lebesguesche Zahl* der Überdeckung $(A_j \mid j \in J)$.

17 Fourier-Reihen

Die Theorie der Fourier-Reihen ist ein wichtiges Instrument der Analysis. Sie beschäftigt sich mit der Entwicklung sehr allgemeiner Funktionen in Reihen aus trigonometrischen Funktionen. Anders als bei den Taylorschen Reihen, die ja unendlich oft differenzierbare Funktionen voraussetzen, kann man in dieser Theorie stetige und noch allgemeinere Funktionen untersuchen.

1 Definition der Fourier-Reihen

Eine Funktion $f\colon \mathbb{R} \to \mathbb{C}$ hat die *Periode* $c > 0$, wenn für alle $x \in \mathbb{R}$ die Gleichung $f(x) = f(x + c)$ gilt. Hat ein stetige Funktion f die Periode c, so können wir die Integrationsgrenzen verschieben

$$\int_0^c f(t)\,dt = \int_a^{a+c} f(t)\,dt.$$

Wir bezeichnen mit P den komplexen Vektorraum der stetigen Funktionen $f\colon \mathbb{R} \to \mathbb{C}$ mit der Periode 2π. Mit der hermiteschen Form

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t)\overline{g(t)}\,dt$$

wird daraus ein *unitärer Raum*. Wir definieren eine Norm auf P durch

$$\|f\|_2 = \sqrt{\langle f, f \rangle}.$$

In der Analysis heißt diese Norm auch L^2-Norm. Eine Familie $(v_j \mid j \in J)$ aus Elementen $v_j \in P$ heißt *Orthonormalsystem*, wenn gilt

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

(1.1) Notiz. *Die Funktionen* $e_k\colon x \mapsto \exp(ikx)$, $k \in \mathbb{Z}$ *bilden ein Orthonormalsystem.*

Beweis. Es ist $2\pi\langle e_k, e_l \rangle = \int_0^{2\pi} \exp(i(k-l)t)\,dt$, und das Integral haben wir früher schon ausgerechnet. □

Ist V ein endlich-dimensionaler unitärer Raum mit Orthonormalbasis $(v_1, \dots, v_n)$, so wird die Komponentendarstellung eines Vektors $v \in V$ durch

$$v = \sum_{i=1}^{n} \langle v, v_j \rangle v_j$$

gegeben. Unter Ausnutzung von Konvergenzbegriffen wollen wir ein analoges Resultat auch für unendlich-dimensionale Vektorräume gewinnen, in unserem Fall für den unitären Raum P. In diesem Kontext verwendet man traditionell eine andere Terminologie.

Die *Fourier-Koeffizienten* von $f \in P$ bezüglich des Orthonormalsystems $(e_k \mid k \in \mathbb{Z})$ sind die Zahlen

$$\textbf{(1.2)} \qquad c_k = c_k(f) = \langle f, e_k \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ikt}\,dt, \quad k \in \mathbb{Z}.$$

Andere typische und wichtige Funktionen der Periode 2π sind $\sin(nx)$ und $\cos(nx)$ mit $n \in \mathbb{N}$. Auch diese Funktionen bilden ein Orthogonalsystem. Wir setzen

$$\textbf{(1.3)} \qquad \begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt)\,dt & n \in \mathbb{N}_0 \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt)\,dt & n \in \mathbb{N}. \end{aligned}$$

Mit Hilfe der Eulerschen Formeln können wir die c_k durch die a_k, b_k ausrechnen, und umgekehrt:

$$\begin{array}{lll} a_0 = 2c_0, & a_k = c_k + c_{-k}, & b_k = i(c_k - c_{-k}) \\ c_0 = \frac{1}{2}a_0, & c_k = \frac{1}{2}(a_k - ib_k), & c_{-k} = \frac{1}{2}(a_k + ib_k). \end{array}$$

Eine Funktion T der Form

$$T(x) = \sum_{k=-n}^{n} c_k e^{ikx}, \qquad x \in \mathbb{R},\ c_k \in \mathbb{C}$$

heißt *trigonometrisches Polynom*. Wir können es in die reelle Form umrechnen

$$T(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx).$$

Das trigonometrische Polynom

$$S_n(f) = \sum_{k=-n}^{n} c_k e^{ikx}$$

mit den durch (1.2) definierten Koeffizienten heißt n-tes *Fourier-Polynom* von f. Die Folge $(S_n f)$ dieser Polynome heißt *Fourier-Reihe* $S_\infty f$ von f. Wir schreiben dafür auch

$$\sum_{-\infty}^{\infty} c_k e^{ikx}.$$

Konvergiert diese Reihe an der Stelle x, so bezeichnet (wir auch sonst üblich bei Reihen) dieses Symbol auch den Grenzwert. Die *reelle Form der Fourrier-Reihe* besteht aus den Partialsummen

$$S_n(f) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

mit den Koeffizienten, die durch (1.3) definiert sind. Aus der Integraldarstellung der a_k und b_k liest man ab:

(1.4) Notiz. *Ist f ungerade, so sind alle $a_k = 0$; ist f gerade, so sind alle $b_k = 0$.* □

Wir werden uns im weiteren der Frage widmen, unter welchen Voraussetzungen die Fourier-Reihe konvergiert und was dann ihr Grenzwert ist. Als eine Konsequenz des Weierstraßschen Approximationssatzes hatten wir hergeleitet:

(1.5) Satz. *Zu jeder Funktion $f \in P$ gibt es eine Folge (T_n) trigonometrischer Polynome T_n, die gleichmäßig gegen f konvergiert.* □

Mit diesem Satz zeigen wir, daß eine Funktion aus P durch ihre Fourier-Koeffizienten eindeutig bestimmt ist:

(1.6) Notiz. *Seien alle Fourier-Koeffizienten von $f \in P$ gleich Null. Dann ist f die Nullfunktion.*

Beweis. Aus der Voraussetzung folgt zunächst, daß für jedes trigonometrische Poynom T das Integral $\int_0^{2\pi} h(t)T(t)\,dt = 0$ ist. Wir wählen eine Folge (T_n) trigonometrischer Polynome, die gleichmäßig gegen f konvergiert. Wegen der gleichmäßigen Konvergenz dürfen wir die Integration mit der Grenzwertbildung vertauschen

$$\int_0^{2\pi} |f(t)|^2\,dt = \lim_{n\to\infty} \int_0^{2\pi} f(t)\overline{T_n(t)}\,dt = 0.$$

Daraus folgt $f = 0$. □

2 Die Parsevalsche Gleichung

Die Norm $\| - \|_2$ führt, wie jede Norm, zu einem Konvergenzbegriff. Während bei endlich-dimensionalen Vektorräumen die Wahl der Norm für den Konvergenzbegriff keine Rolle spielt, sind jetzt die VerHältnisse komplizierter. Die Supremumsnorm $\| - \|_\infty$ führt zum Begriff der gleichmäßigen Konvergenz. Da man eine Intergration auch in manchen Fällen als eine Mittelwertbildung ansieht und da die L^2-Norm durch ein Integral über das Quadrat des Absolutbetrages gegeben ist, bezeichnet man den zugehörigen Konvergenzbegriff auch als *Konvergenz im quadratischen Mittel.* Wir werden zeigen, daß die Fourier-Reihe einer Funktion im quadratischen Mittel immer gegen diese Funktion konvergiert. Dazu vergleichen wir zunächst nach den Regeln der linearen Algebra eine Funktion und ihre Fourier-Polynome

$$\begin{aligned} \|f - \sum_{-n}^{n} c_k e_k\|_2^2 &= \langle f - \sum c_k e_k, f - \sum c_k e_k \rangle \\ &= \langle f, f \rangle - \sum c_k \langle e_k, f \rangle - \sum \overline{c}_k \langle f, e_k \rangle + \sum |c_k|^2 \\ &= \|f\|_2^2 - \sum |c_k|^2. \end{aligned}$$

Durch Grenzübergang erhalten wir daraus

(2.1) Besselsche Ungleichung. *Die Reihe* $\sum_{n=-\infty}^{\infty} |c_k|^2$ *konvergiert und ihr Grenzwert ist kleinergleich* $\langle f, f \rangle$. □

Die Konvergenz der Fourrier-Reihe gegen f in der L^2-Norm besagt $\lim_{n\to\infty} \|f - S_n(f)\|_2 = 0$. Aus der soeben hergeleiteten Identität sehen wir also:

(2.2) Parsevalsche Gleichung. *Genau dann konvergiert die Fourier-Reihe von f im quadratischen Mittel gegen f, wenn die sogenannte Parsevalsche Gleichung* $\|f\|_2^2 = \sum_{k=-\infty}^{\infty} |c_k|^2$ *gilt.* □

(2.3) Satz. *Sei $f \in P$ stetig differenzierbar. Dann konvergiert die Fourier-Reihe von f gleichmäßig.*

Beweis. Partielle Integration liefert

$$\frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int}\, dt = \frac{1}{in} \frac{1}{2\pi} \int_0^{2\pi} f'(t) e^{-int}\, dt,$$

also die Relation $in \cdot c_n(f) = c_n(f')$ zwischen den Fourier-Koeffizienten von f und f'. Wegen

$$|c_n(f)| = \frac{1}{|n|} |c_n(f')| \le \frac{1}{2} \left(\frac{1}{n^2} + |c_n(f')|^2 \right)$$

und der Besselschen Ungleichung erkennen wir die absolute Konvergenz der Reihe $\sum c_n(f)$ und damit die absolute und gleichmäßige Konvergenz der Fourier-Reihe. □

Eine stetige Funktion $f\colon [a,b] \to \mathbb{C}$ heiße stückweise stetig differenzierbar, wenn es eine Zerlegung $a = t_0 < t_1 < \ldots < t_m = b$ so gibt, daß alle Einschränkungen $f|[t_r, t_{r+1}]$ stetig differenzierbar sind. Durch Aufteilung der Integrale liefert die Beweismethode des vorigen Satzes, daß die Fourier-Reihe einer stückweise stetig differenzierbaren Funktion gleichmäßig konvergiert.

(2.4) Satz. *Konvergiert die Fourier-Reihe von f gleichmäßig, so ist ihr Grenzwert f.*

Beweis. Sei $g(x) = S_\infty(f)$ der gleichmäßige Limes der Fourier-Reihe. Die Fourier-Koeffizienten von g errechnen sich dann durch gliedweise Integration

$$c_k(g) = \frac{1}{2\pi} \lim_{n\to\infty} \sum_{\nu=-n}^{n} c_\nu(f) \int_0^{2\pi} e^{i\nu x} e^{-ikx}\, dx = c_k(f).$$

Wegen der Eindeutigkeit (1.6) ist $g = f$. □

(2.5) Satz. *Für jede Funktion $f \in P$ gilt die Parsevalsche Gleichung.*

Beweis. Für trigonometrische Polynome gilt die Parsevalsche Gleichung. Daraus leiten wir die allgemeine Gültigkeit durch Grenzübergang her. Sei $\varepsilon > 0$ und T ein trigonometrisches Polynom, das der Gleichung $\|f-T\|_\infty < \varepsilon$ genügt (1.5). Werde $g = f - T$ gesetzt. Es gilt

$$\|g\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |g(t)|^2\, dt \le \sup |g()|^2 = \|g\|_\infty^2.$$

Sei N so groß gewählt, daß $T = S_n(T)$ für $n > N$ gilt. Damit folgt

$$\begin{aligned} \|f - S_n(f)\|_2 &= \|g + T - S_n(g) - S_n(T)\|_2 \\ &\le \|T - S_n(T)\|_2 + \|g - S_n(g)\|_2 \\ &= \|g - S_n(g)\|_2 \le \|g\|_2 \le \|g\|_\infty < \varepsilon. \end{aligned}$$

Die Ungleichung $\|g - S_n(g)\|_2 \le \|g\|_2$ folgt darin aus der Rechnung vor der Besselschen Ungleichung. □

3 Der Raum l^2

Sei l^2 die Menge der Folgen $(z_k \mid k \in \mathbb{N})$ komplexer Zahlen, für die $\sum_k |z_k|^2$ konvergiert. Wegen $|z+w|^2 \le 2(|z|^2 + |w|^2)$ ist mit (z_k) und (w_k) auch $(z_k + w_k)$ in l^2. Wir erkennen, daß l^2 mit der komponentenweisen Addition und Skalarmultiplikation ein $\mathbb{C}$-Vektorraum ist. Wegen $2|zw| \le |z|^2 + |w|^2$ ist für $z = (z_k)$ und $w = (w_k)$ aus l^2 die Reihe $\langle z, w \rangle = \sum_k z_k \overline{w}_k$ absolut konvergent. Mit der hermiteschen Form $(z,w) \mapsto \langle z, w \rangle$ wird l^2 zu einem unitären Vektorraum. Ein orthogonaler oder unitärer Vektorraum heißt *Hilbert-Raum*, wenn er bezüglich der zugehörigen L^2-Norm vollständig ist.

(3.1) Satz. *Der Raum l^2 ist ein Hilbert-Raum.*

Beweis. Sei $(z^{(k)} \mid k \in \mathbb{N})$ eine Cauchy-Folge in l^2. Wir schreiben $z^{(k)} = (z_n^{(k)} \mid n \in \mathbb{N})$. Sei $\varepsilon > 0$ gegeben und N so gewählt, daß $\|z^{(k)} - z^{(l)}\|_2^2 = \sum_n |z_n^{(k)} - z_n^{(l)}|^2 \leq \varepsilon$ für $k, l \geq N$. Dann folgt für diese k, l natürlich auch $|z_n^{(k)} - z_n^{(l)}|^2 \leq \varepsilon$ für alle n. Wir sehen also, daß $(z_n^{(k)} \mid k \in \mathbb{N})$ eine gewöhnliche Cauchy-Folge ist. Sie konvergiere gegen y_n. Sei $y = (y_n \mid n \in \mathbb{N})$. Wir behaupten, daß y in l^2 liegt und $z^{(k)}$ gegen y konvergiert. Für jedes $M \in \mathbb{N}$ gilt

$$\sum_{n=1}^{M} |z_n^{(k)} - z_n^{(l)}|^2 \leq \varepsilon,$$

sofern $k, l \geq N$. Durch Grenzübergang folgt

$$\sum_{n=1}^{M} |z_n^{(k)} - y_n|^2 \leq \varepsilon.$$

Das gilt für alle M. Also ist

$$\sum_{n=1}^{\infty} |z_n^{(k)} - y_n|^2 \leq \varepsilon$$

für alle $k \geq N$. Daraus folgt zunächst, daß $z^{(k)} - y$ in l^2 liegt und deshalb auch y. Ferner folgt dann daraus, daß $(z^{(k)})$ gegen y konvergiert. □

Wir erhalten eine lineare Abbildung $\iota\colon P \to l^2$, indem wir jeder stetigen Funktion f die Folge $c_0(f), c_1(f), c_{-1}(f), c_2(f), c_{-2}(f), \ldots$ zuordnen. Wegen der Besselschen Ungleichung ist das Bild von f wirklich ein Element in l^2. Wegen des Eindeutigkeitssatzes (1.6) ist ι injektiv. Die Parsevalsche Gleichung besagt, daß ι die Norm erhält, $\|\iota(f)\|_2 = \|f\|_2$. Nach den Rechenregeln der linearen Algebra haben wir die Identität

$$4\langle z, w \rangle = \|z + w\|_2^2 - \|z - w\|_2^2 + i\|z + iw\|_2^2 - i\|z - iw\|_2^2,$$

die in jedem unitären Raum gilt. Also respektiert ι auch die hermitesche Form, das heißt es gilt

$$\langle f, g \rangle = \sum_{-\infty}^{\infty} c_k(f)\overline{c_k(g)}.$$

Wir können also P bis auf Isomorphie als unitären Unterraum von l^2 auffassen. Es stellt sich heraus, daß P nicht der ganze Raum l^2 ist, wohl aber ein dichte Untermenge. Insbesondere ist der unitäre Raum P nicht vollständig. Es erhebt sich die interessante Frage, ob man die fehlenden Elemente auch als Fourier-Reihen geeigneter Funktionen interpretieren kann.

18 Differenzierbarkeit

Für Funktionen von mehreren Veränderlichen gibt es verschiedene Begriffe von Differenzierbarkeit und Ableitung: Die totale Ableitung, die Richtungsableitungen und die partiellen Ableitungen. Die totale Ableitung dient der theoretischen Begriffsbildung, die partiellen Ableitungen werden für konkrete Rechnungen verwendet.

1 Differenzierbarkeit im $\mathbb{R}^n$

Wir untersuchen Funktionen $f: U \to \mathbb{R}^p$, die auf einer offenen Menge U des $\mathbb{R}^n$ definiert sind. Zur Definition der Ableitung verwenden wir den Vektorraum $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ der linearen Abbildungen $\mathbb{R}^n \to \mathbb{R}^p$. Wir identifizieren diesen Raum mit dem Raum $M(p \times n)$ der reellen $p \times n$-Matrizen, indem wir einer linearen Abbildung ihre Matrix bezüglich der Standardbasen (e_k) zuordnen. Dabei ist e_k der Vektor, dessen k-te Komponente 1 ist und dessen andere Komponenten 0 sind. Einer Abbildung $\varphi: U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ entspricht somit eine matrizenwertige Abbildung $\varphi: U \to M(p \times n)$, und φ ist genau dann stetig, wenn alle Matrixeinträge stetig sind. Der Wert von $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ an der Stelle $v \in \mathbb{R}^n$ wird auch mit $A \cdot v$ bezeichnet, um an das Matrizenprodukt zu erinnern; in der Matrizenschreibweise ist dann v als „Spaltenvektor“ zu denken. Auf dem $\mathbb{R}^n$ verwenden wir, wenn nichts anderes gesagt wird, die euklidische Norm $\|v\|$ eines Vektors $v \in \mathbb{R}^n$. Wie erinnerlich ist aber der Begriff der Stetigkeit unabhängig von der verwendeten Norm; er hängt nur von den offenen Mengen, also von der zugrundeliegenden Topologie ab. Im theoretischen Kontext ist die Norm $\|A\|$ von $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ eine Zahl, die für alle v die Ungleichung $\|A \cdot v\| \leq \|A\| \, \|v\|$ erfüllt und durch ?? definiert ist. Wir notieren:

(1.1) Notiz. *Die Abbildung* $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p) \times \mathbb{R}^n \to \mathbb{R}^p$, $(A, v) \mapsto A \cdot v$ *ist stetig.* □

(1.2) Definition. Sei $f: U \to \mathbb{R}^n$ auf einer offenen Menge U des $\mathbb{R}^n$ definiert. Wir nennen f *differenzierbar* im Punkt $a \in U$, wenn es eine stetige Abbildung $\Phi_a: U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ gibt, mit der für alle $x \in U$ die Gleichung

$$f(x) = f(a) + \Phi_a(x) \cdot (x - a)$$

besteht. Wir nennen $A = \Phi_a(a) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ die *Ableitung* oder das *Differential* von f an der Stelle a und bezeichnen es mit $Df(a)$. ◇

Als Ergänzung zur Definition überlegen wir uns, daß die Abbildung A durch die genannten Bedingungen eindeutig bestimmt ist (über die Eindeutigkeit von Φ_a behaupten wir dagegen nichts). Sei also $f(x) = f(a) + \Psi_a(x) \cdot (x - a)$ eine zweite Darstellung mit den in (1.2) genannten Eigenschaften. Sei $v \in \mathbb{R}^n$ beliebig. Da U offen ist, gibt es ein $\varepsilon > 0$, so daß $x = a + tv \in U$ für alle $0 < |t| < \varepsilon$. Es ist dann $(\Phi_a(a + tv) - \Psi_a(a + tv)) \cdot tv = 0$. Nach Division durch t liefert der

Grenzübergang $t \to 0$ wegen der Stetigkeit von Φ_a und Ψ_a bei a die Relation $(\Phi_a(a) - \Psi_a(a)) \cdot v = 0$. Sie gilt für alle v und besagt deshalb, daß die linearen Abbildungen $\Phi_a(a)$ und $\Psi_a(a)$ übereinstimmen.

(1.3) Notiz. *Ist f bei a differenzierbar, so auch stetig.*

Beweis. Das folgt unmittelbar aus der Definition zusammen mit (1.1). □

(1.4) Beispiel. Sei $f: \mathbb{R}^n \to \mathbb{R}^p$, $v \mapsto A \cdot v + b$ eine affine Abbildung ($A \in M(p \times n)\, b \in \mathbb{R}^p$). Dann ist f überall differenzierbar, und es gilt $Df(a) = A$, denn $\Phi_a(x) = A$ erfüllt offenbar die Bedingungen (1.1). ◇

(1.5) Beispiel. Sei $n = 1$. ◇

(1.6) Beispiel. Schreiben wir f in Komponentenform $f = (f_1, \dots, f_p): U \to \mathbb{R}^p$, so entnimmt man der Matrizenform von (1.1) sofort, daß f genau dann bei a differenzierbar ist, wenn alle f_j bei a differenzierbar sind. ◇

(1.7) Linearität.*Seien $f, : U \to \mathbb{R}^p$ bei a differenzierbar und seien α und β reelle Zahlen. Dann ist auch $\alpha f + \beta g$ bei a differenzierbar und es gilt $D(\alpha f + \beta g)(a) = \alpha Df(a) + \beta Dg(a)$.* □

(1.8) Kettenregel.*Sei $f: U \to \mathbb{R}^n$ bei $a \in U \subset \mathbb{R}^m$ differenzierbar, sei $g: V \to \mathbb{R}^p$ bei $b = f(a) \in V \subset \mathbb{R}^n$ differenzierbar und sei $f(U) \subset V$. Dann ist auch $g \circ f$ bei a differenzierbar und es gilt $D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$.*

Beweis. □

Wir wollen jetzt das Differential $Df(a)$ durch eine Matrix bezüglich der Standardbasen beschreiben. Das geschieht durch die partiellen Ableitungen. Sei weiterhin $f: U \to \mathbb{R}^p$ wie in (1.2). Ist $v \in \mathbb{R}^n$ und existiert die Ableitung von $t \mapsto f(a + tv)$ bei $t = 0$, so wird sie mit $D_v f(a)$ bezeichnet und *Richtungsableitung* von f in Richtung v genannt. Ist $v = e_i$, so schreiben wir stattdessen $D_i f(a)$ und sprechen von der *i-ten partiellen Ableitung* von f bei a. Werden die Koordinaten in $\mathbb{R}^n$ mit $(x_1, \dots, x_n)$ bezeichnet, das heißt, wird f in der Form $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$ notiert, so ist für die partielle Ableitung das Symbol

$$D_i f(a) = \frac{\partial f}{\partial x_i}(a)$$

in Gebrauch.

(1.9) Satz. *Sei f bei a differenzierbar. Dann existieren alle Richtungsableitungen von f bei a, und $D_v f(a)$ ist der Wert der linearen Abbildung $Df(a)$ an der Stelle v.*

Beweis. Der Limes $t \to 0$ in

$$\frac{f(a + tv) - f(a)}{t} = \Phi(a + tv) \cdot v$$

liefert die Behauptung wegen (1.1). □

(1.10) Folgerung. *Sei* $f = (f_1, \ldots, f_p)\colon U \to \mathbb{R}^p$. *Die Matrix von* $Df(a)$ *bezüglich der Standardbasen ist*

$$Df(a) = \begin{pmatrix} D_1 f_1(a) & \cdots & D_n f_1(a) \\ & & \\ D_1 f_p(a) & \cdots & D_n f_p(a) \end{pmatrix}.$$

Diese Matrix heißt die Jacobi–Matrix *von* f *an der Stelle* a. *Wir bezeichnen sie meist, wie eben schon, mit demselben Symbol wie das Differential.* ◇

(1.11) Folgerung. *Sei* $v = \sum_i v_i e_i$. *Dann gilt im Falle der Differenzierbarkeit* $D_v f(a) = \sum_i v_i D_i f(a)$. □

Die Kettenregel nimmt natürlich unter Verwendung der Jacobi–Matrizen die Form einer Matrizenmultiplikation an. Wir schreiben hier einen Spezialfall ausdrücklich auf: Sei $w\colon J \to U$ eine differenzierbare Kurve und $f\colon U \to \mathbb{R}$ eine differenzierbare Abbildung, dann hat

$$f \circ w\colon J \to \mathbb{R}, \quad x \mapsto f(w_1(x), \ldots, w_n(x))$$

die Ableitung

$$\textbf{(1.12)} \qquad \frac{d(f \circ w)}{dx}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(w(t)) w_i'(t).$$

Wir erkennen, daß die allgemeine Kettenregel für Funktionen von mehreren Veränderlichen nützlich sein kann, um die Ableitung eine ganz gewöhnlichen Funktion einer Veränderlichen zu berechnen.

Die affine Abbildung $v \mapsto f(a) + Df(a) \cdot v$ ist im Falle der Differenzierbarkeit die beste lineare Approximation an f in einer Umgebung von a (tangentiale Abbildung). Das führt zu einer anderen Charakterisierung der Differenzierbarkeit:

(1.13) Satz. *Genau dann ist* $f\colon U \to \mathbb{R}^p$ *bei* $a \in U$ *differenzierbar, wenn es eine Matrix* $A \in M(p \times n)$ *gibt, mit der die Aussage*

$$\lim_{x \to a} \frac{\|f(x) - f(a) - A \cdot (x-a)\|}{\|x - a\|} = 0$$

gilt. Die Matrix A *ist in diesem Fall eindeutig bestimmt und gleich* $Df(a)$.

Beweis. Sei f bei a differenzierbar und $A = \Phi_a(a)$. Es gilt dann

$$\frac{\|f(x) - f(a) - A \cdot (x-a)\|}{\|x - a\|} = \frac{\|(\Phi(x) - \Phi(a)) \cdot (x-a)\|}{\|x - a\|} \leq \|\Phi(x) - \Phi(a)\|.$$

Die Stetigkeit von Φ bei a liefert $\lim_{x-a} \|\Phi(x) - \Phi(a)\| = 0$.

Sei umgekehrt die Bedingung des Satzes erfüllt. Wir kürzen ab

$$f(x) - f(a) - A \cdot (x-a) = \Delta(x).$$

Als $\Phi(x) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^p)$ definieren wir:

$$v \mapsto \begin{cases} A \cdot v + \Delta(x) \dfrac{\langle x - a, v \rangle}{\langle x - a, x - a \rangle} & x \neq a \\ A \cdot v & x = a \end{cases}$$

Damit gilt dann für $x \neq a$

$$\|(\Phi(x) - \Phi(a)) \cdot v\| = \|\Delta(x)\| \frac{\|\langle x - a, v \rangle\|}{\|x - a\|^2} \leq \frac{\|\Delta(x)\|}{\|x - a\|} \|v\|.$$

In der Rechnung haben wir die Cauchy–Schwarzsche Ungleichung benutzt. Das liefert die Abschätzung $\|\Phi(x) - \Phi(a)\| \leq \|x - a\|^{-1} \|\Delta(x)\|$. Nach Voraussetzung ist der Limes der rechten Seite bei $x \to a$ Null. Das liefert die Stetigkeit von Φ bei a. $\square$

(1.14) Satz. *Die Abbildung $f\colon U \to \mathbb{R}^p$ habe überall partielle Ableitungen. Diese seien bei a stetig. Dann ist f an der Stelle a differenzierbar.*

Beweis. Wegen (1.??) genügt es, den Fall $p = 1$ zu behandeln. $\square$

19 Integration

Ein Integral ordnet einer Funktion einen Zahlenwert zu. Wir formulieren axiomatisch, unter welchen Bedingungen eine derartige Zuordnung Integral heißen soll. Wir entwickeln die Integrationstheorie aus den Axiomen. Ein erstes Ziel ist die Konstruktion von Integralen aus geeigneten Vorstufen. Ein zweites Ziel sind die Konvergenzsätze der Lebesgueschen Theorie. Zur praktischen Berechnung von Integralen über Teilmengen euklidischer Räume dienen der Satz von Fubini und der Transformationssatz.

1 Der Integralbegriff

Sei X eine nichtleere Menge. Ein *Integrationsraum* auf X ist ein Untervektorraum V des Vektorraumes aller Funktionen $X \to \mathbb{R}$, der mit f auch $|f|$ enthält. Mit f und g sind dann auch $\max(f, g) = 2^{-1}(|f - g| + f + g)$, $\min(f, g)$, $f^+ = \max(f, 0)$ und $f^- = f^+ - f$ in V enthalten. Es ist $f = f^+ - f^-$ die kanonische Zerlegung von f. Wir setzen $V^+ = \{f \in V \mid f = f^+\}$.

Eine lineare Abbildung (auch lineares *Funktional* genannt) $I\colon V \to \mathbb{R}$ heißt *positiv*, wenn aus $f \geq 0$ immer $I(f) \geq 0$ folgt; und *stetig*, wenn für jede absteigende Folge $f_1 \geq f_2 \geq f_3 \geq \ldots$ von $f_i \in V$, die punktweise gegen die Nullfunktion konvergiert, der Limes $\lim_{n\to\infty} I(f_n) = 0$ ist.

Ein *Präintegral* ist ein Tripel (X, V, I), das aus einer nichtleeren Menge X, einem Integrationsraum V auf X und einem positiven stetigen linearen Funktional $I\colon V \to \mathbb{R}$ besteht. Wir schreiben dann auch $I(f) = \int_X f = \inf_X f(x) dx$

und nennen diesen Wert das *Integral* von f über X. Die Funktionen in V heißen *integrierbar* oder *integrabel*.

Für ein Integral werden wir noch weitere Axiome hinzufügen, die insbesondere sicherstellen sollen, daß genügend viele Funktionen integrierbar sind. Der Deutlichkeit halber sei gesagt: V ist ein Untervektorraum des Vektorraums aller Funktionen $X \to \mathbb{R}$ mit den punktweise definerten algebraischen Verknüpfungen. Eine Folge (f_n) konvergiert *punktweise*, wenn die Folge reeller Zahlen $(f_n(x))$ für jedes x aus X konvergiert. Im allgemeinen kann man nicht sagen, genau welche Funktionen integrierbar sind. Man möchte, daß zugängliche Funktionen, wie etwa (geeignete) stetige, dazugehören und daß (geeignete) Grenzprozesse nicht aus V herausführen und mit der Integration verträglich sind.

Ist ein Präintegral gegeben, so setzen wir

(1.1)
$$N(f) = \|f\|_1 = I(|f|)$$

und nennen diesen Wert die L^1-*Norm* von f.

(1.2) Notiz. *Die Abbildung* $N\colon V \to \mathbb{R}$ *hat die Eigenschaften einer* Seminorm *auf* V, *das heißt, es gilt:*

(1) $N(f) \geq 0$ *für alle* $f \in V$.
(2) $|\lambda| N(f) = N(\lambda f)$ *für* $\lambda \in \mathbb{R}$ *und* $f \in V$ *(Homogenität).*
(3) $N(f+g) \leq N(f) + N(g)$ *für alle* $f, g \in V$ *(Dreiecksungleichung).* □

Zur Eigenschaft einer Norm fehlt also lediglich: Ist $N(f) = 0$, so ist f die Nullfunktion. Trotzdem können wir auch mit einer Seminorm statt einer Norm Begriffe der Analysis und der Topologie in bekannter Weise formulieren. So heißt (f_n) eine N-Cauchy–Folge, wenn zu jedem $\varepsilon > 0$ ein $k \in \mathbb{N}$ existiert mit $N(f_n - f_k) < \varepsilon$ für alle $n \geq k$. Ferner ist f der N-Limes einer Folge (f_n), in Zeichen $f = N\text{-}\lim f_n$, wenn $\lim_{n\to\infty} N(f - f_n) = 0$ ist. Allerdings ist jetzt der Limes einer Folge im allgemeinen nicht mehr eindeutig bestimmt: Sei $f = N\text{-}\lim f_n$ und $g = N\text{-}\lim f_n$; aus $N(f-g) \leq N(f-f_n) + N(g-f_n)$ folgt durch Grenzübergang $N(f-g) = 0$. Sei $f = N\text{-}\lim f_n$ und $N(f-g) = 0$; aus $N(g-f_n) \leq N(g-f) + N(f-f_n) = N(f-f_n)$ folgt durch Grenzübergang $g = N\text{-}\lim f_n$. Wir nennen den Raum (V, N) *vollständig* oder N*-vollständig*, wenn in diesem Sinne jede N-Cauchy–Folge konvergiert. In diesem Kontext arbeiten wir übrigens mit verschiedenen Konvergenzbegriffen. Insbesondere haben wir neben der N-Konvergenz auch die punktweise Konvergenz von Folgen. Wir geben nun die grundlegende Definition.

(1.3) Definition. Ein *Integral* ist ein Präintegral (X, V, I), bei dem der Raum V bezüglich der L^1-Norm vollständig ist. ◇

(1.4) Notiz. *Sei* I *ein positives lineares Funktional. Es ist genau dann stetig, wenn es die folgende Eigenschaft hat: Für Funktionen* $0 \leq u$, $0 \leq u_1 \leq u_2 \leq \dots$ *aus* V *mit* $u \leq \sup u_n$ *folgt* $I(u) \leq \sup I(u_n)$.

Beweis. Sei $u'_n = \min(|u|, u_n)$ und $u''_n = |e| - e'_n$. Dann ist $u''_n \in V$, es gilt $u''_n \geq u''_{n+1} \geq 0$, und die u''_n konvergieren punktweise gegen Null. Also ist $0 = \lim I(u''_n) = I(u) - \lim I(u'_n)$. Wegen $I(u'_n) \leq I(U_n)$ folgt damit die Behauptung.

Sei umgekehrt f_n eine monoton fallend gegen Null konvergierende Folge. Dann ist $f_1 - f_n$ eine monoton steigende Folge nichtnegativer Funktionen mit $f_1 \leq \sup(f_1 - f_n)$. Es folgt $I(f_1) \geq \sup I(f_1 - f_n) \geq \lim I(f_1 - f_n) = I(f_1) - \lim I(f_n)$ und nach der vorausgesetzen Bedingung $I(f_1) \leq \sup I(f_1 - f_n)$. Also ist $\lim I(f_n) = 0$. □

Ein Ziel der Theorie wird es sein, ein Präintegral zu einem Integral zu vervollständigen. Die Norm bestimmt übrigens das Integral. Wegen $f = f^+ - f^-$ und der Linearität des Funktionals I gilt nämlich $I(f) = I(f^+) - I(f^-) = N(f^+) - N(f^-)$. Manchmal ist es leichter, die Seminorm N zunächst nur für nichtnegative Funktionen zu konstruieren. Ist dann $f \mapsto N(f^+) - N(f^-)$ linear, so ist dieses Funktional auch positiv.

Das Funktional ist auch immer eine stetige Abbildung im Sinne der normierten Vektorräume. Das folgt aus der Ungleichung:

(1.5) Aufgaben und Ergänzungen.

1. Sei $\| - \|_1$ eine Seminorm auf dem Vektorraum V. Die Menge V_0 der L^1-*Nullfunktionen*, das heißt der $f \in V$ mit $\|f\|_1 = 0$, ist ein Untervektorraum von V. Ist $p: V \to V/V_0$ die kanonische Abbildung auf den Faktorraum, so gibt es genau eine Norm auf V, die wieder mit $\| - \|_1$ bezeichnet werden, so daß immer $\|p(v)\|_1 = \|v\|_1$ gilt.

2. Sei $a \in X$. Die Evaluation $e_a: V \to \mathbb{R}$, $f \mapsto f(a)$ an der Stelle a ist ein positives stetiges lineares Funktional (*Dirac–Funktional* an der Stelle a). Es gilt $N(f) = |f(a)|$. Es handelt sich um ein Integral.

3. Seien I_1 und I_2 lineare Funktionale. Seien $a_i \geq 0$. Sind I_1 und I_2 positiv (stetig), so ist auch $a_1 I_1 + a_2 I_2$ positiv (stetig).

2 Die Obernorm

Sei (X, V, I) ein Präintegral. Wir setzen für $f: X \to \overline{\mathbb{R}}$

$$N(f) = \inf(\sup(I(e_n) \mid n \in \mathbb{N})) \in [0, \infty].$$

Das Infimum wird über alle Folgen $(e_n \in V \mid n \in \mathbb{N})$ gebildet, die aufsteigen $0 \leq e_1 \leq e_2 \leq \ldots$ und $|f| \leq \sup(e_n \mid n \in \mathbb{N})$ erfüllen; wir nennen sie *eingrenzende Folgen für f*. Das Infimum ist als ∞ zu lesen, wenn es keine solchen Folgen gibt. Wegen des nächsten Satzes bezeichnen wir N auch als Seminorm oder *Obernorm* zu (X, V, I). Wegen (2.1) ist die Bezeichnung N mit derjenigen des letzten Abschnittes verträglich.

(2.1) Satz. *Die Abbildung N hat die folgenden Eigenschaften:*

(1) $0 \leq N(f) = N(-f) = N(|f|) \leq \infty;\ N(0) = 0.$

(2) $N(\lambda f) = |\lambda| N(f)$.

(3) *Sind f, f_n Funktionen und gilt $|f| \leq \sum_n |f_n|$, so folgt $N(f) \leq \sum_n N(f_n)$.*

(4) *Für $e \in V$ gilt $N(e) = I(|e|)$.*

Beweis. (1) und (2) folgen unmittelbar aus der Definition.
(3) Falls $\sum N(f_n) = \infty$ ist, so ist nichts zu beweisen. Anderfalls ist $N(f_n) < \infty$ für alle n. Nach Definition von N gibt es eine eingrenzende Folge $(e_k^n \mid k \in \mathbb{N})$ mit $|f_n| \leq \sup(e_k^n \mid k \in \mathbb{N})$ und $\sup(I(e_k^n) \mid k \in \mathbb{N}) < N(f_n) + \varepsilon_n$. Wir wählen $\varepsilon = \varepsilon_1 + \varepsilon_2 + \cdots$. Dann ist (mit der Zusatzvereinbarung $e_0^n = 0$)

$$\sum_n N(f_n) + \varepsilon = \sum_n (N(f_n) + \varepsilon_n) \leq \sum_n (\sup(I(e_k^n) \mid k \in \mathbb{N}) \quad =$$

$$\sum_n \sum_k I(e_k^n - e_{k-1}^n) = \sup_\mu \sum_{l=1}^{\mu} \sum_{k=1}^{\mu} I(e_k^n - e_{k-1}^n) = \sup_\mu I(d_\mu)$$

wobei $d_\mu = \sum_{l=1}^{\mu} \sum_{k=1}^{\mu} I(e_k^n - e_{k-1}^n)$ gesetzt wurde. Ganz analog haben wir

$$|f| \leq \sum |f_n| \leq \sum (e_k^n \mid k \in \mathbb{N}) = \sup d_\mu$$

und $0 \leq d_1 \leq d_2 \leq \ldots$, so daß (d_n) eine eingrenzende Folge für f ist. Demnach ist $N(f) \leq \sup_\mu I(d_\mu)$.
(4) Unter Verwendung der konstanten Folge $(e_n = e)$ sehen wir $N(e) \leq I(|e|) \leq \infty$. Zu $\varepsilon > 0$ gibt es eine eingrenzende Folge (e_n) mit $|e| \leq \sup e_n$ und $\sup I(e_n) \leq N(e) + \varepsilon$. Wegen der Stetigkeit des Funktionals I ist aber $I(|e|) \leq \sup I(e_n)$. Also folgt die behauptete Gleichheit. □

3 Nullfunktionen und Nullmengen

Wir nennen $f \colon X \to \overline{\mathbb{R}}$ *Nullfunktion*, wenn $N(f) = 0$ ist, und $A \subset X$ *Nullmenge*, wenn ihre charakteristische Funktion χ_A eine Nullfunktion ist. Nullmengen sind die Mengen, die wir vernachlässigen dürfen oder müssen. Wir verwenden dazu die folgende Terminologie: Ist für jedes $x \in X$ eine Aussage $E(x)$ gegeben, so sagen wir, sie gelte *fast überall*, wenn $E(x)$ für alle x außerhalb einer Nullmenge richtig ist.

Aus den Eigenschaften von N folgt sofort:

(3.1) Notiz. *Eine abzählbare Vereinigung von Nullmengen ist eine Nullmenge. Eine Teilmenge einer Nullmenge ist eine Nullmenge.* □

(3.2) Notiz. *Folgende Aussagen über eine Funktion $f \colon X \to \overline{\mathbb{R}}$ sind äquivalent:*
(1) *f ist eine Nullfunktion.*
(2) *$|f|$ ist eine Nullfunktion.*
(3) *Außerhalb einer Nullmenge ist f gleich Null.*

Beweis. (1) und (2) sind wegen $N(f) = N(|f|)$ äquivalent.
(2) $\Rightarrow$ (3). Sei $N(f) = 0$ und c_k die charakteristische Funktion von $\{|f| \geq k^{-1}\}$. Dann ist $0 \leq c_k \leq k|f|$ und folglich $N(c_k) \leq kN(f) = 0$. Ferner ist $c_k \leq c_{k+1}$ und $c_0 = \chi_{\{|f|>0\}} = \sup(c_k \mid k \in \mathbb{N}) \leq \sum_{k\geq 1} c_k$, und deshalb gilt $0 \leq N(c_0) \leq \sum N(c_k) = 0$.

(3) $\Rightarrow$ (2). Aus $N(c_0) = 0$ folgt $N(c_k) = 0$ und wegen $|f| \le \sum c_k$ gilt $N(f) \le \sum N(c_k) = 0$. □

Wir nennen f und g *nulläquivalent* oder *N-äquivalent*, in Zeichen $f \sim_N g$, wenn fast überall $f(x) = g(x)$ ist. Das ist wegen (3.1) eine Äquivalenzrelation auf der Menge aller Funktionen $X \to \overline{\mathbb{R}}$.

(3.3) Notiz.
(1) *$f \sim_N g \Rightarrow N(f) = N(g)$.*
(2) *$|f| \le_N |g| \Rightarrow N(f) \le N(g)$.*
(3) *$f \sim_N g$, $f, g \in V \Rightarrow I(f) = I(g)$.*
(4) *Sei $N(f) < \infty$. Dann ist $f(x)$ fast überall endlich.*

Beweis. (1) Es genügt, nichtnegative $f$und g zu betrachten. Wir definieren $n\colon X \to \overline{\mathbb{R}}$ durch $n(x) = 0$, falls $f(x) = g(x)$ und $n(x) = \infty$ sonst. Dann ist $f \le g+n$ und n ist eine Nullfunktion. Es folgt $N(f) \le N(g+n) \le N(g)+N(n) = N(g)$. Analog $N(g) \le N(f)$.

(2) Es gibt eine zu f N-äquivalente Funktion f' mit $|f'| \le g$ überall. Mit (1) folgt also $N(f) = N(f') \le N(g)$.

(3) Es ist $0 \le |I(f) - I(g)| \le I(|f-g|) = N(f-g)$. Da $f-g$ eine Nullfunktin ist, folgt $N(f-g) = 0$ und damit $I(f) = I(g)$.

(4) Sei $d_k = \chi_{\{|f|>k\}}$, $k \in \mathbb{N}$. Dann ist $d_\infty = \chi_{\{|f|=\infty\}} \le d_k$, also $N(d_\infty) \le N(d_k)$. Ferner ist $kd_k \le |f|$, also $N(d_\infty) \le N(d_k) \le k^{-1}N(f)$. Das gilt für jedes k, also ist $N(d_\infty) = 0$. □

4 Vollständigkeit

Sei V^* die Menge aller Funktionen $f\colon X \to \mathbb{R}$ mit endlicher Obernorm $N(f)$.

(4.1) Satz. *Der Raum (V^*, N) ist vollständig. Jede N-Cauchy–Folge besitzt eine fast überall konvergente Teilfolge, die gegen den N-Limes der Cauchy–Folge konvergiert.*

Beweis. Sei (f_n) eine N-Cauchy–Folge in V^*. Zu $\varepsilon > 0$ wählen wir $Q(\varepsilon) \in \mathbb{N}$, so daß $N(f_n - f_k) < \varepsilon$ für $n, k \ge Q(\varepsilon)$. Sei $1 = \varepsilon_1 + \varepsilon_2 + \cdots$ mit $\varepsilon_i > 0$ gewählt. Sei $q(n) = Q(\varepsilon_n)$ und $g_n = f_{q(n)}$. Dann ist (g_n) eine Teilfolge von (f_n), und es gilt $N(g_{k+1} - g_k)\varepsilon_k$ für $k \in \mathbb{N}$. Wir setzen

$$g = |g_1| + \sum_{k=1}^{\infty} |g_{k+1} - g_k|.$$

Dann gilt

$$N(g) \le N(g_1) + \sum N(g_{k+1} - g_k) < 1 + N(g_1) < \infty.$$

Nach (3.3) ist g fast überall endlich. Die Reihe für g konvergiert daher fast überall zu einem Wert in $\mathbb{R}$ und deshalb nach dem Majorantenkriterium auch die Reihe $g_1 + \sum_{k=1}^{\infty}(g_{k+1} - g_k)$. Sei h dieser fast überall gebildete Limes. Wegen $|h - g_k| \le |g_{k+1} - g_k| + |g_{k+2} - g_{k+1} + \cdots$ folgt $N(h - g_k) \le \varepsilon_k + \varepsilon_{k+1} + \cdots < \varepsilon$ für schließlich alle k. Also ist $h = N\text{-}\lim g_n = N\text{-}\lim f_n$. Damit ist die Vollständigkeit gezeigt. Da N-Limites eindeutig bestimmt sind, bis auf eine Nullfunktion, so folgt auch die zweite Aussage. □

5 Das Integral zu einem Präintegral

Sei W die abgeschlossene Hülle von V im Raum (V^*, N). Die Elemente von W sind also N-Limites von Folgen in V. Aus dem Rechnen mit N-Limites folgt, daß W ein Integrationsraum ist. Als abgeschlossene Hülle in einem vollständigen Raum ist (W, N) wieder vollständig. Wir setzen

$$J\colon W \to \mathbb{R}, \quad f \mapsto N(f^+) - N(f^-).$$

Dann gilt

(5.1) $$|J(f) - J(g)| \le 2N(f - g).$$

Zum Beweis verwenden wir

$$|J(f) - J(g)| = |(N(f^+) - N(g^+)) - (N(f^-) - N(g^-))| \le N(f^+ - g^+) + N(f^- - g^-)$$

und

$$|f^+ - g^+| = \frac{1}{2}\Big||f| + f - |g| - g\Big| \le \frac{1}{2}\Big(\Big||f| - |g|\Big| + |f - g|\Big) \le |f - g|$$

und entsprechend für f^-, g^-. Insgesamt errechnet man daraus (5.1). Als Konsequenz aus dieser Formel erhalten wir:

(5.2) Notiz. *$J\colon W \to \mathbb{R}$ ist eine stetige Abbildung, wenn wir W mit der N-Norm versehen.* □

(5.3) Monotone Konvergenz. (Beppo Levi) *Sei $(f_n \in W \mid n \in \mathbb{N})$ und gelte fast überall $f_n \le f_{n+1}$. Die Folge $(J(f_n))$ sei beschränkt. Dann konvergiert (f_n) fast überall gegen f, und es gilt $J(f) = \lim J(f_n)$. Außerdem ist $f = N\text{-}\lim f_n$.*

Beweis. Es gilt für $k \ge n$

$$N(f_k - f_n) = J(f_k - f_n) = J(f_k) - J(f_n) < \varepsilon$$

für genügend große n, weil die Folge $(J(f_n))$ monoton und beschränkt, also eine gewöhnliche Cauchy-Folge ist. Mithin ist (f_n) eine N-Cauchy–Folge. Es gibt also eine fast überall konvergente Teilfolge; wegen der Monotonie ist dann auch die gesamte Folge fast überall konvergent. Nun wende man (5.??) an. □

(5.4) Dominierte Konvergenz. (Lebesgue) *Die Folge $(f_n \in W)$ konvergiere fast überall gegen $f\colon X \to \mathbb{R}$. Es sei (f_n) durch eine integrable Funktion g dominiert, das heißt es gelte $|f_n(x)| \le g(x)$ für alle n und für fast alle x. Dann ist $f \in W$, es ist $f = N\text{-}\lim f_n$ und $\lim J(f_n) = J(f)$.*

Beweis. Es genügt wiederum zu zeigen, daß (f_n) eine N-Cauchy–Folge ist. Ohne wesentliche Einschränkung gilt $|f_n| \leq g$ überall und Konvergenz überall. Für $k \geq 1$ ist

$$h_k(x) = \sup\{|f_n(x) - f_m(x)| \;\Big|\; n, m \geq k\} \leq 2g(x).$$

Wir zeigen, daß h_k in W liegt. Da $|f_n - f_m|$ in W liegt, so auch $v_l = \max\{|f_n - f_m| \;\Big|\; k \leq n, m \leq l\}$. Die Folge (v_l) steigt monoton und wird von $2g$ dominiert. Nach dem Satz über monotone Konvergenz ist das Supremum integrabel, liegt also in W. Die Funktionen h_k bilden eine monoton fallende Folge, die nach der vorausgesetzten Konvergenz von $(f_n(x))$ punktweise gegen Null konvergiert. Also ist $\lim J(h_k) = 0$. Wegen $N(f_n - f_k) \leq N(h_k) = J(h_k)$ sehen wir, daß (f_n) eine N-Cauchy-Folge ist. □

6 Maße

Eine *σ-Algebra* auf einer nichtleeren Menge X ist ein System $\mathfrak{M}$ von Teilmengen, das die leere Menge enthält, mit jeder Menge A ihr Komplement $X \setminus A$ und mit jeder Familie $(A_i \in \mathfrak{M} \mid i \in \mathbb{N})$ die Vereinigung $\bigcup_i A_i$. Durch Komplementbildung sieht man, daß dann $X \in \mathfrak{M}$ ist und mit jeder Familie $(A_i \in \mathfrak{M} \mid i \in \mathbb{N})$ auch der Schnitt $\bigcap_i A_i$ zu $\mathfrak{M}$ gehört. Ferner liegen mit A und B auch $A \cup B$, $A \cap B$ und $A \setminus B$ in $\mathfrak{M}$. Ein Paar $(X, \mathfrak{M})$ aus einer nichtleeren Menge X und einer σ-Algebra $\mathfrak{M}$ daraus heißt *Meßraum*. Die Elemente von $\mathfrak{M}$ werden die *meßbaren Mengen* des Meßraumes genannt.

Die Menge aller Teilmengen ist eine σ-Algebra. Der Schnitt einer beliebigen Familie von σ-Algebren ist wieder eine. Zu jeder Menge $\mathfrak{S}$ von Teilmengen von X gibt es deshalb den Schnitt $\mathfrak{S}^\#$ aller σ-Algebren, die $\mathfrak{S}$ umfassen, genannt die von $\mathfrak{S}$ *erzeugte σ-Algebra.*

Ist X eine topologischer Raum, so heißt die von allen offenen Mengen erzeugte σ-Algebra $\mathcal{B}(X)$ die *Borel–Algebra* des Raumes X. Für die elementare Integrationstheorie sind insbesondere die Borel-Algebren der euklidischen Räume relevant.

Ein *Maß* μ auf einer σ-Algebra $\mathfrak{M}$ ist eine Abbildung $\mu\colon \mathfrak{M} \to [0, \infty]$, die der leeren Menge den Wert Null zuordnet und die *σ-additiv* ist. Letzteres soll heißen: Ist $(A_i \mid i \in \mathbb{N})$ eine Familie paarweise disjunkter Mengen aus $\mathfrak{M}$, so gilt $\sum_i \mu(A_i) = \mu(\bigcup_i A_i)$. Da ein Maß auch den Wert ∞ annehmen kann, so treffen wir die Vereinbarungen $\infty + a = a + \infty = \infty$ für $a \in [0, \infty]$, $\infty \cdot a = a \cdot \infty = \infty$ für $a > 0$ und $0 \cdot \infty = \infty \cdot 0 = 0$. Eine Reihe $\sum_i a_i$ mit Werten $a_i \in [0, \infty]$ hat dann immer eine Summe $\sum_i a_i \in [0, \infty]$. Additition und Multiplikation in $[0, \infty]$ sind weiterhin assoziativ und kommutativ. Wir notieren einige unmittelbare Konsequenzen aus den Maßaxiomen.

(6.1) Notiz. *Gegeben sei ein Maßraum $(X, \mathfrak{M}, \mu)$. Dann gilt:*

(1) *Sind $A, B \in \mathfrak{M}$, so ist $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.*

(2) *Sind $A \subset B$ aus $\mathfrak{M}$, so gilt $\mu(A) \leq \mu(B)$.*

(3) *Ist $(A_n \in \mathfrak{M} \mid n \in \mathbb{N})$, so gilt $\mu(\bigcup_j A_j) \leq \sum_j \mu(A_j)$.*

(4) *Ist* $A_1 \subset A_2 \subset \ldots$ *eine aufsteigende Folge in* $\mathfrak{M}$, *so gilt* $\mu(\bigcup_j A_j) = \lim_{n\to\infty} \mu(A_n)$.

Beweis. □

Sei $\mathfrak{X} = (X, V, I)$ ein Integral. Wir nennen $A \subset X$ *integrierbar*, wenn $\chi_A \in V$ ist. Wir setzen dann $\mu(A) = I(\chi_A)$. Wir sagen, X ist σ-*endlich* bezüglich $\mathfrak{X}$, wenn es eine aufsteigende Folge $S_1 \subset S_2 \subset \cdots$ integrierbarer Mengen gibt, deren Vereinigung X ist. Wir nennen dann $A \subset X$ *meßbar* bezüglich $\mathfrak{X}$, wenn alle $A \cap S_i$ integrierbar sind. In dem Fall setzen wir $\mu(A) = \sup I(A \cap E_i)$. Wegen des Satzes über die monotone Konvergenz liefert das für integrierbare A nichts neues. Sei $\mathfrak{M}(\mathfrak{X})$ das System der bezüglich $\mathfrak{X}$ meßbaren Mengen. Aus Rechenregeln des Integrals und dem Satz über monotone Konvergenz folgt dann:

(6.2) Satz. $(X, \mathfrak{M}(\mathfrak{X}), \mu_{\mathfrak{X}})$ *ist ein Maßraum.* □

7 Integration stetiger Funktionen

Eine stetige Funktion $f\colon [a,b] \to \mathbb{R}$ hat das Integral $\int_a^b f(x)\,dx$ der elementaren Analysis. Wir befreien uns von den Definitionsbereichen. Wir sagen, eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ habe *kompakten Träger*, wenn die abgeschlossene Hülle der Menge $\{x \mid f(x) \neq 0\}$, der sogenannte *Träger* von f, kompakt ist. Es gibt in einem solchen Fall also ein Intervall $[a,b]$, das den Träger enthält. Wir setzten $I(f) = \int_{\mathbb{R}} f(x)\,dx = \int_a^b f(x)\,dx$, und diese Festlegung hängt nicht von der Wahl von $[a,b]$ ab. Sei $C_c(\mathbb{R})$ der Vektorraum der Funktionen $\mathbb{R} \to \mathbb{R}$ mit kompaktem Träger. Das Integral I ist dann eine positive lineare Abbildung $\int\colon C_c(\mathbb{R}) \to \mathbb{R}$. Konvergiert (f_n) gleichmäßig gegen f, so gilt $\int f = \lim_n \int f_n$. Manchmal folgt aus der punktweisen Konvergenz die gleichmäßige; es gilt nämlich der *Satz von Dini*:

(7.1) Satz. *Sei* X *ein kompakter Raum und* $(f_n\colon X \to \mathbb{R})$ *eine monoton wachsende Folge stetiger Funktionen, die punktweise gegen die stetige Funktion* f *konvergiert. Dann ist die Konvergenz gleichmäßig.* □

Mit Hilfe des Satzes von Dini sehen wir also, daß $(\mathbb{R}, C_c(\mathbb{R}), I)$ ein Präintegral ist. Das zugehörige Integral $(\mathbb{R}, L, J)$ nennen wir das *Lebesguesche Integral.*

8 Integration Riemannscher Treppenfunktionen

Das Riemannsche Integral (oder ein anderes elementares Integral) wird über Treppenfunktionen definiert. Eine Funktion $t\colon \mathbb{R} \to \mathbb{R}$ heiße *Riemannsche Treppenfunktion*, wenn es eine *Zerlegung* $a = x_0 \leq x_1 \leq \ldots \leq x_k = b$ so gibt, daß f außerhalb $[a,b]$ gleich Null ist und auf den offenen Intervallen $]x_{i-1}, x_i[$ konstant gleich c_i. Über die Funktionswerte in den Punkten x_i legen wir nichts fest. Wir definieren als das Integral $\int_{\mathbb{R}} t(x)\,dx$ der Treppenfunktion t den Wert $\sum_{i=1}^k c_i(x_i - x_{i-1})$. Dieser Wert ist unabhängig von der Wahl der Zerlegung zu t. Diese Treppenfunktionen bilden einen Vektorraum TR, und das Integral ist eine lineare Abbildung $\int_{\mathbb{R}}: TR \to \mathbb{R}$. So einfach diese Begriffsbildungen sind, es gibt einen wichtigen Satz darüber, eine Stetigkeitseigenschaft des Integrals.

(8.1) Satz. *Sei* $t_1 \geq t_2 \geq \dots$ *eine Folge von Treppenfunktionen, die punktweise gegen die Nullfunktion konvergiert. Dann ist* $\lim_n \int_{\mathbb{R}} t_n(x)\,dx = 0$.

Beweis. Die Endpunkte der Intervalle, auf denen t_n konstant ist, bilden eine endliche Menge. Die Menge aller dieser Endpunkte für alle t_n ist eine abzählbare Menge; sie wird deshalb durch eine Folge offene Intervalle J_k überdeckt, deren Längensumme $\sum_{k=1}^{\infty} l(J_k) < \varepsilon$ ist. Sei $U = \bigcup_k J_k$. Wir wählen ein kompaktes Intervall I, außerhalb dessen die Funktionen t_n alle gleich Null sind. Ist $x \in I \setminus U$, so gibt es ein n_x und ein offenes Intervall V_x um x, so daß $t_{n_x}(s) < \varepsilon$ ist für $s \in V_x$, weil die t_n punktweise gegen Null konvergieren und in $I \setminus U$ lokal konstant sind. Wegen der Monotonie der Folge gilt dann $t_m(s) < \varepsilon$ für alle $m \geq n_x$ und $s \in V_x$. Die Familie der offenen Mengen $(J_k, V_x \mid k \in \mathbb{N}, x \in I \setminus U$ überdeckt I. Wir wählen eine endliche Teilüberdeckung

$$J_{k_1}, \dots, J_{k_p}, V_{x_1}, \dots, V_{x_q}.$$

Sei $N = \max(n_{x_1}, \dots, n_{x_q})$. Ist $n \geq N$, so gilt

$$f_n(s) < \varepsilon \qquad \text{für} \qquad s \in W = V_{x_1} \cup \dots \cup V_{x_q}.$$

Das Integral von t_n über W ist also höchstens $l(I)\varepsilon$, und das Integral über die Vereinigung der ausgewählten J_k höchstens gleich dem Maximum C von f_1 mal der Gesamtlänge der J_k, also höchstens gleich $C\varepsilon$. Insgesamt ist also $\int f(x)\,dx \leq C\varepsilon + l(I)\varepsilon$. □

(8.2) Aufgaben und Ergänzungen.

1. Eine stetige Funktion $f: [a, b] \to \mathbb{R}$ ist gleichmäßiger Limes von Treppenfunktionen.

9 Präintegral für Treppenfunktionen

Wir betrachten Treppenfunktionen axiomatisch. Zunächst legen wir ihre Definitionsbereiche fest.

Ein System von Teilmengen $\mathfrak{A}$ einer nichtleeren Menge X heißt *Algebra*, wenn $\emptyset \in \mathfrak{A}$ ist und mit A und B auch $A \cup B$, $A \cap B$ und $A \setminus B$ Elemente aus $\mathfrak{A}$ sind. Eine Mengenfunktion $\mu: \mathfrak{A} \to [0, \infty]$ heißt *Inhalt*, wenn $\mu(\emptyset) = 0$ ist und ferner μ *additiv* ist. Letzteres heißt: Sind A und B disjunkte Mengen aus $\mathfrak{A}$, so ist $\mu(A \cup B) = \mu(A) + \mu(B)$. Ein Inhalt wird *Prämaß* genannt, wenn er außerdem σ-*additiv* ist, das heißt, wenn für eine Familie $(A_j \mid j \in \mathbb{N})$ paarweise disjunkter Mengen $A_j \in \mathfrak{A}$ mit einer in $\mathfrak{A}$ liegenden Vereinigung die Gleichheit $\mu(\bigcup_j A_j) = \sum_j \mu(A_j)$ gilt. Ein *Prämaßraum* ist ein Tripel $(X, \mathfrak{A}, \mu)$ aus einer Algebra $\mathfrak{A}$ auf X zusammen mit einem Prämaß. Ein Prämaßraum heißt σ-*endlich*, wenn es eine Folge (S_n) in $\mathfrak{A}$ gibt mit $\mu(S_n) < \infty$ und $X = \bigcup_n S_n$.

Sei nun ein Inhalt μ auf einer Algebra $(X, \mathfrak{A})$ vorgegeben. Eine Funktion $\varphi: X \to \mathbb{R}$ heißt *Treppenfunktion*, wenn sie nur endlich Werte annimmt und wenn die *Stufe* $\varphi^{-1}(c)$ für alle $c \neq 0$ ein endliches Maß hat. Eine *Zerlegung* zur Treppenfunktion φ ist eine disjunkte Zerlegung $X = A_1 \cup A_2 \cup \ldots \cup A_r$ von X in meßbare Mengen A_j, so daß φ auf A_j konstant gleich c_j ist (sofern $A_j \neq \emptyset$). Als Integral der Treppenfunktion definieren wir

(9.1)
$$\int_X \varphi \, d\mu = \sum_{j=1}^{r} c_r \mu(A_j).$$

Wegen der Vereinbarung $0 \cdot \infty = 0$ sind die A_j von unendlichem Maß schadlos; ist $A_j = \emptyset$, so vereinbaren wir „undefiniert mal Null = Null“. Durch Übergang zu einer gemeinsamen Verfeinerung bestätigt man mittels der Additivität des Maßes, daß der Wert (9.1) nicht von der Zerlegung zu φ abhängt. Sei $T(\mu)$ die Menge der Treppenfunktionen. Als kleine Aufgabe stellen wir:

(9.2) Notiz. *$T(\mu)$ ist ein Vektorraum. Er enthält mit φ auch $|\varphi|$. Das Integral $\varphi \mapsto I(\varphi) = \int_X \varphi \, d\mu$ ist eine lineare Abbildung $I \colon T(\mu) \to \mathbb{R}$.* □

(9.3) Notiz. *Genau dann ist $(X, T(\mu), \int_X)$ ein Präintegral, wenn μ σ-additiv ist.*

Beweis. Sei μ σ-additiv. Seien v und u_n nicht negative Treppenfunktionen, es gelte $v \leq \sup u_n$ und $u_n \leq u_{n+1}$. Wir haben zu zeigen $I(v) \leq \sup I(u_n)$. Sei dazu $v = \sum_{j=1}^{m} \alpha_j \chi(A_j)$ mit $A_j \in \mathfrak{A}$ und $\alpha_j \geq 0$. Wir fixieren eine positive Zahl $\alpha < 1$. Die Mengen $B_n = \{u_n \geq \alpha v\}$ sind in $\mathfrak{A}$ enthalten, und wegen $u_n \geq \alpha v \chi(B_n)$ gilt

$$I(u_n) \geq \alpha I(v \chi(B_n).$$

Aus den Voraussetzungen über v und u_n folgt $B_n \subset B_{n+1}$ und $\bigcup_n B_n = X$, sowie $\bigcup_n A_j \cap B_n = A_j$. Damit folgt

$$I(v) = \sum_{i=1}^{m} \alpha_j \mu(A_j) = \lim_n \sum_{j=1}^{m} \alpha_j \mu(A_j \cap B_n = \lim_n I(v \chi(B_n).$$

Aus den abgesetzten Ungleichungen erhalten wir

$$\sup I(u_n) \geq \alpha \sup I(v \chi(B_n)) = \alpha I(v).$$

Da $\alpha < 1$ beliebig war, folgt die Behauptung.

Sei I stetig. Die σ-Additivität ist ein Spezialfall der Stetigkeit, angewendet auf charakteristische Funktionen. □

(9.4) Beispiel. Das System aller endlichen Vereinigungen von endlichen Intervallen der Form $[a, b[$ ist eine Mengenalgebra auf $\mathbb{R}$. Ebenso das System aller endlichen Vereinigungen von endlichen Intervallen (offen, abgeschlossen, halboffen). Setzen wir mit dem Integral für Riemannsche Treppenfunktionen $\mu(A) = \int \xi_A(x) \, dx$, so wird dadurch ein Prämaß auf diesen Algebren definiert. Das damit auf der zweiten Algebra definierte Integral ist das Integral aus dem zweiten Abschnitt. ◇

10 Äußeres Maß

Sei ein Prämaßraum $(X, \mathfrak{A}, \mu)$ vorgegeben. Wir benutzen ihn dazu, um beliebige Teilmengen von X durch Approximation von oben dem Maße nach abzuschätzen. Für eine beliebige Teilmenge $Y \subset X$ definieren wir

$$\mu^*(Y) = \inf \sum_{i=1}^{\infty} \mu(A_i),$$

wobei das Infimum über alle Folgen (A_n) in $\mathfrak{A}$ mit $Y \subset \cup_{n=1}^{\infty} A_n$ gebildet wird. Das Infimum ist als ∞ zu lesen, wenn es keine solchen Folgen gibt. Im σ-endlichen Fall gibt es immer solche Folgen. Damit erhalten wir $\mu^* \colon \mathfrak{P}(X) \to [0, \infty]$ auf der Menge $\mathfrak{P}(X)$ aller Teilmengen von X.

(10.1) Satz. $\mu^* \colon \mathfrak{P}(X) \to [0, \infty]$ *hat die folgenden Eigenschaften:*
(1) $\mu^*(\emptyset) = 0$.
(2) *Ist* $A \subset B$, *so gilt* $\mu^*(A) \leq \mu^*(B)$.
(3) *Für jede Folge* (A_n) *gilt* $\mu^*(\cup_j A_j) \leq \sum \mu^*(A_n)$.
(4) *Für* $A \in \mathfrak{A}$ *gilt* $\mu(A) = \mu^*(A)$.

Beweis. □

Wir nennen μ^* das *äußere Maß* oder das *Obermaß* zum gegebenen Prämaßraum. Mit dem äußeren Maß definieren wir: $N \subset X$ heißt *Nullmenge*, wenn $\mu^*(N) = 0$ ist. Aus (10.5) folgt, daß eine abzählbare Vereinigung von Nullmengen wieder eine Nullmenge ist. Diese Tatsache verwenden wir oft stillschweigend. Nullmengen spielen in der Integrationstheorie eine große Rolle.

11 Das Integral zu einem Prämaß

Einen Raum mit einer Seminorm kann man immer durch einen formalen Prozess vervollständigen. Die Elemente des vervollständigten Raumes werden zum Beispiel als Äquivalenzklassen von Cauchy–Folgen definiert. Für die Zwecke der Analysis möchte man aber doch statt dieser geisterhaften Elemente lieber richtige Funktionen haben. Soll die hypothetische Grenzfunktion einer L^1-Cauchy–Folge durch einen Grenzprozess aus den schon vorhandenen Funktionen gewonnen werden, so braucht man punktweise Konvergenz. Wir definieren deshalb:

Eine Funktion $f \colon X \to \mathbb{R}$ heiße *integrabel*, wenn es eine L^1-Cauchy–Folge $(\varphi_n \mid n \in \mathbb{N})$ von Treppenfunktionen gibt, die fast überall punktweise gegen f konvergiert. In diesem Fall ist wegen

$$\left| \int \varphi_m - \int \varphi_n \right| \leq \int |\varphi_m - \varphi_n| = \|\varphi_m - \varphi_n\|_1$$

die Folge der reellen Zahlen $(\int \varphi_n \mid n \in \mathbb{N})$ eine gewöhnliche Cauchy–Folge, also konvergent, und deshalb ist es möglich,

$$\int_X f \, d\mu = \lim_{n \to \infty} \int_X \varphi_n \, d\mu \qquad \textbf{(11.1)}$$

zu setzen. Damit wir auf diese Weise zu einem wohldefinierten Integralwert kommen, müssen wir zeigen, daß der letzte Grenzwert nicht von der Auswahl der approximierenden Folge (φ_n) abhängt. Das geschieht durch das nächste Lemma.

(11.2) Lemma. *Seien (φ_n) und (ψ_n) zwei L^1-Cauchy–Folgen, die beide fast überall gegen dieselbe Funktion konvergieren. Dann ist $\lim_{n\to\infty} \|\varphi_n - \psi_n\|_1 = 0$ und wegen $|\int \varphi_n - \int \psi_n| \le \|\varphi_n - \psi_n\|_1$ gilt $\lim_n \int \varphi_n = \lim_n \int \psi_n$.*

Zum Beweis von (11.??) brauchen wir ein weiteres Lemma. Es klärt den Zusammenhang zwischen L^1-Konvergenz und punktweiser Konvergenz und macht die Definition der integrablen Funktion im Hinblick auf die angestrebte L^1-Vollständigkeit plausibler.

(11.3) Lemma. *Sei (φ_n) eine L^1-Cauchy–Folge von Treppenfunktionen. Dann gibt es eine Teilfolge (ψ_n) mit den Eigenschaften:*

(1) *(ψ_n) konvergiert fast überall.*

(2) *Zu jedem $\varepsilon > 0$ gibt es eine Menge Z vom äußeren Maß $\mu^*(Z) < \varepsilon$, so daß (ψ_n) auf $X \setminus Z$ gleichmäßig konvergiert.*

Wir nehmen (11.2) und (11.3) einmal an und zeigen, daß wir mit unseren Definitionen das Ziel erreicht haben.

(11.4) Satz. *Die Menge $\mathcal{L}^1(\mu)$ der integrablen Funktionen ist ein Vektorraum, und $f \mapsto \int_X f\,d\mu$ ist eine lineare Abbildung. Mit f ist auch $|f|$ integrabel. Das Tripel $(X, \mathcal{L}^1(\mu), \int_X)$ ist ein Integral.*

Beweis. Konvergiert die L^1-Cauchy–Folge (φ_n) f.ü. gegen f und die L^1-Cauchy–Folge (ψ_n) f.ü. gegen g, so ist $(\varphi_n + \psi_n)$ wieder eine L^1-Cauchy–Folge, sie konvergiert f.ü. gegen $f+g$, und es gilt $\int(f+g) = \lim \int(\varphi_n+\psi_n) = \lim \int \varphi_n + \lim \int \psi_n = \int f + \int g$. Ebenso für die Skalarmultiplikation. Mit (φ_n) ist auch $(|\varphi_n|)$ eine L^1-Cauchy–Folge, denn

$$\Big\| |\varphi_m| - |\varphi_n| \Big\|_1 = \int \Big| |\varphi_m| - |\varphi_n| \Big| \le \int |\varphi_m - \varphi_n| = \|\varphi_m - \varphi_n\|_1.$$

Konvergiert (φ_n) f.ü. gegen f, so konvergiert $(|\varphi_n|)$ f.ü. gegen $|f|$. Wegen $\pm \int \varphi_n \le \int |\varphi_n|$ folgt durch Grenzübergang $|\int f| \le \int |f|$. Ist $0 \le f$, so ist deshalb wegen $f = |f|$ auch $\int f \ge 0$, und aus $f \le g$ folgt damit durch Betrachtung von $g - f$ die Ungleichung $\int f \le \int g$. Es bleibt zu zeigen, daß die lineare Abbildung $\int_X$ stetig ist und der Raum L^1-vollständig. □

12 Beweis der Hilfssätze

Beweis. Sei (ψ_n) die gegebene L^1-Cauchy–Folge. Wir wählen eine Teilfolge (φ_n) rekursiv, so daß $\| - \varphi_n - \varphi_k\|_1 \le 2^{-2k}$ für $n \ge k, k \in \mathbb{N}$; wegen der Cauchy–Bedingung ist das möglich. Die Menge

$$Y(k) = \{|\varphi_{k+1} - \varphi_k| \ge 2^{-k}\}$$

liegt in $\mathfrak{A}$. Es also ist $2^{-k}\chi_{Y(k)}$ eine Treppenfunktion, die kleinergleich $|\varphi_{k+1}-\varphi_k|$ ist. Aus der Monotonie des Integrals folgt

$$2^{-k}\mu(Y(k)) = \int 2^{-k}\chi_{Y(k)} \leq \int |\varphi_{k+1}-\varphi_k| = \|\varphi_{k+1}-\varphi_k\|_1 \leq 2^{-2k},$$

also $\mu(Y(k)) \leq 2^{-k}$. Die Menge $Z_k = \bigcup_{n\geq k} Y(n)$ hat also das äußere Maß $\mu^*(Z_k) \leq \sum_{n\geq k}\mu(Y(n)) = 2\cdot 2^{-k}$. Für $x \notin Z_k$ ist $|\varphi_{n+1}-\varphi_n| \leq 2^{-n}$ für alle $n \geq k$, nach Definition von $Y(n)$). Nach dem Majorantenkriterium konvergiert die Reihe $\sum_n(\varphi_{n+1}-\varphi_n)$ auf $X \setminus Z_k$ gleichmaßig, und $\mu^+(Z_k) \leq 2\cdot 2^{-k}$ wird für große k beliebig klein. Insbesondere konvergiert die Reihe für $x \not{i}n \bigcup_{k\leq 1} Z_k = Z$, und $\mu^*(Z) = 0$. □

Beweis. Mit (φ_n) und (ψ_n) ist auch $(\tau_n = \varphi_n - \psi_n)$ eine L^1-Cauchy–Folge. Sie konvergiert fast überall gegenNull. Wir müssen $\lim\|\tau_n\|_1 = 0$ ziegen. Es genügt, dies für eine Teilfolge zu zeigen. Diese Teilfolge sei gemäß (??) gewählt und werde wieder mit (τ_n) bezeichnet. Sei $\varepsilon > 0$ gegeben. Wir wählen k so größ, daß $\|\tau_n - \tau_k\|_1 < \varepsilon$ für alle $n \geq k$; ferner wählen wir zu diesem festen k eine Menge Z, so daß (τ_n) auf $X \setminus Z$ gleichmäßig konvergiert und $\|\tau\|\mu^*(Z) < \varepsilon$ ist. Hier ist $\|\tau_k\|$ die Sup-Norm, also das Maximum von $|\tau_k(x)|$; das ist möglich, da τ_k als Treppenfunktion nur endlich viele Werte annimmt. Nach der Definition der Treppenfunktion liegt $M = \{x \in X \mid \tau_k(x) \neq 0\}$ in $\mathfrak{A}$ und hat endliches Maß. Wir schätzen $\|\tau\|_1 = \int_X |\tau_n$ in drei Teilen ab: Durch die Integrale über Z , $M \setminus Z$ und $X \setminus M$. □

Da es leider nicht offensichtlich ist, wie man ein Maß (12.2) konstruiert, betrachtet man auch einige Vorstufen zu den Maßräumen, bei denen nur ein abgeschwächtes Axiomensystem postuliert wird.

Was bei einem Prämaßraum zu einem Maßraum also noch fehlt, ist die Meßbarkeit beliebiger abzählbarer Vereinigungen — und es ist diese letzte Eigenschaft, die Maße nützlich, aber auch nichttrivial und kompliziert macht.

Eine bedeutende Rolle spielen im weiteren die Nullmengen, Mengen, die man vernachlässigen kann und die man vernachlässigen muß. Schon in den elementarsten Überlegungen zur Inhaltsmessung sind sie unvermeidlich, werden dort allerdings meist unter den Teppich gekehrt[1].

Sei $(X, \mathfrak{M}, \mu)$ ein Maßraum. Wir nennen $A \subset X$ eine *Nullmenge* des Maßraums, wenn es $N \in \mathfrak{M}$ gibt mit $A \subset N$ und $\mu(N) = 0$. Es wird nicht $A \in \mathfrak{M}$ verlangt. Ohne Schaden, in der Tat mit Vorteil, kann man $\mathfrak{M}$ um die zugehörigen Nullmengen zu einer σ-Algebra ergänzen und μ entsprechend erweitern. Wir wollen einen Maßraum *nullvollständig* nennen, wenn seine σ-Algebra schon alle Nullmengen enthält. Auch ohne Maß- und Integrationstheorie sind die Nullmengen nützlich.

[1] Zum Beispiel in der „Schulmathematik“ — bekanntlich eine Subkultur, mit den üblichen Erscheinungen einer solchen: Abkehr von den wichtigen Erfordernissen, ...

Ist für jedes $x \in X$ eine Aussage $E(x)$ gegeben, so sagen wir, $E(x)$ gilt *fast überall* (= f.ü.), wenn $E(x)$ für alle x außerhalb einer Nullmenge (12.2) richtig ist.

(12.1) Aufgaben und Ergänzungen.

1. Mit der Definition (12.2.2) zeige man, daß eine abzählbare Vereinigung von Nullmengen wieder eine ist. (Aus den Maßaxiomen folgt das ja sowieso.)

2. Man grübele, ob man mittels (12.2.2) beweisen kann, daß $[a, b]$ für $a < b$ keine Nullmenge ist?

3. *Das Cantorsche Diskontinuum.* Wir konstruieren eine überabzählbare Nullmenge.

4. Jedes Maß hat eine eindeutig bestimmte nullvollständige Erweiterung. Die Mengen in der σ-Algebra $\mathcal{L}(\mathbb{R}^n)$ zur nullvollständigen Erweiterung von (12.1) heißen *Lebesgue–meßbar* und das zugehörige Maß darauf heißt *Lebesgue–Maß.*

13 Maßerweiterung

Das Lebesguesche Maß auf $\mathbb{R}$ wird konstruiert, indem das elementare Prämaß auf der Algbra der Intervalle (siehe ??) durch einen Erweiterungs- und Limesprozess auf die Borel-Algebra ausgedehnt wird. Diese Ausdehnung ist im axiomatischen Kontext möglich und wird durch den folgenden Hauptsatz, den *Maßerweiterungssatz von Hahn*, geliefert.

(13.1) Satz. *Sei $(X, \mathfrak{A}, \mu)$ ein σ-endlicher Prämaßraum. Dann läßt sich μ auf genau eine Weise zu einem Maß auf der von $\mathfrak{A}$ erzeugten σ-Algebra $\mathfrak{M} = \mathfrak{M}(\mathfrak{A})$ machen. Das Maß von M wird dabei durch*

$$\mu(M) = \inf \sum_{n=1}^{\infty} \mu(A_n)$$

definiert, und das Infimum ist über alle Folgen (A_n) in $\mathfrak{A}$ mit $M \subset \cup_{i=1}^{\infty} a_n$ zu bilden.

Zum Beweis des Satzes bemerken wir zunächst, daß das genannte Infimum für jeder Teilmenge von X gebildet werden kann. Die resultierende Funktion ist dann zwar im allgemeinen kein Maß auf der Menge aller Teilmengen. Da durch die Infimumbildung eine Menge von oben durch Vereinigungen meßbarer Mengen approximiert wird, so entsteht ein *Obermaß* oder, wie man meist sagt, ein äußeres Maß.

Ist μ^* ein äußeres Maß auf der Potenzmenge $\mathfrak{P}(X)$ von X, so nennen wir $A \subset X$ *μ^*-meßbar*, wenn für jede Teilmenge Z von X die Zerlegungsformel

$$\mu^*(Z) = \mu^*(Z \cap A) + \mu^*(Z \setminus A)$$

gilt. Diese Definition, die nicht so leicht zu motivieren ist, ist eine beweistechnische großartige Entdeckung. Um sie zu rechtfertigen, zeigen wir:

(13.2) Notiz. *Sei μ^* das in* (13.??) *konstruierte äußere Maß. Die Mengen aus $\mathfrak{A}$ sind dann μ^*-meßbar.*

Beweis. $\square$

Die Nützlichkeit der μ^*-Meßbarkeit wird durch den folgenden Satz von Carathéodory belegt.

(13.3) Satz. *Sei μ^* ein äußeres Maß auf der Potenzmenge von X. Das System $\mathfrak{M}$ der μ^*-meßbaren Teilmengen ist dann eine σ-Algebra und μ^* ein Maß auf $\mathfrak{M}$.*

Beweis. $\square$

Beweis von (13.??). Der Satz von Carathéodory liefert uns eine σ-Algebra $\mathfrak{M}$, die nach (13.??) $\mathfrak{A}$ enthält. Die Maßfunktion μ^* stimmt nach (13.??) mit μ überein. Damit ist die Existenz einer Erweiterung gezeigt.

Eindeutigkeit. Sei μ das soeben konstruierte Maß und ν ein weiteres, daß auf $\mathfrak{A}$ mit μ übereinstimmt. Habe $S \in \mathfrak{A}$ endliches Maß. Wir zeigen für $Y \in \mathfrak{M}$ und $Y \subset S$ die Gleichheit $\mu(Y) = \nu(Y)$. Nach Definition des äußeren Maßes ist

$$\mu(Y) = \inf \sum_n \mu(A_n) = \inf \sum_n \nu(A_n).$$

Wegen $\nu(Y) \le \nu(\bigcup_i A_i) \le \sum_i \nu(A_i)$ ist deshalb $\nu(Y) \le \mu(Y)$. Dieselbe Überlegung dürfen wir auch auf das Komplement $A \setminus Y$ anwenden und bekommen $\nu(A \setminus Y) \le \mu(A \setminus Y)$. Die Additivität der Maße liefert zusammen mit diesen Ungleichungen

$$\mu(A) = \nu(A) = \nu(A \setminus Y) + \nu(A) \le \mu(A \setminus Y) + \mu(Y) = \mu(A).$$

Das ist nach allem nur mit Gleichheiten verträglich, $\mu(Y) = \nu(Y)$. Ist nun das Prämaß σ-endlich und (S_n) eine Folge wie in der Definition dieses Begriffes, so wissen wir also, daß für $Y \in \mathfrak{M}$ die μ- und ν-Maße von $Y \cap (\bigcup_{i=1}^n S_i)$ gleich sind, und durch Übergang zum Limes (??) folgt $\mu(Y) = \nu(Y)$. $\square$

14 Das Lebesguesche Integral

Ein Hauptsatz für die Entwicklung der Integrationstheorie und ein Hauptsatz der Maßtheorie ist:

(14.1) Satz. *Es gibt genau ein Maß λ^n auf der Borel–Algebra $\mathcal{B}(\mathbb{R}^n)$, das auf allen Quadern $\prod_{i=1}^n [a_i, b_i]$ den Wert $\prod_{i=1}^n (b_i - a_i)$ annimmt.*

Das durch (14.1) zur Verfügung gestellte Maß heißt das *Lebesguesche Maß* auf dem $\mathbb{R}^n$. Wir nennen $\lambda^n(A)$ das *(n-dimensionale) Volumen* von $A \subset \mathbb{R}^n$ (auch *Fläche* im Fall $n = 2$). Es hat sich herausgestellt, daß man nicht allen Teilmengen des $\mathbb{R}^n$ sinnvoll ein Volumen zuordnen kann. Die Borel–Algebra und die sogleich besprochene Erweiterung durch die Nullmengen liefert einen größtmöglichen allgemeinen Rahmen für das Volumenproblem. Allerdings weiß niemand zu sagen, genau welche Mengen in $\mathcal{B}(\mathbb{R}^n)$ liegen. Man möchte lediglich sicherstellen, daß einigermaßen zugängliche Mengen, wie offene und abgeschlossene, darin enthalten sind, und übliche mengentheoretische Operationen nicht aus dem System herausführen[2].

Im Fall des Maßes (14.1) lassen sie sich wie folgt charakterisieren:

(14.2) Satz. *Folgende Aussagen sind äquivalent:*

(1) *A ist eine Nullmenge für das Maß* (14.1).

(2) *Zu jedem $\varepsilon > 0$ gibt es eine Familie $(W_i \mid i \in \mathbb{N})$ von Quadern mit $A \subset \bigcup_{i=1}^{\infty} W_i$ und $\sum_i \lambda^n(W_i) < \varepsilon$.*

Wie schon in (14.1) ist auch in (14.2) das Volumen eines Quaders das elementargeometrische: Produkt der Kantenlängen. Wegen dieses Satzes kann man mittels der Eigenschaft (14.2) Nullmengen direkt definieren, ohne ein Maß zu haben.

(14.3) Beispiel. Das verstaubte, aber als Beschäftigungstherapie immer noch benutzte Riemann-Integral hat als Definition oder Satz: Eine Funktion $f\colon \mathbb{R}^n \to \mathbb{R}$ ist *Riemann–integrierbar*, wenn sie beschränkt ist, außerhalb einer kompakten Menge gleich Null und außerhalb einer Nullmenge stetig. $\diamond$

20 Anhang: Mengen. Abbildungen. Relationen

1 Die Mengensprache

Eine *Menge* ist eine Zusammenfassung von *Elementen* zu einem neuen Objekt. Das Symbol $x \in A$ besagt, x ist ein Element der Menge A. Wir sagen dafür auch, x ist in der Menge A enthalten. Wie wird eine Menge mitgeteilt? Eine Möglichkeit ist, ihre Elemente in geschweiften Klammern einfach aufzuschreiben; so ist $\{1, 2, 3\}$ die Menge mit den Elementen 1,2,3. Bei Mengen mit unendlich vielen Elementen ist das natürlich unmöglich. Dann kann man etwa folgende Mitteilungsart wählen: $\{x \mid x \text{ hat die Eigenschaft } E\}$ ist die Menge aller Elemente mit der Eigenschaft E. Jedes Gedankending kann Element einer Menge werden. So können Mengen selbst wieder Elemente einer Menge sein. Einige Zahlenmengen werden durch Standardsymbole notiert:

(1) $\mathbb{N} = \{1, 2, 3, \dots\}$ Menge der natürlichen Zahlen.

(2) $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ Menge der natürlichen Zahlen einschließlich der Null.

(3) $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ Menge der ganzen Zahlen.

[2]Genauso wie die reellen Zahlen eine unauslotbare „ideale Welt“ bilden, in der zum Beispiel alle wünschbaren Grenzprozesse durchführbar sind, so bilden auch die Borel- oder Lebesguemeßbaren Mengen ein „ideales Universum“ für die Zwecke der Maß- und Integrationstheorie. Man muß nicht verstehen wollen, was man nicht verstehen kann.

(4) $\mathbb{Q}$ Menge der rationalen Zahlen (Brüche).
(5) $\mathbb{R}$ Menge der reellen Zahlen.
(6) $\mathbb{C}$ Menge der komplexen Zahlen.

Die Kenntnis dieser Mengen sei unterstellt; zu den komplexen Zahlen jedoch siehe auch den nächsten Abschnitt.

Ist jedes Element von A auch Element von B, so sagt man, A sei in B *enthalten*, A sei *Teilmenge, Untermenge* von B und schreibt $A \subset B$ oder $B \supset A$ dafür. Zwei Mengen sind genau dann gleich, wenn sie dieselben Elemente enthalten, wenn also sowohl $A \subset B$ als auch $B \subset A$ gilt. Auf diesem Wege beweist man auch die Gleichheit zweier Mengen.

Aus vorgegebenen Mengen A und B lassen sich neue Mengen bilden. Der *Durchschnitt* $A \cap B$ besteht aus den Elementen, die sowohl in A als auch in B enthalten sind. Die *Vereinigung* $A \cup B$ besteht aus den Elementen, die entweder in A oder in B oder in beiden enthalten sind. Das *cartesische Produkt* $A \times B$ hat als Elemente alle geordneten Paare (a, b), wobei $a \in A$ und $b \in B$ ist. Der Terminus „geordnet“ bedeutet, daß es auf die Reihenfolge ankommt, das heißt, (a, b) ist von (b, a) zu unterscheiden.

Die vorstehenden Bildungen lassen sich für ein beliebiges System $(A_j \mid j \in J)$ von Mengen A_j durchführen. Der Index j dient hier zur Kennzeichnung und Unterscheidung der Mengen und kann zum Beispiel eine Nummer sein. Man nennt in diesem Kontext J eine *Indexmenge* und $(A_j \mid j \in J)$ eine durch J indizierte *Familie* von Mengen.

Der *Durchschnitt* $\bigcap_{j \in J} A_j$ besteht aus den Elementen, die in sämtlichen A_j liegen. Die *Vereinigung* $\bigcup_{j \in J} A_j$ besteht aus den Elementen, die in mindestens einer der Mengen A_j liegen. Damit man den Durchschnitt immer bilden kann, ist es nützlich, die *leere Menge* $\emptyset$ zuzulassen, die überhaupt kein Element hat und folglich Teilmenge jeder Menge ist. Gilt $A \cap B = \emptyset$, so heißen A und B *disjunkt*. Natürlich verwendet man Bezeichungen wie $A_1 \cap A_2 \cap A_3$ für den Durchschnitt der drei Mengen A_1, A_2, A_3. Analog für Vereinigungen und bei mehreren Mengen.

Sind A und B Mengen, so bezeichnet $A \setminus B$ die Menge der Elemente von A, die nicht in B enthalten sind (*Differenzmenge*). Bei dieser Bezeichung wird nicht unterstellt, daß B Teilmenge von A ist. Besteht B nur aus dem Element b, so schreiben wir auch kurz $A \setminus b$.

Das *cartesische Produkt* $\prod_{j \in J} A_j$ besteht aus allen Familien $(a_j \mid j \in J)$, worin $a_j \in A_j$ ist. Auch hier gibt es Bezeichnungen wie $A_1 \times A_2 \times A_3$.

Seien A und B Mengen. Eine *Abbildung von A nach B* ist eine Vorschrift, die jedem $a \in A$ ein $b \in B$ zuordnet. Die Vorschrift wird meist durch ein Funktionssymbol $b = f(a)$ geschrieben. Man sagt auch, a werde auf $f(a)$ abgebildet. In Symbolen schreiben wir dann

$$f \colon A \to B, \quad a \mapsto f(a) = fa.$$

Manchmal schreiben wir auch f neben oder über den Pfeil. Dabei heißt A der *Definitionsbereich* oder die *Quelle* der Abbildung, B der *Bildbereich* oder das *Ziel* der Abbildung. Die „Vorschrift“ ist durch die Teilmenge

$$\{(a,b) \mid b = f(a),\ a \in A\} \subset A \times B$$

bestimmt und kann mengentheoretisch mit ihr gleichgesetzt werden. Zu einer Abbildung gehören also die drei Daten A, B und f, wenn man präzise sein will. In der Praxis erlaubt man sich allerdings gewisse Freiheiten. Statt Abbildung sagt man auch *Funktion*, insbesondere dann, wenn die Zielmenge aus Zahlen besteht. In einigen Fällen sind auch Worte wie *Operator* oder *Funktional* gebräuchlich. Die Abbildung $A \times B \to A$, $(a,b) \mapsto a$ heißt *Projektion auf den Faktor* A; analog hat man für ein beliebiges cartesisches Produkt die Projektionen auf die Faktoren. Die Abbildung $\mathrm{id}(A) = \mathrm{id}\colon A \to A$, $a \mapsto a$ heißt die *identische Abbildung* von A. Zu einer Inklusion $A \subset B$ gehört die Abbildung $i\colon A \to B$, $a \mapsto a$, die wir ebenfalls *Inklusion* nennen und auch mit $i\colon A \subset B$ mißbräuchlich bezeichnen. Eine gegebene Abbildung $f\colon B \to Y$ besitzt die *Einschränkung* $f \circ i\colon A \to Y$ auf die Teilmenge $A \subset B$, die auch $f|A$ notiert wird.

Die *Verkettung* der Abbildungen $f\colon A \to b$ und $g\colon B \to C$ ist die Abbildung $A \to C$, die a auf $g(f(a))$ abbildet; sie wird mit gf oder $g \circ f$ bezeichnet. Oft hat man eine größere Anzahl von Abbildungen und Verkettungen zu notieren. Das geschieht übersichtlich in einem *Diagramm*, welches aus den Abbildungspfeilen, ihren Quellen und Zielen besteht. Führen alle möglichen Verkettungen zwischen festen Stellen in einem Diagramm immer zum gleichen Ergebnis, so sagen wir, das Diagramm sei *kommutativ* oder es *kommutiere*. Zum Beispiel ist ein Rechteck

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"] & B \arrow[d, "h"] \\ C \arrow[r, "i"] & D \end{tikzcd}$$

kommutativ genau dann, wenn $h \circ f = i \circ g$ gilt.

Zu $f\colon A \to B$ und $f'\colon A' \to B'$ gehört das *cartesische Produkt*

$$f \times f'\colon A \times A' \to B \times B', \quad (a,a') \mapsto (f(a), f'(a')).$$

Analog haben wir das cartesische Produkt

$$\prod_{j \in J} f_j\colon \prod_{j \in J} A_j \to \prod_{j \in J} B_j$$

einer beliebigen Familie $(f_j\colon A_j \to B_j \mid j \in J)$ von Abbildungen.

Eine Abbildung $f\colon A \to B$ heißt *injektiv*, wenn aus $a_1 \neq a_2$ immer $f(a_1) \neq f(a_2)$ folgt, und *surjektiv*, wenn zu jedem $b \in B$ ein $a \in A$ mit $b = f(a)$ existiert. Sie heißt *bijektiv*, wenn sie sowohl surjektiv als auch injektiv ist. Eine Abbildung $f\colon A \to B$ ist genau dann bijektiv, wenn sie eine *Umkehrabbildung* $g\colon B \to A$ hat, das heißt, wenn eine Abbildung g existiert mit den Eigenschaften $gf = \mathrm{id}(A)$ und $fg = \mathrm{id}(B)$. Ist $C \subset A$, so heißt $f(C) = \{b \mid \text{es gibt } c \in C \text{ mit } b = f(c)\}$ das

Bild von C bei f. Ist $D \subset B$, so heißt $f^{-1}(D) = \{a \mid f(a) \in D\}$ das *Urbild* von D bei f. Das Urbild der einelementigen Teilmenge $\{c\}$ wird $f^{-1}(c)$ geschrieben.

Zwei Mengen heißen *gleichmächtig*, wenn es eine Bijektion (= eine bijektive Abbildung) zwischen ihnen gibt. Die *Mächtigkeit* einer Menge M ist die „Anzahl" ihrer Elemente; sie wird mit $|M|$ bezeichnet. Eine Menge heißt *abzählbar unendlich*, wenn es eine Bijektion zur Menge $\mathbb{N}$ der natürlichen Zahlen gibt.

(1.1) Satz. *Die Menge der Abbildungen* $\mathbb{N} \to 0, 1$*, die Menge der sogenannten Null-Eins-Folgen ist nicht abzählbar.* □

Eine *Äquivalenzrelation* auf einer Menge A ist eine Teilmenge $R \subset A \times A$ mit den folgenden Eigenschaften:

(1) Für jedes $a \in A$ ist $(a, a) \in R$.
(2) Ist $(a, b) \in R$, so ist auch $(b, a) \in R$.
(3) Ist $(a, b) \in R$ und $(b, c) \in R$, so ist auch $(a, c) \in R$.

Ist R solch eine Äquivalenzrelation, so schreibt man $a \sim b$ für $(a, b) \in R$ und sagt in diesem Fall, a ist *äquivalent* zu b bezüglich oder modulo dieser Relation. Die Menge $K(a) = \{b \mid a \sim b\}$ heißt die *Äquivalenzklasse* von a. Genau dann gilt $K(a) = K(b)$, wenn $a \sim b$ ist. Die Menge der Äquivalenzklassen wird mit A/R bezeichnet und A modulo R gelesen. Die Abbildung $p\colon A \to A/R$, $a \mapsto K(a)$ heißt die *Quotientabbildung* der Äquivalenzrelation. Es ist eine surjektive Abbildung. Die Menge A ist die disjunkte Vereinigung der Äquivalenzklassen. Ist irgendeine disjunkte Zerlegung von A in nichtleere Teilmengen gegeben, so gibt es genau eine Äquivalenzrelation auf A, deren Äquivalenzklassen genau die Mengen dieser Zerlegung sind. Ist $f\colon A \to B$ eine surjektive Abbildung, so bilden die Urbilder der Elemente von B solch eine disjunkte Zerlegung. Ein Element einer Äquivalenzklasse wird *Repräsentant* dieser Klasse genannt. Ist $p\colon A \to A/R$ die Quotientabbildung einer Äquivalenzrelation und $f\colon A \to B$ eine Abbildung, so gibt es genau dann eine Abbildung $g\colon A/R \to B$ mit der Eigenschaft $g \circ p = f$, wenn f die Äquivalenzklassen von R auf einelementige Mengen abbildet. Man definiert g durch $g(K(a)) = f(a)$ und sagt, die Definition sei *unabhängig* von der Auswahl des Repräsentanten oder kurz und salopp: g ist *wohldefiniert*[3]. Es gibt also drei im wesentlichen gleichwertige Erscheinungsformen von Äquivalenzrelationen: Die Relation R, die disjunkte Zerlegung in Klassen, die surjektive Abbildung auf die Menge der Klassen.

Der Durchschnitt von Äquivalenzrelationen auf A, aufgefaßt als Teilmengen von $A \times A$, ist wieder eine. Zu jeder beliebigen Teilmenge $S \subset A \times A$ gibt es deshalb eine kleinste S umfassende Äquivalenzrelation, genannt die von S *erzeugte*. Sind R_1 und R_2 Äquivalenzrelationen auf A und gilt $R_1 \subset R_2$, so heißt R_1 *feiner* als R_2 (und R_2 *gröber* als R_1. Die feinste Äquivalenzrelation ist die *Gleichheit*; die zugehörige Teilmenge von $A \times A$ ist die *Diagonale* $\{(a, a) \mid a \in A\}$ von A.

Allgemein wird eine Teilmenge $R \subset A \times A$ als eine (zweistellige) *Relation* auf A bezeichnet. Die Idee dabei ist, daß R diejenigen Paare (a, b) aussondert, die in

[3]Damit ist „wohldefiniert" definiert.

einer jeweils festgelegten Beziehung (Relation) zueinander stehen. Meist haben Relationen ihre eigene Notation, zum Beispiel $a < b$, a kleiner als b. In diesem Fall besteht R aus den Paaren (a, b), für die $a < b$ gilt.

2 Das Zornsche Lemma

Sei X eine Menge. Eine Teilmenge $H \subset X \times X$ (eine „Relation") heißt *Halbordnung* oder *teilweise Ordnung* von X, wenn gilt:

(1) $(a, a) \in H$ für jedes $a \in X$.
(2) $(a, b) \in H$ und $(b, a) \in H \Rightarrow a = b$.
(3) $(a, b) \in H$, $(b, c) \in H \Rightarrow (a, c) \in H$.

Statt $(a, b) \in H$ schreiben wir meist $a \leq b$. Eine Halbordnung heißt *Ordnung*, wenn für je zwei Elemente $a, b \in X$ gilt: $a \leq b$ oder $b \leq a$. Eine Teilmenge $K \subset X$ heißt *Kette* bzgl. der Halbordnung, wenn für je zwei Elemente a, b aus K gilt: $a \leq b$ oder $b \leq a$. Ist $A \subset X$ und $s \in X$, so heißt s *obere (untere) Schranke* von A, wenn für alle a $a \in A$ die Relation $a \leq s$ $(s \leq a)$ gilt. Ein $a \in A$ heißt *größtes (kleinstes) Element* von A, wenn a obere (untere) Schranke von A ist. Ein $m \in A$ heißt *maximales (minimales) Element* von A, wenn für jedes $a \in A$ mit $m \leq a$ $(a \leq m)$ sogar $m = a$ gilt. Eine nichtleere Menge X mit Halbordnung heißt *induktiv geordnet*, wenn jede Kette in M eine obere Schranke hat. Das *Zornsche Lemma* lautet nun:

(2.1) Satz. *Jede induktiv geordnete Menge besitzt mindestens ein maximales Element.* □