

**Michael Růžička**

# **Analysis**

aufbauend auf den Skripten von  
Prof. em. Dr. Dr. h.c. Rolf Schneider



# Inhaltsverzeichnis

---

## Analysis I

---

<b>Sprech- und Schreibweisen</b> .....	3
<b>1 Die reellen Zahlen</b> .....	5
1.1 Die Körperaxiome .....	5
1.2 Die Anordnungsaxiome .....	8
1.3 Das Vollständigkeitsaxiom .....	12
1.4 Natürliche Zahlen und vollständige Induktion .....	13
<b>2 Abbildungen</b> .....	23
2.1 Der Funktionsbegriff .....	23
2.2 Abzählbarkeit .....	27
<b>3 Konvergenz</b> .....	31
3.1 Konvergente Folgen .....	31
3.2 Reihen .....	42
3.3 Die Exponentialreihe .....	52
<b>4 Topologie in <math>\mathbb{R}</math> und Stetigkeit</b> .....	57
4.1 Topologische Eigenschaften .....	57
4.2 Grenzwerte von Funktionen und Stetigkeit .....	61
4.3 Eigenschaften stetiger Funktionen .....	69
<b>5 Spezielle Funktionen</b> .....	75
5.1 Logarithmus und allgemeine Potenz .....	75
5.2 Die Exponentialfunktion im Komplexen .....	78
5.3 Die trigonometrischen Funktionen .....	84
<b>6 Differenzierbare Funktionen</b> .....	91
6.1 Die Ableitung .....	91
6.2 Eigenschaften differenzierbarer Funktionen .....	98
6.3 Höhere Ableitungen und Taylorformel .....	104

<b>7</b>	<b>Integration</b> .....	115
	7.1 Regelfunktionen .....	115
	7.2 Das Integral einer Regelfunktion .....	119
	7.3 Integration und Differentiation .....	125
	7.4 Berechnung von Integralen .....	129
	7.5 Parameterabhängige Integrale .....	140
	7.6 Uneigentliche Integrale .....	145
<b>8</b>	<b>Funktionenreihen</b> .....	151
	8.1 Konvergenz von Funktionenfolgen .....	151
	8.2 Potenzreihen .....	157
	8.3 Fourierreihen .....	166

---

**Analysis II**


---

<b>9</b>	<b>Metrische Räume</b> .....	189
	9.1 Metrische und topologische Grundbegriffe .....	190
	9.2 Konvergenz und Vollständigkeit .....	195
	9.3 Der Banachsche Fixpunktsatz .....	199
	9.4 Stetigkeit und Zusammenhang .....	205
	9.5 Kompaktheit .....	208
<b>10</b>	<b>Der euklidische Raum</b> .....	213
	10.1 Der $n$ -dimensionale euklidische Vektorraum .....	213
	10.2 Abbildungen und Koordinatenfunktionen .....	221
<b>11</b>	<b>Differentiation</b> .....	227
	11.1 Differenzierbarkeit .....	229
	11.2 Partielle Ableitungen .....	237
	11.3 Höhere Ableitungen und Anwendungen .....	243
	11.4 Differenzierbare Abbildungen .....	252
<b>12</b>	<b>Gewöhnliche Differentialgleichungen</b> .....	269
	12.1 Motivation .....	269
	12.2 Existenztheorie .....	273
	12.3 Spezialfälle für Gleichungen 1. Ordnung .....	292
	12.3.1 Ortsunabhängige rechte Seiten $f = f(t)$ .....	292
	12.3.2 Zeitunabhängige rechte Seiten $f = f(y)$ .....	293
	12.3.3 Separierbare rechte Seiten $f = h(t)g(y)$ .....	297
	12.3.4 Lineare Gleichungen .....	301

**13 Systeme linearer Differentialgleichungen** ..... 305

13.1 Grundlagen ..... 305

13.2 Homogene Systeme ..... 308

13.3 Inhomogene Systeme ..... 312

13.4 Systeme mit konstantem **A** ..... 313

    13.4.1 Symmetrische Matrizen **A** ..... 314

    13.4.2 Matrizen **A** mit nur reellen Eigenwerten ..... 314

    13.4.3 Matrizen **A** mit komplexen Eigenwerten ..... 319

    13.4.4 Reelle Systeme für  $n = 2$  ..... 322

13.5 Exponentialfunktion für Matrizen ..... 328

**A Anhang** ..... 335

A.1 Die Jordan'sche Normalform ..... 335



## Analysis I



## Sprech- und Schreibweisen

Mathematische Sachverhalte werden mithilfe von Aussagesätzen (Aussagen) formuliert und benutzen oft die Mengenschreibweise. Dies gilt analog für das mathematische Argumentieren und Beweisen. Einige der dabei üblichen Ausdrucksweisen wollen wir jetzt zusammenstellen, weitere folgen bei Bedarf.

Unter einer Menge versteht man eine Zusammenfassung von unterscheidbaren mathematischen Objekten; diese Gesamtheit wird dann als neues Objekt angesehen. Diese plausibel klingende Erklärung pflegt man in der Mathematik als „naiv“ zu bezeichnen. Sie bedarf nämlich der Präzisierung. Eine strenge Begründung der Mengenlehre ist aber zum jetzigen Zeitpunkt weder notwendig noch angebracht. Erfahrungsgemäß kann man in der elementaren Analysis mit diesem naiven Gebrauch von Mengen sehr gut arbeiten. Wir werden also nicht genau definieren, was eine Menge sein soll, sondern lediglich gewisse Mengen als gegeben ansehen und gewisse Vorschriften, neue Mengen zu bilden, als zulässig betrachten. Mit anderen Worten, wir verwenden die Mengenlehre lediglich als eine bequeme und klare Sprechweise und nicht als das axiomatisch begründbare Fundament der Mathematik, als das sie sich in einem späteren Stadium darstellt.

Folgende Abkürzungen werden verwendet:

$x \in M$  bedeutet:  $x$  ist Element von  $M$

$x \notin M$  bedeutet:  $x$  ist nicht Element von  $M$

Nach Definition werden zwei Mengen  $M, N$  genau dann (dies heißt: dann und nur dann) als gleich angesehen, geschrieben  $M = N$ , wenn sie dieselben Elemente enthalten. Die Menge, die kein Element enthält, heißt *leere Menge*, sie wird mit  $\emptyset$  bezeichnet. Für konkrete Mengen, die im Verlauf einer mathematischen Darstellung auftreten, gibt es verschiedene Möglichkeiten diese anzugeben, zum Beispiel:

$\{a_1, \dots, a_n\}$  ist die Menge, deren Elemente genau die Objekte  $a_1, \dots, a_n$  sind (sie dürfen in beliebiger Reihenfolge hingeschrieben werden).

$\{x \in M \mid A(x)\}$  ist die Menge aller  $x \in M$  ( $M$  eine gegebene Menge), die die Eigenschaft  $A$  besitzen (manchmal wird die Grundmenge  $M$  weggelassen).

Aus gegebenen Mengen kann man neue bilden nach den folgenden Vorschriften:

$$M_1 \cap M_2 := \{x \mid x \in M_1 \text{ und } x \in M_2\}$$

ist der *Durchschnitt* von  $M_1$  und  $M_2$ ,

$$M_1 \cup M_2 := \{x \mid x \in M_1 \text{ oder } x \in M_2\}$$

ist die *Vereinigung* von  $M_1$  und  $M_2$ ,

$$M_1 \setminus M_2 := \{x \mid x \in M_1 \text{ und } x \notin M_2\}$$

ist die (*mengentheoretische*) *Differenz* von  $M_1$  und  $M_2$ .

$M_1$  heißt *Teilmenge* von  $M_2$ , geschrieben  $M_1 \subset M_2$ , wenn jedes Element von  $M_1$  auch Element von  $M_2$  ist. Offenbar gilt  $M_1 = M_2$  genau dann, wenn  $M_1 \subset M_2$  und  $M_2 \subset M_1$  ist. Für eine Teilmenge  $N$  einer Grundmenge  $M$  nennt man  $N^c := M \setminus N$  das *Komplement* von  $N$ .

Mathematische Aussagen sind sprachliche Gebilde, von denen es sinnvoll ist zu fragen, ob sie wahr oder falsch sind. Aussagen sind entweder wahr (W) oder falsch (F). Aus gegebenen Aussagen  $A, B$  kann man mit den folgenden Verknüpfungen neue bilden:

$\neg A$	(nicht A)	Negation
$A \wedge B$	(A und B)	Konjunktion
$A \vee B$	(A oder B)	Adjunktion
$A \Rightarrow B$	(A impliziert B)	Implikation
$A \Leftrightarrow B$	(A und B sind äquivalent)	Äquivalenz

Die folgende Tafel legt fest, wie diese Verknüpfungen gebraucht werden.

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$
W	W	F	W	W	W	W
W	F	F	F	W	F	F
F	W	W	F	W	W	F
F	F	W	F	F	W	W

Folgenden Bestandteile von Aussagen werden als Quantoren bezeichnet:

$\forall x$  bedeutet: für alle  $x$   
 $\exists x$  bedeutet: es gibt ein  $x$

Beide Quantoren werden immer nur in Verbindung mit Aussagen verwendet, die von einer Variablen abhängen, z.B.  $\exists x \in M : A(x)$  bedeutet: Es gibt ein Element  $x$  aus der Menge  $M$ , so dass die Aussage  $A(x)$  wahr ist. Aussagen mit Quantoren können auch negiert werden. Diese Negationen sind wie folgt definiert:

$\neg(\forall x \in M : A(x))$  wird zu  $\exists x \in M : \neg A(x)$   
 $\neg(\exists x \in M : A(x))$  wird zu  $\forall x \in M : \neg A(x)$

# 1 Die reellen Zahlen

Grundlage der Analysis sind die reellen Zahlen. Was sind Zahlen? Natürlich kann jeder mit Zahlen umgehen und hat genügend Erfahrung, um von ihrer Nützlichkeit überzeugt zu sein. Die Frage, was eine Zahl „wirklich“ ist, wird man schwerlich vernünftig beantworten können; das ist aber auch nicht notwendig, sondern es genügt zu wissen, wie man mit Zahlen umgeht. Wir wollen hier nicht den Versuch machen, explizit zu sagen, was reelle Zahlen sind, sondern lediglich gewisse Regeln festlegen, nach denen man mit reellen Zahlen rechnen und anderweitig umgehen kann. Wir werden dabei bemüht sein, so wenige Regeln wie möglich an den Anfang zu stellen und aus diesen alle anderen Regeln als Folgerungen herzuleiten. Dieses Vorgehen ist heute typisch für den Aufbau einer mathematischen Theorie: Gewisse Grundbegriffe werden undefiniert an den Anfang gestellt und gewisse Aussagen über diese Grundbegriffe als wahr angenommen. Weitere Begriffe dürfen dann nur durch Rückgriff auf bereits gegebene definiert werden, und weitere Aussagen müssen aus bereits bekannten hergeleitet werden. Die an den Anfang gestellten Aussagen bezeichnet man als *Axiome*. Sie haben in gewissem Sinn den Charakter von „Spielregeln“. In diesem Sinn wollen wir also jetzt ein Axiomensystem für reelle Zahlen angeben und erste Folgerungen daraus herleiten. Wir teilen das Axiomensystem in mehrere Axiomgruppen auf, die wir einzeln betrachten.

## 1.1 Die Körperaxiome

Gegeben sei eine nichtleere Menge  $\mathbb{R}$ . Ihre Elemente werden wir später *reelle Zahlen* nennen. Je zwei Elementen  $a, b \in \mathbb{R}$  sei ein Element  $a+b$ , genannt ihre *Summe*, und ein Element  $a \cdot b$  (meist  $ab$  geschrieben), ihr *Produkt*, eindeutig zugeordnet. Dabei sollen folgende Aussagen gelten:

**Die Körperaxiome**

$$(1.1) \quad \forall a, b, c \in \mathbb{R} : a + (b + c) = (a + b) + c$$

„Assoziativgesetz der Addition“

(Bedeutung der Beklammerung:  $a + (b + c) := a + d$  mit  $d := b + c$ , und analog in anderen Fällen)

$$(1.2) \quad \forall a, b \in \mathbb{R} : a + b = b + a$$

„Kommutativgesetz der Addition“

$$(1.3) \quad \exists 0 \in \mathbb{R} \quad \forall a \in \mathbb{R} : a + 0 = a$$

„Existenz eines neutralen Elements der Addition“

$$(1.4) \quad \forall a \in \mathbb{R} \quad \exists -a \in \mathbb{R} : a + (-a) = 0$$

„Existenz eines inversen Elements der Addition“

.....

$$(1.5) \quad \forall a, b, c \in \mathbb{R} : a(bc) = (ab)c$$

$$(1.6) \quad \forall a, b \in \mathbb{R} : ab = ba$$

$$(1.7) \quad \exists 1 \in \mathbb{R} \setminus \{0\} \quad \forall a \in \mathbb{R} : a \cdot 1 = a$$

$$(1.8) \quad \forall a \in \mathbb{R} \setminus \{0\} \quad \exists a^{-1} \in \mathbb{R} : aa^{-1} = 1$$

.....

$$(1.9) \quad \forall a, b, c \in \mathbb{R} : a(b + c) = ab + ac$$

„Distributivgesetz“

**Bemerkung.** Wegen der Assoziativgesetze dürfen wir schreiben:

$$a + (b + c) = (a + b) + c =: a + b + c,$$

$$a(bc) = (ab)c =: abc.$$

**Folgerungen aus den Axiomen**

**1.10 Behauptung.** Die Zahl 0 ist eindeutig bestimmt.

*Beweis.* Sei  $0' \in \mathbb{R}$  ein weiteres Element mit  $a + 0' = a$  für alle  $a \in \mathbb{R}$ . Speziell gilt dann  $0 + 0' = 0$ . Nach (1.3) gilt  $0' + 0 = 0'$ . Nach (1.2) ist  $0 + 0' = 0' + 0$ . Somit folgt aus den letzten drei Gleichungen  $0 = 0'$ . ■

**1.11 Behauptung.** Zu  $a \in \mathbb{R}$  ist  $-a$  eindeutig bestimmt.

*Beweis.* Sei  $a' \in \mathbb{R}$  ein Element mit  $a + a' = 0$ . Addition von  $-a$  ergibt  $-a + (a + a') = -a$ . Aus (1.1) folgt  $(-a + a) + a' = -a$ . Wegen  $-a + a \stackrel{(1.2)}{=} a + (-a) \stackrel{(1.4)}{=} 0$  und  $0 + a' \stackrel{(1.2)}{=} a' + 0 \stackrel{(1.3)}{=} a'$  folgt  $a' = -a$ . ■

**1.12 Behauptung.**  $-0 = 0$ .

*Beweis.* Es gilt  $0 \stackrel{(1.4)}{=} 0 + (-0) \stackrel{(1.2)}{=} -0 + 0 \stackrel{(1.3)}{=} -0$ . ■

**Bezeichnungen.** Statt  $a + (-b)$  schreibt man  $a - b$ , genannt *Differenz* von  $a$  und  $b$ .

**1.13 Behauptung.** Sei  $a, b \in \mathbb{R}$ . Die Gleichung  $a + x = b$  hat eine eindeutig bestimmte Lösung  $x \in \mathbb{R}$ .

*Beweis.* Wegen  $a + (b - a) \stackrel{(1.2)}{=} a + (-a + b) \stackrel{(1.1)}{=} (a + (-a)) + b = 0 + b = b + 0 = b$  ist  $x = b - a$  eine Lösung. Sei  $y \in \mathbb{R}$  ein Element mit  $a + y = b$ . Dann ist  $b - a = -a + b = -a + (a + y) = (-a + a) + y = 0 + y = y$ . ■

**1.14 Behauptung.** Für alle  $a \in \mathbb{R}$  gilt  $-(-a) = a$ .

*Beweis.* Nach Definition ist  $-(-a)$  das Inverse von  $-a$ , d.h. es gilt  $(-a) + (-(-a)) = 0$ . Andererseits ist  $(-a) + a = a + (-a) = 0$ . Aus Behauptung 1.11 folgt  $-(-a) = a$ . ■

**1.15 Behauptung.** Für alle  $a, b \in \mathbb{R}$  gilt  $-(a + b) = -a - b$ .

*Beweis.* Es gilt  $(a + b) - (a + b) = 0$  und  $(a + b) + (-a - b) = a + (b - a - b) = a + (-a + b - b) = a + (-a) + 0 = 0$ . Aus Behauptung 1.11 folgt die Behauptung. ■

Völlig analog wie die vorstehenden Behauptungen beweist man die folgenden Aussagen über die Multiplikation anstelle der Addition:

**1.16 Behauptung.** Die Zahl 1 ist eindeutig bestimmt.

**1.17 Behauptung.** Zu  $a \in \mathbb{R} \setminus \{0\}$  ist  $a^{-1}$  eindeutig bestimmt.

**1.18 Behauptung.** Sei  $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ . Die Gleichung  $ax = b$  hat eine eindeutig bestimmte Lösung  $x$ . Die Lösung ist  $a^{-1}b =: \frac{b}{a} =: b/a$ .

**1.19 Behauptung.** Für alle  $a \in \mathbb{R} \setminus \{0\}$  gilt  $(a^{-1})^{-1} = a$ .

**1.20 Behauptung.** Für alle  $a, b \in \mathbb{R} \setminus \{0\}$  gilt  $(ab)^{-1} = b^{-1}a^{-1} = a^{-1}b^{-1}$ .

Erst die nachstehenden Folgerungen benutzen auch das Distributivgesetz.

**1.21 Behauptung.** Für alle  $a, b, c \in \mathbb{R}$  gilt  $(a + b)c = ac + bc$ .

*Beweis.* Es ist  $(a + b)c \stackrel{(1.6)}{=} c(a + b) \stackrel{(1.9)}{=} ca + cb \stackrel{(1.6)}{=} ac + bc$ . ■

**1.22 Behauptung.** Für alle  $a \in \mathbb{R}$  gilt  $a \cdot 0 = 0$ .

*Beweis.* Wir haben  $a \cdot 0 + a \cdot 0 \stackrel{(1.9)}{=} a(0 + 0) = a \cdot 0$ . Da auch  $a \cdot 0 + 0 = a \cdot 0$  ist, folgt  $a \cdot 0 = 0$  aus Behauptung 1.13. ■

**1.23 Behauptung.** Für alle  $a, b \in \mathbb{R}$  gilt  $ab = 0$  genau dann, wenn  $a = 0$  oder  $b = 0$  ist.

*Beweis.* „ $\Rightarrow$ “ Sei  $ab = 0$  und etwa  $a \neq 0$ . Dann folgt  $0 = a^{-1}(ab) \stackrel{(1.5)}{=} (a^{-1}a)b = 1 \cdot b = b$ , also  $b = 0$ .

„ $\Leftarrow$ “ Sei  $a = 0$  oder  $b = 0$ , etwa  $b = 0$ . Dann ist  $ab = a \cdot 0 = 0$  nach Behauptung 1.22. ■

**1.24 Behauptung.** Für alle  $a, b \in \mathbb{R}$  gilt  $(-a)b = -(ab)$ .

*Beweis.* Es ist  $ab + (-a)b \stackrel{1.21}{=} (a + (-a))b = 0 \cdot b \stackrel{1.22}{=} 0$  und  $ab + (-ab) = 0$ . Aus der Eindeutigkeit des Inversen folgt die Behauptung. ■

Speziell ist also  $(-1)b = -b$  für alle  $b \in \mathbb{R}$ .

**1.25 Behauptung.** Für alle  $a, b \in \mathbb{R}$  gilt  $(-a)(-b) = ab$ .

*Beweis.* Es ist  $(-a)(-b) \stackrel{1.24}{=} -(a(-b)) \stackrel{1.24}{=} -(-ab) \stackrel{1.14}{=} ab$ . ■

**Bemerkung.** Ein Tripel  $(M, +, \cdot)$ , bestehend aus einer nichtleeren Menge  $M$  und zwei Verknüpfungen  $+$  und  $\cdot$  auf  $M$ , für die die Axiome (1.1)–(1.9) (mit  $M$  anstelle von  $\mathbb{R}$ ) gelten, wird als (*kommutativer*) *Körper* bezeichnet.

Wie man leicht nachweist, gibt es einen Körper  $(M, +, \cdot)$  mit  $M = \{0, 1\}$  und der Eigenschaft  $1 + 1 = 0$ . Aus den Körperaxiomen läßt sich also noch nicht herleiten, dass  $1 + 1 \neq 0$  ist.

## 1.2 Die Anordnungsaxiome

Zusätzlich nehmen wir nun an, dass auf der Menge  $\mathbb{R}$  eine Beziehung  $<$  (gelesen „kleiner als“) gegeben ist, die zwischen zwei Zahlen bestehen kann. Für je zwei Zahlen  $a, b \in \mathbb{R}$  ist also  $a < b$  eine Aussage, die wahr oder falsch sein kann. Wir fordern die folgenden Regeln, die sogenannten „Anordnungsaxiome“.

(2.1)  $\forall a, b \in \mathbb{R} : a \neq b \Leftrightarrow (a < b \vee b < a)$   
„Vergleichbarkeit“

(2.2)  $\forall a, b, c \in \mathbb{R} : (a < b \wedge b < c) \Rightarrow a < c$   
„Transitivität“

(2.3)  $\forall a, b, c \in \mathbb{R} : a < b \Rightarrow a + c < b + c$   
 „Monotoniegesetz der Addition“

(2.4)  $\forall a, b, c \in \mathbb{R} : (a < b \wedge 0 < c) \Rightarrow ac < bc$   
 „Monotoniegesetz der Multiplikation“

**Bezeichnungen.**

$a > b \Leftrightarrow b < a$  ( $>$  wird gelesen: „größer als“)

$a \leq b \Leftrightarrow a < b$  oder  $a = b$

$a \geq b \Leftrightarrow a > b$  oder  $a = b$

Statt  $a < b$  und  $b < c$  schreibt man kurz  $a < b < c$ . Hierfür sagt man auch, „ $b$  liegt zwischen  $a$  und  $c$ “.

**Definition.**  $a \in \mathbb{R}$  heißt *positiv* (*negativ*), wenn  $a > 0$  (bzw.  $a < 0$ ) ist.

Wir leiten nun aus den Anordnungsaxiomen (und den Körperaxiomen) einige Folgerungen her.

**2.5 Behauptung.**  $\forall a, b \in \mathbb{R} : a < b \Rightarrow -b < -a$ .

*Beweis.* Gelte  $a < b$ . Aus (2.3) (mit  $c = -a - b$ ) folgt  $a - a - b < b - a - b$ , also  $-b < -a$ . ■

**2.6 Behauptung.**  $\forall a, b, a', b' \in \mathbb{R} : (a < b \wedge a' < b') \Rightarrow a + a' < b + b'$ .

*Beweis.* Aus  $a < b$  und  $a' < b'$  folgt nach (2.3)  $a + a' < b + a' = a' + b < b' + b = b + b'$ , nach (2.2) ist also  $a + a' < b + b'$ . ■

**2.7 Behauptung.**  $\forall a, b, a', b' \in \mathbb{R} : (0 \leq a < b \wedge 0 \leq a' < b') \Rightarrow aa' < bb'$ .

*Beweis.* Ist  $a' = 0$ , so ist  $aa' = 0$  und  $0 < bb'$  nach (2.4), also  $aa' < bb'$ . Sei jetzt  $0 < a'$ . Nach (2.4) ist  $aa' < ba' < bb'$ , und nach (2.2) also  $aa' < bb'$ . ■

**2.8 Behauptung.**  $\forall a, b, c \in \mathbb{R} : (a < b \wedge c < 0) \Rightarrow ac > bc$ .

*Beweis.* Aus  $c < 0$  folgt nach Behauptung 2.5  $0 < -c$  und damit nach (2.4)  $a(-c) < b(-c)$ , also  $-(ac) < -(bc)$  und somit nach Behauptung 2.5  $ac > bc$ . ■

**2.9 Behauptung.**  $\forall a \in \mathbb{R} \setminus \{0\} : aa > 0$ .

*Beweis.* Ist  $a > 0$ , so folgt die Behauptung aus (2.4). Andernfalls ist nach (2.1)  $a < 0$ , also  $-a > 0$  und daher nach Behauptung 1.25 und (2.4)  $aa = (-a)(-a) > 0$ . ■

Insbesondere gilt  $1 > 0$ , da wegen (1.7) gilt  $1 \neq 0$  und wegen 2.9 gilt  $1 = 1 \cdot 1 > 0$ . Aus  $1 > 0$  und Behauptung 2.6 folgt  $1 + 1 > 0 + 0 = 0$ ; wegen  $1 + 1 > 0$  ist  $1 + 1 \neq 0$ , was allein aus den Körperaxiomen nicht gefolgert werden konnte.

**2.10 Behauptung.**  $\forall a \in \mathbb{R} : a > 0 \Rightarrow a^{-1} > 0$ .

*Beweis.* Sei  $a > 0$ . Es gilt  $a^{-1} = (aa^{-1})a^{-1} = a(a^{-1}a^{-1}) > 0$  nach Behauptung 2.9 und (2.4) (denn wegen  $aa^{-1} = 1$  und Behauptung 1.23 ist  $a^{-1} \neq 0$ ). ■

Analog ergibt sich:  $\forall a \in \mathbb{R} : a < 0 \Rightarrow a^{-1} < 0$ .

**2.11 Behauptung.**  $\forall a, b \in \mathbb{R} : 0 < a < b \Rightarrow a^{-1} > b^{-1}$ .

*Beweis.* Aus  $a > 0$  und  $b > 0$  folgt  $ab > 0$  und daraus  $a^{-1}b^{-1} = (ab)^{-1} > 0$  nach Behauptung 2.10. Hieraus und aus  $a < b$  folgt nach (2.4)  $aa^{-1}b^{-1} < ba^{-1}b^{-1}$ , also  $b^{-1} < a^{-1}$ . ■

**Bezeichnungen.**  $\mathbb{R}^+ := \{x \in \mathbb{R} \mid x > 0\}$

Für  $a, b \in \mathbb{R}$  mit  $a \leq b$  wird definiert:

$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$	abgeschlossenes Intervall
$(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$	offenes Intervall
$[a, b) := \{x \in \mathbb{R} \mid a \leq x < b\}$	halboffenes Intervall
$(a, b] := \{x \in \mathbb{R} \mid a < x \leq b\}$	halboffenes Intervall

In jedem Fall heißt  $b - a$  auch die *Länge* des Intervalls. Ferner wird definiert:

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

Das Symbol  $\infty$  ist dabei nur zur Vereinfachung der Schreibweise gewählt; es ist *kein* Element von  $\mathbb{R}$ . Alle vorstehend definierten Mengen, dazu  $\emptyset$  und  $\mathbb{R}$ , heißen *Intervalle*.

**Definition.** Für  $a \in \mathbb{R}$  definieren wir

$$|a| := \begin{cases} a & \text{falls } a \geq 0, \\ -a & \text{falls } a < 0. \end{cases}$$

Die Zahl  $|a|$  heißt der *Absolutbetrag* oder kurz der *Betrag* der Zahl  $a$ .

**Folgerung (Eigenschaften des Absolutbetrages).** Unmittelbar aus der Definition folgt für alle  $a \in \mathbb{R}$ :

$$\begin{aligned} |a| &\geq 0, \\ |a| &= 0 \Leftrightarrow a = 0, \\ -|a| &\leq a \leq |a|. \end{aligned}$$

**2.12 Behauptung.**  $\forall a \in \mathbb{R} : |-a| = |a|$

*Beweis.* Fallunterscheidung:

1. Fall:  $a \geq 0$ . Dann ist  $|a| = a$  und  $|-a| = -(-a) = a$ .
2. Fall:  $a < 0$ . Dann ist  $|a| = -a$  und  $|-a| = -a$ . ■

**2.13 Behauptung.** Für alle  $a, b \in \mathbb{R}$  gilt  $|ab| = |a||b|$ .

*Beweis.* Fallunterscheidung:

1. Fall:  $a \geq 0$  und  $b \geq 0$ . Dann ist  $ab \geq 0$ , also  $|ab| = ab$ , und es ist  $|a| = a$ ,  $|b| = b$ , also  $ab = |a||b|$ .
2. Fall:  $a \geq 0$  und  $b < 0$ . Dann ist  $ab \leq 0$ , also  $|ab| = -ab$ , und es ist  $|a| = a$ ,  $|b| = -b$ , also  $-ab = |a||b|$ .
3. Fall:  $a < 0$  und  $b \geq 0$ . Analog.
4. Fall:  $a < 0$  und  $b < 0$ . Analog. ■

**2.14 Behauptung.**  $\forall a \in \mathbb{R} \setminus \{0\} : \left| \frac{1}{a} \right| = \frac{1}{|a|}$ .

*Beweis.* Aus  $1 = |1| = \left| a \cdot \frac{1}{a} \right| = |a| \cdot \left| \frac{1}{a} \right|$  folgt die Behauptung. ■

**2.15 Behauptung (Dreiecksungleichung).**  $\forall a, b \in \mathbb{R} : |a + b| \leq |a| + |b|$ .

*Beweis.* Wegen  $a \leq |a|$  und  $b \leq |b|$  ist nach Behauptung 2.6  $a + b \leq |a| + |b|$ . Wegen  $-a \leq |a|$  und  $-b \leq |b|$  ist ebenso  $-(a + b) \leq |a| + |b|$ . Nach Definition des Absolutbetrages folgt  $|a + b| \leq |a| + |b|$ . ■

**2.16 Behauptung.**  $\forall a, b \in \mathbb{R} : \left| |a| - |b| \right| \leq |a + b|$ .

*Beweis.* Es ist  $|a| = |(a + b) - b| \leq |a + b| + |b|$ , also  $|a| - |b| \leq |a + b|$ . Vertauschung von  $a$  und  $b$  ergibt  $-(|a| - |b|) = |b| - |a| \leq |a + b|$ , insgesamt folgt die Behauptung. ■

**Bemerkung.** Ein Quadrupel  $(M, +, \cdot, <)$ , wobei  $M$  eine nichtleere Menge ist,  $+$  und  $\cdot$  zwei Verknüpfungen und  $<$  eine Beziehung auf  $M$  sind derart, dass die Axiome (1.1)–(1.9) und (2.1)–(2.4) (mit  $M$  statt  $\mathbb{R}$ ) erfüllt sind, heißt *angeordneter Körper*.

Wir bemerken ohne Beweis, dass bis jetzt zum Beispiel die folgenden Aussagen noch nicht beweisbar sind:

- (a) Es gibt ein  $x \in \mathbb{R}$  mit  $x^2 = 2$  (nach Definition ist  $x^2 := x \cdot x$  und  $2 := 1+1$ ).
- (b) Für gegebenes  $a \in \mathbb{R}$  ist schließlich  $1 + 1 + \cdots + 1 > a$ , wenn genügend oft 1 addiert wird.

### 1.3 Das Vollständigkeitsaxiom

Wie die Anmerkung am Ende des vorigen Paragraphen zeigt, reichen die Körper- und Anordnungsaxiome noch nicht aus, um alle Eigenschaften zu erfassen, die wir von den reellen Zahlen erwarten. Zur Formulierung des letzten Axioms dient die folgende Definition:

**Definition.** Sei  $M \subseteq \mathbb{R}$ . Eine Zahl  $s \in \mathbb{R}$  heißt *obere Schranke* für  $M$ , wenn

$$x \leq s \quad \text{für alle } x \in M$$

gilt.  $M$  heißt *nach oben beschränkt*, wenn es in  $\mathbb{R}$  eine obere Schranke für  $M$  gibt.

Eine Zahl  $s_0 \in \mathbb{R}$  heißt *kleinste obere Schranke* oder *Supremum* der Menge  $M$ , wenn  $s_0$  obere Schranke für  $M$  ist und jede Zahl  $s < s_0$  nicht obere Schranke für  $M$  ist.

Enthält die Menge  $M$  ein größtes Element  $m_0$  (also ein  $m_0 \in M$  mit  $x \leq m_0$  für alle  $x \in M$ ), so heißt  $m_0$  *Maximum* der Menge  $M$ .

Ein Maximum einer Menge  $M \subseteq \mathbb{R}$  ist also stets auch Supremum von  $M$ , aber im Allgemeinen nicht umgekehrt, denn ein Supremum von  $M$  braucht nicht Element von  $M$  zu sein.

Völlig analog definiert man die Begriffe untere Schranke, nach unten beschränkt, größte untere Schranke (Infimum), Minimum.

Die Menge  $M \subseteq \mathbb{R}$  heißt *beschränkt*, wenn sie nach oben und nach unten beschränkt ist.

**Beispiele.** Für die Intervalle  $[a, b]$ ,  $[a, b)$  ist jede Zahl  $s$  mit  $s \geq b$  eine obere Schranke, und  $b$  ist kleinste obere Schranke. Für  $[a, b]$  ist  $b$  das Maximum und  $a$  das Minimum, aber  $[a, b)$  besitzt kein Maximum.

**Bezeichnungen.**

$\sup M :=$  Supremum von  $M$

$\inf M :=$  Infimum von  $M$

$\max M :=$  Maximum von  $M$

$\min M :=$  Minimum von  $M$

Nun formulieren wir das **Vollständigkeitsaxiom**:

- (3.1) Zu jeder nichtleeren, nach oben beschränkten Teilmenge von  $\mathbb{R}$  gibt es in  $\mathbb{R}$  eine kleinste obere Schranke.

Als unmittelbare Folgerung haben wir:

**3.2 Behauptung.** *Zu jeder nichtleeren, nach unten beschränkten Teilmenge von  $\mathbb{R}$  gibt es in  $\mathbb{R}$  eine größte untere Schranke.*

*Beweis.* Sei  $\emptyset \neq M \subseteq \mathbb{R}$ . Setze  $-M := \{x \in \mathbb{R} \mid -x \in M\}$ . Für  $s \in \mathbb{R}$  zeigt man leicht:

$s$  untere Schranke für  $M \Leftrightarrow -s$  obere Schranke für  $-M$

$s$  größte untere Schranke für  $M \Leftrightarrow -s$  kleinste obere Schranke für  $-M$ .

Die Behauptung ergibt sich jetzt durch Anwendung des Vollständigkeitsaxioms auf  $-M$ . ■

Die Bedeutung des Axioms liegt darin, dass es die Existenz von Zahlen mit bestimmten Eigenschaften sichert. Zum Beispiel folgt die Existenz der Zahl  $\sqrt{2}$  aus dem Vollständigkeitsaxiom.

Das Quadrupel  $(\mathbb{R}, +, \cdot, <)$  mit (1.1)–(1.9), (2.1)–(2.4), (3.1) heißt *Körper der reellen Zahlen*. Eine solche Struktur ist „im wesentlichen“ eindeutig bestimmt, was wir hier aber nicht näher erläutern wollen. Die Elemente einer solchen Struktur nennen wir *reelle Zahlen*. Wenn wir im folgenden nur von „Zahlen“ sprechen, sind stets reelle Zahlen gemeint.

## 1.4 Natürliche Zahlen und vollständige Induktion

Im Körper  $\mathbb{R}$  der reellen Zahlen sind insbesondere die Zahlen

$$\begin{aligned}
&1 \\
&2 := 1 + 1 \\
&3 := 2 + 1 \\
&4 := 3 + 1 \\
&\text{u.s.w.}
\end{aligned}$$

enthalten. Sie heißen *natürliche Zahlen*. Das obenstehende „u.s.w.“ kann aber kaum als Definition dienen. Wir müssen also eine Definition der Menge der natürlichen Zahlen geben, und wir können das jetzt auch tun. Zur Menge der natürlichen Zahlen sollen selbstverständlich genau die Zahlen gehören, die man in der obigen Weise erhalten kann, daher die folgende

**Definition.**  $\mathbb{N}$  sei die kleinste Teilmenge von  $\mathbb{R}$ , die die Zahl 1 und mit  $x$  auch  $x + 1$  enthält. Die Elemente von  $\mathbb{N}$  heißen *natürliche Zahlen*.

Die „kleinste“ Teilmenge von  $\mathbb{R}$  mit einer Eigenschaft  $E$  ist dabei zu verstehen als Durchschnitt aller Teilmengen von  $\mathbb{R}$  mit der Eigenschaft  $E$ . Der Durchschnitt einer Familie von Mengen ist völlig analog erklärt wie früher der Durchschnitt von zwei Mengen. Der Deutlichkeit halber wollen wir vorübergehend (d.h. nur in diesem Paragraphen) eine Teilmenge  $M \subseteq \mathbb{R}$  *induktiv* nennen, wenn  $1 \in M$  ist und mit  $x \in M$  auch  $x + 1 \in M$  gilt. Zum Beispiel sind  $\mathbb{R}$  und  $\mathbb{R}^+$  induktive Teilmengen. Das System (synonym für „die Menge“) aller induktiven Teilmengen von  $\mathbb{R}$  sei mit  $I$  bezeichnet. Dann haben wir also nach Definition

$$\begin{aligned}
\mathbb{N} &:= \bigcap_{M \in I} M := \{a \in \mathbb{R} \mid a \in M \text{ für alle } M \in I\} \\
&= \{a \in \mathbb{R} \mid a \text{ ist Element jeder induktiven Teilmenge von } \mathbb{R}\}.
\end{aligned}$$

**Definition.**

$$\begin{aligned}
\mathbb{N}_0 &:= \mathbb{N} \cup \{0\} := \{0, 1, 2, 3, \dots\}, \\
\mathbb{Z} &:= \{a \in \mathbb{R} \mid a \in \mathbb{N}_0 \text{ oder } -a \in \mathbb{N}_0\}.
\end{aligned}$$

Die Elemente von  $\mathbb{Z}$  heißen *ganze Zahlen*.

$$\mathbb{Q} := \left\{ a \in \mathbb{R} \mid \exists m \in \mathbb{Z} \exists n \in \mathbb{N} : a = \frac{m}{n} \right\}.$$

Die Elemente von  $\mathbb{Q}$  heißen *rationale Zahlen* (oder *Brüche*). Die Elemente von  $\mathbb{R} \setminus \mathbb{Q}$  heißen *irrationale Zahlen*.

**Bemerkung.**  $\mathbb{Q}$  mit den Verknüpfungen  $+$  und  $\cdot$  und der Beziehung  $<$  ist ebenfalls ein angeordneter Körper. In diesem Körper ist die Gleichung  $x^2 = 2$  unlösbar.

**Bemerkung.** Dass es überhaupt irrationale Zahlen gibt, wird erst später bewiesen.

Wir zeigen nun, wie sich einige an sich wohlvertraute Eigenschaften natürlicher Zahlen aus dieser Definition streng herleiten lassen.

**4.1 Behauptung.** *Die Menge  $\mathbb{N}$  der natürlichen Zahlen ist nicht nach oben beschränkt.*

*Beweis.* [Beweis durch Widerspruch: Wir zeigen, dass aus der Negation (dem logischen Gegenteil) der Behauptung ein Widerspruch folgt.] Angenommen, die Behauptung wäre falsch. Nach dem Vollständigkeitsaxiom gibt es dann für  $\mathbb{N}$  eine kleinste obere Schranke  $s$ . Da  $s - 1$  keine obere Schranke für  $\mathbb{N}$  ist, existiert ein  $n \in \mathbb{N}$  mit  $s - 1 < n$ , woraus  $s < n + 1$  folgt. Da  $\mathbb{N}$  induktiv ist, ist mit  $n \in \mathbb{N}$  auch  $n + 1 \in \mathbb{N}$ , also ist  $s$  nicht obere Schranke für  $\mathbb{N}$ , ein Widerspruch. ■

**4.2 Behauptung.** *Zu jedem  $a \in \mathbb{R}^+$  und jedem  $b \in \mathbb{R}$  existiert ein  $n \in \mathbb{N}$  mit  $na > b$ .*

*Beweis.* Angenommen, das wäre falsch. Dann gibt es ein  $a \in \mathbb{R}^+$  und ein  $b \in \mathbb{R}$  mit  $na \leq b$  für alle  $n \in \mathbb{N}$ . Es ist also  $n \leq \frac{b}{a}$  für alle  $n \in \mathbb{N}$ , im Widerspruch zu Behauptung 4.1. ■

Behauptung 4.2 drückt eine besondere Eigenschaft der durch  $<$  gegebenen Anordnung aus, die man auch die *archimedische* Eigenschaft nennt.

Die folgenden Konsequenzen werden später bei der Behandlung des Grenzwertbegriffes ständig benutzt.

**4.3 Behauptung.**  $\forall a \in \mathbb{R}^+ \exists n \in \mathbb{N} : \frac{1}{n} < a$ .

*Beweis.* Nach Behauptung 4.2 existiert ein  $n \in \mathbb{N}$  mit  $na > 1$ , also mit  $\frac{1}{n} < a$ . ■

**4.4 Behauptung.** *Sei  $a \in \mathbb{R}$ ,  $a \geq 0$ . Ist  $a \leq \frac{1}{n}$  für alle  $n \in \mathbb{N}$ , so ist  $a = 0$ .*

*Beweis.* Andernfalls wäre  $\frac{1}{a}$  eine obere Schranke für  $\mathbb{N}$ . ■

**4.5 Behauptung.**  $\forall n \in \mathbb{N} : (n > 1 \Rightarrow n - 1 \in \mathbb{N})$ .

*Beweis.* Die Menge  $\{n \in \mathbb{N} \mid n = 1 \vee n - 1 \in \mathbb{N}\}$  ist induktiv, also gleich  $\mathbb{N}$ . ■

Hier wurde benutzt: Ist  $M \subseteq \mathbb{N}$  eine induktive Teilmenge, so muß nach Definition von  $\mathbb{N}$  auch  $\mathbb{N} \subseteq M$  gelten, also ist  $M = \mathbb{N}$ . Hierauf stützt sich auch das häufig verwendete

### Beweisprinzip der vollständigen Induktion

Dieses läßt sich folgendermaßen beschreiben. Es bezeichne  $A$  eine Aussage über natürliche Zahlen mit folgenden Eigenschaften:

- (a)  $A$  gilt für die Zahl 1,
- (b) Wenn  $A$  für die natürliche Zahl  $n$  gilt, dann auch für  $n + 1$ .

Dann gilt  $A$  für alle  $n \in \mathbb{N}$ .

Um dies einzusehen, betrachtet man die Menge

$$M := \{n \in \mathbb{N} \mid A \text{ gilt für } n\}.$$

Sie ist induktiv, muß also nach der obigen Bemerkung gleich  $\mathbb{N}$  sein.

Um also zu beweisen, dass eine Behauptung für alle natürlichen Zahlen zutrifft, verfährt man folgendermaßen:

- (a) „Induktionsanfang“: Man beweist, dass die Behauptung für die Zahl 1 zutrifft.
- (b) „Induktionsschluß“: Unter der „Induktionsannahme“, die Behauptung treffe für ein  $n \in \mathbb{N}$  zu, zeigt man, dass die Behauptung auch für  $n + 1$  zutrifft.

Hierfür nun einige Beispiele.

**4.6 Behauptung.**  $\forall n \in \mathbb{N} \forall x \in \mathbb{R} : (n < x < n + 1 \Rightarrow x \notin \mathbb{N})$ .

*Beweis.* durch vollständige Induktion:

**Induktionsanfang:** Für  $n = 1$  lautet die Behauptung:

$$\forall x \in \mathbb{R} : (1 < x < 2 \Rightarrow x \notin \mathbb{N})$$

Sie ist richtig, weil die Menge  $M := \{x \in \mathbb{R} \mid x = 1 \vee x \geq 2\}$  induktiv ist und daher  $\mathbb{N} \subseteq M$  gilt. Für alle  $n \in \mathbb{N}$  ist also  $n = 1$  oder  $n \geq 2$ .

**Induktionsschluß:** Wir machen die *Induktionsannahme*, die Behauptung gelte für die natürliche Zahl  $n$ , also

$$\forall x \in \mathbb{R} : (n < x < n + 1 \Rightarrow x \notin \mathbb{N}).$$

Sei jetzt  $x \in \mathbb{R}$  eine Zahl mit  $n + 1 < x < n + 2$ . Dann folgt  $n < x - 1 < n + 1$ , nach Induktionsannahme also  $x - 1 \notin \mathbb{N}$ . Wegen  $x > 1$  folgt aus Behauptung 4.5 jetzt  $x \notin \mathbb{N}$ . Die Behauptung ist also richtig für  $n + 1$ . Damit ist gezeigt, dass sie für alle  $n \in \mathbb{N}$  gilt. ■

Hieraus können wir den folgenden, häufig nützlichen Satz herleiten:

**4.7 Satz** (Wohlordnung der natürlichen Zahlen). *In jeder nichtleeren Menge von natürlichen Zahlen gibt es ein kleinstes Element.*

*Beweis.* Sei  $\emptyset \neq A \subseteq \mathbb{N}$ . Setze  $M := \{n \in \mathbb{N} \mid \forall a \in A : n \leq a\}$ .

**Behauptung.**  $\exists k \in M : k + 1 \notin M$ .

*Beweis.* Wäre das falsch, so wäre  $M$  induktiv, also  $M = \mathbb{N}$ . Wegen  $A \neq \emptyset$  existiert ein  $a \in A$ , wegen  $a + 1 \in \mathbb{N}$  folgt  $a + 1 \leq a$ , ein Widerspruch. ■

**Behauptung.**  $k$  ist kleinstes Element von  $A$ .

*Beweis.* Wegen  $k \in M$  ist  $k \leq a$  für alle  $a \in A$ . Es bleibt  $k \in A$  zu zeigen. Wegen  $k + 1 \notin M$  existiert ein  $a_0 \in A$  mit  $k + 1 > a_0$ . Aus  $k \leq a_0 < k + 1$  folgt nach Behauptung 4.6 aber  $k = a_0$ . ■

Damit ist Satz 4.7 bewiesen. ■

Als wichtiges Anwendungsbeispiel haben wir den folgenden Satz:

**4.8 Satz.** *Zu beliebigen reellen Zahlen  $a, b \in \mathbb{R}$  mit  $a < b$  existiert eine rationale Zahl  $r \in \mathbb{Q}$  mit  $a < r < b$ .*

*Beweis.* O.B.d.A. sei  $a \geq 0$  (andernfalls ersetze man  $a, b$  durch  $a + m, b + m$  mit genügend großem  $m \in \mathbb{N}$ ). Nach Behauptung 4.3 gibt es eine Zahl  $n \in \mathbb{N}$  mit  $\frac{1}{n} < b - a$ . Wähle ein solches  $n$  und setze  $M := \{m \in \mathbb{N} \mid \frac{m}{n} > a\}$ . Wegen Behauptung 4.2 ist  $M \neq \emptyset$ . Nach Satz 4.7 enthält  $M$  ein kleinstes Element  $k$ . Es ist also  $a < \frac{k}{n}$ . Wäre  $\frac{k}{n} \geq b$ , so wäre

$$\frac{k-1}{n} \geq b - \frac{1}{n} > b + a - b = a,$$

also  $k - 1 \in M$ , entgegen der Definition von  $k$ . Somit ist  $a < \frac{k}{n} < b$ . ■

Für die Aussage von Satz 4.8 sagt man auch: „ $\mathbb{Q}$  liegt dicht in  $\mathbb{R}$ “.

Neben Beweisen durch Induktion sind auch Definitionen durch Induktion möglich. Wir erläutern dies an einem Beispiel:

**Definition.** Sei  $a \in \mathbb{R}$ . Setze  $a^0 := 1$  und  $a^1 := a$ . Ist  $a^n$  für ein  $n \in \mathbb{N}$  schon definiert, so setze  $a^{n+1} := a \cdot a^n$ . Damit ist  $a^n$  für alle  $n \in \mathbb{N}_0$  definiert.

Man nennt ein solches Vorgehen „Definition durch vollständige Induktion“ oder „rekursive Definition“. Damit ist also folgendes gemeint. Sei  $k \in \mathbb{N}$ . Um für jedes  $n \in \mathbb{N}$  mit  $n \geq k$  eine Zahl  $a_n$  zu definieren, definiert man zuerst  $a_k$ . Dann gibt man an, wie  $a_n$  zu definieren ist, wenn  $n > k$  ist und die Zahlen  $a_k, \dots, a_{n-1}$  schon definiert sind. Dass damit  $a_n$  in der Tat für alle  $n \in \mathbb{N}$  mit  $n \geq k$  definiert ist, ist plausibel. Die Möglichkeit einer rekursiven Definition läßt sich durch Ausnutzung der Eigenschaften von  $\mathbb{N}$  streng begründen, was hier aber nicht geschehen soll. (Statt Zahlen können die  $a_n$  auch Elemente irgendwelcher Mengen sein.)

Über die so definierten *Potenzen* beweisen wir nun einige einfache Aussagen.

**4.9 Behauptung.**  $\forall a \in \mathbb{R} \forall m, n \in \mathbb{N} : a^m a^n = a^{m+n}$ .

*Beweis.* Die Behauptung ist richtig für  $n = 1$  (nach Definition). Sei sie bewiesen für ein  $n \geq 1$ . Für  $m \in \mathbb{N}$  gilt dann

$$a^m a^{n+1} \stackrel{\text{Def.}}{=} a^m a^n a \stackrel{\text{Ind. - Ann.}}{=} a^{m+n} a \stackrel{\text{Def.}}{=} a^{m+n+1}. \quad \blacksquare$$

**4.10 Behauptung** (Bernoullische Ungleichung).

$$\forall a \in \mathbb{R} \forall n \in \mathbb{N}_0 : (a > -1 \Rightarrow (1+a)^n \geq 1+na).$$

*Beweis.* Für  $n = 0$  und  $n = 1$  ist die Behauptung richtig. Sei sie bewiesen für ein  $n \geq 1$ . Dann folgt  $(1+a)^{n+1} = (1+a)(1+a)^n \geq (1+a)(1+na) = 1 + (n+1)a + na^2 \geq 1 + (n+1)a$ .  $\blacksquare$

**4.11 Behauptung.** Sei  $b \in \mathbb{R}$ ,  $a \in \mathbb{R}^+$ .

(1)  $b > 1 \Rightarrow \exists n \in \mathbb{N} : b^n > a$ .

(2)  $0 < b < 1 \Rightarrow \exists n \in \mathbb{N} : b^n < a$ .

*Beweis.* (1) Nach Behauptung 4.2 existiert ein  $n \in \mathbb{N}$  mit  $n(b-1) > a-1$ . Nach Behauptung 4.10 folgt

$$b^n = (1 + (b - 1))^n \geq 1 + n(b - 1) > 1 + a - 1 = a.$$

(2) Aus  $0 < b < 1$  folgt  $\frac{1}{b} > 1$ , nach (1) existiert also ein  $n \in \mathbb{N}$  mit  $\left(\frac{1}{b}\right)^n > \frac{1}{a}$ . Also gilt  $\frac{1}{b^n} > \frac{1}{a}$  (denn  $\left(\frac{1}{b}\right)^n = \frac{1}{b^n}$ , wie sofort durch Induktion folgt). Behauptung 2.11 liefert also  $b^n < a$ . ■

Es folgen nun noch einige Aussagen über

### Summen und Produkte

**4.12 Behauptung.**  $\forall n \in \mathbb{N} \forall m \in \mathbb{N} : n + m \in \mathbb{N}$ .

*Beweis.* Wir beweisen dies durch Induktion nach  $n$ :

**Induktionsanfang:** Zu zeigen ist die Behauptung

$$\forall m \in \mathbb{N} : 1 + m \in \mathbb{N}.$$

Sie ist richtig nach Definition von  $\mathbb{N}$ .

**Induktionsschluß:** Wir machen die Annahme

$$\forall m \in \mathbb{N} : n + m \in \mathbb{N}.$$

Für beliebiges  $m \in \mathbb{N}$  gilt dann  $(n + 1) + m = n + (m + 1) \in \mathbb{N}$  nach Induktionsannahme. Die Behauptung gilt also auch für  $n + 1$ . ■

**4.13 Behauptung.**  $\forall n \in \mathbb{N} \forall m \in \mathbb{N} : nm \in \mathbb{N}$

*Beweis.* Analog, unter Verwendung von Behauptung 4.12. ■

**4.14 Behauptung.**  $\forall n \in \mathbb{N} \forall m \in \mathbb{N} : (m > n \Rightarrow m - n \in \mathbb{N})$ .

*Beweis.* Ähnlich; zum Induktionsanfang siehe Behauptung 4.5. ■

Durch rekursive Definition können wir auch Summen und Produkte von  $n$  reellen Zahlen erklären. Für  $a_1, a_2 \in \mathbb{R}$  sind  $a_1 + a_2$  und  $a_1 \cdot a_2$  bereits erklärt. Nehmen wir an, für gegebenes  $n \in \mathbb{N}$  und beliebige Zahlen  $a_1, \dots, a_n \in \mathbb{R}$  seien die Summe  $a_1 + \dots + a_n$  und das Produkt  $a_1 \cdot \dots \cdot a_n$  bereits definiert. Dann definieren wir für  $a_1, \dots, a_{n+1} \in \mathbb{R}$

$$\begin{aligned} a_1 + \dots + a_{n+1} &:= (a_1 + \dots + a_n) + a_{n+1}, \\ a_1 \cdot \dots \cdot a_{n+1} &:= (a_1 \cdot \dots \cdot a_n) \cdot a_{n+1}. \end{aligned}$$

Man kann nun leicht aus dem Assoziativgesetz durch vollständige Induktion herleiten, dass

$$(a_1 + \cdots + a_p) + (a_{p+1} + \cdots + a_n) = a_1 + \cdots + a_n$$

ist. Daraus kann man dann, ebenfalls durch Induktion, herleiten, dass man bei einem Ausdruck aus Summenzeichen und Klammern überhaupt die Klammern weglassen kann (allgemeines Assoziativgesetz). Aus dem Kommutativgesetz leitet man durch Induktion leicht her, dass es bei einer Summe  $a_1 + \cdots + a_n$  nicht auf die Reihenfolge der Summanden ankommt (allgemeines Kommutativgesetz). Analoges gilt für die Multiplikation. Aus dem Distributivgesetz leitet man durch Induktion das allgemeine Distributivgesetz her:

$$a(a_1 + \cdots + a_n) = aa_1 + \cdots + aa_n.$$

Beim Arbeiten mit derartigen Ausdrücken ist die Verwendung von Summen- und Produktzeichen praktisch:

### Bezeichnungen.

$$\sum_{i=k}^n a_i := a_k + a_{k+1} + \cdots + a_n, \quad \text{falls } n \geq k$$

$$\prod_{i=k}^n a_i := a_k a_{k+1} \cdots a_n, \quad \text{falls } n \geq k.$$

Aus Assoziativ-, Kommutativ- und Distributivgesetz folgen einige Rechenregeln für Summen, z.B.

$$\sum_{i=k}^l \sum_{j=m}^n a_{ij} = \sum_{j=m}^n \sum_{i=k}^l a_{ij},$$

$$\left( \sum_{i=k}^l a_i \right) \left( \sum_{j=m}^n b_j \right) = \sum_{i=k}^l \sum_{j=m}^n a_i b_j.$$

Zum Abschluß dieses Paragraphen wollen wir noch zwei häufig auftretende Summen bzw. Produkte berechnen:

**4.15 Behauptung** (Summe einer endlichen geometrischen Reihe). *Für  $q \in \mathbb{R} \setminus \{0, 1\}$  und  $n \in \mathbb{N}$  gilt*

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

*Beweis.* Setze  $\sum_{k=0}^n q^k =: x$ . Dann ist

$$qx = q + q^2 + \cdots + q^{n+1} = x - 1 + q^{n+1},$$

also

$$x = \frac{1 - q^{n+1}}{1 - q}. \quad \blacksquare$$

Zur nun folgenden Berechnung der Potenz  $(a + b)^n$  sind zunächst einige Definitionen erforderlich.

**Definition.**

$$\begin{aligned} n! &:= 1 \cdot 2 \cdot \cdots \cdot n = \prod_{k=1}^n k && \text{für } n \in \mathbb{N} \\ 0! &:= 1 \\ \binom{n}{k} &:= \frac{n(n-1)\cdots(n-k+1)}{1 \cdot 2 \cdot \cdots \cdot k} && \text{für } n, k \in \mathbb{N} \\ \binom{n}{0} &:= 1 && \text{für } n \in \mathbb{N}_0 \end{aligned}$$

Man liest  $n!$  als „ $n$  Fakultät“ und  $\binom{n}{k}$  als „ $n$  über  $k$ “. Die Größen  $\binom{n}{k}$  heißen *Binomialkoeffizienten*. Offenbar gilt

$$\begin{aligned} \binom{n}{k} &= \frac{n!}{k!(n-k)!} && \text{für } k, n \in \mathbb{N}_0, n \geq k \\ \binom{n}{k} &= 0 && \text{für } k > n \\ \binom{n}{k} &= \binom{n}{n-k} \end{aligned}$$

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

Beweis der letzten Gleichung:

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

Jetzt können wir zeigen:

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

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

*Beweis.* (Induktion): Für  $n = 1$  gilt die Formel. Sei sie bewiesen für ein  $n \geq 1$ . Dann folgt

$$\begin{aligned} (a + b)^{n+1} &= (a + b)(a + b)^n \stackrel{\text{Ind. - Ann.}}{=} (a + b) \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^n \binom{n}{k} a^k b^{n-k+1} \\ &= a^{n+1} + \sum_{k=0}^{n-1} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=1}^n \binom{n}{k} a^k b^{n-k+1} + b^{n+1} \\ &= a^{n+1} + \sum_{j=1}^n \left[ \binom{n}{j-1} + \binom{n}{j} \right] a^j b^{n+1-j} + b^{n+1} \\ &= \sum_{j=0}^{n+1} \binom{n+1}{j} a^j b^{n+1-j}. \quad \blacksquare \end{aligned}$$

## 2 Abbildungen

### 2.1 Der Funktionsbegriff

In diesem Paragraphen geht es hauptsächlich um die Vereinbarung einiger Sprechweisen und die Festlegung von Bezeichnungen. Es kommen also vorwiegend Definitionen vor und fast keine Sätze.

Der Begriff der Funktion, der im folgenden präzisiert werden soll, wird auch im täglichen Leben verwendet. Was meinen wir etwa, wenn wir sagen, beim Autofahren sei der Luftwiderstand eine Funktion der Geschwindigkeit? Gemeint ist doch, dass zu jedem Geschwindigkeitswert ein wohlbestimmter Wert des Luftwiderstandes gehört. Es ist aber nicht unbedingt gemeint, dass wir diesen durch eine explizite Formel oder Berechnungsvorschrift angeben können. Historisch mag es wohl so gewesen sein, dass man unter einer Funktion zunächst eine konkrete Berechnungsvorschrift verstanden hat, aber dies hat sich als zu eng erwiesen. Wesentlich an einer Funktion, wie man sie heute versteht, ist lediglich, dass durch jeden „Argumentwert“ der zugehörige „Funktionswert“ eindeutig bestimmt ist. Unter einer Funktion wird man also eine eindeutige Zuordnung verstehen. Zum Beispiel beschreibt auch jedes Telefonbuch eine Funktion: Jedem Inhaber eines Telefonanschlusses innerhalb eines gewissen Gebietes wird darin seine Telefonnummer zugeordnet. Wenn man dieses Beispiel im Sinn behält, wird man nicht in die Gefahr geraten, eine Funktion für eine explizite Berechnungsvorschrift zu halten.

Wir wollen eine formale Definition einer Funktion geben. Dazu müssen also eine Menge  $M_1$ , die „Definitionsmenge“, und eine Menge  $M_2$ , die „Zielmenge“ oder „Wertemenge“ der zu definierenden Funktion gegeben sein. Jedem  $x \in M_1$  soll ein eindeutig bestimmter Funktionswert  $y \in M_2$  zugeordnet sein. Die Zuordnung ist vollständig festgelegt, wenn wir zu jedem  $x \in M_1$  und jedem  $y \in M_2$  wissen, ob  $y$  dem Element  $x$  zugeordnet ist oder nicht. Wir können daher die Zuordnungsvorschrift geradezu identifizieren mit der Angabe aller geordneten Paare  $(x, y)$ , für die  $y$  dem Element  $x$  zugeordnet ist.

Hierzu zunächst eine Vorbemerkung. Sei  $n \geq 2$  eine natürliche Zahl. Zu  $n$  Objekten  $x_1, \dots, x_n$  (die durch die Numerierung von 1 bis  $n$  mit einer Reihenfolge versehen sind) bilden wir ein neues Objekt  $(x_1, \dots, x_n)$ , das wir das (geordnete)  $n$ -Tupel von  $x_1, \dots, x_n$  nennen (2-Tupel heißen auch *Paare*, 3-Tupel *Tripel*). Wir wollen hierfür keine formale Definition geben (obwohl dies unter alleiniger Verwendung des Mengenbegriffs möglich wäre), sondern lediglich eine Gleichheitsdefinition festlegen:

$$(x_1, \dots, x_n) = (y_1, \dots, y_n) \\ \Leftrightarrow x_1 = y_1 \quad \text{und} \quad x_2 = y_2 \quad \text{und} \quad \dots \quad \text{und} \quad x_n = y_n.$$

**Bemerkung.** Mit  $(a, b)$  haben wir sowohl das offene Intervall von  $a$  bis  $b$  als auch das geordnete Paar mit Einträgen  $a$  und  $b$  bezeichnet. Diese (an sich unkorrekte) Mehrdeutigkeit gibt aber nicht Anlaß zu Verwechslungen, da aus dem Zusammenhang stets klar sein wird, was gemeint ist. In Zweifelsfällen wird angegeben, ob das geordnete Paar oder das Intervall gemeint ist.

Sind nun  $M_1, \dots, M_n$  Mengen, so nennt man die Menge

$$M_1 \times M_2 \times \dots \times M_n := \{(x_1, \dots, x_n) \mid x_1 \in M_1 \wedge \dots \wedge x_n \in M_n\}$$

das (*kartesische*) *Produkt* der Mengen  $M_1, \dots, M_n$  (in dieser Reihenfolge). Man schreibt auch

$$\underbrace{M \times \dots \times M}_{n \text{ - mal}} =: M^n.$$

Jetzt können wir definieren ( $M_1, M_2$  seien Mengen):

**Definition.** Eine *Funktion* oder *Abbildung* von  $M_1$  nach  $M_2$  ist eine Teilmenge  $f$  der Produktmenge  $M_1 \times M_2$  derart, dass zu jedem  $x \in M_1$  genau ein  $y \in M_2$  existiert mit  $(x, y) \in f$ .  $M_1$  heißt *Definitionsbereich* und  $M_2$  *Wertebereich* von  $f$ .

**Bezeichnungen.** Statt  $(x, y) \in f$  schreibt man  $y = f(x)$  und nennt  $f(x)$  das *Bild von  $x$  unter  $f$* , auch den *Wert von  $f$  an der Stelle  $x$* .

Die Abkürzungen

$$f : M_1 \rightarrow M_2 \quad \text{oder} \quad M_1 \xrightarrow{f} M_2$$

bedeuten, dass  $f$  eine Funktion von  $M_1$  nach  $M_2$  ist. In konkreten Fällen schreibt man auch

$$f : M_1 \rightarrow M_2 : x \mapsto f(x) \quad \text{oder} \quad f : x \mapsto f(x) \quad \text{für } x \in M_1.$$

**Bemerkung.** Am Anfang von Abschnitt 1.1 hatten wir gesagt: „Je zwei Elementen  $a, b \in \mathbb{R}$  sei ein Element  $a+b$ , ihre Summe, eindeutig zugeordnet.“ Wir können das jetzt so präzisieren, dass eine Abbildung  $+$  :  $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  gegeben sein soll. Ebenso ist die Multiplikation  $\cdot$  eine Abbildung von  $\mathbb{R} \times \mathbb{R}$  in  $\mathbb{R}$ .

**Bezeichnungen.** Sei  $f : M_1 \rightarrow M_2$  eine Abbildung. Für  $A \subseteq M_1$  heißt

$$\begin{aligned} f(A) &:= \{f(x) \mid x \in A\} \\ &= \{y \in M_2 \mid \exists x \in A : f(x) = y\} \end{aligned}$$

das *Bild* von  $A$  unter  $f$ . Die Menge  $f(M_1)$  heißt *Bild* oder *Wertemenge* von  $f$  und wird auch mit Bild  $f$  bezeichnet. Für  $B \subseteq M_2$  heißt

$$f^{-1}(B) := \{x \in M_1 \mid f(x) \in B\}$$

das *Urbild* von  $B$  bezüglich  $f$ .

$f$  heißt *injektiv* (Injektion)  $\Leftrightarrow \forall x, y \in M_1 : (x \neq y \Rightarrow f(x) \neq f(y))$

$f$  heißt *surjektiv* (Surjektion) auf  $M_2 \Leftrightarrow f(M_1) = M_2$

$f$  heißt *bijektiv* (Bijektion) auf  $M_2 \Leftrightarrow f$  injektiv und surjektiv auf  $M_2$ .

Ist  $f : M_1 \rightarrow M_2$  injektiv, so ist

$$f^{-1} := \{(f(x), x) \mid x \in M_1\}$$

eine bijektive Abbildung von Bild  $f = f(M_1)$  auf  $M_1$ ;  $f^{-1}$  heißt die *Umkehrabbildung* von  $f$ . Es gilt

$$\begin{aligned} f^{-1}(f(x)) &= x && \text{für alle } x \in M_1, \\ f(f^{-1}(y)) &= y && \text{für alle } y \in \text{Bild } f. \end{aligned}$$

Sind  $f : M_1 \rightarrow M_2$  und  $g : M_3 \rightarrow M_4$  Funktionen mit Bild  $f \subseteq M_3$ , so definiert man

$$g \circ f := \{(x, g(f(x))) \mid x \in M_1\}.$$

$g \circ f$  ist also eine Funktion von  $M_1$  nach  $M_4$ . Es gilt  $(g \circ f)(x) = g(f(x))$  für alle  $x \in M_1$ .  $g \circ f$  heißt *Verkettung* oder *Komposition* von  $g$  und  $f$ .

Für eine beliebige Menge  $M$  ist die *Identität* auf  $M$  erklärt als die Abbildung  $\text{id}_M : M \rightarrow M$  mit  $\text{id}_M(x) := x$  für alle  $x \in M$ . Ist  $f : M_1 \rightarrow M_2$  eine Bijektion auf  $M_2$  und  $f^{-1}$  die Umkehrabbildung, so hat man offenbar

$$f^{-1} \circ f = \text{id}_{M_1}, \quad f \circ f^{-1} = \text{id}_{M_2}.$$

Ist  $f : M_1 \rightarrow M_2$  eine Funktion und  $M \subseteq M_1$ , so heißt

$$f|_M := \{(x, f(x)) \mid x \in M\}$$

die *Einschränkung* oder *Restriktion* von  $f$  auf  $M$ .

Im weiteren Verlauf der Analysis werden wir es zunächst mit speziellen Funktionen, den Folgen, zu tun haben.

**Definition.** Sei  $M$  eine Menge. Eine *Folge in  $M$*  ist eine Abbildung  $f: \mathbb{N} \rightarrow M$  (an die Stelle von  $\mathbb{N}$  kann auch  $\{m \in \mathbb{Z} \mid m \geq m_0\}$  treten).

**Bezeichnungen.** Ist  $f : \mathbb{N} \rightarrow M$  eine Folge, so schreibt man statt  $f(n)$  meist  $f_n$  und statt  $f$  oft

$$(f_n)_{n \in \mathbb{N}} \quad \text{oder} \quad (f_1, f_2, f_3, \dots).$$

Man nennt  $f_n$  auch das  *$n$ -te Folgenglied*.

**Bemerkung.** Achtung! Man verwechsle nicht die Folge  $(f_1, f_2, f_3, \dots)$  mit der Menge  $\{f_1, f_2, f_3, \dots\}$ . Zum Beispiel ist  $(-1, 1, -1, 1, -1, \dots)$  die Abbildung, die der Zahl  $n \in \mathbb{N}$  die Zahl  $(-1)^n$  zuordnet, und  $\{-1, 1, -1, 1, \dots\}$  ist die Menge  $\{1, -1\}$ .

Später werden wir es vorwiegend mit reellen Funktionen zu tun haben.

**Definition.** Eine *reelle Funktion* ist eine Abbildung  $f : D \rightarrow \mathbb{R}$  mit  $D \subseteq \mathbb{R}$ .

Einige mögliche Eigenschaften reeller Funktionen:

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine reelle Funktion. Sie heißt *beschränkt*, wenn ihr Wertebereich beschränkt (d.h. nach oben und nach unten beschränkt) ist, also:

$$f \text{ heißt beschränkt} \Leftrightarrow \exists c \in \mathbb{R}^+ \quad \forall x \in D : |f(x)| \leq c.$$

Ferner:

$f$  heißt *monoton wachsend* (*streng monoton wachsend*)  $\Leftrightarrow \forall x, y \in D : (x < y \Rightarrow f(x) \leq f(y))$  (bzw.  $<$ ).

Analog werden definiert: *monoton fallend*, *streng monoton fallend*, *monoton* (= wachsend oder fallend), *streng monoton*.

Um in diesem Abschnitt wenigstens etwas zu beweisen, zeigen wir:

**1.1 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  eine streng monotone reelle Funktion. Dann ist  $f$  injektiv und die Umkehrfunktion  $f^{-1}$  streng monoton.

*Beweis.* Sei  $f$  etwa streng monoton wachsend (im anderen Fall schließt man analog). Sei  $x, y \in D, x \neq y$ . Aus  $x < y$  folgt  $f(x) < f(y)$ , also  $f(x) \neq f(y)$ , analog für  $x > y$ . Also ist  $f$  injektiv. Somit existiert die Umkehrfunktion  $f^{-1} : W \rightarrow D$  mit  $W := \text{Bild } f$ . Sei  $u, v \in W, u < v$ . Dann ist  $u = f(x), v = f(y)$  mit geeigneten  $x, y \in D$ . Wäre  $y \leq x$ , so folgte  $f(y) \leq f(x)$ , also  $v \leq u$ , was ein Widerspruch ist. Somit ist  $f^{-1}(u) = x < y = f^{-1}(v)$ ; also ist auch  $f^{-1}$  streng monoton wachsend. ■

## 2.2 Abzählbarkeit

Die nun zur Verfügung stehenden Begriffe der bijektiven und surjektiven Abbildung erlauben eine Unterscheidung der „Größe“ von Mengen durch Vergleich mit  $\mathbb{N}$  oder Teilmengen davon.

**Bezeichnungen.** Für  $n \in \mathbb{N}_0$  sei  $A_n := \{k \in \mathbb{N} \mid k \leq n\}$ , also  $A_0 = \emptyset, A_1 = \{1\}, A_2 = \{1, 2\}$ , u.s.w.  $A_n$  heißt der  $n$ -te Abschnitt von  $\mathbb{N}$ .

**2.1 Satz** (Schubfachprinzip). Für  $n, k \in \mathbb{N}$  mit  $n > k$  gibt es keine injektive Abbildung  $f : A_n \rightarrow A_k$ .

*Beweis.* Angenommen, das wäre falsch. Dann gibt es eine Zahl  $n \in \mathbb{N}$  und eine injektive Abbildung  $f : A_{n+1} \rightarrow A_n$  (denn jede Abbildung  $f : A_{n+1} \rightarrow A_k$  mit  $k < n$  ist auch Abbildung von  $A_{n+1}$  nach  $A_n$ ). Sei  $n$  die (nach dem Wohlordnungssatz 4.7 aus Kapitel 1 existierende) kleinste derartige Zahl.

1. Fall:  $n \notin \text{Bild } f$  oder  $n = f(n+1)$ .

Setze  $g := f|_{A_n}$  (Einschränkung von  $f$  auf  $A_n$ .)

2. Fall:  $n = f(k)$  für ein  $k < n+1$ .

Setze

$$g(m) := \begin{cases} f(m) & \text{für } m \in A_n \setminus \{k\} \\ f(n+1) & \text{für } m = k. \end{cases}$$

In beiden Fällen ist  $g$  eine injektive Abbildung von  $A_n$  in  $A_{n-1}$ . Das widerspricht der Wahl von  $n$  als kleinster Zahl mit einer derartigen Abbildung. ■

**Definition.** Eine Menge  $M$  heißt *n-elementig*, wenn es eine Bijektion von  $A_n$  auf  $M$  gibt.  $M$  heißt *endlich*, wenn  $M$   $n$ -elementig für ein  $n \in \mathbb{N}$  oder  $M = \emptyset$  ist. Andernfalls heißt  $M$  *unendlich*.

Aus Satz 2.1 folgt insbesondere, dass  $M$   $n$ -elementig nur für ein  $n$  sein kann; ferner folgt, dass es für eine endliche Menge keine Bijektion auf eine echte Teilmenge geben kann. Bei unendlichen Mengen ist das anders, zum Beispiel ist

$$f : \mathbb{N} \rightarrow \mathbb{N} \\ n \mapsto n + 1$$

eine Bijektion von  $\mathbb{N}$  auf die echte Teilmenge  $\{m \in \mathbb{N} \mid m \geq 2\}$ .

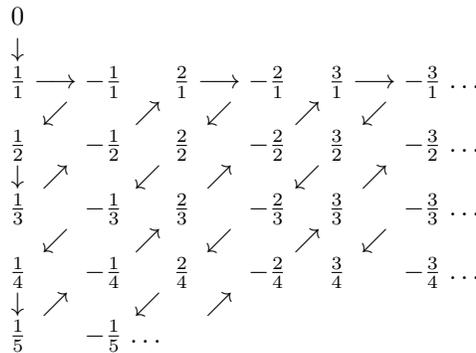
**Definition.** Eine Menge  $M$  heißt *abzählbar*, wenn es eine Surjektion von  $\mathbb{N}$  auf  $M$  gibt, und *überabzählbar*, wenn sie nicht abzählbar ist.

Man beachte, dass eine abzählbare Menge endlich oder unendlich sein kann. Im letzteren Fall heißt sie *abzählbar unendlich*.

Der folgende Satz ist in Anbetracht von Satz 4.8 aus Kapitel 1 zunächst vielleicht etwas überraschend:

**2.2 Satz.** Die Menge  $\mathbb{Q}$  der rationalen Zahlen ist abzählbar.

*Beweis.* Betrachte das untenstehende Schema. Jede rationale Zahl  $\neq 0$  ist von der Form  $\frac{m}{n}$  oder  $-\frac{m}{n}$  mit  $m, n \in \mathbb{N}$ . Daher kommt im Schema jede rationale Zahl vor (in der Tat öfter, was aber nicht stört). Durchläuft man das Schema, indem man den Pfeilen folgt, so wird dadurch offenbar eine fortlaufende Numerierung erklärt, also eine Surjektion von  $\mathbb{N}$  auf  $\mathbb{Q}$  gegeben. (Man könnte die Abbildung explizit hinschreiben.)



■

**Bemerkung.** Mit derselben Beweisidee läßt sich offenbar zeigen: Die Vereinigung einer abzählbaren Familie (= Menge) von abzählbaren Mengen ist abzählbar.

Im Gegensatz hierzu ist die Menge  $\mathbb{R}$  der reellen Zahlen überabzählbar. Um das zu beweisen, zeigen wir zunächst das auch später sehr wichtige „Prinzip der Intervallschachtelung“:

**2.3 Satz** (Intervallschachtelungsprinzip). *Sei  $(J_n)_{n \in \mathbb{N}}$  eine Folge abgeschlossener, beschränkter Intervalle (also  $J_n = [a_n, b_n]$  mit  $a_n, b_n \in \mathbb{R}$ ,  $a_n \leq b_n$  für  $n \in \mathbb{N}$ ) mit der Eigenschaft*

$$J_0 \supseteq J_1 \supseteq J_2 \supseteq \dots$$

*Dann gibt es eine reelle Zahl  $s \in \mathbb{R}$  mit  $s \in \bigcap_{n \in \mathbb{N}} J_n$ .*

*Beweis.* Nach Voraussetzung gilt  $a_k \leq a_n \leq b_n \leq b_k$  für  $k < n$ , also ist die Menge  $A := \{a_n \mid n \in \mathbb{N}\}$  nach oben beschränkt. Nach dem Vollständigkeitsaxiom besitzt sie in  $\mathbb{R}$  eine kleinste obere Schranke  $s$ . Gäbe es ein  $k \in \mathbb{N}$  mit  $s \notin J_k$ , so wäre  $s < a_k$  oder  $s > b_k$ . Im ersten Fall ist  $s$  nicht obere Schranke für  $A$ , im zweiten Fall ist  $s$  nicht kleinste obere Schranke, da  $b_k$  obere Schranke für  $A$  ist, was ein Widerspruch ist. ■

**Bemerkung.** In Satz 2.3 ist die Voraussetzung, dass die Intervalle  $J_n$  abgeschlossen und beschränkt sind, nicht entbehrlich, z.B. gilt

$$\bigcap_{n \in \mathbb{N}} [n, \infty) = \emptyset, \quad \bigcap_{n \in \mathbb{N}} (0, \frac{1}{n}) = \emptyset.$$

**2.4 Satz.** *Die Menge  $\mathbb{R}$  der reellen Zahlen ist überabzählbar.*

*Beweis.* Angenommen, das wäre falsch. Dann gibt es eine Surjektion  $f$  von  $\mathbb{N}$  auf  $\mathbb{R}$ . Setze  $f(n) = x_n$ , also  $\mathbb{R} = \{x_n \mid n \in \mathbb{N}\}$ . Wähle ein Intervall  $J_1 = [a_1, b_1]$  mit  $a_1 < b_1$  und  $x_1 \notin J_1$ . Ist  $J_n$  schon definiert, so wähle ein Intervall  $J_{n+1} = [a_{n+1}, b_{n+1}]$  mit  $a_{n+1} < b_{n+1}$  und  $J_{n+1} \subseteq J_n$ , so dass  $x_{n+1} \notin J_{n+1}$  ist. Damit ist rekursiv eine Folge  $(J_n)_{n \in \mathbb{N}}$  definiert, die die Voraussetzungen von Satz 2.3 erfüllt. Es gibt also eine Zahl  $s \in \mathbb{R}$  mit  $s = \bigcap_{n \in \mathbb{N}} J_n$ . Wegen  $\mathbb{R} = \{x_n \mid n \in \mathbb{N}\}$  gibt es eine Zahl  $k \in \mathbb{N}$  mit  $s = x_k$ . Nach Konstruktion ist aber  $x_k \notin J_k$ , q.e.a. ■

Aus den Sätzen 2.2 und 2.4 folgt, dass die Menge der irrationalen Zahlen überabzählbar ist (wäre  $\mathbb{R} \setminus \mathbb{Q}$  abzählbar, so wäre  $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$  abzählbar). Insbesondere ist damit die Existenz von irrationalen Zahlen gezeigt.



## 3 Konvergenz

### 3.1 Konvergente Folgen

Unter Folgen sollen, solange nichts anderes gesagt ist, Folgen in  $\mathbb{R}$  verstanden werden.

**Definition.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge in  $\mathbb{R}$  und sei  $a \in \mathbb{R}$ . Die Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *konvergent gegen  $a$* , und  $a$  heißt *Grenzwert* oder *Limes* dieser Folge, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $n_0 \in \mathbb{N}$  existiert mit

$$|a_n - a| < \varepsilon \quad \text{für alle } n \geq n_0.$$

In Kurzschreibweise lautet die Definition:

$$(a_n)_{n \in \mathbb{N}} \text{ konvergiert gegen } a \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \in \mathbb{N} : (n \geq n_0 \Rightarrow |a_n - a| < \varepsilon).$$

Eine Folge heißt *konvergent*, oder sie *konvergiert*, wenn sie konvergent gegen ein  $a \in \mathbb{R}$  ist. Eine gegen 0 konvergente Folge heißt *Nullfolge*. Eine Folge heißt *divergent*, oder sie *divergiert*, wenn sie nicht konvergiert.

**1.1 Satz.** *Der Grenzwert einer konvergenten Folge ist eindeutig bestimmt.*

*Beweis.* Die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergiere gegen  $a$  und gegen  $b$ . Angenommen, es wäre  $a \neq b$ . Zu  $\varepsilon := \frac{1}{2}|a - b|$  existieren ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < \varepsilon$  für  $n \geq n_0$  und ein  $n_1 \in \mathbb{N}$  mit  $|a_n - b| < \varepsilon$  für  $n \geq n_1$ . Für beliebiges  $n \geq \max\{n_0, n_1\}$  gilt dann nach Behauptung 2.15 aus Kapitel 1  $|a - b| = |a - a_n - (b - a_n)| \leq |a - a_n| + |b - a_n| < \varepsilon + \varepsilon = |a - b|$ , was ein Widerspruch ist. ■

**Bezeichnungen.** Für die Aussage „ $(a_n)_{n \in \mathbb{N}}$  konvergiert gegen  $a$ “ schreibt man auch

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

Zur Vereinfachung der Sprechweise sind ferner die folgenden Bezeichnungen zweckmäßig:

**Definition.** Sei  $a \in \mathbb{R}$ ,  $\varepsilon \in \mathbb{R}^+$ . Das Intervall  $(a - \varepsilon, a + \varepsilon)$  heißt  $\varepsilon$ -Umgebung von  $a$ . Eine Teilmenge  $U \subset \mathbb{R}$  heißt *Umgebung von  $a$* , wenn sie eine  $\varepsilon$ -Umgebung von  $a$  für geeignetes  $\varepsilon \in \mathbb{R}^+$  enthält.

„Fast alle“ heißt: alle, bis auf endlich viele.

Damit können wir auch formulieren:

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n = a &\Leftrightarrow \text{Zu jeder Umgebung } U \text{ von } a \text{ existiert } n_0 \in \mathbb{N} \\ &\text{mit } a_n \in U \text{ für alle } n \geq n_0 \\ &\Leftrightarrow \text{Jede Umgebung von } a \text{ enthält } a_n \text{ für fast alle } n. \end{aligned}$$

**Bemerkungen.** (a) Aus  $\lim_{n \rightarrow \infty} a_n = a$  und  $b_n = a_n$  für fast alle  $n$  folgt  $\lim_{n \rightarrow \infty} b_n = a$ .

(b) Aus  $\lim_{n \rightarrow \infty} a_n = a$  folgt  $\lim_{n \rightarrow \infty} a_{n+k} = a$  für  $k \in \mathbb{N}$ , und umgekehrt.

Manchen Folgen kann man schon aufgrund des folgenden Satzes ansehen, dass sie nicht konvergieren:

**1.2 Satz.** *Jede konvergente Folge ist beschränkt.*

*Beweis.* Sei  $(a_n)_{n \in \mathbb{N}}$  eine konvergente Folge mit Grenzwert  $a$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < 1$  für  $n \geq n_0$ . Für  $n \geq n_0$  gilt also

$$|a_n| = |a_n - a + a| \leq |a_n - a| + |a| < 1 + |a|.$$

Mit  $c := \max\{|a_1|, \dots, |a_{n_0-1}|, 1 + |a|\}$  gilt  $|a_n| \leq c$  für alle  $n \in \mathbb{N}$ . ■

**Beispiele.** (1) Sei  $a_n = 1/n$  für  $n \in \mathbb{N}$ .

**Behauptung.**  $\lim_{n \rightarrow \infty} a_n = 0$ .

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Nach Behauptung 4.3 in Kapitel 1 gibt es ein  $n_0 \in \mathbb{N}$  mit  $1/n_0 < \varepsilon$ . Für  $n \geq n_0$  gilt dann

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon. \quad \blacksquare$$

(2) Sei  $a_n = (-1)^n$  für  $n \in \mathbb{N}$ .

**Behauptung.**  $(a_n)_{n \in \mathbb{N}}$  divergiert.

*Beweis.* Angenommen,  $\lim_{n \rightarrow \infty} a_n = a$ . Dann existiert ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < 1$  für  $n \geq n_0$ . Für beliebiges  $n \geq n_0$  gilt dann

$$2 = |a_{n+1} - a_n| = |a_{n+1} - a - (a_n - a)| \leq |a_{n+1} - a| + |a_n - a| < 2,$$

was ein Widerspruch ist. ■

(3) Sei  $a_n = \sqrt{1 + \frac{2}{n}}$  für  $n \in \mathbb{N}$ .

**Behauptung.**  $\lim_{n \rightarrow \infty} a_n = 1$ .

*Beweis.* Zu gegebenem  $\varepsilon \in \mathbb{R}^+$  wähle  $n_0 \in \mathbb{N}$  mit  $n_0 > \frac{2}{\varepsilon}$ . Mithilfe der Binomischen Formel  $(a+b)(a-b) = a^2 - b^2$  gilt dann für  $n \geq n_0$

$$\left| \sqrt{1 + \frac{2}{n}} - 1 \right| = \frac{\frac{2}{n}}{\sqrt{1 + \frac{2}{n}} + 1} < \frac{2}{n} < \varepsilon. \quad \blacksquare$$

(4) Sei  $a_n = b^n$  für  $n \in \mathbb{N}$  ( $b \in \mathbb{R}$  gegeben).

1. Fall:  $|b| < 1$ . Aus Behauptung 4.11 aus Kapitel 1 folgt  $\lim_{n \rightarrow \infty} b^n = 0$ .
2. Fall:  $|b| > 1$ . Die Folge  $(b^n)_{n \in \mathbb{N}}$  ist wegen Behauptung 4.11 aus Kapitel 1 nicht beschränkt, also nach Satz 1.2 divergent.
3. Fall:  $b = 1$ . trivial.
4. Fall:  $b = -1$ . divergent (s.o.).

Der folgende Satz faßt einige wichtige Rechenregeln für Grenzwerte zusammen.

**1.3 Satz.** Seien  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  konvergente Folgen mit  $\lim_{n \rightarrow \infty} a_n = a$  und  $\lim_{n \rightarrow \infty} b_n = b$ . Dann sind die Folgen  $(a_n + b_n)_{n \in \mathbb{N}}$ ,  $(a_n b_n)_{n \in \mathbb{N}}$ ,  $(\lambda a_n)_{n \in \mathbb{N}}$  mit  $\lambda \in \mathbb{R}$  konvergent, und es gilt

$$\lim_{n \rightarrow \infty} (a_n + b_n) = a + b, \quad \lim_{n \rightarrow \infty} (a_n b_n) = ab, \quad \lim_{n \rightarrow \infty} (\lambda a_n) = \lambda a.$$

Ist  $a \neq 0$ , so gibt es ein  $m \in \mathbb{N}$  mit  $a_n \neq 0$  für  $n \geq m$ , und die Folge  $(1/a_n)_{n \in \mathbb{N}, n \geq m}$  ist konvergent gegen  $1/a$ .

*Beweis.* (1)  $a_n + b_n$

Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < \varepsilon/2$  für  $n \geq n_0$  und ein  $n_1 \in \mathbb{N}$  mit  $|b_n - b| < \varepsilon/2$  für  $n \geq n_1$ . Für  $n \geq \max\{n_0, n_1\}$  gilt also

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

(2)  $a_n b_n$

Sei  $\varepsilon \in \mathbb{R}^+$ . Nach Satz 1.2 existiert ein  $c \in \mathbb{R}^+$  mit  $|b_n| \leq c$  für  $n \in \mathbb{N}$ . Ferner existiert ein  $n_0 \in \mathbb{N}$  mit

$$|a_n - a| < \frac{\varepsilon}{c + |a|}, \quad |b_n - b| < \frac{\varepsilon}{c + |a|} \quad \text{für } n \geq n_0.$$

Für alle  $n \geq n_0$  gilt also

$$\begin{aligned} |a_n b_n - ab| &= |a_n b_n - ab_n + ab_n - ab| \leq |a_n - a||b_n| + |a||b_n - b| \\ &< \frac{\varepsilon}{c + |a|}c + |a|\frac{\varepsilon}{c + |a|} = \varepsilon. \end{aligned}$$

(3)  $\lambda a_n$ . Folgt aus (2), wenn  $b_n = \lambda$  gesetzt wird.

(4)  $1/a_n$

Sei also  $a \neq 0$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < \min\{\frac{1}{2}|a|, \frac{1}{2}|a|^2\varepsilon\}$  für  $n \geq n_0$ . Für  $n \geq n_0$  gilt also

$$|a| = |a - a_n + a_n| \leq |a - a_n| + |a_n| < \frac{1}{2}|a| + |a_n|,$$

folglich  $|a_n| > \frac{1}{2}|a|$ , insbesondere  $a_n \neq 0$ , und daher

$$\left| \frac{1}{a_n} - \frac{1}{a} \right| = \frac{|a - a_n|}{|a_n||a|} < \frac{\frac{1}{2}|a|^2\varepsilon}{\frac{1}{2}|a|^2} = \varepsilon. \quad \blacksquare$$

**Beispiel.**  $\lim_{n \rightarrow \infty} \frac{5n^3 + 2n^2 + 1}{2n^3 - 15n} = ?$  Es gilt

$$\frac{5n^3 + 2n^2 + 1}{2n^3 - 15n} = \frac{5 + \frac{2}{n} + \frac{1}{n^3}}{2 - \frac{15}{n^2}} \rightarrow \frac{5 + 0 + 0}{2 - 0} = \frac{5}{2}$$

für  $n \rightarrow \infty$ .

**1.4 Satz.** Sind  $(a_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \in \mathbb{N}}$  konvergente Folgen mit  $a_n \leq b_n$  für  $n \in \mathbb{N}$ , so ist  $\lim a_n \leq \lim b_n$ .

*Beweis.* Angenommen, es wäre  $a := \lim a_n > \lim b_n =: b$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit

$$|a_n - a| < \frac{1}{2}(a - b) \quad \text{und} \quad |b_n - b| < \frac{1}{2}(a - b) \quad \text{für } n \geq n_0.$$

Für solches  $n$  folgt

$$\left. \begin{array}{l} a - a_n < \frac{1}{2}(a - b) \Rightarrow a_n > \frac{1}{2}(a + b) \\ b_n - b < \frac{1}{2}(a - b) \Rightarrow b_n < \frac{1}{2}(a + b) \end{array} \right\} \Rightarrow a_n > b_n, \text{ q.e.a.}$$

■

**Bemerkung.** Achtung! Auch aus der schärferen Voraussetzung  $a_n < b_n$  für  $n \in \mathbb{N}$  folgt nur  $\lim a_n \leq \lim b_n$ .

Wir ergänzen die Definition der Konvergenz noch durch die folgende, manchmal bequeme Verabredung.

**Definition.** Die Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *bestimmt divergent gegen*  $\infty$ , geschrieben

$$\lim_{n \rightarrow \infty} a_n = \infty,$$

wenn zu jedem  $c \in \mathbb{R}$  ein  $n_0 \in \mathbb{N}$  existiert mit

$$a_n > c \quad \text{für alle } n \geq n_0.$$

**Beispiele.** (1) Aus  $\lim a_n = 0$  und  $a_n > 0$  für  $n \in \mathbb{N}$  folgt  $\lim 1/a_n = \infty$ .

(2) Sei  $a_n = \sqrt{1 + \frac{2}{n}} - 1$  und  $b_n = n$ . Dann wissen wir, dass  $b_n$  bestimmt divergent gegen  $\infty$  ist, d.h.  $\lim_{n \rightarrow \infty} b_n = \infty$ , und aus dem vorletzten Beispiel, dass  $\lim_{n \rightarrow \infty} a_n = 0$ . Also können wir Satz 1.3 auf die Folge  $(a_n b_n)_{n \in \mathbb{N}}$  nicht anwenden, insbesondere wissen wir auch nicht was  $0 \cdot \infty$  ist. Allerdings gilt, mithilfe der Rechnung aus obigen Beispiel,

$$a_n b_n = \frac{\frac{2}{n}}{\sqrt{1 + \frac{2}{n}} + 1} n = \frac{2}{\left(\sqrt{1 + \frac{2}{n}} - 1\right) + 2} \rightarrow \frac{2}{0 + 2} = 1$$

für  $n \rightarrow \infty$ .

### Kriterien für Konvergenz

Häufig wird eine Folge vorliegen, von der man Konvergenz vermutet, ohne den eventuell existierenden Grenzwert zu kennen. Man benötigt daher Kriterien für Konvergenz, in denen nicht explizit der Grenzwert vorkommt.

**1.5 Satz.** *Jede monotone, beschränkte Folge in  $\mathbb{R}$  konvergiert.*

*Beweis.* Sei  $(a_n)_{n \in \mathbb{N}}$  eine beschränkte Folge, etwa monoton wachsend. Nach dem Vollständigkeitsaxiom hat die Menge  $\{a_n | n \in \mathbb{N}\}$  eine kleinste obere Schranke  $s$ . Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Es gibt ein  $n_0 \in \mathbb{N}$  mit  $a_{n_0} > s - \varepsilon$ , denn andernfalls wäre  $s - \varepsilon$  obere Schranke. Da  $(a_n)_{n \in \mathbb{N}}$  monoton wachsend ist, gilt  $a_n \geq a_{n_0} > s - \varepsilon$  für  $n \geq n_0$ . Da  $s$  obere Schranke ist, gilt außerdem  $a_n \leq s$ , also  $|a_n - s| = s - a_n < \varepsilon$  für  $n \geq n_0$ . Somit ist  $\lim_{n \rightarrow \infty} a_n = s$ . ■

Diese hinreichende Konvergenzbedingung ist sehr nützlich, aber natürlich nur begrenzt anwendbar. Für feinere Konvergenzbetrachtungen sind die folgenden Begriffsbildungen zweckmäßig.

**Definition.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge und  $(n_k)_{k \in \mathbb{N}}$  eine streng monoton wachsende Folge natürlicher Zahlen. Dann heißt die Folge

$$(a_{n_k})_{k \in \mathbb{N}} = (a_{n_1}, a_{n_2}, a_{n_3}, \dots)$$

Teilfolge der Folge  $(a_n)_{n \in \mathbb{N}}$ .

Aus der Definition folgt unmittelbar: Konvergiert eine Folge gegen  $a$ , so konvergiert auch jede ihrer Teilfolgen gegen  $a$ .

**Definition.** Eine Zahl heißt *Häufungswert* (oder *Häufungspunkt*) einer Folge, wenn sie Grenzwert einer Teilfolge ist.

**Beispiel.**  $a_n := (-1)^n + 1/n$ . Die Teilfolge  $(a_{2k})_{k \in \mathbb{N}}$  konvergiert gegen 1, die Teilfolge  $(a_{2k+1})_{k \in \mathbb{N}}$  gegen  $-1$ . Diese beiden Zahlen sind also Häufungswerte der Folge; andere Häufungswerte gibt es offenbar nicht.

Der folgende Satz (das Beste, was man als „Umkehrung“ von Satz 1.2 aussagen kann), ist sehr wichtig und wird an vielen Stellen benutzt.

**1.6 Satz** (von Bolzano-Weierstraß). *Jede beschränkte Folge in  $\mathbb{R}$  besitzt eine konvergente Teilfolge.*

*Beweis.* Sei  $(x_n)_{n \in \mathbb{N}}$  eine beschränkte Folge. Wir definieren zunächst rekursiv eine Folge  $(J_n)_{n \in \mathbb{N}_0}$  abgeschlossener Intervalle  $J_n = [a_n, b_n]$  derart, dass für  $n \in \mathbb{N}$  gilt:

- (a)  $J_n \subseteq J_{n-1}$ ,
- (b) für unendlich viele  $k \in \mathbb{N}$  gilt  $x_k \in [a_n, b_n]$ ,
- (c)  $b_n - a_n = 2^{-n}(b_0 - a_0)$ .

Hierzu werde  $[a_0, b_0]$  so gewählt, dass  $x_k \in [a_0, b_0]$  für alle  $k \in \mathbb{N}$  gilt (das ist möglich, da die Folge beschränkt ist). Sei  $[a_n, b_n]$  schon definiert, so dass (a), (b), (c) erfüllt sind. Wir halbieren das Intervall  $[a_n, b_n]$  durch seinen Mittelpunkt  $z = (a_n + b_n)/2$ . Wenigstens eines der Intervalle  $[a_n, z]$ ,  $[z, b_n]$  enthält  $x_k$  für unendlich viele  $k$ ; ein solches Teilintervall nehmen wir als  $[a_{n+1}, b_{n+1}] = J_{n+1}$ . Dann sind (a), (b), (c) auch für  $J_{n+1}$  erfüllt. Damit ist die Folge  $(J_n)_{n \in \mathbb{N}}$  rekursiv definiert.

Nun definieren wir rekursiv eine Teilfolge  $(x_{k_n})_{n \in \mathbb{N}}$  mit  $x_{k_n} \in [a_n, b_n]$  für  $n \in \mathbb{N}$ . Wähle  $k_1$  so, dass  $x_{k_1} \in [a_1, b_1]$  gilt. Sei  $k_n$  mit  $x_{k_n} \in [a_n, b_n]$  schon definiert. Das Intervall  $[a_{n+1}, b_{n+1}]$  enthält  $x_k$  für unendlich viele  $k$ , also ist die Menge  $\{k \in \mathbb{N} \mid k > k_n \text{ und } x_k \in [a_{n+1}, b_{n+1}]\}$  nicht leer. Ihr kleinstes Element nehmen wir als  $k_{n+1}$ . Damit ist die Folge rekursiv definiert.

Nach dem Intervallschachtelungsprinzip 2.3 aus Kapitel 2 gibt es eine reelle Zahl  $x \in \bigcap_{n \in \mathbb{N}} J_n$ . Für alle  $n \in \mathbb{N}$  haben wir  $x_{k_n} \in [a_n, b_n]$  und  $x \in [a_n, b_n]$ , woraus folgt

$$|x_{k_n} - x| \leq b_n - a_n = \frac{b_0 - a_0}{2^n}.$$

Daraus folgt  $\lim_{n \rightarrow \infty} x_{k_n} = x$ . ■

Konvergiert die Folge  $(a_n)_{n \in \mathbb{N}}$  gegen  $a$ , so auch (wie früher bemerkt) jede Teilfolge, also ist  $a$  der einzige Häufungswert. Für beschränkte Folgen läßt sich dies umkehren:

**1.7 Satz.** *Hat eine beschränkte Folge genau einen Häufungswert, so konvergiert sie gegen diesen Häufungswert.*

*Beweis.* Sei  $(a_n)_{n \in \mathbb{N}}$  eine beschränkte Folge und  $a$  ihr einziger Häufungswert. Angenommen, die Folge konvergiere nicht gegen  $a$ . Das bedeutet:

$$\exists \varepsilon \in \mathbb{R}^+ \quad \forall n_0 \in \mathbb{N} \quad \exists n \geq n_0 : |a_n - a| \geq \varepsilon.$$

Wir nehmen ein solches  $\varepsilon > 0$ . Es gibt dann ein  $n_1 \in \mathbb{N}$  mit  $|a_{n_1} - a| \geq \varepsilon$ . Ist  $n_k$  schon definiert, so ist die Menge  $\{n \in \mathbb{N} \mid n \geq n_k + 1 \text{ und } |a_n - a| \geq \varepsilon\}$

nicht leer; sei  $n_{k+1}$  ihr kleinstes Element. Damit ist rekursiv eine Teilfolge  $(a_{n_k})_{k \in \mathbb{N}}$  definiert mit  $|a_{n_k} - a| \geq \varepsilon$  für alle  $k \in \mathbb{N}$ . Da sie beschränkt ist, besitzt sie nach Satz 1.6 eine konvergente Teilfolge. Für deren Grenzwert  $b$  gilt  $|b - a| \geq \varepsilon$ , also hat die ursprüngliche Folge zwei verschiedene Häufungswerte, was ein Widerspruch ist. ■

Nun sind wir in der Lage, ein notwendiges und hinreichendes Kriterium für die Konvergenz einer Folge aufzustellen, das nicht explizit auf den Grenzwert Bezug nimmt. Eine notwendige Bedingung für Konvergenz liegt nahe, indem man den Grenzwert eliminiert: Die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergiere gegen  $a$ . Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|a_n - a| < \varepsilon/2$  für  $n \geq n_0$ . Für alle  $m, n \geq n_0$  gilt also

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

Diese für die Konvergenz notwendige Eigenschaft erweist sich auch als hinreichend. Zwecks größerer Klarheit und vor allem im Hinblick auf spätere Verallgemeinerungen ist eine Definition angebracht:

**Definition.** Die Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *Cauchy-Folge*, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $n_0 \in \mathbb{N}$  existiert mit  $|a_m - a_n| < \varepsilon$  für alle  $m, n \geq n_0$ .

**1.8 Satz** (Konvergenzkriterium von Cauchy). *Eine Folge in  $\mathbb{R}$  ist genau dann konvergent, wenn sie eine Cauchy-Folge ist.*

*Beweis.* Dass jede konvergente Folge eine Cauchy-Folge ist, wurde bereits gezeigt. Sei nun  $(a_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge.

**Behauptung.** 1.  $(a_n)_{n \in \mathbb{N}}$  ist beschränkt.

*Beweis.* Da  $(a_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge ist, existiert ein  $n_0 \in \mathbb{N}$  so, dass  $|a_m - a_n| < 1$  für  $m, n \geq n_0$ . Für alle  $n \geq n_0$  gilt also  $|a_n| = |a_n - a_{n_0} + a_{n_0}| \leq |a_n - a_{n_0}| + |a_{n_0}| < 1 + |a_{n_0}|$ . Mit  $c := \max\{|a_1|, \dots, |a_{n_0-1}|, 1 + |a_{n_0}|\}$  gilt also  $|a_n| \leq c$  für alle  $n \in \mathbb{N}$ . Somit ist  $(a_n)_{n \in \mathbb{N}}$  beschränkt. ■

Nun folgt aus dem Satz von Bolzano-Weierstraß 1.6, dass eine konvergente Teilfolge  $(a_{n_k})_{k \in \mathbb{N}}$  existiert. Sei  $a$  ihr Grenzwert.

**Behauptung.** 2.  $\lim_{n \rightarrow \infty} a_n = a$ .

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Da  $(a_n)_{n \in \mathbb{N}}$  Cauchy-Folge ist, existiert ein  $n_0 \in \mathbb{N}$  mit

$$|a_n - a_m| < \frac{\varepsilon}{2} \quad \text{für } m, n \geq n_0.$$

Wegen  $\lim_{k \rightarrow \infty} a_{n_k} = a$  existiert ein  $j \in \mathbb{N}$  mit  $n_j \geq n_0$  und

$$|a_{n_j} - a| < \frac{\varepsilon}{2}.$$

Für alle  $n \geq n_0$  gilt also

$$\begin{aligned} |a_n - a| &= |a_n - a_{n_j} + a_{n_j} - a| \leq |a_n - a_{n_j}| + |a_{n_j} - a| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Damit ist  $a_n \rightarrow a$  für  $n \rightarrow \infty$  gezeigt. ■

Damit ist Satz 1.8 gezeigt. ■

Wir gehen noch auf eine Begriffsbildung ein, die bei feineren Konvergenzuntersuchungen bzw. bei nicht konvergenten Folgen von Nutzen ist. Zunächst sei daran erinnert, dass wir für eine Teilmenge  $M \subset \mathbb{R}$  die kleinste obere Schranke, falls sie existiert, auch mit

$$\sup M = \textit{Supremum} \text{ von } M$$

und die größte untere Schranke mit

$$\inf M = \textit{Infimum} \text{ von } M$$

bezeichnet haben.

**Definition.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine beschränkte Folge in  $\mathbb{R}$ . Dann heißt

$$\limsup_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} (\sup\{a_k \mid k \geq n\})$$

der *Limes superior* der Folge, und

$$\liminf_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} (\inf\{a_k \mid k \geq n\})$$

heißt *Limes inferior* der Folge.

Zu dieser Definition ist eine Erläuterung erforderlich. Setzt man

$$b_n := \sup\{a_k \mid k \geq n\} \quad \text{für } n \in \mathbb{N},$$

so ist die Folge  $(b_n)_{n \in \mathbb{N}}$  monoton fallend; da sie beschränkt ist, existiert nach Satz 1.5 der Grenzwert  $\lim_{n \rightarrow \infty} b_n$ . Man kann also in der Tat wie oben definieren. Entsprechendes gilt für  $\liminf$ .

**Beispiel.**  $a_n = (-1)^n(1 + \frac{1}{n})$ . Es ist

$$\sup\{a_k | k \geq n\} = \begin{cases} 1 + \frac{1}{n}, & \text{falls } n \text{ gerade,} \\ 1 + \frac{1}{n+1}, & \text{falls } n \text{ ungerade.} \end{cases}$$

Also ist  $\limsup_{n \rightarrow \infty} a_n = 1$ . Ferner ist

$$\inf\{a_k | k \geq n\} = \begin{cases} -(1 + \frac{1}{n}), & \text{falls } n \text{ ungerade,} \\ -(1 + \frac{1}{n+1}), & \text{falls } n \text{ gerade,} \end{cases}$$

also  $\liminf_{n \rightarrow \infty} a_n = -1$ .

Der folgende Satz enthält eine äquivalente Definition, die oft besser zu handhaben ist als die ursprüngliche.

**1.9 Satz.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine beschränkte Folge und  $a \in \mathbb{R}$ . Dann gilt:

(1)  $\limsup a_n \leq a \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+$  ist die Menge  $\{k \in \mathbb{N} | a_k \geq a + \varepsilon\}$  endlich.

(2)  $\limsup a_n \geq a \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+$  ist die Menge  $\{k \in \mathbb{N} | a_k \geq a - \varepsilon\}$  unendlich.

Entsprechendes gilt für  $\liminf$ .

*Beweis.* (1) „ $\Rightarrow$ “: Sei  $\limsup a_n \leq a$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Wäre die Menge  $\{k \in \mathbb{N} | a_k \geq a + \varepsilon\}$  unendlich, so wäre

$$\sup\{a_k | k \geq n\} \geq a + \varepsilon \quad \text{für alle } n \in \mathbb{N},$$

also  $\limsup a_n \geq a + \varepsilon > a$ , was ein Widerspruch ist.

„ $\Leftarrow$ “: Sei  $\{k \in \mathbb{N} | a_k \geq a + \varepsilon\}$  endlich für jedes  $\varepsilon \in \mathbb{R}^+$ . Dann gibt es zu gegebenem  $\varepsilon \in \mathbb{R}^+$  ein  $n_0 \in \mathbb{N}$  so, dass  $a_k < a + \varepsilon$  für  $n \geq n_0$ , also

$$\sup\{a_k | k \geq n\} \leq a + \varepsilon \quad \text{für } n \geq n_0.$$

Es folgt  $\limsup a_n \leq a + \varepsilon$ ; da  $\varepsilon \in \mathbb{R}^+$  beliebig war, folgt  $\limsup a_n \leq a$ .

(2) „ $\Rightarrow$ “: Sei  $\limsup a_n \geq a$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Wäre die Menge  $\{k \in \mathbb{N} | a_k \geq a - \varepsilon\}$  endlich, so existierte ein  $n_0 \in \mathbb{N}$  so, dass  $a_k < a - \varepsilon$  für  $n \geq n_0$ , also

$$\sup\{a_k | k \geq n\} \leq a - \varepsilon \quad \text{für } n \geq n_0.$$

Daraus folgt  $\limsup a_n \leq a - \varepsilon < a$ , was ein Widerspruch ist.

„ $\Leftarrow$ “: Sei  $\{k \in \mathbb{N} \mid a_k \geq a - \varepsilon\}$  unendlich für jedes  $\varepsilon \in \mathbb{R}^+$ . Dann gilt für jedes  $\varepsilon \in \mathbb{R}^+$

$$\sup\{a_k \mid k \geq n\} \geq a - \varepsilon \quad \text{für alle } n \in \mathbb{N},$$

also ist  $\limsup a_n \geq a - \varepsilon$ . Da  $\varepsilon \in \mathbb{R}^+$  beliebig war, folgt  $\limsup a_n \geq a$ . ■

Der folgende Satz macht die Bedeutung von Limes superior und Limes inferior deutlicher und erklärt zugleich die Wahl dieser Bezeichnungen.

**1.10 Satz.** *Sei  $(a_n)_{n \in \mathbb{N}}$  eine beschränkte Folge. Dann ist  $\limsup a_n$  der größte und  $\liminf a_n$  der kleinste Häufungswert dieser Folge.*

*Beweis.* Es genügt, die Behauptung für  $\limsup$  zu beweisen. Sei  $H$  die Menge der Häufungswerte der Folge  $(a_n)_{n \in \mathbb{N}}$ . Nach Satz 1.6 ist  $H \neq \emptyset$ , und  $H$  ist beschränkt. Setze  $s := \sup H$  und  $\limsup a_n =: a$ .

Nach Satz 1.9 (1) ist für jedes  $\varepsilon \in \mathbb{R}^+$  die Menge  $\{k \in \mathbb{N} \mid a_k \geq a + \varepsilon\}$  endlich. Also kann keine Zahl  $\geq a + \varepsilon$  Grenzwert einer Teilfolge (d.h. Häufungswert) sein. Für alle  $h \in H$  gilt also  $h \leq a$ , folglich ist  $s \leq a$ .

**Behauptung.**  $a \in H$  [Daraus folgt dann  $a \leq s$ , also  $a = s$ .]

*Beweis.* Wir konstruieren rekursiv eine Teilfolge, die gegen  $a$  konvergiert. Setze  $k_1 = 1$ . Sei  $k_{n-1}$  schon definiert. Nach Satz 1.9 (2) ist die Menge

$$\left\{ k \in \mathbb{N} \mid a_k \geq a - \frac{1}{n} \right\}$$

unendlich, also existiert ein  $k_n \in \mathbb{N}$  mit  $k_n > k_{n-1}$  und  $a_{k_n} \geq a - 1/n$ . Damit ist die Teilfolge  $(a_{k_n})_{n \in \mathbb{N}}$  definiert. Für alle  $n \in \mathbb{N}$  gilt (wegen  $k_n \geq n$ )

$$\sup\{a_k \mid k \geq n\} \geq a_{k_n} \geq a - \frac{1}{n}.$$

Für  $n \rightarrow \infty$  konvergieren beide Seiten gegen  $a$ , woraus  $\lim_{n \rightarrow \infty} a_{k_n} = a$  folgt. ■

Damit ist Satz 1.10 gezeigt. ■

Wir können nun ein weiteres notwendiges und hinreichendes Kriterium für die Konvergenz einer Folge aufstellen.

**1.11 Satz.** *Die Folge  $(a_n)_{n \in \mathbb{N}}$  konvergiert genau dann, wenn sie beschränkt ist und wenn  $\limsup a_n = \liminf a_n$  ist.*

Das folgt sofort aus Satz 1.2, 1.7, 1.10.

### 3.2 Reihen

Unter Verwendung des Grenzwertbegriffes kann man auch für gewisse Folgen  $(a_n)_{n \in \mathbb{N}}$  die „unendliche Summe“  $a_1 + a_2 + \dots$  erklären. Solche unendlichen Reihen sind wichtig, da man spezielle reelle Zahlen oder Funktionen häufig durch Reihen ausdrücken kann.

**Definition.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge in  $\mathbb{R}$ . Man nennt die Zahl

$$s_n := \sum_{k=1}^n a_k \quad (n \in \mathbb{N})$$

die  $n$ -te *Partialsumme* dieser Folge. Die Folge  $(s_n)_{n \in \mathbb{N}}$  heißt *unendliche Reihe* und wird mit

$$\sum_{k=1}^{\infty} a_k$$

bezeichnet. Ist die Folge  $(s_n)_{n \in \mathbb{N}}$  konvergent, so wird ihr Grenzwert ebenfalls mit  $\sum_{k=1}^{\infty} a_k$  bezeichnet. In diesem Fall sagt man, dass die unendliche Reihe *konvergiert*.

Hiernach bedeutet  $\sum_{k=1}^{\infty} a_k$  also zweierlei; das führt erfahrungsgemäß aber nicht zu Mißverständnissen. Nach Definition ist

$$\sum_{k=1}^{\infty} a_k = s \Leftrightarrow \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k = s.$$

**Bemerkung.** Es ist klar, was unter  $\sum_{k=m}^{\infty} a_k$  (mit  $m \in \mathbb{Z}$ ) zu verstehen ist.

**Beispiel (unendliche geometrische Reihe).** Sei  $q \in \mathbb{R}$ ,  $|q| < 1$ . Nach Behauptung 4.15 aus Kapitel 1 gilt

$$\sum_{k=0}^n q^k = \frac{1 - q^{n+1}}{1 - q},$$

also

$$\left| \sum_{k=0}^n q^k - \frac{1}{1 - q} \right| = \frac{|q|^{n+1}}{1 - q} \rightarrow 0 \quad \text{für } n \rightarrow \infty,$$

wobei wir Behauptung 4.11 aus Kapitel 1 benutzt haben. Folglich ist

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1 - q}.$$

Konvergenz einer Reihe bedeutet nach Definition nichts anderes als die Konvergenz der zugehörigen Partialsummen-Folge. Die bisher für Folgenkonvergenz gewonnenen Ergebnisse übertragen sich also unmittelbar auf Ergebnisse über Konvergenz von Reihen. Wir wollen diese Resultate im folgenden zusammenstellen. Da es sich nur um Umformulierungen handelt, erübrigt es sich, auf die Beweise einzugehen.

Aus Satz 1.3 und Satz 1.4 folgt:

**2.1 Satz.** Seien  $\sum_{k=1}^{\infty} a_k$ ,  $\sum_{k=1}^{\infty} b_k$  konvergente Reihen. Dann ist die Reihe  $\sum_{k=1}^{\infty} (a_k + b_k)$  konvergent, und es gilt

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

Gilt  $a_k \leq b_k$  für  $k \in \mathbb{N}$ , so ist  $\sum_{k=1}^{\infty} a_k \leq \sum_{k=1}^{\infty} b_k$ .

Aus Satz 1.5 folgt:

**2.2 Satz.** Jede Reihe  $\sum_{k=1}^{\infty} a_k$  mit  $a_k \geq 0$  für alle  $k \in \mathbb{N}$  ist konvergent oder bestimmt divergent.

**Beispiel.** An dieser Stelle können wir die gewohnte Praxis rechtfertigen, reelle Zahlen durch (im Prinzip unendliche) Dezimalbrüche darzustellen. Zum Beispiel bedeutet die Schreibweise

$$\pi = 3,141592653\dots$$

ja verabredungsgemäß nichts anderes, als dass

$$\pi = 3 + \frac{1}{10} + \frac{4}{100} + \frac{1}{1000} + \frac{5}{10000} + \dots$$

ist, also dass  $\pi$  durch eine gewisse Reihe mit nichtnegativen Gliedern dargestellt werden kann.

Insbesondere stellt also jeder „unendliche Dezimalbruch“ eine reelle Zahl dar. Es ist nicht schwer zu zeigen, dass umgekehrt jede reelle Zahl als Dezimalbruch dargestellt werden kann (man bestimme die  $a_k$  rekursiv).

Wir fahren fort, Ergebnisse für Folgen auf Reihen umzuformulieren. Ist

$$s_n = \sum_{k=1}^n a_k,$$

so ist für  $m < n$

$$|s_n - s_m| = \left| \sum_{k=m+1}^n a_k \right|.$$

Aus Satz 1.8 erhalten wir also sofort die folgende wichtige Aussage.

**2.3 Satz** (Konvergenzkriterium von Cauchy). *Die Reihe  $\sum_{k=1}^{\infty} a_k$  ist genau dann konvergent, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $n_0 \in \mathbb{N}$  existiert mit*

$$\left| \sum_{k=m}^{m+p} a_k \right| < \varepsilon \quad \text{für } m \geq n_0 \text{ und } p \in \mathbb{N}_0.$$

Da die Bedingung insbesondere für  $p = 0$  erfüllt sein muß, haben wir:

**2.4 Satz.** *Notwendig für die Konvergenz der Reihe  $\sum_{k=1}^{\infty} a_k$  ist die Bedingung  $\lim_{k \rightarrow \infty} a_k = 0$ .*

Diese Bedingung ist aber nicht hinreichend.

**Beispiel.** Die „harmonische Reihe“  $\sum_{k=1}^{\infty} \frac{1}{k}$  ist bestimmt divergent. Es ist nämlich

$$1 + \left(\frac{1}{2}\right) + \underbrace{\left(\frac{1}{3} + \frac{1}{4}\right)}_{> 2 \cdot \frac{1}{4} = \frac{1}{2}} + \underbrace{\left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right)}_{> 4 \cdot \frac{1}{8} = \frac{1}{2}} + \underbrace{\left(\frac{1}{9} + \dots + \frac{1}{16}\right)}_{> 8 \cdot \frac{1}{16} = \frac{1}{2}} + \dots$$

u.s.w., also  $s_{2^n} \geq 1 + n \cdot \frac{1}{2}$ , woraus die Behauptung folgt.

**Beispiel.** Für  $\alpha > 1$  ist  $\sum_{k=1}^{\infty} \frac{1}{k^\alpha}$  konvergent. [Dabei sei  $\alpha \in \mathbb{N}$ , denn sonst ist bis jetzt  $k^\alpha$  noch nicht erklärt. Wenn aber später  $k^\alpha$  für  $\alpha \in \mathbb{R}$  erklärt sein wird, bleiben Behauptung und Beweis für  $\alpha > 1$  richtig.] Sei  $n \in \mathbb{N}$ . Wähle  $m \in \mathbb{N}$  mit  $n \leq 2^{m+1} - 1$ . Dann gilt

$$\begin{aligned} s_n &= \sum_{k=1}^n \frac{1}{k^\alpha} \leq \sum_{k=1}^{2^{m+1}-1} \frac{1}{k^\alpha} = 1 + \left(\frac{1}{2^\alpha} + \frac{1}{3^\alpha}\right) + \dots + \left(\sum_{k=2^m}^{2^{m+1}-1} \frac{1}{k^\alpha}\right) \\ &\leq \sum_{\ell=0}^m 2^\ell \frac{1}{(2^\ell)^\alpha} = \sum_{\ell=0}^m (2^{1-\alpha})^\ell < \sum_{\ell=0}^{\infty} (2^{1-\alpha})^\ell = \frac{1}{1-2^{1-\alpha}}. \end{aligned}$$

Dabei wurde  $2^{1-\alpha} < 1$  (wegen  $\alpha > 1$ ) benutzt. Die Konvergenz folgt jetzt aus Satz 2.2.

In Fällen, wo die Voraussetzung von Satz 2.2 nicht erfüllt ist, ist häufig das folgende Kriterium von Nutzen.

**2.5 Satz** (Leibniz-Kriterium für alternierende Reihen). *Ist  $(a_n)_{n \in \mathbb{N}_0}$  eine monoton fallende Nullfolge, so ist die Reihe*

$$\sum_{k=0}^{\infty} (-1)^k a_k$$

*konvergent.*

*Beweis.* Setze  $s_n := \sum_{k=0}^n (-1)^k a_k$ . Es gilt  $s_{2n+2} - s_{2n} = -a_{2n+1} + a_{2n+2} \leq 0$ , also ist die Folge  $(s_{2n})_{n \in \mathbb{N}}$  monoton fallend. Analog ist die Folge  $(s_{2n+1})_{n \in \mathbb{N}}$  monoton wachsend. Wegen  $s_{2n+1} - s_{2n} = -a_{2n+1} \leq 0$  gilt

$$s_{2n} \geq s_{2n+1} \geq s_1 \quad \text{und} \quad s_{2n+1} \leq s_{2n} \leq s_0.$$

Nach Satz 1.5 existieren also

$$a := \lim_{n \rightarrow \infty} s_{2n} \quad \text{und} \quad b := \lim_{n \rightarrow \infty} s_{2n+1}.$$

Es ist  $a - b = \lim_{n \rightarrow \infty} (s_{2n} - s_{2n+1}) = \lim_{n \rightarrow \infty} a_{2n+1} = 0$ , also  $a = b$ .

Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Es gibt ein  $n_0 \in \mathbb{N}$  mit

$$|s_{2n} - a| < \varepsilon \quad \text{für } 2n \geq n_0,$$

und es gibt ein  $n_1 \in \mathbb{N}$  mit

$$|s_{2n+1} - a| < \varepsilon \quad \text{für } 2n + 1 \geq n_1.$$

Für  $m \geq \max\{n_0, n_1\}$  gilt also

$$|s_m - a| < \varepsilon.$$

Damit ist  $\sum_{k=0}^{\infty} (-1)^k a_k = a$  gezeigt. ■

### Absolute Konvergenz

Feinere Konvergenzbetrachtungen für Reihen erfordern eine neue Begriffsbildung. Hierauf werden wir beispielshalber bei der folgenden Fragestellung geführt. Gilt für unendliche Reihen auch ein „Kommutativgesetz“? Mit anderen Worten: Hängen Konvergenz und Grenzwert einer Reihe von der Reihenfolge der Summanden ab, oder darf man die Summanden beliebig umordnen? Mit einer „Umordnung“ ist folgendes gemeint:

**Definition.** Sei  $\sum_{k=1}^{\infty} a_k$  eine Reihe und  $\tau : \mathbb{N} \rightarrow \mathbb{N}$  eine Permutation (bijektive Abbildung auf sich) von  $\mathbb{N}$ . Dann heißt die Reihe  $\sum_{k=1}^{\infty} a_{\tau(k)}$  eine *Umordnung* der ursprünglichen.

**Beispiel.** Die Reihe  $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n}$  ist nach dem Leibniz-Kriterium konvergent. Da  $\sum_{n=1}^{\infty} \frac{1}{n}$  divergiert, läßt sich leicht eine Umordnung angeben, die divergiert: Für  $n \in \mathbb{N}$  gilt

$$\frac{1}{2^n + 1} + \frac{1}{2^n + 3} + \cdots + \frac{1}{2^{n+1} - 1} > 2^{n-1} \cdot \frac{1}{2^{n+1}} = \frac{1}{4},$$

also

$$\begin{aligned} & \frac{1}{1} - \frac{1}{2} \\ & + \frac{1}{3} - \frac{1}{4} \\ & + \left( \frac{1}{5} + \frac{1}{7} \right) - \frac{1}{6} \\ & + \left( \frac{1}{9} + \frac{1}{11} + \frac{1}{13} + \frac{1}{15} \right) - \frac{1}{8} \\ & + \dots \\ & + \left( \frac{1}{2^n + 1} + \cdots + \frac{1}{2^{n+1} - 1} \right) - \frac{1}{2n + 2} \\ & \rightarrow \infty \text{ für } n \rightarrow \infty. \end{aligned}$$

Wie sich zeigt, kann etwas derartiges nicht passieren, wenn neben  $\sum a_k$  auch die Reihe  $\sum |a_k|$  konvergiert.

**Definition.** Die Reihe  $\sum_{k=1}^{\infty} a_k$  heißt *absolut konvergent*, wenn die Reihe  $\sum_{k=1}^{\infty} |a_k|$  konvergiert.

**2.6 Satz.** *Jede absolut konvergente Reihe ist konvergent.*

*Beweis.* Sei  $\sum |a_k|$  konvergent. Wir zeigen, dass  $\sum a_k$  die Voraussetzung des Konvergenzkriteriums von Cauchy erfüllt. Sei  $\varepsilon \in \mathbb{R}^+$ . Da  $\sum |a_k|$  konvergiert, existiert ein  $n_0 \in \mathbb{N}$  mit

$$\sum_{k=m}^{m+p} |a_k| < \varepsilon \quad \text{für } m \geq n_0, p \in \mathbb{N}_0.$$

Für diese  $m, p$  gilt also erst recht

$$\left| \sum_{k=m}^{m+p} a_k \right| \leq \sum_{k=m}^{m+p} |a_k| < \varepsilon.$$

Nach Satz 2.3 ist die Reihe  $\sum_{k=1}^{\infty} a_k$  konvergent. ■

Hier wurde die Ungleichung  $|b_1 + \dots + b_n| \leq |b_1| + \dots + |b_n|$  benutzt, die natürlich aus der Dreiecksungleichung durch Induktion herzuleiten ist.

**2.7 Satz** (Umordnungssatz). *Sei  $\sum_{k=1}^{\infty} a_k$  eine absolut konvergente Reihe mit Grenzwert  $a$ . Dann konvergiert auch jede Umordnung dieser Reihe gegen  $a$ .*

*Beweis.* Sei  $\tau : \mathbb{N} \rightarrow \mathbb{N}$  eine Permutation von  $\mathbb{N}$ . Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Da  $\sum |a_k|$  konvergiert, gibt es ein  $m \in \mathbb{N}$  mit

$$\sum_{k=m+1}^{\infty} |a_k| < \frac{\varepsilon}{2}.$$

Wählen wir ein  $n_0 \in \mathbb{N}$  mit  $\{1, \dots, m\} \subset \{\tau(1), \dots, \tau(n_0)\}$ , so gilt für alle  $n \geq n_0$

$$\begin{aligned} \left| \sum_{k=1}^n a_{\tau(k)} - a \right| &\leq \left| \sum_{k=1}^n a_{\tau(k)} - \sum_{k=1}^m a_k \right| + \left| \sum_{k=1}^m a_k - a \right| \\ &\leq \sum_{k=m+1}^{\infty} |a_k| + \sum_{k=m+1}^{\infty} |a_k| \\ &< \varepsilon. \end{aligned}$$

Hierbei haben wir benutzt, dass für  $\ell > m$  gilt

$$\left| \sum_{k=1}^m a_k - a \right| = \left| \sum_{k=1}^{\ell} a_k - a - \sum_{k=m+1}^{\ell} a_k \right| \leq \left| \sum_{k=1}^{\ell} a_k - a \right| + \sum_{k=m+1}^{\infty} |a_k|.$$

Der Grenzwert  $\ell \rightarrow \infty$  liefert die obige Abschätzung des zweiten Terms. ■

Nachdem schon hierdurch (und erst recht im weiteren Verlauf) die Bedeutung von absoluter Konvergenz unterstrichen wird, wollen wir Kriterien für absolute Konvergenz zusammenstellen. Am häufigsten anwendbar ist wohl das folgende:

**2.8 Satz** (Majorantenkriterium). *Seien  $\sum a_k, \sum c_k$  Reihen. Gilt*

$$|a_k| \leq c_k \quad \text{für alle } k \in \mathbb{N}$$

*und ist die Reihe  $\sum c_k$  konvergent, dann ist die Reihe  $\sum a_k$  absolut konvergent.*

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$ . Da  $\sum c_k$  konvergiert, existiert ein  $n_0 \in \mathbb{N}$  mit

$$\sum_{k=m}^{m+p} c_k < \varepsilon \quad \text{für } m \geq n_0, p \in \mathbb{N}_0.$$

Für diese  $m, p$  gilt also

$$\sum_{k=m}^{m+p} |a_k| \leq \sum_{k=m}^{m+p} c_k < \varepsilon.$$

Aus Satz 2.3 folgt die Behauptung. ■

Gilt  $|a_k| \leq c_k$  für  $k \in \mathbb{N}$ , so nennt man die Reihe  $\sum c_k$  eine *Majorante* der Reihe  $\sum a_k$ . Eine Reihe ist also absolut konvergent, wenn sie eine konvergente Majorante besitzt. Häufig brauchbare Vergleichsreihen sind die geometrische Reihe ( $c_k = q^k, q < 1$ ) und die  $\zeta$ -Reihen ( $c_k = \frac{1}{k^\alpha}$  mit einem  $\alpha > 1$ ).

Gilt dagegen  $a_k \geq c_k \geq 0$  für  $k \in \mathbb{N}_0$ , so nennt man die Reihe  $\sum c_k$  eine *Minorante* der Reihe  $\sum a_k$ . Besitzt die Reihe  $\sum a_k$  eine bestimmt divergente Minorante, so ist sie selbst bestimmt divergent. Eine manchmal brauchbare Vergleichsreihe ist die harmonische Reihe.

Wenn sich keines der obigen Kriterien unmittelbar anwenden läßt, versuche man es mit einem der folgenden.

**2.9 Satz (Quotientenkriterium).** Sei  $\sum a_k$  eine Reihe mit  $a_k \neq 0$  für  $k \in \mathbb{N}$ . Es gebe eine Zahl  $q \in \mathbb{R}$  mit  $0 < q < 1$  und

$$\left| \frac{a_{n+1}}{a_n} \right| \leq q \quad \text{für } n \in \mathbb{N}.$$

Dann ist die Reihe  $\sum a_n$  absolut konvergent.

*Beweis.* Aus der Voraussetzung folgert man durch Induktion sofort

$$|a_{n+1}| \leq q^n |a_1|.$$

Da die Reihe  $\sum q^n |a_1| = |a_1| \sum q^n$  wegen  $|q| < 1$  konvergiert, folgt die Behauptung aus dem Majorantenkriterium. ■

Natürlich genügt es, wenn beim Quotientenkriterium die Voraussetzungen  $a_n \neq 0$  und  $\left| \frac{a_{n+1}}{a_n} \right| \leq q$  für fast alle  $n$  erfüllt sind.

Besonders bequem ist das Quotientenkriterium anzuwenden, wenn es gelingt,  $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| =: a$  zu bestimmen, falls dieser Grenzwert existiert. Ist

$a < 1$ , so gilt für beliebiges  $q$  mit  $a < q < 1$  jedenfalls  $\left| \frac{a_{n+1}}{a_n} \right| < q$  für fast alle  $n$ ; die Reihe ist also absolut konvergent. Gilt  $a > 1$ , so gilt

$$\left| \frac{a_{n+1}}{a_n} \right| \geq 1 \quad \text{für fast alle } n.$$

In diesem Fall ist die Reihe divergent, da  $(a_n)_{n \in \mathbb{N}}$  keine Nullfolge ist. Im Fall  $a = 1$  kann sowohl absolute Konvergenz als auch Divergenz vorliegen (z.B.  $\sum \frac{1}{n^2}$  und  $\sum \frac{1}{n}$ ).

Die Formulierung eines weiteren nützlichen Kriteriums erfordert eine Vorbemerkung. Sind Zahlen  $n \in \mathbb{N}$  und  $a \in \mathbb{R}^+$  gegeben, so gibt es eine Zahl  $b \in \mathbb{R}^+$  mit  $b^n = a$ . Man schreibt dann  $b = \sqrt[n]{a}$  und nennt  $b$  die  $n$ -te Wurzel aus  $a$ . Sie ist eindeutig bestimmt, denn aus  $0 < b < c$  folgt  $b^n < c^n$ . Die Existenz der  $n$ -ten Wurzel könnten wir bereits an dieser Stelle leicht zeigen; wir verzichten aber darauf, da sie sich später in allgemeinerem Rahmen ergibt.

Ferner sei daran erinnert, dass wir in Abschnitt 3.1 für eine beschränkte Folge  $(a_n)_{n \in \mathbb{N}}$  den Limes superior  $\limsup_{n \rightarrow \infty} a_n$  definiert hatten. Wir ergänzen diese Definition noch:

**Definition.** Ist  $(a_n)_{n \in \mathbb{N}}$  nicht nach oben beschränkt, so schreibt man

$$\limsup_{n \rightarrow \infty} a_n = \infty.$$

Man vereinbart noch  $\infty > a$  für  $a \in \mathbb{R}$ .

**2.10 Satz (Wurzelkriterium).** Die Reihe  $\sum_{n=1}^{\infty} a_n$  ist absolut konvergent, wenn

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1$$

ist, und divergent, wenn

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1$$

ist.

*Beweis.* Sei  $\limsup \sqrt[n]{|a_n|} =: a < 1$ . Dann gibt es ein  $q < 1$  mit  $\sqrt[n]{|a_n|} \leq q$  und daher  $|a_n| \leq q^n$  für fast alle  $n$ . Aus dem Majorantenkriterium folgt die absolute Konvergenz der Reihe  $\sum a_n$ .

Ist  $\limsup \sqrt[n]{|a_n|} > 1$ , so gilt  $|a_n| \geq 1$  für unendlich viele  $n$ ; also ist  $(a_n)_{n \in \mathbb{N}}$  keine Nullfolge und daher  $\sum a_n$  divergent. ■

**Bemerkung.** Besonders bequem ist die Anwendung des Wurzelkriteriums, wenn  $\lim \sqrt[n]{|a_n|}$  existiert und bestimmt werden kann.

Man beachte, dass das Wurzelkriterium im Fall  $\limsup \sqrt[n]{|a_n|} = 1$  keine Aussage über Konvergenz oder Divergenz der Reihe  $\sum a_n$  ermöglicht.

**Beispiele.**

$$\sum \frac{n^2}{2^n} \quad (\text{Quot.-Krit.}), \quad \sum \frac{n!}{3 \cdot 5 \cdot 7 \cdots (2n+1)} \quad (\text{Quot.-Krit.}),$$

$$\sum (\sqrt[n]{n} - 1)^n \quad (\text{Wurzel-Krit. und}$$

$$b_n := (\sqrt[n]{n} - 1) \Rightarrow n = (b_n + 1)^n \geq 1 + \binom{n}{2} b_n^2$$

$$\sum \left( \frac{n}{n+1} \right)^{n^2} \quad (\text{Wurzel-Krit. und Bernoullische Ungl.})$$

Absolute Konvergenz spielt auch eine Rolle, wenn wir versuchen, zwei konvergente Reihen gliedweise zu multiplizieren. Betrachten wir zunächst endliche Summen. Es ist

$$(a_1 + \cdots + a_n)(b_1 + \cdots + b_m) = \sum a_i b_j,$$

wo über alle Paare  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$  (in beliebiger Reihenfolge) zu summieren ist. Wenn wir dies auf unendliche Reihen übertragen wollen, müssen wir für die Paare  $(i, j) \in \mathbb{N} \times \mathbb{N}$  eine Reihenfolge festlegen. Es ist übersichtlich, die Paare  $(i, j)$  nach wachsendem  $i + j$  anzuordnen (und die mit festem  $i + j$  beliebig). Man definiert daher folgendermaßen:

**Definition.** Seien  $\sum_{k=0}^{\infty} a_k$ ,  $\sum_{k=0}^{\infty} b_k$  Reihen. Das *Cauchy-Produkt* dieser Reihen ist die Reihe

$$\sum_{k=0}^{\infty} c_k \quad \text{mit} \quad c_k := \sum_{j=0}^k a_j b_{k-j} = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0.$$

**2.11 Satz.** Sind die Reihen  $\sum_{k=0}^{\infty} a_k$  und  $\sum_{k=0}^{\infty} b_k$  absolut konvergent, so ist auch ihr Cauchy-Produkt absolut konvergent, und es gilt

$$\left( \sum_{k=0}^{\infty} a_k \right) \left( \sum_{k=0}^{\infty} b_k \right) = \sum_{k=0}^{\infty} \left( \sum_{j=0}^k a_j b_{k-j} \right).$$

*Beweis.* Setze  $\sum_{j=0}^k a_j b_{k-j} =: c_k$ . Partialsummen bezeichnen wir mit den entsprechenden großen Buchstaben, also

$$A_n := \sum_{k=0}^n a_k, \quad B_n := \sum_{k=0}^n b_k, \quad C_n := \sum_{k=0}^n c_k.$$

**Behauptung.** 1.  $\lim_{n \rightarrow \infty} (A_n B_n - C_n) = 0$ .

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Es ist

$$A_n B_n - C_n = \sum_{\substack{i \leq n \\ j \leq n}} a_i b_j - \sum_{i+j \leq n} a_i b_j = \sum_{(i,j) \in \Delta_n} a_i b_j$$

mit

$$\Delta_n := \{(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0 \mid i \leq n, j \leq n, i + j > n\}.$$

Setze

$$D_n := \left( \sum_{k=0}^n |a_k| \right) \left( \sum_{k=0}^n |b_k| \right) = \sum_{\substack{i \leq n \\ j \leq n}} |a_i b_j|.$$

Nach Voraussetzung und Satz 1.3 ist die Folge  $(D_n)_{n \in \mathbb{N}}$  konvergent, also existiert nach Satz 1.8 ein  $n_0 \in \mathbb{N}$  mit

$$|D_n - D_{n_0}| < \varepsilon \quad \text{für } n \geq n_0.$$

Es ist

$$D_n - D_{n_0} = \sum_{(i,j) \in \Gamma_n} |a_i b_j|$$

mit

$$\Gamma_n := \{(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0 \mid i \leq n, j \leq n\} \setminus \{(i, j) \in \mathbb{N}_0 \times \mathbb{N}_0 \mid i \leq n_0, j \leq n_0\}.$$

Für alle  $n \geq 2n_0$  gilt  $\Delta_n \subset \Gamma_n$  und daher

$$|A_n B_n - C_n| \leq \sum_{(i,j) \in \Delta_n} |a_i b_j| \leq \sum_{(i,j) \in \Gamma_n} |a_i b_j| = |D_n - D_{n_0}| < \varepsilon.$$

Damit ist die 1. Behauptung bewiesen. ■

Aus der 1. Behauptung folgt nach Satz 1.3

$$\lim_{n \rightarrow \infty} C_n = \left( \lim_{n \rightarrow \infty} A_n \right) \left( \lim_{n \rightarrow \infty} B_n \right).$$

**Behauptung.** 2.  $\sum_{k=0}^{\infty} c_k$  konvergiert absolut.

*Beweis.* Setze  $c_k^* := \sum_{j=0}^k |a_j| |b_{k-j}|$ . Da die Reihen  $\sum |a_k|$ ,  $\sum |b_k|$  absolut konvergieren, konvergiert nach dem Vorstehenden die Reihe  $\sum_{k=0}^{\infty} c_k^*$ . Es gilt

$$|c_k| \leq \sum_{j=0}^k |a_j b_{k-j}| = c_k^*.$$

Nach dem Majorantenkriterium ist  $\sum c_k$  absolut konvergent. ■

Damit ist Satz 2.11 gezeigt. ■

### 3.3 Die Exponentialreihe

Als erste Anwendung des Vorstehenden behandeln wir nun die Exponentialreihe. Viele für die Analysis und ihre Anwendungen wichtige reelle Funktionen werden durch unendliche Reihen definiert. Als erstes und besonders wichtiges Beispiel hierfür wollen wir jetzt die Exponentialfunktion einführen. Warum sie zum Beispiel bei der Beschreibung physikalischer Vorgänge, aber auch in anderen Anwendungsgebieten, oft auftritt, läßt sich allerdings erst gut erklären, wenn wir die Differentiation von Funktionen eingeführt haben. Einen Hinweis gibt aber schon Satz 3.4.

**3.1 Satz.** Für jedes  $x \in \mathbb{R}$  ist die Reihe

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}$$

absolut konvergent.

*Beweis.* Wegen

$$\left| \frac{x^{n+1} n!}{(n+1)! x^n} \right| = \frac{|x|}{n+1} < \frac{1}{2} \quad \text{für } n \geq 2|x|$$

folgt dies aus dem Quotientenkriterium 2.9. ■

**Definition.** Die Reihe  $\sum_{n=0}^{\infty} \frac{x^n}{n!}$  heißt *Exponentialreihe*. Die durch

$$\exp(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad \text{für } x \in \mathbb{R}$$

definierte Funktion  $\exp : \mathbb{R} \rightarrow \mathbb{R}$  heißt *Exponentialfunktion*. Man setzt

$$\sum_{n=0}^{\infty} \frac{1}{n!} = \exp(1) = e \quad (\text{Eulersche Zahl}).$$

Später werden wir  $\exp(x) = e^x$  schreiben, aber zuvor muß gezeigt werden, dass in der Tat  $\exp(n) = e^n$  für  $n \in \mathbb{Z}$  gilt.

Will man  $\exp(x)$  näherungsweise aus der Exponentialreihe berechnen, so muß man eine „Fehlerabschätzung“ haben.

**3.2 Satz.** Wird das „Restglied“  $R_{m+1}(x)$  durch

$$\exp(x) = \sum_{n=0}^m \frac{x^n}{n!} + R_{m+1}(x)$$

definiert, so gilt

$$|R_{m+1}(x)| \leq 2 \frac{|x|^{m+1}}{(m+1)!} \quad \text{für } m \geq 2|x| - 2.$$

*Beweis.* Es ist

$$\begin{aligned} |R_{m+1}(x)| &= \left| \sum_{n=m+1}^{\infty} \frac{x^n}{n!} \right| \leq \sum_{n=m+1}^{\infty} \frac{|x|^n}{n!} \\ &= \frac{|x|^{m+1}}{(m+1)!} \left( 1 + \frac{|x|}{m+2} + \frac{|x|^2}{(m+2)(m+3)} + \dots \right) \\ &\leq \frac{|x|^{m+1}}{(m+1)!} \left( 1 + \frac{|x|}{m+2} + \left( \frac{|x|}{m+2} \right)^2 + \dots \right). \end{aligned}$$

Wird also  $\frac{|x|}{m+2} \leq \frac{1}{2}$  vorausgesetzt, so folgt

$$|R_{m+1}(x)| \leq \frac{|x|^{m+1}}{(m+1)!} \left( 1 + \frac{1}{2} + \frac{1}{4} + \dots \right) = 2 \frac{|x|^{m+1}}{(m+1)!}. \quad \blacksquare$$

Wenn also eine Genauigkeitsschranke  $\alpha > 0$  vorgeschrieben ist, so kann man aus der Abschätzung in Satz 3.2 zu gegebenem  $x$  eine Zahl  $m$  errechnen, so dass

$$\left| \exp(x) - \sum_{n=0}^m \frac{x^n}{n!} \right| < \alpha$$

ist. Zum Beispiel berechnet man so (man nehme  $m = 14$ ):

$$e = 2,718281828459 \pm 2 \cdot 10^{-12}.$$

**Bemerkung.** Für  $x = 1$  können wir auch wie folgt abschätzen:

$$\begin{aligned} 0 < R_{m+1}(1) &< \frac{1}{(m+1)!} \left( 1 + \frac{1}{m+1} + \left( \frac{1}{m+1} \right)^2 + \dots \right) \\ &= \frac{1}{(m+1)!} \frac{1}{1 - \frac{1}{m+1}} = \frac{1}{m \cdot m!}. \end{aligned}$$

Hieraus können wir folgern, dass die Zahl  $e$  irrational ist: Andernfalls wäre nämlich  $e = \frac{k}{m}$  mit passenden  $k, m \in \mathbb{N}$ , also

$$0 < \frac{k}{m} - \sum_{n=0}^m \frac{1}{n!} < \frac{1}{m \cdot m!},$$

folglich

$$0 < km! - m \sum_{n=0}^m \frac{m!}{n!} < 1,$$

was nicht sein kann, da es sich um eine ganze Zahl handelt.

Wir benutzen jetzt die im vorigen Abschnitt behandelte Multiplikation von Reihen, um die wichtige Funktionalgleichung der Exponentialfunktion herzuleiten.

**3.3 Satz.** Für alle  $x, y \in \mathbb{R}$  gilt  $\exp(x+y) = \exp(x) \exp(y)$ .

*Beweis.* Da die Reihen  $\sum \frac{x^n}{n!}$  und  $\sum \frac{y^n}{n!}$  absolut konvergieren, ist das Produkt ihrer Grenzwerte nach Satz 2.11 gleich dem Grenzwert ihres Cauchy-Produktes, also ist

$$\begin{aligned} \exp(x) \exp(y) &= \left( \sum_{n=0}^{\infty} \frac{x^n}{n!} \right) \left( \sum_{n=0}^{\infty} \frac{y^n}{n!} \right) = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{x^k}{k!} \frac{y^{n-k}}{(n-k)!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = \sum_{n=0}^{\infty} \frac{1}{n!} (x+y)^n = \exp(x+y), \end{aligned}$$

wobei die binomische Formel benutzt wurde. ■

### Folgerungen

Für  $n \in \mathbb{N}$  ist

$$\exp(n) = \exp(1 + \dots + 1) = \exp(1) \cdots \exp(1) = \exp(1)^n = e^n$$

nach Definition der Zahl  $e$ . Aus der Reihendarstellung ergibt sich  $\exp(0) = 1 = e^0$ .

Für  $x \in \mathbb{R}$  ist

$$1 = \exp(0) = \exp(-x + x) = \exp(-x) \exp(x),$$

also

$$\exp(-x) = \frac{1}{\exp x}.$$

Von nun an schreiben wir

$$\exp(x) = e^x \quad \text{für } x \in \mathbb{R};$$

da wir  $\exp(k) = e^k$  für  $k \in \mathbb{Z}$  gezeigt haben, ist das in Einklang mit der bereits festgelegten Bedeutung von  $e^x$  für  $x \in \mathbb{Z}$ . Mit dieser Schreibweise ist also

$$e^{x+y} = e^x e^y \quad \text{für } x, y \in \mathbb{R}, \quad e^{-x} = \frac{1}{e^x}.$$

Für  $x > 0$  ist

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots > 1.$$

Für  $x < 0$  ist  $-x > 0$ , also  $e^x = \frac{1}{e^{-x}} > 0$ . Also ist  $e^x > 0$  für alle  $x \in \mathbb{R}$ . Für  $x > y$  gilt ferner

$$\frac{e^x}{e^y} = e^{x-y} > 1, \quad \text{also } e^x > e^y.$$

Die  $e$ -Funktion ist also streng monoton wachsend.

Zum Schluß geben wir noch eine Darstellung der Exponentialfunktion durch einen Grenzwert. Sie gibt einen ersten Hinweis auf das Auftreten der Exponentialfunktion in verschiedenen Anwendungen (Interpretation: Kontinuierliche Verzinsung, radioaktiver Zerfall).

**3.4 Satz.** Für alle  $x \in \mathbb{R}$  gilt

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

*Beweis.* Sei  $x \in \mathbb{R}$  gegeben. Wir setzen  $a_n := \left(1 + \frac{x}{n}\right)^n$  (für gegebenes  $x \in \mathbb{R}$ ). Nach der binomischen Formel ist

$$a_n = \sum_{k=0}^n \binom{n}{k} \frac{x^k}{n^k}.$$

Zunächst sei  $x \geq 0$ . Für  $k \geq 1$  gilt

$$\binom{n}{k} \frac{1}{n^k} = \frac{1}{k!} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \leq \frac{1}{k!},$$

also

$$a_n \leq \sum_{k=0}^n \frac{x^k}{k!} \leq \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x.$$

Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $m \in \mathbb{N}$  mit

$$\sum_{k=0}^m \frac{x^k}{k!} > e^x - \frac{\varepsilon}{2}.$$

Für  $n \geq m$  gilt

$$a_n \geq \sum_{k=0}^m \binom{n}{k} \frac{x^k}{n^k} = \sum_{k=0}^m \frac{x^k}{k!} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}.$$

Für  $n \rightarrow \infty$  konvergiert der  $k$ -te Summand der rechten Seite gegen  $\frac{x^k}{k!}$ , also die rechte Seite gegen  $\sum_{k=0}^m \frac{x^k}{k!}$ . Daher gibt es ein  $n_0 \geq m$  mit

$$a_n > \sum_{k=0}^m \frac{x^k}{k!} - \frac{\varepsilon}{2} \quad \text{für } n \geq n_0.$$

Für alle  $n \geq n_0$  gilt also

$$e^x \geq a_n > e^x - \varepsilon.$$

Somit gilt  $\lim_{n \rightarrow \infty} a_n = e^x$ .

Sei nun  $x < 0$ . Es gilt

$$\left(1 - \frac{x}{n}\right)^n \left(1 + \frac{x}{n}\right)^n = \left(1 - \left(\frac{x}{n}\right)^2\right)^n \leq 1.$$

Nach der Bernoullischen Ungleichung ist für  $n > |x|$

$$\left(1 - \left(\frac{x}{n}\right)^2\right)^n \geq 1 - \frac{x^2}{n},$$

daraus folgt

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n \left(1 + \frac{x}{n}\right)^n = 1.$$

Wegen  $-x > 0$  ist

$$\lim_{n \rightarrow \infty} \left(1 - \frac{x}{n}\right)^n = e^{-x},$$

also nach Satz 1.3

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

## 4 Topologie in $\mathbb{R}$ und Stetigkeit

### 4.1 Topologische Eigenschaften von Mengen reeller Zahlen

Die genauere Untersuchung von reellen Funktionen, wie sie in den folgenden Abschnitten beabsichtigt ist, erfordert zunächst einige Kenntnisse über mögliche Eigenschaften ihrer Definitionsbereiche. Eigenschaften der Art, wie sie hier betrachtet werden sollen, bezeichnet man als „topologische“ Eigenschaften (von griechisch *topos* = Ort), da es bei ihnen weniger um Größeneigenschaften als um Lage- und gestaltliche Eigenschaften geht. Im folgenden sprechen wir (zur Unterstützung der Anschauung) gerne von „Punktmengen“ anstatt von Zahlenmengen. Gemeint sind aber immer Teilmengen von  $\mathbb{R}$ .

Zunächst sei an den Umgebungsbegriff erinnert. Für  $x \in \mathbb{R}$  und  $\varepsilon \in \mathbb{R}^+$  hatten wir die Menge

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

als die  $\varepsilon$ -Umgebung von  $x$  bezeichnet. Eine Menge  $U \subseteq \mathbb{R}$  heißt *Umgebung von  $x$* , wenn  $U$  eine passende  $\varepsilon$ -Umgebung von  $x$  enthält. Wir bemerken:

**1.1 Behauptung.** *Der Durchschnitt von endlich vielen Umgebungen von  $x$  ist Umgebung von  $x$ .*

*Beweis.* Seien  $U_1, \dots, U_n$  Umgebungen von  $x$ . Es gibt also Zahlen  $\varepsilon_1, \dots, \varepsilon_n \in \mathbb{R}^+$  mit  $U_{\varepsilon_i}(x) \subseteq U_i$  ( $i = 1, \dots, n$ ). Mit  $\varepsilon := \min\{\varepsilon_1, \dots, \varepsilon_n\}$  gilt  $U_\varepsilon(x) \subseteq U_{\varepsilon_i}(x) \subseteq U_i$  für  $i = 1, \dots, n$ , also  $U_\varepsilon(x) \subseteq U_1 \cap \dots \cap U_n$ , folglich ist  $U_1 \cap \dots \cap U_n$  Umgebung von  $x$ . ■

Nun kommen wir zu zwei grundlegenden Eigenschaften, die Mengen reeller Zahlen haben können.

**Definition.** Die Menge  $M \subseteq \mathbb{R}$  heißt *offen*, wenn sie Umgebung jedes ihrer Punkte ist.

Mit anderen Worten:

$$M \text{ ist offen} \Leftrightarrow \forall x \in M \exists \varepsilon \in \mathbb{R}^+ : U_\varepsilon(x) \subseteq M.$$

**Beispiele.** offene Intervalle (Beweis trivial),  $\emptyset, \mathbb{R}$ . Weitere Beispiele ergeben sich aus dem folgenden Satz:

**1.2 Satz.** *Die Vereinigung von (beliebig vielen) offenen Mengen ist offen. Der Durchschnitt von endlich vielen offenen Mengen ist offen.*

*Beweis.* trivial bzw. Behauptung 1.1. ■

**Definition.** Sei  $M \subseteq \mathbb{R}$  eine Teilmenge und  $a \in \mathbb{R}$ . Dann heißt  $a$  *Häufungspunkt* von  $M$ , wenn jede Umgebung von  $a$  einen von  $a$  verschiedenen Punkt von  $M$  enthält.

Mit anderen Worten:

$$\begin{aligned} a \text{ ist Häufungspunkt von } M \\ \Leftrightarrow \text{Für jede Umgebung } U \text{ von } a \text{ gilt } (U \cap M) \setminus \{a\} \neq \emptyset \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists x \in M : 0 < |x - a| < \varepsilon. \end{aligned}$$

Man muß also unterscheiden zwischen einem Häufungspunkt einer Folge und einem Häufungspunkt einer Menge. Nicht jeder Häufungspunkt einer Folge  $(a_n)_{n \in \mathbb{N}}$  ist notwendig auch Häufungspunkt der Menge  $\{a_n \mid n \in \mathbb{N}\}$ . (Beispiel:  $a_n = (-1)^n$ ).

**Bemerkung.** Ist  $a$  Häufungspunkt von  $M$ , so liegen in jeder Umgebung von  $a$  unendlich viele Elemente von  $M$ .

Analog zum Satz von Bolzano-Weierstraß können wir folgende Variante herleiten, die ebenfalls als Satz von Bolzano-Weierstraß bezeichnet wird:

**1.3 Satz** (Bolzano-Weierstraß). *Zu jeder beschränkten unendlichen Menge  $M \subseteq \mathbb{R}$  gibt es in  $\mathbb{R}$  einen Häufungspunkt.*

*Beweis.* Man zeigt dies in offensichtlicher Weise völlig analog zum Beweis von Satz 1.6 aus Kapitel 3 durch Intervallhalbierungsverfahren und Intervallschachtelung. ■

**Definition.** Die Menge  $M \subseteq \mathbb{R}$  heißt *abgeschlossen*, wenn sie alle ihre Häufungspunkte enthält.

**Beispiele.** Abgeschlossene Intervalle,  $\emptyset, \mathbb{R}$ .

**Bemerkung.** Achtung! Eine Menge braucht weder offen noch abgeschlossen zu sein; eine Menge kann sowohl offen als auch abgeschlossen sein.

Es besteht jedoch ein einfacher Zusammenhang zwischen offenen und abgeschlossenen Mengen, siehe Satz 1.4.

**Bezeichnungen.** Ist  $M \subseteq \mathbb{R}$ , so nennt man die Menge

$$M^c := \mathbb{R} \setminus M$$

das *Komplement* von  $M$ . Für  $x \in \mathbb{R}$  gilt also:  $x \in M^c \Leftrightarrow x \notin M$ . Es ist  $(M^c)^c = M$ .

**1.4 Satz.** Für  $M \subseteq \mathbb{R}$  gilt:

$$M \text{ offen} \Leftrightarrow M^c \text{ abgeschlossen.}$$

*Beweis.* „ $\Rightarrow$ “: Sei  $M$  offen. Sei  $x$  Häufungspunkt von  $M^c$ . Angenommen, es wäre  $x \notin M^c$ , also  $x \in M$ . Dann ist  $M$  Umgebung von  $x$  mit  $M \cap M^c = \emptyset$ , was ein Widerspruch ist. Also ist  $x \in M^c$ .

„ $\Leftarrow$ “: Sei  $M^c$  abgeschlossen. Sei  $x \in M$ . Dann ist  $x \notin M^c$ , also  $x$  nicht Häufungspunkt von  $M^c$ . Folglich existiert eine Umgebung  $U$  von  $x$  mit  $U \cap M^c = \emptyset$ , also mit  $U \subseteq M$ . Da  $x \in M$  beliebig war, ist  $M$  offen. ■

Für die Komplementbildung gelten die folgenden sogenannten *de Morgan'schen Regeln*, die leicht zu beweisen sind (hier ist  $\mathcal{M}$  ein beliebiges System von Mengen):

$$\begin{aligned} \left( \bigcup_{M \in \mathcal{M}} M \right)^c &= \bigcap_{M \in \mathcal{M}} M^c \\ \left( \bigcap_{M \in \mathcal{M}} M \right)^c &= \bigcup_{M \in \mathcal{M}} M^c. \end{aligned}$$

Berücksichtigt man sie, so folgert man aus Satz 1.4 und Satz 1.2:

**1.5 Satz.** Der Durchschnitt von (beliebig vielen) abgeschlossenen Mengen ist abgeschlossen. Die Vereinigung von endlich vielen abgeschlossenen Mengen ist abgeschlossen.

Unter den abgeschlossenen Mengen sind die beschränkten besonders wichtig, daher haben sie einen besonderen Namen:

**Definition.** Eine Menge  $M \subseteq \mathbb{R}$  heißt *kompakt*, wenn sie beschränkt und abgeschlossen ist.

Aus Satz 1.3 sowie Satz 1.6 aus Kapitel 3 ergibt sich unmittelbar:

**1.6 Satz.** Sei  $M \subseteq \mathbb{R}$  kompakt. Dann gilt:

- (a) Jede unendliche Teilmenge von  $M$  besitzt einen Häufungspunkt in  $M$ .
- (b) Jede Folge in  $M$  besitzt eine Teilfolge, die gegen ein Element von  $M$  konvergiert.

Eine weitere wichtige Eigenschaft kompakter Mengen ist die folgende:

**1.7 Satz.** Jede nichtleere kompakte Menge  $M \subseteq \mathbb{R}$  enthält ein Maximum (größtes Element) und ein Minimum (kleinstes Element).

*Beweis.* Sei  $M \subseteq \mathbb{R}$  kompakt,  $M \neq \emptyset$ . Dann existiert  $s := \sup M$ . Zu jedem  $\varepsilon \in \mathbb{R}^+$  existiert ein  $x \in M$  mit  $x > s - \varepsilon$ . Folglich gilt entweder  $s \in M$ , oder  $s$  ist Häufungspunkt von  $M$ , also gilt ebenfalls  $s \in M$ . Somit ist  $s = \max M$ . Analog schließt man für das Minimum. ■

Eine etwas weniger einfach zu formulierende, aber bei vielen feineren Untersuchungen besonders wichtige Eigenschaft kompakter Mengen bringt der „Überdeckungssatz von Heine-Borel“ zum Ausdruck.

**Definition.** Ein System (= Menge)  $\mathcal{M}$  von Teilmengen von  $\mathbb{R}$  heißt *Überdeckung* der Menge  $A \subseteq \mathbb{R}$ , wenn  $A \subseteq \bigcup_{M \in \mathcal{M}} M$  gilt.  $\mathcal{M}$  heißt *offene Überdeckung*, wenn alle Elemente von  $\mathcal{M}$  offene Mengen sind.

**1.8 Satz** (Überdeckungssatz von Heine-Borel). Sei  $A \subseteq \mathbb{R}$  kompakt und  $\mathcal{M}$  eine offene Überdeckung von  $A$ . Dann enthält  $\mathcal{M}$  eine endliche Überdeckung von  $A$ .

*Beweis.* Angenommen, die Behauptung wäre falsch. Dann gibt es eine offene Überdeckung  $\mathcal{M}$  von  $A$  ohne endliche Teilüberdeckung, d.h. je endlich viele Mengen von  $\mathcal{M}$  haben die Eigenschaft, dass ihre Vereinigung nicht ganz  $A$  enthält. Wir definieren rekursiv eine Intervallschachtelung  $(J_n)_{n \in \mathbb{N}}$  mit folgender Eigenschaft:  $J_n \cap A$  kann nicht durch endlich viele Mengen von  $\mathcal{M}$  überdeckt werden. Da  $A$  beschränkt ist, existiert ein Intervall  $J_1 = [a_1, b_1]$  mit  $A \subseteq J_1$ . Sei  $J_n = [a_n, b_n]$  schon definiert, so dass obige Eigenschaft besteht. Mindestens eine der Mengen  $[a_n, \frac{1}{2}(a_n + b_n)] \cap A$ ,  $[\frac{1}{2}(a_n + b_n), b_n] \cap A$  benötigt zu ihrer Überdeckung unendlich viele Mengen aus  $\mathcal{M}$ , sei  $J_{n+1}$  das

erste Intervall mit dieser Eigenschaft. Damit ist die Folge  $(J_n)_{n \in \mathbb{N}}$  definiert. Aus der Konstruktion folgt, dass alle  $J_n$  unendlich viele Elemente von  $A$  enthalten müssen. Nach Satz 2.3 aus Kapitel 2 existiert ein  $x \in \bigcap_{n \in \mathbb{N}} J_n$ . Jede Umgebung von  $x$  enthält fast alle  $J_n$ , insbesondere unendlich viele Punkte aus  $A$ . Also ist  $x$  Häufungspunkt von  $A$  und daher  $x \in A$ . Da  $\mathcal{M}$  Überdeckung von  $A$  ist, existiert ein  $M \in \mathcal{M}$  mit  $x \in M$ . Da  $M$  offen ist, existiert ein  $\varepsilon \in \mathbb{R}^+$  mit  $U_\varepsilon(x) \subseteq M$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $b_{n_0} - a_{n_0} < \varepsilon$ . Für alle  $y \in J_{n_0}$  gilt dann (wegen  $x \in J_{n_0}$ )  $|x - y| < \varepsilon$ , also  $y \in U_\varepsilon(x)$ . Somit ist  $J_{n_0} \subseteq U_\varepsilon(x) \subseteq M$ .  $J_{n_0}$  wird also sogar durch die eine Menge  $M \in \mathcal{M}$  überdeckt, ein Widerspruch. ■

Satz 1.8 gestattet auch eine Umkehrung:

**1.9 Satz.** *Sei  $A \subseteq \mathbb{R}$  eine Menge mit der Eigenschaft, dass jede offene Überdeckung von  $A$  eine endliche Überdeckung von  $A$  enthält. Dann ist  $A$  kompakt.*

*Beweis.* Da die Überdeckung  $\{U_1(a) \mid a \in A\}$  eine endliche Überdeckung von  $A$  enthält, ist  $A$  beschränkt. Sei  $x \in A^c$ . Dann ist das System  $\{M_n \mid n \in \mathbb{N}\}$  mit  $M_n := \{y \in \mathbb{R} \mid |x - y| > \frac{1}{n}\}$  eine offene Überdeckung von  $A$ . Es gibt also ein  $k \in \mathbb{N}$  mit

$$A \subseteq \bigcup_{n=1}^k M_n = M_k,$$

folglich mit  $U_{\frac{1}{k}}(x) \subseteq A^c$ . Also ist  $A^c$  offen und damit  $A$  abgeschlossen. ■

## 4.2 Grenzwerte von Funktionen und Stetigkeit

Wir betrachten im folgenden reelle Funktionen, worunter wir Funktionen  $f : D \rightarrow \mathbb{R}$  mit einem nichtleeren Definitionsbereich  $D \subseteq \mathbb{R}$  verstehen wollen. Von den Funktionen, die einfache physikalische Vorgänge beschreiben (z.B. „Weg als Funktion der Zeit“ bei der Bewegung eines Massenpunktes) ist man es gewöhnt, dass „kleinen“ Änderungen des Argumentes nur „kleine“ Änderungen der Funktionswerte entsprechen. Der Versuch, diese Vorstellung mathematisch zu präzisieren, führt zur Begriffsbildung der Stetigkeit, die für den weiteren Aufbau der Analysis von grundlegender Bedeutung ist. Zur bequemeren Behandlung dieser Eigenschaft von Funktionen führen wir zunächst einen Grenzwertbegriff für Funktionen ein, der in Analogie zum Grenzwert von Folgen gebildet ist.

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine reelle Funktion ( $D \subseteq \mathbb{R}$ ),  $a$  ein Häufungspunkt von  $D$  und  $b \in \mathbb{R}$ . Man schreibt

$$\lim_{x \rightarrow a} f(x) = b, \quad (2.1)$$

wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta \in \mathbb{R}^+$  existiert mit

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

Gilt (2.1), so sagt man, dass  $\lim_{x \rightarrow a} f(x)$  (der Grenzwert von  $f$  bei Annäherung an  $a$ ) existiert.

Kurz gefaßt lautet die Definition also:

$$\begin{aligned} \lim_{x \rightarrow a} f(x) = b \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D : (0 < |x - a| < \delta \Rightarrow |f(x) - b| < \varepsilon). \end{aligned}$$

Bei der Anwendung des Grenzwertbegriffes für Funktionen ist es oft nützlich, dass man die Untersuchung von  $\lim_{x \rightarrow a} f(x)$  zurückführen kann auf die Untersuchung von Folgen-Grenzwerten:

**2.2 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a$  ein Häufungspunkt von  $D$ . Dann gilt:

$$\lim_{x \rightarrow a} f(x) = b \Leftrightarrow (*) \text{ Für jede Folge } (x_n)_{n \in \mathbb{N}} \text{ in } D \setminus \{a\} \text{ mit } \lim_{n \rightarrow \infty} x_n = a \text{ gilt } \lim_{n \rightarrow \infty} f(x_n) = b.$$

*Beweis.* „ $\Rightarrow$ “: Gelte  $\lim_{x \rightarrow a} f(x) = b$ . Sei  $(x_n)_{n \in \mathbb{N}}$  eine Folge in  $D \setminus \{a\}$  mit  $\lim_{n \rightarrow \infty} x_n = a$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Wegen  $\lim_{x \rightarrow a} f(x) = b$  gibt es ein  $\delta \in \mathbb{R}^+$  mit

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

Wegen  $\lim_{n \rightarrow \infty} x_n = a$  gibt es ein  $n_0 \in \mathbb{N}$  mit

$$|x_n - a| < \delta \quad \text{für alle } n \geq n_0.$$

Für  $n \geq n_0$  gilt also  $|f(x_n) - b| < \varepsilon$ . Damit ist  $\lim_{n \rightarrow \infty} f(x_n) = b$  gezeigt.

„ $\Leftarrow$ “: Gelte (\*). Angenommen, es wäre *nicht*  $\lim_{x \rightarrow a} f(x) = b$ . Das bedeutet:

$$\exists \varepsilon \in \mathbb{R}^+ \forall \delta \in \mathbb{R}^+ \exists x \in D \setminus \{a\} : (|x - a| < \delta \wedge |f(x) - b| \geq \varepsilon).$$

Insbesondere können wir zu jedem  $n \in \mathbb{N}$  ein  $x_n \in D \setminus \{a\}$  auswählen mit  $|x_n - a| < \frac{1}{n}$  und  $|f(x_n) - b| \geq \varepsilon$ . Es gibt also eine Folge  $(x_n)_{n \in \mathbb{N}}$  in  $D \setminus \{a\}$  mit  $\lim_{n \rightarrow \infty} x_n = a$ , für die nicht  $\lim_{n \rightarrow \infty} f(x_n) = b$  gilt, ein Widerspruch. ■

**Bemerkung.** Im letzten Teil des vorstehenden Beweises haben wir Gebrauch gemacht von der folgenden Tatsache:

(A) Ist  $(M_n)_{n \in \mathbb{N}}$  eine Folge nichtleerer Mengen, so gibt es eine Abbildung  $\alpha : \mathbb{N} \rightarrow \bigcup_{n \in \mathbb{N}} M_n$  mit  $\alpha(n) \in M_n$  für alle  $n \in \mathbb{N}$ .

Obwohl wir in dieser Vorlesung die Mengenlehre nur in der sogenannten „naiven“ Weise benutzen, also ohne axiomatische Begründung, sei doch darauf hingewiesen, dass es sich hier um ein Axiom der Mengenlehre handelt (das abzählbare Auswahlaxiom). Es ist in der elementaren Analysis üblich, dieses Axiom stillschweigend zu benutzen.

Für die Existenz des Grenzwertes einer Funktion haben wir auch ein Cauchy-Kriterium zur Verfügung:

**2.3 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a$  ein Häufungspunkt von  $D$ . Der Grenzwert  $\lim_{x \rightarrow a} f(x)$  existiert genau dann, wenn gilt:

$$\begin{aligned} \forall \varepsilon \in \mathbb{R}^+ \quad \exists \delta \in \mathbb{R}^+ \quad \forall x, y \in D \setminus \{a\} : \\ (|x - a| < \delta \text{ und } |y - a| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon). \end{aligned} \quad (2.4)$$

*Beweis.* „ $\Rightarrow$ “: Sei  $\lim_{x \rightarrow a} f(x) = b$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $\delta \in \mathbb{R}^+$  mit

$$|f(x) - b| < \frac{\varepsilon}{2} \quad \text{für alle } x \in D \setminus \{a\} \text{ mit } |x - a| < \delta.$$

Für alle  $x, y \in D \setminus \{a\}$  mit  $|x - a| < \delta$ ,  $|y - a| < \delta$  gilt also

$$|f(x) - f(y)| \leq |f(x) - b| + |f(y) - b| < \varepsilon.$$

(2.4) ist also erfüllt.

„ $\Leftarrow$ “: Gelte (2.4). Sei  $(x_n)_{n \in \mathbb{N}}$  eine Folge in  $D \setminus \{a\}$  mit  $\lim_{n \rightarrow \infty} x_n = a$ . Sei  $\varepsilon \in \mathbb{R}^+$  gegeben; sei  $\delta$  gemäß (2.4) gewählt. Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|x_n - a| < \delta$  für  $n \geq n_0$ . Für  $m, n \geq n_0$  gilt also  $|f(x_m) - f(x_n)| < \varepsilon$ . Somit ist die Folge  $(f(x_n))_{n \in \mathbb{N}}$  eine Cauchyfolge und daher nach Satz 1.8 aus Kapitel 3 konvergent gegen eine Zahl  $b$ . Diese Zahl  $b$  hängt (wegen (2.4)) offenbar nicht von der speziellen Folge ab. Aus Satz 2.2 folgt also  $\lim_{x \rightarrow a} f(x) = b$ . ■

Zu dem eingeführten Grenzwertbegriff für Funktionen kann man einige naheliegende Varianten erklären, die manchmal nützlich sind.

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a \in \mathbb{R}$ . Ist  $a$  Häufungspunkt von  $D \cap (-\infty, a)$ , so schreibt man

$$\lim_{x \nearrow a} f(x) = b \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D : (a - \delta < x < a \Rightarrow |f(x) - b| < \varepsilon).$$

Ist  $a$  Häufungspunkt von  $D \cap (a, \infty)$ , so schreibt man

$$\lim_{x \searrow a} f(x) = b \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D : (a < x < a + \delta \Rightarrow |f(x) - b| < \varepsilon).$$

Mit anderen Worten, es ist

$$\lim_{x \nearrow a} f(x) = b \Leftrightarrow \lim_{x \rightarrow a} f|_{(-\infty, a)}(x) = b,$$

wobei  $f|_{(-\infty, a)}$  die Einschränkung von  $f$  auf  $(-\infty, a)$  ist. Entsprechendes gilt für  $\lim_{x \searrow a}$ . Jetzt ist auch klar, dass sich die Sätze 2.2 und 2.3 sinngemäß auf diese „linksseitigen“ und „rechtsseitigen“ Grenzwerte übertragen lassen.

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a$  ein Häufungspunkt von  $D$ . Man schreibt

$$\lim_{x \rightarrow a} f(x) = \infty \\ \Leftrightarrow \forall c \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D \setminus \{a\} : (|x - a| < \delta \Rightarrow f(x) > c).$$

Analog definiert man  $\lim_{x \rightarrow a} f(x) = -\infty$ ,  $\lim_{x \nearrow a} f(x) = \infty$ , u.s.w.

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $D$  nicht nach oben beschränkt. Man schreibt

$$\lim_{x \rightarrow \infty} f(x) = b \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists c \in \mathbb{R} \forall x \in D : (x > c \Rightarrow |f(x) - b| < \varepsilon).$$

Analog definiert man  $\lim_{x \rightarrow -\infty} f(x) = b$ ,  $\lim_{x \rightarrow \infty} f(x) = \infty$ , u.s.w.

**Beispiele.** Sei  $D := \mathbb{R} \setminus \{0\}$ ,  $f : D \rightarrow \mathbb{R}$  definiert durch  $x \mapsto \frac{1}{x}$ . Dann ist

$$\lim_{x \nearrow 0} f(x) = -\infty, \quad \lim_{x \searrow 0} f(x) = \infty.$$

Mit demselben  $D$  sei  $f : D \rightarrow \mathbb{R}$  definiert durch  $x \mapsto \frac{1}{x^2}$ . Dann ist

$$\lim_{x \rightarrow 0} f(x) = \infty.$$

Sei  $p : \mathbb{R} \rightarrow \mathbb{R}$  eine *Polynomfunktion*, das heißt

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \quad \text{für } x \in \mathbb{R}$$

mit gegebenen Zahlen  $a_n, \dots, a_0 \in \mathbb{R}$ . Sei etwa  $a_n > 0$ . Dann gilt (für  $n \geq 1$ )

$$\begin{aligned} \lim_{x \rightarrow \infty} p(x) &= \infty, \\ \lim_{x \rightarrow -\infty} p(x) &= \begin{cases} \infty, & \text{wenn } n \text{ gerade ist,} \\ -\infty, & \text{wenn } n \text{ ungerade ist.} \end{cases} \end{aligned}$$

Dies liest man sofort an der für  $x \neq 0$  gültigen Darstellung

$$p(x) = a_n x^n \left( 1 + \frac{a_{n-1}}{a_n x} + \frac{a_{n-2}}{a_n x^2} + \cdots + \frac{a_0}{a_n x^n} \right)$$

ab.

### Stetigkeit

Die am Anfang dieses Paragraphen intuitiv erläuterte Stetigkeit einer Funktion läßt sich nun bequem mit Hilfe des Grenzwertbegriffes formulieren. Wir ziehen aber zunächst eine äquivalente Definition vor.

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine reelle Funktion und  $a \in D$ . Die Funktion  $f$  heißt *stetig in  $a$* , wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta \in \mathbb{R}^+$  existiert mit

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

Also kurz:  $f$  ist stetig in  $a \Leftrightarrow$

$$\forall \varepsilon \in \mathbb{R}^+ \quad \exists \delta \in \mathbb{R}^+ \quad \forall x \in D : (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon).$$

In gleichwertiger Weise können wir die Stetigkeit nun mit Verwendung des Grenzwertbegriffs ausdrücken, wie sich unmittelbar aus den Definitionen ergibt.

**2.5 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a \in D$  ein Häufungspunkt von  $D$ . Dann gilt

$$f \text{ stetig in } a \Leftrightarrow \lim_{x \rightarrow a} f(x) = f(a).$$

**Bemerkung.** Der Fall, dass  $a \in D$ , aber  $a$  nicht Häufungspunkt von  $D$  ist ( $a$  heißt dann *isolierter Punkt* von  $D$ ), kann außer Betracht bleiben, da offenbar jede reelle Funktion in jedem isolierten Punkt ihres Definitionsbereiches stetig ist.

Unmittelbar aus Satz 2.5 und 2.2 erhalten wir:

**2.6 Satz.** Die Funktion  $f : D \rightarrow \mathbb{R}$  ist in  $a \in D$  genau dann stetig, wenn für jede Folge  $(x_n)_{n \in \mathbb{N}}$  in  $D$  mit  $\lim_{n \rightarrow \infty} x_n = a$  gilt:  $\lim_{n \rightarrow \infty} f(x_n) = f(a)$ .

Hieraus ergibt sich unmittelbar, dass Summen und Produkte von stetigen Funktionen stetig sind. Für Funktionen  $f, g : D \rightarrow \mathbb{R}$  sind  $f+g$  und  $fg$  erklärt durch  $(f+g)(x) := f(x) + g(x)$  und  $(fg)(x) := f(x)g(x)$  für alle  $x \in D$ .

**2.7 Satz.** Seien  $f, g : D \rightarrow \mathbb{R}$  in  $a \in D$  stetig. Dann sind die Funktionen  $f+g$  und  $fg$  in  $a$  stetig. Ist  $f(a) \neq 0$ , so existiert eine Umgebung  $U$  von  $a$  mit  $f(x) \neq 0$  für  $x \in U \cap D$ , und  $1/f|_{U \cap D}$  ist stetig in  $a$ .

*Beweis.* Dies ergibt sich sofort aus Satz 2.6 und den Rechenregeln für Grenzwerte in Satz 1.3 aus Kapitel 3. Die Existenz von  $U$  folgt unmittelbar aus der Definition der Stetigkeit. ■

Bei Stetigkeitsbeweisen kann man wahlweise diejenige der äquivalenten Formulierungen benutzen, die gerade am bequemsten ist.

**Beispiele.** Zum Nachweis der Stetigkeit in  $a$  einer Funktion  $f : D \rightarrow \mathbb{R}$  ist, wenn wir unmittelbar die gewählte Definition verwenden wollen, zu zeigen, dass zu gegebenem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta \in \mathbb{R}^+$  existiert mit

$$x \in D \wedge |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

Wir geben im folgenden, soweit möglich, zu vorgegebenem  $\varepsilon$  ein solches  $\delta$  explizit an.

- 1) Konstante Funktionen:  $\delta$  beliebig.
- 2)  $\text{id}_{\mathbb{R}} : \mathbb{R} \rightarrow \mathbb{R}$  mit  $x \mapsto x$ , z.B.  $\delta = \varepsilon$ .
- 3) Hieraus und aus Satz 2.7 folgt, dass jede *rationale Funktion* in ihren natürlichen Definitionsbereich stetig ist. Hierunter versteht man eine Funktion  $f : D \rightarrow \mathbb{R}$  der Form  $f(x) = \frac{p(x)}{q(x)}$  für  $x \in D$ , wobei  $p, q$  Polynomfunktionen sind und  $D := \{x \in \mathbb{R} \mid q(x) \neq 0\}$  ist.
- 4)  $\sqrt{\cdot} : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ ,  $x \mapsto \sqrt{x}$ ,
  1. Fall:  $a = 0$ . Wähle z.B.  $\delta = \varepsilon^2$ .
  2. Fall:  $a > 0$ . Wähle z.B.  $\delta = \varepsilon\sqrt{a}$ , denn es ist

$$|\sqrt{x} - \sqrt{a}| = \frac{|x - a|}{\sqrt{x} + \sqrt{a}} \leq \frac{|x - a|}{\sqrt{a}}.$$

5)  $\text{abs} : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto |x|$

$$\delta = \varepsilon \text{ leistet das Gewünschte wegen } \left| |x| - |a| \right| \leq |x - a|$$

6) Die Exponentialfunktion: Nach Satz 3.2 aus Kapitel 3 gilt

$$|e^x - 1| \leq 2|x| \quad \text{für } |x| \leq 1, \text{ also}$$

$$|e^x - e^a| = e^a |e^{x-a} - 1| \leq e^a \cdot 2|x - a|.$$

Daher leistet  $\delta := \min\{1, \frac{\varepsilon}{2e^a}\}$  das Gewünschte.

Nun zwei unstetige (d.h. in mindestens einem Punkt nicht stetige) Funktionen:

7) Definiere

$$f(x) := \begin{cases} 0, & \text{wenn } x < 0 \\ 1, & \text{wenn } x \geq 0 \end{cases} \quad \text{für } x \in \mathbb{R}.$$

Diese Funktion ist offenbar genau in 0 unstetig.

8) Die durch

$$f(x) := \begin{cases} 1, & \text{wenn } x \in \mathbb{Q} \\ 0, & \text{wenn } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

definierte sog. *Dirichlet-Funktion* ist offenbar in jedem Punkt  $a \in \mathbb{R}$  unstetig. Denn in jeder Umgebung einer reellen Zahl liegen rationale und irrationale Zahlen.

Verschiedene äquivalente Formulierungen der Stetigkeit sind für die Anwendungen zweckmäßig. Zunächst eine besonders einprägsame unter Verwendung des Umgebungsbegriffes:

**2.8 Satz.** Die Funktion  $f : D \rightarrow \mathbb{R}$  ist in  $a \in D$  genau dann stetig, wenn zu jeder Umgebung  $V$  von  $f(a)$  eine Umgebung  $U$  von  $a$  existiert mit  $f(U \cap D) \subseteq V$ .

*Beweis.* „ $\Rightarrow$ “: Sei  $f$  stetig in  $a$ . Sei  $V$  eine Umgebung von  $f(a)$ . Dann existiert ein  $\varepsilon \in \mathbb{R}^+$  mit  $U_\varepsilon(f(a)) \subseteq V$ . Zu diesem  $\varepsilon$  gibt es nach Voraussetzung ein  $\delta \in \mathbb{R}^+$  mit

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

Setze  $U := U_\delta(a)$ . Dann gilt  $f(U \cap D) \subseteq V$ . In der Tat, sei  $y \in f(U \cap D)$ , also  $y = f(x)$  mit  $x \in U \cap D$ . Wegen  $x \in U_\delta(a)$  gilt  $|x - a| < \delta$ , also  $|f(x) - f(a)| < \varepsilon$ , d.h.  $y = f(x) \in U_\varepsilon(f(a)) \subseteq V$ .

„ $\Leftarrow$ “: Zu jeder Umgebung  $V$  von  $f(a)$  gebe es eine Umgebung  $U$  von  $a$  mit  $f(U \cap D) \subseteq V$ . Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Zu  $V := U_\varepsilon(f(a))$  existiert nach Voraussetzung eine Umgebung  $U$  von  $a$  mit  $f(U \cap D) \subseteq V$ . Es gibt ein  $\delta \in \mathbb{R}^+$  mit  $U_\delta(a) \subseteq U$ . Sei jetzt  $x \in D$  und  $|x - a| < \delta$ . Dann ist  $x \in U_\delta(a) \subseteq U$ , also  $f(x) \in f(U \cap D) \subseteq V = U_\varepsilon(f(a))$ , d.h.  $|f(x) - f(a)| < \varepsilon$ . ■

Mit dieser Formulierung der Stetigkeit ist bequem zu arbeiten. Hierfür ein Beispiel.

**2.9 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  stetig in einem  $a \in D$ , sei  $g : D' \rightarrow \mathbb{R}$  mit  $f(D) \subseteq D'$  stetig in  $f(a)$ . Dann ist  $g \circ f$  stetig in  $a$ .

*Beweis.* Wir benutzen Satz 2.8. Sei  $W$  eine Umgebung von  $(g \circ f)(a) = g(f(a))$ . Da  $g$  stetig in  $f(a)$  ist, existiert eine Umgebung  $V$  von  $f(a)$  mit  $g(V \cap D') \subseteq W$ . Da  $f$  stetig in  $a$  ist, existiert eine Umgebung  $U$  von  $a$  mit  $f(U \cap D) \subseteq V$ . Also gilt  $(g \circ f)(U \cap D) = g(f(U \cap D)) \subseteq g(V \cap D') \subseteq W$ . ■

Schließlich wird noch naheliegenderweise definiert:

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  heißt *stetig*, wenn  $f$  stetig ist in  $a$  für alle  $a \in D$ .

Die Funktion im siebten der obigen Beispiele hat eine Unstetigkeit von besonders einfacher Art:

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  hat in  $a \in D$  eine *Sprungstelle*, wenn die einseitigen Grenzwerte

$$\lim_{x \nearrow a} f(x) \quad \text{und} \quad \lim_{x \searrow a} f(x)$$

existieren und verschieden sind.

Ist  $f$  eine monotone Funktion, so ist leicht zu sehen, dass in jedem Punkt  $a$  des Definitionsbereiches die einseitigen Grenzwerte von  $f(x)$  für  $x \nearrow a$  und  $x \searrow a$  existieren. Bei einer monotonen Funktion sind die Unstetigkeitsstellen also sämtlich Sprungstellen. Hieraus kann man folgern, dass bei einer monotonen Funktion höchstens abzählbar viele Unstetigkeitsstellen auftreten können.

Im übernächsten Paragraphen werden wir einige wichtige spezielle Funktionen als Umkehrfunktionen von gewissen streng monotonen Funktionen einführen. Nach Satz 1.1 aus Kapitel 2 ist eine streng monotone Funktion injektiv, und die daher existierende Umkehrfunktion  $f^{-1}$  ist ebenfalls streng monoton.

**2.10 Satz.** *Sei  $D$  ein Intervall,  $f : D \rightarrow \mathbb{R}$  streng monoton. Dann ist die Umkehrfunktion  $f^{-1} : f(D) \rightarrow \mathbb{R}$  stetig.*

*Beweis.* O.B.d.A. sei  $f$  streng monoton wachsend, und  $D$  enthalte mehr als einen Punkt. Seien  $y \in f(D)$  und  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Es gilt  $y = f(x)$  für ein (eindeutig bestimmtes)  $x \in D$ .

1. Fall:  $x$  ist nicht Endpunkt des Intervalls  $D$ . Dann existiert ein  $\varepsilon' \in \mathbb{R}^+$  mit  $\varepsilon' \leq \varepsilon$  und  $[x - \varepsilon', x + \varepsilon'] \subseteq D$ . Setze

$$\delta := \min\{y - f(x - \varepsilon'), f(x + \varepsilon') - y\}.$$

Wegen der strengen Monotonie von  $f$  ist  $\delta > 0$ . Sei jetzt  $y_1 \in f(D)$  mit  $|y_1 - y| < \delta$ . Dann ist

$$f(x - \varepsilon') < y_1 < f(x + \varepsilon')$$

und daher nach Satz 1.1 aus Kapitel 2

$$x - \varepsilon' = f^{-1}(f(x - \varepsilon')) < f^{-1}(y_1) < f^{-1}(f(x + \varepsilon')) = x + \varepsilon',$$

wegen  $x = f^{-1}(y)$  also  $|f^{-1}(y_1) - f^{-1}(y)| < \varepsilon' \leq \varepsilon$ .

2. Fall:  $x$  ist Endpunkt, etwa Maximum, des Intervalls  $D$ . Dann existiert ein  $\varepsilon' \in \mathbb{R}^+$  mit  $\varepsilon' \leq \varepsilon$  und  $[x - \varepsilon', x] \subseteq D$ . Setze  $\delta := y - f(x - \varepsilon')$ . Für  $y_1 \in f(D)$  mit  $|y_1 - y| < \delta$  gilt  $f(x - \varepsilon') < y_1 \leq y$ , woraus wie oben  $|f^{-1}(y_1) - f^{-1}(y)| < \varepsilon$  folgt. ■

Man beachte, dass in Satz 2.10 nicht die Stetigkeit von  $f$  vorausgesetzt zu werden braucht. Andererseits läßt sich selbst bei stetigem  $f$  nicht auf die Stetigkeit von  $f^{-1}$  schließen, falls der Definitionsbereich von  $f$  kein Intervall ist. Dies läßt sich leicht durch Beispiele belegen.

### 4.3 Eigenschaften stetiger Funktionen

Wir stellen in diesem Abschnitt einige wichtige Sätze zusammen, die sich auf reelle Funktionen beziehen, die in ihrem ganzen Definitionsbereich stetig sind. Diese Sätze haben mannigfache Anwendungen. Aus der Stetigkeit

einer Funktion lassen sich besonders dann interessante Folgerungen ziehen, wenn der Definitionsbereich von spezieller Art ist. Zunächst betrachten wir kompakte Definitionsbereiche.

**3.1 Satz.** *Sei  $D \subseteq \mathbb{R}$  kompakt und  $f : D \rightarrow \mathbb{R}$  stetig. Dann ist  $f(D)$  kompakt.*

*Beweis.* Wir benutzen die durch den Überdeckungssatz von Heine-Borel 1.8 und seine Umkehrung 1.9 gegebene Charakterisierung kompakter Mengen.

Sei  $\mathcal{M}$  eine offene Überdeckung von  $f(D)$ . Sei  $x \in D$ . Es gibt eine Menge  $M \in \mathcal{M}$  mit  $f(x) \in M$ . Da  $M$  Umgebung von  $f(x)$  und  $f$  stetig ist, können wir nach Satz 2.8 eine offene Umgebung  $U$  von  $x$  wählen mit  $f(U \cap D) \subseteq M$ . Das System

$$\{U \mid U \subseteq \mathbb{R} \text{ offen, } \exists M \in \mathcal{M} : f(U \cap D) \subseteq M\}$$

ist also eine offene Überdeckung von  $D$  und enthält daher nach Satz 1.8 eine endliche Teilüberdeckung  $\{U_1, \dots, U_n\}$ . Zu  $i \in \{1, \dots, n\}$  gibt es ein  $M_i \in \mathcal{M}$  mit  $f(U_i \cap D) \subseteq M_i$ . Es gilt  $f(D) \subseteq \bigcup_{i=1}^n M_i$ , also ist  $\{M_1, \dots, M_n\}$  eine endliche Überdeckung von  $f(D)$ . Aus Satz 1.9 folgt die Kompaktheit von  $f(D)$ . ■

**3.2 Satz.** *Jede auf einer (nichtleeren) kompakten Menge stetige reelle Funktion nimmt dort ein Maximum (und analog ein Minimum) an.*

Genauer: Ist  $D \subseteq \mathbb{R}$  kompakt und  $f : D \rightarrow \mathbb{R}$  stetig, so existieren  $a, b \in D$  mit  $f(a) \leq f(x) \leq f(b)$  für alle  $x \in D$ . Der Beweis ergibt sich sofort aus Satz 3.1 und Satz 1.7, angewandt auf  $f(D)$ .

Nun betrachten wir stetige Funktionen, die auf Intervallen definiert sind. Der folgende Satz enthält eine der wichtigsten Eigenschaften stetiger Funktionen.

**3.3 Satz.** *Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig,  $f(a) < 0$  und  $f(b) > 0$ . Dann existiert ein  $x_0 \in (a, b)$  mit  $f(x_0) = 0$ .*

*Beweis.* Wir definieren rekursiv eine Intervallschachtelung  $(J_n)_{n \in \mathbb{N}}$ ,  $J_n = [a_n, b_n]$ , mit  $f(a_n) < 0$  und  $f(b_n) \geq 0$ : Setze  $J_1 := [a, b]$ . Ist  $J_n$  schon definiert, so setze

$$J_{n+1} := \begin{cases} [a_n, \frac{1}{2}(a_n + b_n)], & \text{falls } f(\frac{1}{2}(a_n + b_n)) \geq 0, \\ [\frac{1}{2}(a_n + b_n), b_n], & \text{falls } f(\frac{1}{2}(a_n + b_n)) < 0. \end{cases}$$

Die Folge  $(J_n)_{n \in \mathbb{N}}$  hat die angegebenen Eigenschaften. Nach dem Intervallschachtelungsprinzip existiert ein  $x_0 \in \bigcap_{n \in \mathbb{N}} J_n$ . Offenbar gilt  $\lim_{n \rightarrow \infty} a_n =$

$x_0 = \lim_{n \rightarrow \infty} b_n$ . Aus der Stetigkeit von  $f$ , Satz 2.5 und Satz 1.4 aus Kapitel 3 folgt wegen  $f(a_n) < 0$  und  $f(b_n) \geq 0$

$$0 \geq \lim f(a_n) = f(x_0) = \lim f(b_n) \geq 0,$$

also  $f(x_0) = 0$ . ■

**Bemerkung.** Das Beweisverfahren ist konstruktiv, d.h. es kann zur näherungsweisen Berechnung einer Nullstelle verwendet werden.

**3.4 Satz (Zwischenwertsatz).** *Eine auf einem Intervall definierte stetige reelle Funktion nimmt jeden Wert zwischen zwei Funktionswerten an.*

*Beweis.* Sei  $D$  ein Intervall,  $f : D \rightarrow \mathbb{R}$  stetig. Seien  $u, v$  zwei Werte von  $f$ , o.B.d.A.  $u < v$ . Es gibt also Elemente  $a, b \in D$  mit  $f(a) = u$ ,  $f(b) = v$ . Sei nun  $u < z < v$ . Im Fall  $a < b$  wende man Satz 3.3 an auf die Funktion

$$\begin{aligned} g : [a, b] &\rightarrow \mathbb{R}, \\ x &\mapsto f(x) - z \end{aligned}$$

im Fall  $a > b$  auf die Funktion

$$\begin{aligned} g : [a, b] &\rightarrow \mathbb{R}, \\ x &\mapsto -f(x) + z \end{aligned} \quad \blacksquare$$

Anwendungsbeispiel: Jede Polynomfunktion ungeraden Grades hat eine Nullstelle.

### Gleichmäßige Stetigkeit

Werfen wir noch einmal einen Blick auf die Definition der (globalen) Stetigkeit. Dass  $f : D \rightarrow \mathbb{R}$  stetig ist, bedeutet definitionsgemäß:

$$\forall x \in D \quad \forall \varepsilon \in \mathbb{R}^+ \quad \exists \delta \in \mathbb{R}^+ \quad \forall y \in D : (|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon).$$

Hier ist also zu gegebenem  $x \in D$  und gegebenem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta$  zu finden, so dass die rechts stehende Aussage gilt. Im allgemeinen wird (bei festem  $\varepsilon$ ) dieses  $\delta$  von dem vorgegebenem  $x$  abhängen. Beispiel:

$$\begin{aligned} f : (0, 1] &\rightarrow \mathbb{R} \\ x &\mapsto \frac{1}{x} \end{aligned}$$

Je näher  $x$  bei 0 liegt, umso kleiner muß (bei festem  $\varepsilon$ ) das  $\delta$  gewählt werden; es gibt kein  $\delta \in \mathbb{R}^+$ , mit dem man für alle  $x \in D$  auskommt. Manche Konstruktionen und Beweise der Analysis sind aber nur möglich, wenn sich ein solches  $\delta$  einheitlich für den ganzen Definitionsbereich wählen läßt. Funktionen, für die dies der Fall ist, heißen gleichmäßig stetig.

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  heißt *gleichmäßig stetig*, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta \in \mathbb{R}^+$  existiert mit

$$|f(x) - f(y)| < \varepsilon$$

für alle  $x, y \in D$  mit  $|x - y| < \delta$ . Kurz:  $f$  ist gleichmäßig stetig  $\Leftrightarrow$

$$\forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D \forall y \in D : (|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon).$$

Wie das obige Beispiel zeigt, ist nicht jede stetige Funktion gleichmäßig stetig. Dass diese Beispielfunktion unbeschränkt ist, ist nicht der entscheidende Punkt. Zum Beispiel wird eine beschränkte, stetige, aber nicht gleichmäßig stetige Funktion  $f : (0, 1] \rightarrow \mathbb{R}$  erklärt durch

$$f(x) := \begin{cases} n(n+1)x - n & \text{für } x \in \left[\frac{1}{n+1}, \frac{1}{n}\right], n \in \mathbb{N} \text{ ungerade,} \\ -n(n+1)x + n + 1 & \text{für } x \in \left[\frac{1}{n+1}, \frac{1}{n}\right], n \in \mathbb{N} \text{ gerade.} \end{cases}$$

Wir wollen jetzt zeigen, dass so etwas nicht passieren kann, wenn der Definitionsbereich kompakt ist.

**3.5 Satz.** *Jede stetige Funktion mit kompaktem Definitionsbereich ist gleichmäßig stetig.*

*Beweis.* Sei  $D$  kompakt,  $f : D \rightarrow \mathbb{R}$  stetig. Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Zu jedem  $a \in D$  existiert, da  $f$  in  $a$  stetig ist, ein  $\delta(a) \in \mathbb{R}^+$  mit

$$|f(x) - f(a)| < \frac{\varepsilon}{2} \quad \text{für alle } x \in D \quad \text{mit } |x - a| < \delta(a).$$

Das System  $\{U_{\frac{1}{2}\delta(a)}(a) \mid a \in D\}$  ist eine offene Überdeckung von  $D$ , enthält also nach Satz 1.8 eine endliche Teilüberdeckung  $\{U_{\frac{1}{2}\delta(a_i)}(a_i) \mid i = 1, \dots, n\}$ . Setze  $\delta := \min\{\frac{1}{2}\delta(a_1), \dots, \frac{1}{2}\delta(a_n)\}$ . Seien dann  $x, y \in D$  Punkte mit  $|x - y| < \delta$ . Es gibt ein  $i \in \{1, \dots, n\}$  mit  $x \in U_{\frac{1}{2}\delta(a_i)}(a_i)$ . Wegen  $|a_i - x| < \frac{1}{2}\delta(a_i) < \delta(a_i)$  ist

$$|f(a_i) - f(x)| < \frac{\varepsilon}{2}.$$

Wegen  $|a_i - y| \leq |a_i - x| + |x - y| \leq \frac{1}{2}\delta(a_i) + \delta \leq \delta(a_i)$  ist

$$|f(a_i) - f(y)| < \frac{\varepsilon}{2}.$$

Es folgt

$$|f(x) - f(y)| \leq |f(x) - f(a_i)| + |f(a_i) - f(y)| < \varepsilon.$$

Das angegebene  $\delta$  leistet also das Gewünschte. ■



## 5 Spezielle Funktionen

In diesem Kapitel definieren und untersuchen wir einige spezielle Funktionen, die in der Analysis und ihren Anwendungen häufig verwendet werden.

### 5.1 Logarithmus und allgemeine Potenz

Die Exponentialfunktion  $\exp : \mathbb{R} \rightarrow \mathbb{R}$  ist, wie in Abschnitt 3.3 gezeigt wurde, streng monoton wachsend. Ihre Umkehrfunktion existiert also, sie ist nach Satz 1.1 aus Kapitel 2 ebenfalls streng monoton wachsend und nach Satz 2.10 aus Kapitel 4 stetig. In Abschnitt 4.3 wurde gezeigt, dass  $\exp$  stetig ist. Aus der Reihendarstellung folgt

$$e^x \geq 1 + x \quad \text{für } x \geq 0$$

und daraus, ebenfalls für  $x \geq 0$ ,

$$e^{-x} = \frac{1}{e^x} \leq \frac{1}{1+x},$$

also gilt  $\lim_{x \rightarrow \infty} e^x = \infty$  und  $\lim_{x \rightarrow -\infty} e^x = 0$ . Aus dem Zwischenwertsatz 3.4 aus Kapitel 4 folgt, dass das Bild der Exponentialfunktion gleich  $\mathbb{R}^+$  ist (beachte  $e^x > 0$  für alle  $x$ ). Wir können daher formulieren:

**1.1 Definition.** Die Umkehrfunktion der Exponentialfunktion ist auf  $\mathbb{R}^+$  definiert, stetig und streng monoton wachsend; sie heißt (natürliche) Logarithmusfunktion und wird mit  $\exp^{-1} = \ln$  bezeichnet.

**1.2 Satz** (Funktionalgleichung der  $\ln$ -Funktion). Für  $x, y \in \mathbb{R}^+$  gilt

$$\ln xy = \ln x + \ln y.$$

*Beweis.* Setze  $\ln x =: a$ ,  $\ln y =: b$ . Dann ist  $x = e^a$ ,  $y = e^b$ , also nach Satz 3.3 aus Kapitel 3  $xy = e^a e^b = e^{a+b}$  und daher  $\ln xy = a + b = \ln x + \ln y$ . ■

**Bemerkung.** Für  $x \in \mathbb{R}^+$  folgt

$$\ln \frac{1}{x} + \ln x = \ln \left( \frac{1}{x} \cdot x \right) = \ln 1 = 0$$

(wegen  $e^0 = 1$ ), also

$$\ln \frac{1}{x} = -\ln x.$$

**Definition.** Für  $a \in \mathbb{R}^+$  schreibt man

$$a^x := \exp_a x := e^{x \ln a} \quad \text{für } x \in \mathbb{R}.$$

Die Funktion  $\exp_a : \mathbb{R} \rightarrow \mathbb{R}^+$  heißt *Exponentialfunktion* oder *Potenz zur Basis a*.

Für die allgemeine Potenz  $a^x$  stellen wir einige Rechenregeln zusammen:

**1.3 Satz.** Für  $a, b \in \mathbb{R}^+$  und  $x, y \in \mathbb{R}$  gilt

(a)  $a^x a^y = a^{x+y}$ ,

(b)  $(a^x)^y = a^{xy}$ ,

(c)  $a^x b^x = (ab)^x$ ,

(d)  $\left(\frac{1}{a}\right)^x = a^{-x} = \frac{1}{a^x}$ .

*Beweis.* (a)  $a^x a^y = e^{x \ln a} e^{y \ln a} = e^{(x+y) \ln a} = a^{x+y}$ .

(b) Wegen  $a^x = e^{x \ln a}$  ist  $\ln a^x = x \ln a$ , also  $(a^x)^y = e^{y \ln a^x} = e^{yx \ln a} = a^{xy}$ .

(c)  $a^x b^x = e^{x \ln a} e^{x \ln b} = e^{x(\ln a + \ln b)} = e^{x \ln ab} = (ab)^x$ .

(d)  $\left(\frac{1}{a}\right)^x = e^{x \ln \frac{1}{a}} = e^{-x \ln a} = a^{-x}$  und  $a^{-x} = e^{-x \ln a} = \frac{1}{e^{x \ln a}} = \frac{1}{a^x}$ , wobei wir  $\ln \frac{1}{a} = -\ln a$  und  $e^{-y} = \frac{1}{e^y}$  benutzt haben. ■

**Bemerkung.** Für  $a \in \mathbb{R}^+$  und  $n \in \mathbb{N}$  schreibt man auch

$$a^{\frac{1}{n}} =: \sqrt[n]{a}$$

( $n$ -te Wurzel aus  $a$ ). Nach Satz 1.3 (b) ist

$$\left(\sqrt[n]{a}\right)^n = \left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n} \cdot n} = a.$$

Die  $n$ -te Wurzel  $\sqrt[n]{\phantom{x}}$  ist also die Umkehrfunktion der  $n$ -ten Potenz  $a \mapsto a^n$ . (Wir haben die  $n$ -te Wurzel schon mehrfach benutzt, aber erst jetzt definiert.)

Offenbar ist auch die Exponentialfunktion zur Basis  $a$  stetig. Im Fall  $a > 1$  ist sie (wegen  $\ln a > 0$ ) streng monoton wachsend, im Fall  $0 < a < 1$  ist sie (wegen  $\ln a < 0$ ) streng monoton fallend. Für  $a \neq 1$  existiert also ihre Umkehrfunktion; sie heißt *Logarithmusfunktion zur Basis  $a$*  und wird mit  $\log_a$  bezeichnet. Sie unterscheidet sich nur um einen festen Faktor von der natürlichen Logarithmusfunktion: Setzen wir  $\log_a x =: y$ , so ist

$$x = a^y = e^{y \ln a},$$

also  $\ln x = y \ln a$  und daher

$$\log_a x = \frac{\ln x}{\ln a}.$$

Es genügt daher, die natürliche Logarithmusfunktion (also die zur Basis  $e$ ) zu kennen.

**Bemerkung.** Die Schreibweisen für die Logarithmusfunktionen sind in der Literatur nicht einheitlich. So wird für  $\ln$  gelegentlich auch  $\log$  geschrieben; oft ist mit  $\log$  auch  $\log_{10}$  gemeint.

### Einige Grenzwerte

Wir wollen einige mit der Exponential- und Logarithmusfunktion gebildete Grenzwerte bestimmen, die häufig auftreten.

**1.4 Behauptung.**  $\forall a \in \mathbb{R}^+ : \lim_{n \rightarrow \infty} \sqrt[n]{a} = 1.$

*Beweis.*  $\sqrt[n]{a} = e^{\frac{1}{n} \ln a} \rightarrow e^0 = 1$  für  $n \rightarrow \infty$  wegen der Stetigkeit der  $e$ -Funktion und Satz 2.6 aus Kapitel 4. ■

**1.5 Behauptung.**  $\forall k \in \mathbb{N} : \lim_{x \rightarrow \infty} \frac{e^x}{x^k} = \infty.$

*Beweis.* Für alle  $x > 0$  gilt

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} > \frac{x^{k+1}}{(k+1)!},$$

woraus die Behauptung folgt. ■

**1.6 Behauptung.**  $\forall k \in \mathbb{N} : \lim_{x \rightarrow \infty} x^k e^{-x} = 0.$

*Beweis.*  $x^k e^{-x} = \left(\frac{e^x}{x^k}\right)^{-1}$  und Behauptung 1.5. ■

**1.7 Behauptung.**  $\forall k \in \mathbb{N} : \lim_{x \searrow 0} x^k e^{\frac{1}{x}} = \infty.$

*Beweis.*

$$\lim_{x \searrow 0} x^k e^{\frac{1}{x}} = \lim_{y \rightarrow \infty} \left(\frac{1}{y}\right)^k e^y = \lim_{y \rightarrow \infty} \frac{e^y}{y^k} = \infty. \quad \blacksquare$$

**1.8 Behauptung.**  $\lim_{x \rightarrow \infty} \ln x = \infty, \quad \lim_{x \rightarrow 0} \ln x = -\infty.$

*Beweis.* Zu gegebenem  $c \in \mathbb{R}$  setze  $c_0 := e^c$ . Für alle  $x \geq c_0$  gilt dann  $\ln x \geq c$ . Also ist  $\lim_{x \rightarrow \infty} \ln x = \infty$ . Ferner gilt

$$\lim_{x \rightarrow 0} \ln x = \lim_{y \rightarrow \infty} \ln \frac{1}{y} = - \lim_{y \rightarrow \infty} \ln y = -\infty. \quad \blacksquare$$

**1.9 Behauptung.**  $\forall \alpha \in \mathbb{R}^+ : \lim_{x \searrow 0} x^\alpha = 0.$

*Beweis.* Zu gegebenem  $\varepsilon \in \mathbb{R}^+$  existiert wegen Behauptung 1.8 ein  $\delta \in \mathbb{R}^+$  mit  $\alpha \ln x < \ln \varepsilon$  für  $0 < x < \delta$ . Aus  $0 < x < \delta$  folgt dann

$$0 < x^\alpha = e^{\alpha \ln x} < e^{\ln \varepsilon} = \varepsilon. \quad \blacksquare$$

**1.10 Behauptung.**  $\forall \alpha \in \mathbb{R}^+ : \lim_{x \rightarrow \infty} \frac{\ln x}{x^\alpha} = 0.$

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Wegen Behauptung 1.5 existiert ein  $c_0 \in \mathbb{R}$  mit

$$0 < \frac{1}{\alpha} \frac{y}{e^y} < \varepsilon \quad \text{für } y > c_0.$$

Wegen Behauptung 1.8 existiert ein  $c \in \mathbb{R}$  mit  $\alpha \ln x > c_0$  für  $x > c$ . Für alle  $x > c$  gilt also

$$0 < \frac{\ln x}{x^\alpha} = \frac{1}{\alpha} \frac{\alpha \ln x}{e^{\alpha \ln x}} < \varepsilon. \quad \blacksquare$$

**1.11 Behauptung.**  $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1.$

*Beweis.*  $\sqrt[n]{n} = e^{\frac{1}{n} \ln n} \rightarrow e^0 = 1$  für  $n \rightarrow \infty$  wegen Behauptung 1.10.  $\blacksquare$

## 5.2 Die Exponentialfunktion im Komplexen

Um bequem die trigonometrischen Funktionen einführen und untersuchen zu können, ist es zweckmäßig, die Exponentialfunktion auch für komplexe

Argumente zur Verfügung zu haben. Wir werden daher jetzt die hierzu erforderlichen Grundbegriffe und Sätze der Analysis unter Zugrundelegung des Körpers der komplexen Zahlen entwickeln.

Ein Anlaß zur Einführung der komplexen Zahlen ist zum Beispiel die Tatsache, dass im Bereich der reellen Zahlen nicht jede „quadratische Gleichung“ lösbar ist. So gibt es etwa keine reelle Zahl  $x$  mit  $x^2 = -1$  (denn es ist  $-1 < 0$ , und nach Behauptung 2.9 aus Kapitel 1 gilt  $x^2 \geq 0$  für alle  $x \in \mathbb{R}$ ). Man erweitert daher den Bereich der reellen Zahlen in geeigneter Weise, was man formal folgendermaßen tun kann.

$\mathbb{C}$  sei die Menge  $\mathbb{R} \times \mathbb{R}$  der geordneten Paare reeller Zahlen. Addition und Multiplikation der Elemente von  $\mathbb{C}$  werden erklärt durch

$$\begin{aligned}(a, b) + (c, d) &:= (a + c, b + d) \\ (a, b) \cdot (c, d) &:= (ac - bd, ad + bc)\end{aligned}$$

für  $(a, b), (c, d) \in \mathbb{C}$ . Das Tripel  $(\mathbb{C}, +, \cdot)$  erfüllt die in Abschnitt 1.1 zusammengestellten Körperaxiome, wie man durch elementare Rechnungen leicht nachweist. Man bezeichnet  $(\mathbb{C}, +, \cdot)$  oder kurz  $\mathbb{C}$  als den *Körper der komplexen Zahlen* und die Elemente als komplexe Zahlen.

Wegen

$$\begin{aligned}(a, 0) + (c, 0) &= (a + c, 0) \\ (a, 0) \cdot (c, 0) &= (ac, 0)\end{aligned}$$

kann man die Teilmenge  $\{(a, 0) \mid a \in \mathbb{R}\}$  mit den darauf eingeschränkten Verknüpfungen identifizieren mit dem Körper der reellen Zahlen. Statt  $(a, 0)$  schreibt man daher  $a$ . Ferner setzt man zur Abkürzung

$$(0, 1) =: i$$

und nennt diese komplexe Zahl die *imaginäre Einheit*. Dann ist  $i^2 = -1$ .

Mit diesen Vereinbarungen ist jetzt für  $(a, b) \in \mathbb{C}$

$$(a, b) = (a, 0)(1, 0) + (b, 0)(0, 1) = a + bi.$$

Jede komplexe Zahl  $z \in \mathbb{C}$  kann also in der Form  $z = a + ib$  mit  $a, b \in \mathbb{R}$  dargestellt werden, und zwar eindeutig, denn aus  $a + ib = a' + ib'$  mit  $a, b, a', b' \in \mathbb{R}$ , also  $(a, b) = (a', b')$ , folgt  $a = a'$  und  $b = b'$  nach der Gleichheitsdefinition für geordnete Paare.

**Definition.** Sei  $z \in \mathbb{C}$ ,  $z = a + ib$  mit  $a, b \in \mathbb{R}$ . Dann heißt

$$\begin{aligned}\operatorname{Re}(z) &:= a && \text{der Realteil und} \\ \operatorname{Im}(z) &:= b && \text{der Imaginärteil}\end{aligned}$$

von  $z$ . Die komplexe Zahl

$$\bar{z} := a - ib$$

heißt die zu  $z$  *konjugiert-komplexe* Zahl.

**2.1 Behauptung.** Für  $z, z_1, z_2 \in \mathbb{C}$  gilt

$$(a) \operatorname{Re}(z) = \frac{1}{2}(z + \bar{z}), \operatorname{Im}(z) = \frac{1}{2i}(z - \bar{z}),$$

$$(b) \overline{\bar{z}} = z$$

$$(c) \overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$$

$$(d) \overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$$

*Beweis.* trivial (nachrechnen) ■

Nachdem wir den Körper der reellen Zahlen ausgedehnt haben zum Körper der komplexen Zahlen, stellt sich die Frage, ob auch die Beziehung  $<$  von den reellen Zahlen auf die komplexen Zahlen fortgesetzt werden kann. Dabei würde man natürlich fordern, dass die Anordnungsaxiome aus Abschnitt 1.2, die die Verträglichkeit dieser Kleinerbeziehung mit den Körperverknüpfungen regeln, erfüllt sind. Dies ist aber nicht möglich, da in einem angeordneten Körper nach Behauptung 2.9 aus Kapitel 1 für jedes Körperelement  $a$  die Beziehung  $a^2 \geq 0$  gilt, während doch  $i^2 = -1 < 0$  ist.

Eine eingehendere Durchsicht der Konvergenzbetrachtungen in den vorhergehenden Abschnitten zeigt nun aber, dass dabei gar nicht so sehr von der Größerbeziehung Gebrauch gemacht wurde, sondern vorwiegend von den daraus abgeleiteten Eigenschaften des Absolutbetrages. Ein Absolutbetrag mit ähnlichen Eigenschaften läßt sich nun auch für komplexe Zahlen erklären, und dies hat zur Folge, dass man auch im Bereich der komplexen Zahlen eine weitgehend analoge Konvergenztheorie aufbauen kann.

**Definition.** Für  $z = a + ib$  ( $a, b \in \mathbb{R}$ ) heißt

$$|z| := \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$$

der *Betrag* von  $z$ .

Der Betrag einer komplexen Zahl ist also eine nichtnegative reelle Zahl. Für  $z \in \mathbb{R}$  stimmt diese Definition mit der früher für den Absolutbetrag gegebenen überein. Für  $z \in \mathbb{C}$  ist  $|\bar{z}| = |z|$ , ferner  $|\operatorname{Re}(z)| \leq |z|$ ,  $|\operatorname{Im}(z)| \leq |z|$ .

**2.2 Satz.** Für  $z, z_1, z_2 \in \mathbb{C}$  gilt

(a)  $|z| \geq 0$ ;  $|z| = 0 \Leftrightarrow z = 0$ ,

(b)  $|z_1 z_2| = |z_1| |z_2|$ ,

(c)  $|z_1 + z_2| \leq |z_1| + |z_2|$  (Dreiecksungleichung)

*Beweis.* (a) ist trivial.

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

(c) Unter Verwendung von

$$\operatorname{Re}(z_1 \bar{z}_2) \leq |z_1 \bar{z}_2| = |z_1| |\bar{z}_2| = |z_1| |z_2|$$

folgt

$$\begin{aligned} |z_1 + z_2|^2 &= (z_1 + z_2)(\bar{z}_1 + \bar{z}_2) = z_1 \bar{z}_1 + z_1 \bar{z}_2 + z_2 \bar{z}_1 + z_2 \bar{z}_2 \\ &= |z_1|^2 + 2\operatorname{Re}(z_1 \bar{z}_2) + |z_2|^2 \\ &\leq |z_1|^2 + 2|z_1| |z_2| + |z_2|^2 = (|z_1| + |z_2|)^2. \quad \blacksquare \end{aligned}$$

Wir sind nun in der Lage, die Konvergenzbetrachtungen der vorhergehenden Paragraphen weitgehend vom Reellen ins Komplexe zu übertragen. Alle Behauptungen und Beweise, die nicht von der Anordnung der reellen Zahlen, sondern (außer von den Körperaxiomen) nur vom Absolutbetrag und seinen Eigenschaften Gebrauch machten, bleiben ohne Änderungen gültig. Die Resultate, deren Beweise sich wörtlich übertragen lassen, werden wir nur auflisten, ohne die Begründungen zu wiederholen.

**Definition.** Sei  $(z_n)_{n \in \mathbb{N}}$  eine Folge in  $\mathbb{C}$  und  $z \in \mathbb{C}$ . Die Folge  $(z_n)$  heißt *konvergent* gegen  $z$ , und  $z$  heißt *Grenzwert* dieser Folge, geschrieben  $\lim_{n \rightarrow \infty} z_n = z$ , wenn gilt

$$\forall \varepsilon \in \mathbb{R}^+ \quad \exists n_0 \in \mathbb{N} \quad \forall n \in \mathbb{N} : (n \geq n_0 \Rightarrow |z_n - z| < \varepsilon).$$

Die Sätze 1.1 (Eindeutigkeit des Grenzwertes), 1.2 (Beschränktheit konvergenter Folgen;  $(z_n)_{n \in \mathbb{N}}$  heißt *beschränkt*, wenn  $c \in \mathbb{R}$  existiert mit  $|z_n| \leq c \quad \forall n \in \mathbb{N}$ ), 1.3 (Rechenregeln für Grenzwerte) und ihre Beweise aus Kapitel 3 lassen sich wörtlich übertragen.

Andere Aussagen von früher lassen sich nicht verallgemeinern, weil sie in  $\mathbb{C}$  sinnlos sind. Zum Beispiel kann man nicht von einer monotonen Folge komplexer Zahlen sprechen.

Der folgende Satz gestattet es, Konvergenz im Komplexen zurückzuführen auf Konvergenz im Reellen:

**2.3 Satz.** Für jede Folge  $(z_n)_{n \in \mathbb{N}}$  in  $\mathbb{C}$  gilt:

$$(z_n)_{n \in \mathbb{N}} \text{ konvergent} \Leftrightarrow (\operatorname{Re}(z_n))_{n \in \mathbb{N}} \text{ und } (\operatorname{Im}(z_n))_{n \in \mathbb{N}} \text{ konvergent.}$$

Ist  $(z_n)_{n \in \mathbb{N}}$  konvergent, so gilt

$$\lim_{n \rightarrow \infty} z_n = \lim_{n \rightarrow \infty} \operatorname{Re}(z_n) + i \lim_{n \rightarrow \infty} \operatorname{Im}(z_n).$$

*Beweis.* Unter Verwendung der Abschätzungen

$$|a|, |b| \leq |a + ib| \leq |a| + |b|$$

ergibt sich dies folgendermaßen. Setze  $z_n = a_n + ib_n, z = a + ib$ .

„ $\Rightarrow$ “: Sei  $\lim_{n \rightarrow \infty} z_n = z$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|z_n - z| < \varepsilon$  für  $n \geq n_0$ . Für  $n \geq n_0$  gilt also  $|a_n - a| \leq |z_n - z| < \varepsilon$ . Somit gilt  $\lim_{n \rightarrow \infty} a_n = a$ . Analog folgt  $\lim_{n \rightarrow \infty} b_n = b$ .

„ $\Leftarrow$ “: Gelte  $\lim a_n = a$  und  $\lim b_n = b$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_1 \in \mathbb{N}$  mit  $|a_n - a| < \frac{\varepsilon}{2}$  für  $n \geq n_1$  und ein  $n_2 \in \mathbb{N}$  mit  $|b_n - b| < \frac{\varepsilon}{2}$  für  $n \geq n_2$ . Für  $n \geq n_0 := \max\{n_1, n_2\}$  gilt also  $|z_n - z| \leq |a_n - a| + |b_n - b| < \varepsilon$  mit  $z = a + ib$ . Daraus folgt die Behauptung. ■

**Definition.**  $(z_n)_{n \in \mathbb{N}}$  heißt *Cauchy-Folge*

$$\Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall m, n \geq n_0 : |z_m - z_n| < \varepsilon.$$

**2.4 Satz.** Die Folge  $(z_n)_{n \in \mathbb{N}}$  in  $\mathbb{C}$  ist genau dann eine Cauchy-Folge, wenn  $(\operatorname{Re}(z_n))_{n \in \mathbb{N}}$  und  $(\operatorname{Im}(z_n))_{n \in \mathbb{N}}$  Cauchy-Folgen sind.

*Beweis.* Analog zum Beweis von Satz 2.3. ■

Aus den beiden vorhergehenden Sätzen und dem Cauchy-Kriterium 1.8 aus Kapitel 3 im Reellen folgt jetzt sofort:

**2.5 Satz.** Eine Folge in  $\mathbb{C}$  ist genau dann konvergent, wenn sie eine Cauchy-Folge ist.

Reihen komplexer Zahlen und ihre Konvergenz werden genau wie im Reellen erklärt. So bezeichnet  $\sum_{k=1}^{\infty} z_k$  sowohl die Folge  $(s_n)_{n \in \mathbb{N}}$  der Partialsummen  $s_n = \sum_{k=1}^n z_k$  als auch, wenn die Reihe konvergiert, ihren Grenzwert, und  $\sum_{k=1}^{\infty} z_k = z$  bedeutet  $\lim_{n \rightarrow \infty} \sum_{k=1}^n z_k = z$ .

**Definition.** Die Reihe  $\sum_{k=1}^{\infty} z_k$  heißt *absolut konvergent*, wenn die Reihe

$$\sum_{k=1}^{\infty} |z_k|$$

konvergiert.

Die folgenden Sätze über Konvergenz und ihre Beweise aus Kapitel 3 lassen sich nun wörtlich übertragen.

Satz 2.3 Konvergenzkriterium von Cauchy für Reihen

Satz 2.6 Absolute Konvergenz  $\Rightarrow$  Konvergenz

Satz 2.7 Umordnungssatz

Satz 2.8 Majorantenkriterium

Satz 2.9 Quotientenkriterium

Satz 2.10 Wurzelkriterium

Satz 2.11 Cauchy-Produkt

Nun können wir durch wörtliche Übertragung der Beweise die Betrachtungen aus Abschnitt 3.3 über die Exponentialreihe ins Komplexe ausdehnen.

**2.6 Satz.** Für jedes  $z \in \mathbb{C}$  ist die Reihe

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

*absolut konvergent.*

Beweis wie für Satz 3.1 aus Kapitel 3.

**Definition.**  $\exp(z) := e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!}$  für  $z \in \mathbb{C}$ .

**2.7 Satz.** Für alle  $z, w \in \mathbb{C}$  gilt  $e^{z+w} = e^z e^w$ .

Beweis wie für Satz 3.3 aus Kapitel 3.

**Folgerung.**  $e^z \neq 0$  für alle  $z \in \mathbb{C}$ , denn  $e^z e^{-z} = 1$ .

**2.8 Behauptung.**  $e^{\bar{z}} = \overline{e^z}$  für  $z \in \mathbb{C}$ .

*Beweis.* Mit  $s_n(z) := \sum_{k=0}^n \frac{z^k}{k!}$  gilt nach Behauptung 2.1

$$s_n(\bar{z}) = \sum_{k=0}^n \frac{\bar{z}^k}{k!} = \overline{\sum_{k=0}^n \frac{z^k}{k!}} = \overline{s_n(z)},$$

also wegen Satz 2.3

$$\begin{aligned} \overline{e^z} &= \overline{\lim \operatorname{Re}(s_n(z)) + i \lim \operatorname{Im}(s_n(z))} \\ &= \lim \operatorname{Re}(s_n(z)) - i \lim \operatorname{Im}(s_n(z)) \\ &= \lim \overline{s_n(z)} = \lim s_n(\bar{z}) = e^{\bar{z}}. \end{aligned} \quad \blacksquare$$

**Folgerung.** Ist  $z \in \mathbb{C}$  rein imaginär, d.h.  $z = ix$  mit  $x \in \mathbb{R}$ , so ist  $|e^z| = 1$ .  
Denn für  $x \in \mathbb{R}$  gilt  $|e^{ix}|^2 = e^{ix} \overline{e^{ix}} = e^{ix} e^{-ix} = e^{ix} e^{-ix} = e^0 = 1$ .

### 5.3 Die trigonometrischen Funktionen

Das Ziel unseres Ausfluges ins Komplexe war es, jetzt in bequemer Weise die Funktionen Sinus und Cosinus einführen zu können.

**Definition.** Für  $x \in \mathbb{R}$  setzt man

$$\cos x := \operatorname{Re}(e^{ix}), \quad \sin x = \operatorname{Im}(e^{ix}).$$

Es gilt also

$$\begin{aligned} e^{ix} &= \cos x + i \sin x, \\ \cos x &= \frac{1}{2}(e^{ix} + e^{-ix}), \\ \sin x &= \frac{1}{2i}(e^{ix} - e^{-ix}). \end{aligned}$$

Daraus folgt sofort für  $x \in \mathbb{R}$

$$\begin{aligned} \cos(-x) &= \cos x, \\ \sin(-x) &= -\sin x, \\ \sin^2 x + \cos^2 x &= |e^{ix}|^2 = 1. \end{aligned}$$

Aus der Funktionalgleichung der Exponentialfunktion ergeben sich die Additionstheoreme der „trigonometrischen Funktionen“ sin und cos:

**3.1 Satz** (Additionstheoreme). Für  $x, y \in \mathbb{R}$  gilt

$$\begin{aligned}\cos(x+y) &= \cos x \cos y - \sin x \sin y, \\ \sin(x+y) &= \sin x \cos y + \cos x \sin y.\end{aligned}$$

*Beweis.* Es ist

$$\begin{aligned}\cos(x+y) + i \sin(x+y) &= e^{i(x+y)} = e^{ix} e^{iy} \\ &= (\cos x + i \sin x)(\cos y + i \sin y) \\ &= (\cos x \cos y - \sin x \sin y) + i(\sin x \cos y + \cos x \sin y).\end{aligned}$$

Vergleich von Real- und Imaginärteil ergibt die Behauptung. ■

**3.2 Satz.** Für  $x \in \mathbb{R}$  gilt

$$\begin{aligned}\cos x &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \dots, \\ \sin x &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} \dots\end{aligned}$$

Beide Reihen sind absolut konvergent.

*Beweis.* Die absolute Konvergenz folgt nach dem Majorantenkriterium aus der absoluten Konvergenz der Exponentialreihe. Es ist

$$\begin{aligned}\cos x + i \sin x &= e^{ix} = \sum_{n=0}^{\infty} \frac{i^n x^n}{n!} \\ &= \frac{1}{0!} + i \frac{x}{1!} - \frac{x^2}{2!} - i \frac{x^3}{3!} + \frac{x^4}{4!} + i \frac{x^5}{5!} - \frac{x^6}{6!} - i \frac{x^7}{7!} + \dots \\ &= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots\right) + i \left(\frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots\right),\end{aligned}$$

wobei Satz 2.3 benutzt wurde. Vergleich von Real- und Imaginärteil ergibt die Behauptung. ■

**3.3 Satz.** Für die durch

$$\begin{aligned}\cos x &= \sum_{k=0}^n (-1)^k \frac{x^{2k}}{(2k)!} + R_{2n+2}(x), \\ \sin x &= \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{(2k+1)!} + R_{2n+3}(x)\end{aligned}$$

definierten Restglieder gilt die Abschätzung

$$|R_m(x)| \leq \frac{|x|^m}{m!} \quad \text{für } |x| \leq m+1.$$

*Beweis.*

$$\begin{aligned} |R_{2n+2}(x)| &= \frac{|x|^{2n+2}}{(2n+2)!} \left| 1 - \frac{x^2}{(2n+3)(2n+4)} + \frac{x^4}{(2n+3)\cdots(2n+6)} - \cdots \right| \\ &= \frac{|x|^{2n+2}}{(2n+2)!} |1 - a_1 + a_2 - \dots| \end{aligned}$$

mit

$$a_k := \frac{x^{2k}}{(2n+3)\cdots(2n+2k+2)}.$$

Wegen

$$a_k = a_{k-1} \frac{x^2}{(2n+2k+1)(2n+2k+2)}$$

gilt für  $|x| \leq 2n+3$  die Ungleichungskette  $1 \geq a_1 \geq a_2 \geq \dots$ , also  $0 \leq 1 - a_1 + a_2 - a_3 + \dots \leq 1$ . Analog schließt man für  $R_{2n+3}(x)$ . ■

Als Folgerung notieren wir:

### 3.4 Behauptung.

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

*Beweis.* Nach Satz 3.3 ist  $|\sin x - x| \leq \frac{|x|^3}{6}$  für  $|x| \leq 4$ , woraus nach Division durch  $x$  die Behauptung folgt. ■

### 3.5 Satz. $\cos$ und $\sin$ sind stetig.

*Beweis.* Satz 3.2 aus Kapitel 3 gilt mit demselben Beweis auch für  $x \in \mathbb{C}$ . Für  $z, w \in \mathbb{C}$  ist also  $|e^z - e^w| \leq 2|e^w||z - w|$  für  $|z| \leq 1$ . Für  $x, a \in \mathbb{R}$  mit  $|x - a| \leq 1$  folgt also

$$|\cos x - \cos a| = |\operatorname{Re}(e^{ix} - e^{ia})| \leq |e^{ix} - e^{ia}| \leq 2|x - a|,$$

woraus die Stetigkeit von  $\cos$  folgt. Analog für  $\sin$ . ■

## Die Zahl $\pi$

Die Zahl  $\pi$ , klassisch definiert als das Verhältnis des Kreisumfangs zum Kreisdurchmesser, werden wir hier auf ganz andere Weise einführen. Erst wesentlich später ergibt sich der Zusammenhang mit dem Kreisumfang. Die nachfolgenden Behauptungen dienen zur Vorbereitung.

**3.6 Behauptung.**  $\sin x > 0$  für  $x \in (0, 2)$ .

*Beweis.* Nach Satz 3.3 ist

$$\sin x = x \left( 1 + \frac{R_3(x)}{x} \right)$$

mit

$$\left| \frac{R_3(x)}{x} \right| \leq \frac{|x|^2}{6} \quad \text{für } |x| \leq 4.$$

Für  $x \in (0, 2]$  gilt also  $\left| \frac{R_3(x)}{x} \right| \leq \frac{2}{3}$  und daher

$$\sin x = x \left( 1 + \frac{R_3(x)}{x} \right) \geq x \cdot \frac{1}{3} > 0. \quad \blacksquare$$

**3.7 Behauptung.** In  $[0, 2]$  ist  $\cos$  streng monoton fallend.

*Beweis.* Mit  $\frac{x+y}{2} =: u$ ,  $\frac{x-y}{2} =: v$  folgt aus Satz 3.1

$$\begin{aligned} \cos x &= \cos(u+v) = \cos u \cos v - \sin u \sin v \\ \cos y &= \cos(u-v) = \cos u \cos v + \sin u \sin v, \end{aligned}$$

also

$$-\cos x + \cos y = 2 \sin \frac{x+y}{2} \sin \frac{x-y}{2}.$$

Für  $0 \leq y < x \leq 2$  ist nach Behauptung 3.6 die rechte Seite positiv.  $\blacksquare$

**3.8 Behauptung.**  $\cos 2 < 0$ .

*Beweis.* Nach Satz 3.3 ist

$$\cos x = 1 - \frac{x^2}{2} + R_4(x) \quad \text{mit } |R_4(x)| \leq \frac{|x|^4}{4!} \quad \text{für } |x| \leq 5,$$

also insbesondere

$$\cos 2 = 1 - 2 + R_4(2) \quad \text{mit } |R_4(2)| \leq \frac{16}{24} = \frac{2}{3},$$

woraus die Behauptung folgt.  $\blacksquare$

Über die Funktion  $f := \cos|_{[0, 2]}$  wissen wir nun:  $f$  ist stetig,  $f(0) = 1$ ,  $f(2) < 0$ . Nach dem Zwischenwertsatz 3.4 aus Kapitel 4 existiert ein  $x_0 \in (0, 2)$  mit  $f(x_0) = 0$ . Wegen 3.7 gibt es nur ein solches  $x$ . Wir können also definieren:

**Definition.** Die eindeutig bestimmte Nullstelle der Funktion  $\cos$  im Intervall  $(0, 2)$  wird mit  $\frac{\pi}{2}$  bezeichnet.

Näherungsweise Berechnung ergibt

$$\pi = 3,141592653 \pm 10^{-9}.$$

Aus  $\cos \frac{\pi}{2} = 0$  ergeben sich einige spezielle Werte für die Funktionen  $\sin$ ,  $\cos$ ,  $\exp$ : Es ist

$$\sin^2 \frac{\pi}{2} = 1 - \cos^2 \frac{\pi}{2} = 1,$$

wegen Behauptung 3.6 also

$$\sin \frac{\pi}{2} = 1,$$

folglich

$$e^{i\frac{\pi}{2}} = i.$$

Daraus folgt  $e^{\pi i} = -1$ ,  $e^{2\pi i} = 1$ ,  $e^{\frac{3}{2}\pi i} = -i$  und hieraus

$x$	$0$	$ \pi/2 $	$\pi$	$3\pi/2$	$2\pi$
$\sin x$	$0$	$1$	$0$	$-1$	$0$
$\cos x$	$1$	$0$	$-1$	$0$	$1$

Zusammen mit den Additionstheoremen 3.1 liefert das:

**3.9 Satz.** Für alle  $x \in \mathbb{R}$  gilt

$$\begin{aligned} \cos(x + 2\pi) &= \cos x & \sin(x + 2\pi) &= \sin x \\ \cos(x + \pi) &= -\cos x & \sin(x + \pi) &= -\sin x \\ \cos\left(x + \frac{\pi}{2}\right) &= -\sin x & \sin\left(x + \frac{\pi}{2}\right) &= \cos x. \end{aligned}$$

Insbesondere ist

$$\cos\left(\frac{\pi}{2} - x\right) = \sin x, \quad \sin\left(\frac{\pi}{2} - x\right) = \cos x.$$

**3.10 Satz.**

$$\begin{aligned} \{x \in \mathbb{R} \mid \sin x = 0\} &= \{k\pi \mid k \in \mathbb{Z}\}, \\ \{x \in \mathbb{R} \mid \cos x = 0\} &= \left\{ \left(k + \frac{1}{2}\right)\pi \mid k \in \mathbb{Z} \right\}. \end{aligned}$$

*Beweis.*  $\sin k\pi = 0$  für  $k \in \mathbb{Z}$  ist klar. Sei jetzt  $x \in \mathbb{R}$  eine Zahl mit  $\sin x = 0$ . Wir können  $x = k\pi + x_1$  mit  $k \in \mathbb{Z}$  und  $x_1 \in [0, \pi)$  schreiben. Dann ist

$$\sin x_1 = \sin(x - k\pi) = \sin x \cos k\pi - \cos x \sin k\pi = 0.$$

Nun gilt nach Behauptung 3.6:  $\sin x_1 > 0$  für  $x_1 \in (0, \frac{\pi}{2})$ .

Für  $x_1 \in [\frac{\pi}{2}, \pi)$  ist  $\sin x_1 = \cos(\frac{\pi}{2} - x_1) = \cos(x_1 - \frac{\pi}{2}) > 0$  wegen  $x_1 - \frac{\pi}{2} \in [0, \frac{\pi}{2})$  und der Definition von  $\pi/2$ . Also muß  $x_1 = 0$  sein, d.h.  $x = k\pi$ .

Wegen  $\cos x = \sin(x + \frac{\pi}{2})$  folgt die zweite Gleichheit aus der ersten. ■

Nachdem wir jetzt die Nullstellen von  $\sin$  und  $\cos$  kennen, können wir definieren:

**Definition.**

$$\begin{aligned} \text{Tangens :} \quad \tan x &:= \frac{\sin x}{\cos x} && \text{für } x \in \mathbb{R} \setminus \left\{ \left(k + \frac{1}{2}\right)\pi \mid k \in \mathbb{Z} \right\} \\ \text{Cotangens :} \quad \cot x &:= \frac{\cos x}{\sin x} && \text{für } x \in \mathbb{R} \setminus \{k\pi \mid k \in \mathbb{Z}\}. \end{aligned}$$

Weitere wichtige spezielle Funktionen ergeben sich als Umkehrfunktionen der bisher eingeführten Funktionen. Da die trigonometrischen Funktionen aber nicht injektiv sind, existieren Umkehrfunktionen nur für geeignete Einschränkungen.

**3.11 Satz.** (a) Die Einschränkung  $\cos|_{[0, \pi]}$  ist streng monoton fallend und eine bijektive Abbildung auf  $[-1, 1]$ . Die Umkehrfunktion wird mit

$$\arccos : [-1, 1] \rightarrow \mathbb{R} \quad (\text{Arcus Cosinus})$$

bezeichnet.

(b)  $\sin|_{[-\frac{\pi}{2}, \frac{\pi}{2}]}$  ist streng monoton wachsend und bijektiv auf  $[-1, 1]$ . Die Umkehrfunktion wird mit

$$\arcsin : [-1, 1] \rightarrow \mathbb{R} \quad (\text{Arcus Sinus})$$

bezeichnet.

(c)  $\tan|_{(-\frac{\pi}{2}, \frac{\pi}{2})}$  ist streng monoton wachsend und bijektiv auf  $\mathbb{R}$ . Die Umkehrfunktion wird mit

$$\arctan : \mathbb{R} \rightarrow \mathbb{R} \quad (\text{Arcus Tangens})$$

bezeichnet.

(d)  $\cot|_{(0,\pi)}$  ist streng monoton fallend und bijektiv auf  $\mathbb{R}$ . Die Umkehrfunktion wird mit

$$\operatorname{arccot} : \mathbb{R} \rightarrow \mathbb{R} \quad (\text{Arcus Cotangens})$$

bezeichnet.

Wir verzichten auf die Beweise, da sie mit Hilfe der bisher bekannten Ergebnisse sehr leicht zu erbringen sind.

Als Anwendung leiten wir eine besondere Darstellung der komplexen Zahlen her:

**3.12 Satz.** Jede komplexe Zahl  $z \in \mathbb{C}$  ist darstellbar in der Form

$$z = re^{i\varphi}$$

mit  $r = |z|$  und einer reellen Zahl  $\varphi \in [0, 2\pi)$  (genannt das Argument von  $z$ ). Für  $z \neq 0$  ist  $\varphi$  eindeutig bestimmt.

*Beweis.* Für  $z = 0$  ist  $z = 0 \cdot e^{i\varphi}$  mit beliebigem  $\varphi$ . Sei  $z \neq 0$ ,  $r := |z|$  und  $\frac{z}{r} = x + iy$  mit  $x, y \in \mathbb{R}$ . Wegen  $|\frac{z}{r}| = 1$  ist  $x^2 + y^2 = 1$ , also  $x \in [-1, 1]$ . Daher ist  $\alpha := \arccos x$  definiert. Wegen  $\cos \alpha = x$  ist  $\sin^2 \alpha = 1 - \cos^2 \alpha = 1 - x^2 = y^2$ , also  $\sin \alpha = \pm y$ . Setze

$$\varphi := \begin{cases} \alpha, & \text{falls } \sin \alpha = y, \\ 2\pi - \alpha, & \text{falls } \sin \alpha = -y \text{ und } \alpha \neq 0. \end{cases}$$

Dann gilt  $\sin \varphi = y$  und  $\cos \varphi = \cos \alpha = x$ , also

$$e^{i\varphi} = \cos \varphi + i \sin \varphi = x + iy = \frac{z}{|z|}.$$

Aus  $e^{i\varphi} = e^{i\psi}$  folgt  $e^{i(\varphi-\psi)} = 1$ , also  $\cos(\varphi - \psi) = 1$  und  $\sin(\varphi - \psi) = 0$ . Ist o.B.d.A.  $\varphi \geq \psi$ , so folgt aus Satz 3.10  $\varphi - \psi = 0$  oder  $\pi$ , wegen  $\cos \pi = -1$  also  $\varphi - \psi = 0$ . ■

## 6 Differenzierbare Funktionen

### 6.1 Die Ableitung

Ein nützliches allgemeines Prinzip bei der Untersuchung komplizierter Funktionen besteht darin, diese durch einfache, besser bekannte Funktionen zu approximieren (d.h. anzunähern). Der Begriff der Approximation kann auf verschiedene Weisen verstanden werden. Im folgenden interessieren wir uns für Approximation durch affine Funktionen in der Nähe eines festen Punktes. Dieses Verfahren ist heutzutage schon aus dem täglichen Leben geläufig, zum Beispiel wenn man von „Geschwindigkeit“ spricht. Bei einer gleichförmigen Bewegung, das heißt einer solchen, bei der der zurückgelegte Weg proportional zur Zeit ist, versteht man unter der Geschwindigkeit das Verhältnis von Weg zu Zeit. Bei einer nicht gleichförmigen Bewegung kann man nur von einer momentanen Geschwindigkeit reden, und man versteht hierunter die Geschwindigkeit der gleichförmigen Bewegung, die der betrachteten im gegebenen Zeitpunkt besonders nahe kommt. Dass es eine solche gibt, läuft mathematisch auf die Forderung der Differenzierbarkeit hinaus. Eine Funktion soll im Punkt  $x$  differenzierbar heißen, wenn sie in  $x$  von erster Ordnung durch eine affine Funktion approximiert werden kann, d.h. genauer:

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  heißt *differenzierbar in  $x \in D$* , wenn  $x$  Häufungspunkt von  $D$  ist und es eine Zahl  $c \in \mathbb{R}$  gibt mit

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - ch}{h} = 0. \quad (1.1)$$

Gibt es eine solche Zahl  $c$ , so ist sie eindeutig bestimmt; sie wird mit  $f'(x)$  bezeichnet und heißt *Ableitung von  $f$  an der Stelle  $x$* .

Die Gleichung (1.1) ist äquivalent mit

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

Damit ist auch die Eindeutigkeit von  $c$  klar. Die Ableitung von  $f$  an der Stelle  $x$  ist also definiert durch

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{y \rightarrow x} \frac{f(y) - f(x)}{y - x},$$

falls dieser Grenzwert existiert.

**Bemerkung.** Der letzte Limes ergibt die bekannte geometrische Interpretation der Ableitung als „Steigung“ der „Tangente“ an den Graphen („Limes“ von Sekanten) von  $f$ .

**Bezeichnungen.** Üblich (aber anfechtbar) ist auch die Schreibweise

$$f'(x) = \frac{df}{dx}.$$

Besser wäre dann schon die Schreibweise

$$f'(x) = \left. \frac{df(y)}{dy} \right|_{y=x}.$$

**1.2 Satz.** Ist  $f : D \rightarrow \mathbb{R}$  in  $x$  differenzierbar, so ist  $f$  in  $x$  stetig.

*Beweis.* Aus

$$\lim_{y \rightarrow x} \frac{f(y) - f(x)}{y - x} = f'(x)$$

und  $\lim_{y \rightarrow x} (y - x) = 0$  folgt  $\lim_{y \rightarrow x} [f(y) - f(x)] = f'(x) \cdot 0$ . Also ist  $\lim_{y \rightarrow x} f(y) = f(x)$ . Nach Satz 2.6 aus Kapitel 4 ist  $f$  stetig in  $x$ . ■

**Bemerkung.** Es gibt stetige Funktionen  $f : \mathbb{R} \rightarrow \mathbb{R}$ , die in keinem Punkt differenzierbar sind.

**Definition.**  $f : D \rightarrow \mathbb{R}$  heißt *differenzierbar*, wenn  $f$  in jedem  $x \in D$  differenzierbar ist. In diesem Fall heißt die Funktion  $f' : D \rightarrow \mathbb{R} : x \mapsto f'(x)$ , die *Ableitung* von  $f$ .

**Beispiele.** (1) Sei  $c \in \mathbb{R}$  und  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto c$ . Wegen

$$\frac{f(x+h) - f(x)}{h} = 0$$

ist  $f$  differenzierbar und  $f'(x) = 0$  für alle  $x \in \mathbb{R}$ .

(2) Sei  $c \in \mathbb{R}$  und  $f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto cx$ . Wegen

$$\frac{f(x+h) - f(x)}{h} = c$$

ist  $f$  differenzierbar und  $f'(x) = c$  für alle  $x \in \mathbb{R}$ .

(3) exp. Es ist

$$\frac{e^{x+h} - e^x}{h} = e^x \frac{e^h - 1}{h}.$$

Nach Satz 3.2 aus Kapitel 3 ist

$$|e^h - 1 - h| = |R_2(h)| \leq 2 \frac{|h|^2}{2!} \quad \text{für } |h| \leq \frac{3}{2},$$

also

$$\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$$

und folglich

$$\exp'(x) = \exp(x) \quad \text{für alle } x \in \mathbb{R}.$$

Die Exponentialfunktion reproduziert sich also bei Differentiation. Dies ist eine Erklärung (von mehreren) für ihr häufiges Auftreten in Anwendungen.

(4) cos. Im Beweis von Behauptung 3.7 aus Kapitel 5 wurde berechnet

$$\cos y - \cos x = 2 \sin \frac{x+y}{2} \sin \frac{x-y}{2},$$

also ist

$$\frac{\cos(x+h) - \cos x}{h} = \frac{-2 \sin \left(x + \frac{h}{2}\right) \sin \frac{h}{2}}{h} = -\sin \left(x + \frac{h}{2}\right) \frac{\sin \frac{h}{2}}{\frac{h}{2}}.$$

Nun ist  $\lim_{h \rightarrow 0} \sin \left(x + \frac{h}{2}\right) = \sin x$  wegen der Stetigkeit von sin und

$$\lim_{h \rightarrow 0} \frac{\sin \frac{h}{2}}{\frac{h}{2}} = 1 \quad \text{nach Beh. 3.4 aus Kapitel 5,}$$

also

$$\lim_{h \rightarrow 0} \frac{\cos(x+h) - \cos x}{h} = -\sin x.$$

Somit gilt  $\cos' x = -\sin x$  für  $x \in \mathbb{R}$ .

(5) sin. Analog wie in (4) findet man  $\sin' x = \cos x$ .

(6) Die Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $x \mapsto |x|$  ist in 0 nicht differenzierbar, denn es ist

$$\lim_{h \searrow 0} \frac{|0+h| - |0|}{h} = 1 \neq -1 = \lim_{h \nearrow 0} \frac{|0+h| - |0|}{h}.$$

Hier existieren aber einseitige Ableitungen; sie sind naheliegenderweise wie folgt definiert.

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  heißt *in  $x$  von rechts differenzierbar*, wenn

$$f'_r(x) := \lim_{h \searrow 0} \frac{f(x+h) - f(x)}{h}$$

existiert, und *in  $x$  von links differenzierbar*, wenn

$$f'_l(x) := \lim_{h \nearrow 0} \frac{f(x+h) - f(x)}{h}$$

existiert.

Die Berechnung der Ableitung von komplizierteren, „zusammengesetzten“ Funktionen wird ermöglicht durch einige allgemeine Sätze, die sich auf die Differentiation von Summen, Produkten, Quotienten, Umkehrfunktionen und Kompositionen differenzierbarer Funktionen beziehen.

**1.3 Satz.** Die Funktionen  $f, g : D \rightarrow \mathbb{R}$  seien differenzierbar in einem Punkt  $x \in D$ . Dann sind die Funktionen  $f + g$ ,  $fg$ ,  $\lambda f$  ( $\lambda \in \mathbb{R}$ ) in  $x$  differenzierbar, und es gilt

$$\begin{aligned} (f + g)'(x) &= f'(x) + g'(x), \\ (fg)'(x) &= f'(x)g(x) + f(x)g'(x) \quad (\text{Produktregel}), \\ (\lambda f)'(x) &= \lambda f'(x). \end{aligned}$$

Ist  $g(x) \neq 0$ , so ist die Funktion  $\frac{f}{g}$  differenzierbar in  $x$ , und es gilt

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \quad (\text{Quotientenregel}).$$

*Beweis.* Die Behauptung für  $f + g$  folgt unmittelbar aus den Definitionen. Um  $fg$  zu behandeln, nutzen wir

$$\begin{aligned} \frac{(fg)(x+h) - (fg)(x)}{h} &= \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \frac{f(x+h) - f(x)}{h}g(x) + f(x+h)\frac{g(x+h) - g(x)}{h}. \end{aligned}$$

Daraus und aus der Stetigkeit von  $f$  folgt

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

Als Spezialfall ergibt sich  $(\lambda f)'(x) = \lambda f'(x)$ .

Sei  $g(x) \neq 0$ . Dann ist  $g(y) \neq 0$  für alle  $y$  aus einer Umgebung von  $x$ . Es ist

$$\begin{aligned} \frac{\left(\frac{1}{g}\right)(x+h) - \left(\frac{1}{g}\right)(x)}{h} &= \frac{1}{h} \left( \frac{1}{g(x+h)} - \frac{1}{g(x)} \right) \\ &= -\frac{1}{g(x+h)g(x)} \frac{g(x+h) - g(x)}{h}, \end{aligned}$$

woraus

$$\left(\frac{1}{g}\right)'(x) = -\frac{g'(x)}{g(x)^2}$$

folgt. Sodann ergibt sich

$$\begin{aligned} \left(\frac{f}{g}\right)'(x) &= \left(f \cdot \frac{1}{g}\right)'(x) = f'(x)\frac{1}{g}(x) + f(x) \left(-\frac{g'(x)}{g(x)^2}\right) \\ &= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}. \quad \blacksquare \end{aligned}$$

**Beispiele.** (1) Sei  $f(x) = x^n$  für  $x \in \mathbb{R}$  ( $n \in \mathbb{N}$ ).

**Behauptung.**  $f'(x) = nx^{n-1}$ .

*Beweis.* durch Induktion. Für  $n = 1$  ist die Behauptung richtig. Sei sie bewiesen für  $n$ . Sei  $f(x) = x^{n+1}$ , also  $f(x) = x^n \cdot x$  und daher

$$f'(x) = nx^{n-1} \cdot x + x^n \cdot 1 = (n+1)x^n. \quad \blacksquare$$

(2) Sei  $f(x) = x^{-n}$  für  $x \in \mathbb{R}$ , ( $n \in \mathbb{N}$ ). Nach der Quotientenregel und Beispiel (1) folgt

$$f'(x) = \frac{-nx^{n-1}}{(x^n)^2} = -nx^{-n-1}.$$

(3) Aus  $\tan x = \frac{\sin x}{\cos x}$  folgt

$$\begin{aligned} \tan'(x) &= \frac{\sin' x \cos x - \sin x \cos' x}{\cos^2 x} \\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}. \end{aligned}$$

Der folgende Satz zeigt, wie man Umkehrfunktionen differenzierbarer Funktionen differenziert.

**1.4 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  injektiv, in  $a \in D$  differenzierbar, und  $f'(a) \neq 0$ . Ist die Umkehrfunktion  $\varphi : f(D) \rightarrow D$  von  $f$  in  $b = f(a)$  stetig, so ist  $\varphi$  in  $b$  differenzierbar, und es gilt

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

*Beweis.* Setze

$$g(x) := \begin{cases} \frac{f(x) - f(a)}{x - a} & \text{für } x \in D \setminus \{a\}, \\ f'(a) & \text{für } x = a. \end{cases}$$

Dann ist  $g$  an der Stelle  $a$  stetig, also ist nach Satz 2.8 aus Kapitel 4 und Voraussetzung die Funktion  $g \circ \varphi$  an der Stelle  $b$  stetig. Somit gilt

$$\lim_{y \rightarrow b} (g \circ \varphi)(y) = (g \circ \varphi)(b) = g(\varphi(b)) = g(a) = f'(a) \neq 0$$

und daher

$$\lim_{y \rightarrow b} \frac{1}{(g \circ \varphi)(y)} = \frac{1}{f'(a)}.$$

Für  $y \in f(D)$  mit  $y \neq b$  ist  $\varphi(y) \neq a$ , also

$$\frac{1}{(g \circ \varphi)(y)} = \frac{1}{\frac{f(\varphi(y)) - f(a)}{\varphi(y) - a}} = \frac{\varphi(y) - a}{f(\varphi(y)) - f(a)},$$

woraus die Behauptung folgt. ■

**Beispiele.** In den folgenden Beispielen sind Differenzierbarkeit der Funktion  $f$  und Stetigkeit der Umkehrfunktion  $\varphi$  jeweils bereits bekannt. Wir benutzen

$$\varphi'(x) = \frac{1}{f'(\varphi(x))}.$$

(1)  $\ln$  ist die Umkehrfunktion von  $\exp$ , und es ist  $\exp' x = \exp x \neq 0$ , also

$$\ln' x = \frac{1}{\exp(\ln x)} = \frac{1}{x}.$$

- (2)  $\arccos$  ist die Umkehrfunktion von  $\cos$   $|[0, \pi]$ . Für  $y \in (0, \pi)$  gilt  $\cos' y = -\sin y \neq 0$ , also für  $x \in (-1, 1)$

$$\arccos' x = \frac{1}{\cos'(\arccos x)} = \frac{-1}{\sin(\arccos x)}.$$

Setzen wir  $\arccos x = y$ , so ist  $\cos y = x$ , also  $\sin^2 y = 1 - \cos^2 y = 1 - x^2$ , folglich  $\sin y = \sqrt{1 - x^2}$  ( $-\sqrt{1 - x^2}$  kommt nicht in Frage wegen  $\sin y > 0$ ). Es folgt

$$\arccos' x = \frac{-1}{\sqrt{1 - x^2}} \quad \text{für } -1 < x < 1.$$

- (3) Analog findet man

$$\arcsin' x = \frac{1}{\sqrt{1 - x^2}} \quad \text{für } -1 < x < 1.$$

- (4)  $\arctan$  ist die Umkehrfunktion von  $\tan$   $|(-\frac{\pi}{2}, \frac{\pi}{2})$ . Für  $y \in (-\frac{\pi}{2}, \frac{\pi}{2})$  ist  $\tan' y = \frac{1}{\cos^2 y} \neq 0$ , also für  $x \in \mathbb{R}$

$$\arctan' x = \cos^2(\arctan x).$$

Mit  $y := \arctan x$  ist  $x = \tan y = \frac{\sin y}{\cos y}$ . Daraus folgt  $x^2 \cos^2 y = 1 - \cos^2 y$ , was man auch als  $\cos^2 y = \frac{1}{1+x^2}$  schreiben kann. Somit haben wir

$$\arctan' x = \frac{1}{1+x^2}.$$

**1.5 Satz** (Kettenregel). Sei  $f : D \rightarrow \tilde{D}$  differenzierbar in  $a \in D$  und sei  $g : \tilde{D} \rightarrow \mathbb{R}$  differenzierbar in  $f(a)$ . Dann ist  $g \circ f$  differenzierbar in  $a$ , und es gilt

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

*Beweis.* Setze

$$h(y) := \begin{cases} \frac{g(y) - g(f(a))}{y - f(a)} & \text{für } y \in \tilde{D} \setminus \{f(a)\}, \\ g'(f(a)) & \text{für } y = f(a). \end{cases}$$

Dann ist  $h$  an der Stelle  $f(a)$  stetig, also ist die Funktion  $h \circ f$  an der Stelle  $a$  stetig. Für  $x \in D \setminus \{a\}$  gilt

$$\begin{aligned}
& \frac{(g \circ f)(x) - (g \circ f)(a)}{x - a} \\
&= \begin{cases} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \frac{f(x) - f(a)}{x - a}, & \text{falls } f(x) \neq f(a) \\ 0, & \text{falls } f(x) = f(a) \end{cases} \\
&= h(f(x)) \frac{f(x) - f(a)}{x - a}.
\end{aligned}$$

Aus  $\lim_{x \rightarrow a} h(f(x)) = h(f(a)) = g'(f(a))$  und

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

folgt also die Behauptung. ■

**Beispiele.** (1) Sei  $\alpha \in \mathbb{R}$ ,  $f(x) = x^\alpha$  für  $x > 0$ . Es ist  $f(x) = e^{\alpha \ln x} = \exp(g(x))$  mit  $g(x) = \alpha \ln x$ . Es folgt

$$f'(x) = \exp'(g(x))g'(x) = e^{\alpha \ln x} \frac{\alpha}{x} = \alpha x^{\alpha-1}.$$

(2)  $f(x) = \ln \sin x$  für  $0 < x < \pi$ . Es ist  $f'(x) = \ln'(\sin x) \cdot \sin' x = \frac{1}{\sin x} \cdot \cos x = \cotan x$ .

(3)  $f(x) = \cos e^{(x^2)}$ . Es ist

$$f'(x) = -\left(\sin e^{(x^2)}\right) \cdot e^{(x^2)} \cdot 2x.$$

(4)  $f(x) = x^{\frac{1}{x}}$  für  $x > 0$ . Es ist  $f(x) = e^{\frac{1}{x} \ln x}$ , also

$$f'(x) = e^{\frac{1}{x} \ln x} \left(-\frac{1}{x^2} \ln x + \frac{1}{x} \cdot \frac{1}{x}\right) = x^{\frac{1}{x}-2}(1 - \ln x).$$

## 6.2 Eigenschaften differenzierbarer Funktionen

Bei differenzierbaren Funktionen steht man oft vor dem Problem, von Eigenschaften der Ableitung auf Eigenschaften der Funktion selbst schließen zu müssen. Zentrales Hilfsmittel hierbei ist der Mittelwertsatz der Differentialrechnung mit seinen Folgerungen. Der wesentliche Kern dieses Satzes steckt bereits in der folgenden Behauptung.

**2.1 Satz** (von Rolle). Sei  $a < b$ ,  $f : [a, b] \rightarrow \mathbb{R}$  stetig,  $f|_{(a, b)}$  differenzierbar und  $f(a) = f(b)$ . Dann existiert ein  $c \in (a, b)$  mit  $f'(c) = 0$ .

*Beweis.* Nach Satz 3.2 aus Kapitel 4 nimmt  $f$  ein Maximum und ein Minimum an. Mindestens eines von beiden, o.B.d.A. ein Maximum, wird an einer Stelle  $c \in (a, b)$  angenommen (wegen  $f(a) = f(b)$ ). Es ist

$$f'(c) = \lim_{h \searrow 0} \frac{f(c+h) - f(c)}{h} \leq 0$$

und

$$f'(c) = \lim_{h \nearrow 0} \frac{f(c+h) - f(c)}{h} \geq 0,$$

also  $f'(c) = 0$ . ■

**2.2 Satz** (Mittelwertsatz der Differentialrechnung). Sei  $a < b$ ,  $f : [a, b] \rightarrow \mathbb{R}$  stetig und  $f|_{(a, b)}$  differenzierbar. Dann existiert ein  $c \in (a, b)$  mit

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

*Beweis.* Setze

$$g(x) := f(x) - \frac{f(b) - f(a)}{b - a}(x - a) \quad \text{für } x \in [a, b]$$

und wende Satz 2.1 an. ■

**Bemerkung.** Wir können das Ergebnis auch etwas anders schreiben: Ist  $f$  in  $[x - |h|, x + |h|]$  differenzierbar, so gilt

$$f(x+h) = f(x) + f'(x + \vartheta h)h \quad \text{mit einem } \vartheta \in (0, 1).$$

Wir können also den „Zuwachs“  $f(x+h) - f(x)$  nicht nur näherungsweise durch  $f'(x)h$ , sondern exakt durch  $f'(x + \vartheta h)h$  ausdrücken. Allerdings wissen wir von  $\vartheta$  nur, dass es in  $(0, 1)$  liegt.

Wir kommen zu einigen Folgerungen aus dem Mittelwertsatz. Von Interesse in Anwendungen ist statt der schärferen Aussage des Mittelwertsatzes oft nur die folgende unmittelbare Konsequenz.

**2.3 Folgerung.** Sei  $D$  ein Intervall und  $f : D \rightarrow \mathbb{R}$  differenzierbar mit  $|f'(x)| \leq c$  für  $x \in D$  ( $c \in \mathbb{R}$  fest). Dann gilt

$$|f(x) - f(y)| \leq c|x - y| \quad \text{für } x, y \in D. \quad (2.4)$$

**Bemerkung.** Eine Funktion  $f$  mit der Eigenschaft (2.4) nennt man auch “Lipschitz stetig“.

Anwendung mit  $c = 0$  ergibt:

**2.5 Satz.** Sei  $D$  ein Intervall,  $f : D \rightarrow \mathbb{R}$  differenzierbar und  $f'(x) = 0$  für alle  $x \in D$ . Dann ist  $f$  konstant.

**Bemerkung.** Die Voraussetzung, dass der Definitionsbereich von  $f$  ein Intervall ist, ist offenbar wesentlich.

Die folgende Beobachtung ist in den Anwendungen wichtig:

**Behauptung.** Ist  $f : [0, b] \rightarrow \mathbb{R}$  eine differenzierbare Funktion mit

$$f'(x) = cf(x) \quad \text{für } x \in [0, b] \quad \text{und } f(0) = a$$

( $a, b, c \in \mathbb{R}$  gegeben), so ist

$$f(x) = ae^{cx} \quad \text{für } x \in [0, b].$$

*Beweis.* Setze  $g(x) := f(x)e^{-cx}$  für  $x \in [0, b]$ . Dann ist

$$g'(x) = f'(x)e^{-cx} - f(x)ce^{-cx} = (f'(x) - cf(x))e^{-cx} = 0$$

für alle  $x \in [0, b]$ , also ist  $g$  konstant, und zwar gleich  $g(0) = a$ . ■

**2.6 Satz.** Sei  $D$  ein Intervall und  $f : D \rightarrow \mathbb{R}$  differenzierbar. Dann ist  $f$  genau dann monoton wachsend, wenn  $f'(x) \geq 0$  für  $x \in D$  gilt.

*Beweis.* „ $\Rightarrow$ “: Sei  $f$  monoton wachsend. Dann gilt für  $x \in D$ ,  $h \neq 0$  und  $x + h \in D \setminus \{x\}$ :

$$\frac{f(x+h) - f(x)}{h} \geq 0,$$

also  $f'(x) \geq 0$ .

„ $\Leftarrow$ “: Nach Satz 2.2 gilt für  $x, y \in D$  mit  $y > x$

$$f(y) - f(x) = f'(c)(y - x) \geq 0$$

mit einem geeigneten  $c \in (x, y)$ . ■

**Bemerkung.** Aus  $f'(x) > 0$  für alle  $x \in D$  folgt offenbar, dass  $f$  streng monoton wachsend ist (aber nicht umgekehrt).

Wir beweisen noch eine Verallgemeinerung des Mittelwertsatzes, die vor allem bei der Ermittlung von Grenzwerten gute Dienste leistet:

**2.7 Satz** (Zweiter Mittelwertsatz der Differentialrechnung). *Sei  $a < b$ , seien  $f, g : [a, b] \rightarrow \mathbb{R}$  stetig,  $f|_{(a,b)}$  und  $g|_{(a,b)}$  differenzierbar. Dann existiert ein  $c \in (a, b)$  mit*

$$g'(c)(f(b) - f(a)) = f'(c)(g(b) - g(a)).$$

*Beweis.* Setze

$$h(x) := (g(x) - g(a))(f(b) - f(a)) - (f(x) - f(a))(g(b) - g(a))$$

für  $x \in [a, b]$  und wende den Satz von Rolle 2.1 an. ■

Als Anwendung haben wir die folgende, oft nützliche Regel.

**2.8 Satz** (Regel von de l'Hospital). *Seien  $f, g : (a, b] \rightarrow \mathbb{R}$  differenzierbar. Es gelte*

$$\begin{aligned} \lim_{x \searrow a} f(x) = \lim_{x \searrow a} g(x) = 0, \\ g(x) \neq 0, \quad g'(x) \neq 0 \quad \text{für } x \in (a, b], \\ \lim_{x \searrow a} \frac{f'(x)}{g'(x)} = c. \end{aligned}$$

*Dann ist*

$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = c.$$

*Beweis.* Wir setzen noch  $f(a) = g(a) = 0$ ; dann sind  $f$  und  $g$  stetige Funktionen auf  $[a, b]$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Nach Voraussetzung existiert ein  $\delta \in \mathbb{R}^+$  mit

$$\left| \frac{f'(y)}{g'(y)} - c \right| < \varepsilon \quad \text{für } a < y < a + \delta.$$

Sei jetzt  $a < x < a + \delta$ . Nach Satz 2.7 existiert ein  $y \in (a, x)$  mit

$$\frac{f(x)}{g(x)} = \frac{f'(y)}{g'(y)}.$$

Es folgt

$$\left| \frac{f(x)}{g(x)} - c \right| = \left| \frac{f'(y)}{g'(y)} - c \right| < \varepsilon.$$

Damit ist

$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = c$$

gezeigt. ■

**Bemerkung.** Die analoge Aussage für  $\lim_{x \nearrow a}$  liegt auf der Hand. Aus beiden folgt eine entsprechende Aussage über  $\lim_{x \rightarrow a}$ . Durch eine naheliegende Abwandlung des Beweises zeigt man auch eine entsprechende Aussage für  $\lim_{x \rightarrow \infty}$  bzw.  $\lim_{x \rightarrow -\infty}$ .

Nützlich ist oft auch die folgende Version der Regel von de l'Hospital, bei der Zähler und Nenner des fraglichen Quotienten nicht gegen 0 konvergieren, sondern bestimmt divergieren.

**2.9 Satz.** Seien  $f, g : (a, b] \rightarrow \mathbb{R}$  differenzierbar. Es gelte

$$\begin{aligned} \lim_{x \searrow a} f(x) = \infty, \quad \lim_{x \searrow a} g(x) = \infty, \\ g(x) \neq 0, \quad g'(x) \neq 0 \quad \text{für } x \in (a, b], \\ \lim_{x \searrow a} \frac{f'(x)}{g'(x)} = c. \end{aligned}$$

Dann ist

$$\lim_{x \searrow a} \frac{f(x)}{g(x)} = c.$$

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$ . Nach Voraussetzung existiert ein  $d > a$  mit

$$\left| \frac{f'(y)}{g'(y)} - c \right| < \frac{\varepsilon}{4} \quad \text{für } a < y < d. \quad (\star)$$

Es gibt ein  $\delta \in \mathbb{R}^+$  mit  $a + \delta \leq d$ , so dass für alle  $x$  mit  $a < x < a + \delta$  gilt:

$$\begin{aligned} 0 \neq g(x) \neq g(d), \quad 0 \neq f(x) \neq f(d), \\ \left| \frac{1 - \frac{g(d)}{g(x)}}{1 - \frac{f(d)}{f(x)}} \right| < 2, \quad |c| \left| \frac{1 - \frac{g(d)}{g(x)}}{1 - \frac{f(d)}{f(x)}} - 1 \right| < \frac{\varepsilon}{2}. \end{aligned}$$

Für alle  $x$  mit  $a < x < a + \delta$  gilt dann wegen

$$\begin{aligned} \frac{f(x)}{g(x)} - c &= \left( \frac{f(x) - f(d)}{g(x) - g(d)} - c \right) \frac{f(x)}{f(x) - f(d)} \frac{g(x) - g(d)}{g(x)} \\ &\quad + c \left( \frac{f(x)}{f(x) - f(d)} \frac{g(x) - g(d)}{g(x)} - 1 \right) \end{aligned}$$

die Abschätzung

$$\left| \frac{f(x)}{g(x)} - c \right| \leq \left| \frac{f(x) - f(d)}{g(x) - g(d)} - c \right| \underbrace{\left| \frac{1 - \frac{g(d)}{g(x)}}{1 - \frac{f(d)}{f(x)}} \right|}_{< 2} + |c| \underbrace{\left| \frac{1 - \frac{g(d)}{g(x)}}{1 - \frac{f(d)}{f(x)}} - 1 \right|}_{< \frac{\varepsilon}{2}}$$

Nach Satz 2.7 existiert ein  $y \in (x, d)$  mit

$$\frac{f(x) - f(d)}{g(x) - g(d)} = \frac{f'(y)}{g'(y)}.$$

Dies und die Abschätzung (\*) ergeben also

$$\left| \frac{f(x) - f(d)}{g(x) - g(d)} - c \right| < \frac{\varepsilon}{4}.$$

Damit ist

$$\left| \frac{f(x)}{g(x)} - c \right| < \varepsilon$$

gezeigt, woraus die Behauptung folgt. ■

**Bemerkung.** Analoge Aussagen gelten wieder für  $\lim_{x \nearrow a}$  etc.

**Beispiele.** (1)  $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$ , denn  $\lim_{x \rightarrow 0} \sin x = \lim_{x \rightarrow 0} x = 0$  und  $\lim_{x \rightarrow 0} \frac{\cos x}{1} = 1$ .

(2)  $\lim_{x \rightarrow 0} \left( \frac{1}{\sin x} - \frac{1}{x} \right) = ?$  Es ist

$$\frac{x - \sin x}{x \sin x} =: \frac{f(x)}{g(x)} \quad \text{mit} \quad \lim_{x \rightarrow 0} f(x) = \lim_{x \rightarrow 0} g(x) = 0,$$

$$f'(x) = 1 - \cos x, \quad g'(x) = \sin x + x \cos x,$$

$$\lim_{x \rightarrow 0} f'(x) = 0, \quad \lim_{x \rightarrow 0} g'(x) = 0,$$

$$(f')'(x) = \sin x, \quad (g')'(x) = 2 \cos x - x \sin x,$$

$$\lim_{x \rightarrow 0} \frac{(f')'(x)}{(g')'(x)} = 0 \Rightarrow \lim_{x \rightarrow 0} \frac{f'(x)}{g'(x)} = 0 \Rightarrow \lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = 0.$$

(3)  $\lim_{x \searrow 0} x^\alpha \ln x = ?$  ( $\alpha > 0$ ) Es ist

$$-x^\alpha \ln x = \frac{\ln \frac{1}{x}}{x^{-\alpha}} =: \frac{f(x)}{g(x)},$$

$$\lim_{x \searrow 0} f(x) = \infty, \quad \lim_{x \searrow 0} g(x) = \infty,$$

$$f'(x) = -\frac{1}{x}, \quad g'(x) = -\alpha x^{-\alpha-1},$$

$$\frac{f'(x)}{g'(x)} = \frac{1}{\alpha} x^\alpha, \quad \lim_{x \searrow 0} \frac{f'(x)}{g'(x)} = 0 \Rightarrow \lim_{x \searrow 0} x^\alpha \ln x = 0.$$

### 6.3 Höhere Ableitungen und Taylorformel

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion,  $a \in D$ . Falls  $f$  in einer Umgebung von  $a$  (geschnitten mit  $D$ ) differenzierbar und  $f'$  in  $a$  differenzierbar ist, heißt  $f$  *zweimal differenzierbar in  $a$* , und

$$(f')'(a) =: f''(a) =: f^{(2)}(a) =: \left. \frac{d^2 f(x)}{dx^2} \right|_{x=a}$$

heißt *zweite Ableitung von  $f$  in  $a$* .

Rekursive Definition: Falls  $f$  in einer Umgebung von  $a$  (geschnitten mit  $D$ )  $k$ -mal differenzierbar und  $f^{(k)}$  in  $a$  differenzierbar ist, heißt  $f$   *$(k+1)$ -mal differenzierbar in  $a$* , und

$$f^{(k)'}(a) =: f^{(k+1)}(a) =: \left. \frac{d^{k+1} f(x)}{dx^{k+1}} \right|_{x=a}$$

heißt  *$(k+1)$ -te Ableitung von  $f$  in  $a$* . Setze noch  $f^{(0)} = f$ ,  $f^{(1)} = f'$ .

$f$  heißt  *$k$ -mal differenzierbar*, wenn  $f$   $k$ -mal differenzierbar in  $a$  für alle  $a \in D$  ist.

$f$  heißt  *$k$ -mal stetig differenzierbar*, wenn  $f$   $k$ -mal differenzierbar und  $f^{(k)}$  stetig ist.

**Beispiele.** Die bisher betrachteten speziellen Funktionen, also rationale Funktionen, exp, ln, sin, cos, tan, arcsin etc. und alle Funktionen, die aus

solchen durch rationale Rechenoperationen und Komposition zusammengesetzt sind, führen bei Differentiation auf Funktionen vom selben Typ; sie sind daher beliebig oft differenzierbar (im jeweiligen Definitionsbereich).

Wir bringen noch ein weniger triviales Beispiel, das verschiedentlich von Nutzen ist. Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  definiert durch

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

**Behauptung.**  $f$  ist beliebig oft differenzierbar.

*Beweis.* An jeder Stelle  $x \neq 0$  ist  $f$  beliebig oft differenzierbar (trivial für  $x < 0$ , klar nach vorstehender Bemerkung für  $x > 0$ ). Für  $x > 0$  gilt

$$f^{(k)}(x) = e^{-\frac{1}{x}} \frac{p_k(x)}{x^{2k}} \quad (3.1)$$

mit einer gewissen Polynomfunktion  $p_k$ . Dies beweisen wir durch Induktion. Zunächst ist

$$f^{(1)}(x) = f'(x) = e^{-\frac{1}{x}} \cdot \frac{1}{x^2}.$$

Für eine Zahl  $k \geq 1$  sei die Darstellung (3.1) bereits gezeigt. Dann ist

$$\begin{aligned} f^{(k+1)}(x) &= e^{-\frac{1}{x}} \left( \frac{1}{x^2} \frac{p_k(x)}{x^{2k}} + \frac{p_k'(x)x^{2k} - p_k(x)2kx^{2k-1}}{x^{4k}} \right) \\ &= e^{-\frac{1}{x}} \frac{p_k(x) + p_k'(x)x^2 - 2kp_k(x)x}{x^{2(k+1)}}, \end{aligned}$$

woraus die Behauptung folgt.

Nun zeigen wir durch Induktion, dass  $f$  in 0  $k$ -mal differenzierbar ist. Für  $x > 0$  gilt

$$\frac{f(x) - f(0)}{x} = \frac{1}{x} e^{-\frac{1}{x}} \rightarrow 0$$

für  $x \searrow 0$  nach Behauptung 1.7 aus Kapitel 5, also  $f_r'(0) = 0$ . Da auch  $f_\ell'(0) = 0$  ist, ist  $f'(0) = 0$ .

Sei bereits bewiesen, dass  $f$  in 0  $k$ -mal differenzierbar ist. Dann existiert also die  $k$ -te Ableitung  $f^{(k)} : \mathbb{R} \rightarrow \mathbb{R}$ . Für  $x > 0$  gilt

$$\frac{f^{(k)}(x) - f^{(k)}(0)}{x} = \frac{p_k(x)}{x^{2k+1}} e^{-\frac{1}{x}} \rightarrow 0$$

für  $x \searrow 0$  nach Behauptung 1.7 aus Kapitel 5, also  $f_r^{(k)'}(0) = 0$ . Da auch  $f_\ell^{(k)'}(0) = 0$  ist, ist  $f^{(k)'}(0) = 0$ . ■

Differenzierbarkeit einer Funktion bedeutet, grob gesagt, dass sich die Funktion durch eine affine Funktion, also eine Polynomfunktion vom Grad 1, gut approximieren lässt. Analog kann man  $k$ -malige Differenzierbarkeit so interpretieren, dass die Funktion durch eine Polynomfunktion vom Grad  $k$  gut approximiert werden kann, und zwar „von  $k$ -ter Ordnung“. Zunächst zeigen wir unter einer schärferen Voraussetzung eine genauere Aussage.

**3.2 Satz** (Taylorscher Satz). *Sei  $a < b$ ,  $n \in \mathbb{N}_0$ ,  $f : [a, b] \rightarrow \mathbb{R}$   $n$ -mal stetig differenzierbar und  $f$  auf  $(a, b)$   $(n+1)$ -mal differenzierbar. Dann existiert ein  $c \in (a, b)$  mit*

$$f(b) = f(a) + \frac{1}{1!} f'(a)(b-a) + \frac{1}{2!} f''(a)(b-a)^2 + \dots + \frac{1}{n!} f^{(n)}(a)(b-a)^n + \frac{1}{(n+1)!} f^{(n+1)}(c)(b-a)^{n+1}.$$

*Beweis.* Setze

$$g(x) := f(b) - f(x) - f'(x)(b-x) - \dots - \frac{f^{(n)}(x)}{n!} (b-x)^n - \alpha \frac{(b-x)^{n+1}}{(n+1)!}$$

für  $a \leq x \leq b$ . Dann ist  $g(b) = 0$ , und  $\alpha$  sei so gewählt, dass auch  $g(a) = 0$  ist. Dann kann man auf  $g$  den Satz von Rolle 2.1 anwenden. Es folgt die Existenz eines  $c \in (a, b)$  mit  $g'(c) = 0$ . Nun ist

$$\begin{aligned} g'(x) &= 0 - f'(x) - [f''(x)(b-x) - f'(x)] - \dots \\ &\quad - \left[ \frac{f^{(n+1)}(x)}{n!} (b-x)^n - \frac{f^{(n)}(x)}{(n-1)!} (b-x)^{n-1} \right] + \alpha \frac{(b-x)^n}{n!} \\ &= - \frac{f^{(n+1)}(x)}{n!} (b-x)^n + \alpha \frac{(b-x)^n}{n!}. \end{aligned}$$

Einsetzen von  $x = c$  ergibt  $f^{(n+1)}(c) = \alpha$ . Setzt man jetzt  $x = a$  in der Definitionsgleichung für  $g$ , so folgt die Behauptung. ■

**Bemerkung.** Für  $n = 0$  reduziert sich der Taylorsche Satz 3.2 gerade auf den Mittelwertsatz 2.2.

Analog wie Satz 3.2 kann man den Fall  $b < a$  behandeln; aus beiden Fällen ergibt sich Satz 3.3.

**3.3 Satz.** *Sei  $D$  ein Intervall und  $f : D \rightarrow \mathbb{R}$   $(n+1)$ -mal differenzierbar, sei  $a \in D$ . Dann gilt für beliebige  $x \in D$*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + \frac{f^{(n+1)}(a + \vartheta(x-a))}{(n+1)!} (x-a)^{n+1}$$

(„Taylorformel“) mit einem (von  $a$  und  $x$  abhängenden)  $\vartheta \in (0, 1)$ .

Die Funktion

$$x \mapsto \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k$$

nennt man auch das *Taylorpolynom  $n$ -ter Ordnung von  $f$  zur Stelle  $a$* .

Ist  $f^{(n+1)}$  in  $D$  beschränkt, so folgt aus Satz 3.3 unter den dortigen Voraussetzungen insbesondere

$$\lim_{x \rightarrow a} \frac{1}{(x-a)^n} \left| f(x) - \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k \right| = 0.$$

Diese Aussage läßt sich auch schon unter schwächeren Voraussetzungen beweisen:

**3.4 Satz.** *Sei  $a \in \mathbb{R}$ ,  $n \in \mathbb{N}$ . Die Funktion  $f$  sei in einer Umgebung von  $a$   $(n-1)$ -mal differenzierbar und in  $a$   $n$ -mal differenzierbar. Dann gilt*

$$\lim_{x \rightarrow a} \frac{1}{(x-a)^n} \left| f(x) - \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k \right| = 0.$$

*Beweis.* Im Fall  $n = 1$  lautet die Behauptung

$$\lim_{x \rightarrow a} \left| \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} \right| = 0;$$

dies ist wegen der Differenzierbarkeit von  $f$  in  $a$  erfüllt.

Sei jetzt  $n \geq 2$ . Es gibt nach Voraussetzung ein  $\delta \in \mathbb{R}^+$ , so dass  $f$  in  $[a-\delta, a+\delta]$   $(n-1)$ -mal differenzierbar ist. Setze

$$g(x) := f(x) - \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k \quad \text{für } x \in [a-\delta, a+\delta].$$

Sei  $x \in [a-\delta, a+\delta]$ . Nach dem Taylorschen Satz 3.3 (für  $n-2$  statt  $n$ ), angewendet auf die Funktion  $g$ , existiert ein (i.a. von  $x$  abhängendes)  $c$  mit  $|c-a| \leq |x-a|$ , so dass

$$g(x) = \sum_{k=0}^{n-2} \frac{g^{(k)}(a)}{k!} (x-a)^k + \frac{g^{(n-1)}(c)}{(n-1)!} (x-a)^{n-1}$$

ist. Nun ist  $g(a) = g'(a) = \dots = g^{(n-2)}(a) = 0$  und

$$g^{(n-1)}(c) = f^{(n-1)}(c) - f^{(n-1)}(a) - f^{(n)}(a)(c-a),$$

also

$$\begin{aligned} |g(x)| &= \frac{1}{(n-1)!} \left| f^{(n-1)}(c) - f^{(n-1)}(a) - f^{(n)}(a)(c-a) \right| |x-a|^{n-1} \\ &\leq \frac{1}{(n-1)!} \left| \frac{f^{(n-1)}(c) - f^{(n-1)}(a)}{c-a} - f^{(n)}(a) \right| |x-a|^n. \end{aligned}$$

Da  $f^{(n-1)}$  nach Voraussetzung in  $a$  differenzierbar ist, gilt

$$\lim_{c \rightarrow a} \left| \frac{f^{(n-1)}(c) - f^{(n-1)}(a)}{c-a} - f^{(n)}(a) \right| = 0.$$

Wegen  $|c-a| \leq |x-a|$  folgt

$$\lim_{x \rightarrow a} \frac{|g(x)|}{|x-a|^n} = 0,$$

wie behauptet. ■

### Taylorreihen

Ist die Funktion  $f$  in einer Umgebung von  $a$  beliebig oft differenzierbar, so gilt für sie die Taylorformel für jede natürliche Zahl  $n$ . Dies legt es nahe, neben den Taylorpolynomen

$$\sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k$$

auch die unendliche Reihe

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

zu betrachten.

**Definition.** Sei  $D \subset \mathbb{R}$  ein Intervall und  $f : D \rightarrow \mathbb{R}$  beliebig oft differenzierbar, sei  $a \in D$ . Dann heißt die Reihe

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

die *Taylorreihe* von  $f$  an der Stelle  $x$  zur Entwicklungsstelle  $a$ .

**Bemerkung.** Achtung! Die Taylorreihe einer Funktion  $f$  an der Stelle  $x$  braucht nicht zu konvergieren, und wenn sie konvergiert, braucht sie nicht gegen  $f(x)$  zu konvergieren. Zum Beispiel sei

$$f(x) := \frac{1}{1-x} \quad \text{für } x \neq 1.$$

Man berechnet

$$f^{(k)}(x) = \frac{k!}{(1-x)^{k+1}}, \quad \text{also } \frac{f^{(k)}(0)}{k!} = 1.$$

Die Taylorreihe von  $f$  zur Entwicklungsstelle  $a = 0$  ist also gegeben durch

$$\sum_{k=0}^{\infty} x^k.$$

Diese Reihe konvergiert, wie wir wissen, für  $|x| < 1$ , und zwar gegen  $f(x)$ . Aber für  $|x| \geq 1$  ist die Reihe divergent. – Es gibt sogar beliebig oft differenzierbare Funktionen, deren Taylorreihe zur Entwicklungsstelle  $a$  für jedes  $x \neq a$  divergiert.

Als zweites Beispiel betrachten wir

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

Wie früher gezeigt, ist  $f$  beliebig oft differenzierbar und  $f^{(k)}(0) = 0$  für alle  $k \in \mathbb{N}_0$ . Die Taylorreihe dieser Funktion zur Entwicklungsstelle  $a = 0$  konvergiert also trivialerweise, aber für  $x > 0$  nicht gegen  $f(x)$ .

Man muß also in jedem Einzelfall untersuchen, für welche  $x$  die Taylorreihe einer gegebenen Funktion  $f$  an der Stelle  $x$  wirklich gegen  $f(x)$  konvergiert. Definieren wir das  $(n+1)$ -te Restglied  $R_{n+1}$  durch

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + R_{n+1}(x),$$

so ist die Konvergenz der Taylorreihe gegen  $f(x)$  also gleichbedeutend mit

$$\lim_{n \rightarrow \infty} R_{n+1}(x) = 0.$$

Hier ist nun die Taylorformel von Nutzen, die uns eine Darstellung dieses Restgliedes angibt, nämlich

$$R_{n+1}(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1}$$

mit einer Zahl  $c$  zwischen  $a$  und  $x$ . In manchen, aber nicht in allen Fällen kann man hiermit die gewünschte Konvergenz der Taylorreihe gegen  $f$  zeigen.

Als Beispiel betrachten wir die Exponentialfunktion und eine beliebige Entwicklungsstelle  $a$ . Das Restglied lautet

$$R_{n+1}(x) = \frac{f^{(n+1)}(c_n)}{(n+1)!}(x-a)^{n+1} = \frac{e^{c_n}}{(n+1)!}(x-a)^{n+1}.$$

Die Zahl  $c_n$  hängt zwar von  $n$  ab, aber es ist

$$e^{c_n} \leq \max\{e^a, e^x\},$$

also gilt  $\lim_{n \rightarrow \infty} R_{n+1}(x) = 0$ . Daher ist

$$e^x = \sum_{k=0}^{\infty} \frac{e^a}{k!}(x-a)^k = e^a e^{x-a},$$

womit sich wieder die Funktionalgleichung der Exponentialfunktion ergeben hat.

Die genauere Behandlung von Taylorreihen erfolgt erst in Abschnitt 8.2, da es nützlich ist, hierzu die Integralrechnung zur Verfügung zu haben.

Wir beschließen dieses Kapitel mit zwei Anwendungen von Ableitungen zweiter und höherer Ordnung.

### Extremwerte differenzierbarer Funktionen

Die Taylorformel liefert Informationen über das Verhalten hinreichend oft differenzierbarer Funktionen in der Umgebung eines Punktes. Als Anwendung ergeben sich zum Beispiel Aussagen über Extremwerte.

**Definition.** Die Funktion  $f : D \rightarrow \mathbb{R}$  hat im Punkt  $a \in D$  ein *lokales Maximum (starkes lokales Maximum)*, wenn es eine Umgebung  $U$  von  $a$  gibt mit  $f(a) \geq f(x)$  (bzw.  $f(a) > f(x)$ ) für alle  $x \in (U \cap D) \setminus \{a\}$ .

Analog: Minimum, Oberbegriff: Extremum

Der Punkt  $a \in D$  heißt *innerer Punkt* von  $D$ , wenn  $D$  Umgebung von  $a$  ist, also wenn ein  $\varepsilon > 0$  existiert mit  $(a - \varepsilon, a + \varepsilon) \subset D$ .

**3.5 Satz.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion und  $a$  innerer Punkt von  $D$ .

- (a) Ist  $f$  in  $a$  differenzierbar und hat  $f$  in  $a$  ein lokales Extremum, so ist  $f'(a) = 0$ .

(b) Sei  $f$   $n$ -mal stetig differenzierbar und

$$f'(a) = \dots = f^{(n-1)}(a) = 0, \quad f^{(n)}(a) \neq 0.$$

Ist  $n$  ungerade, so hat  $f$  in  $a$  kein lokales Extremum. Sei  $n$  gerade. Im Fall  $f^{(n)}(a) > 0$  hat  $f$  in  $a$  ein starkes lokales Minimum, im Fall  $f^{(n)}(a) < 0$  ein starkes lokales Maximum.

*Beweis.* Die Behauptung (a) wurde bereits im Beweis von Satz 2.1 gezeigt.

(b) Da  $a$  innerer Punkt von  $D$  ist, folgt aus der Taylorformel und den Voraussetzungen

$$f(x) = f(a) + \frac{f^{(n)}(c)}{n!}(x-a)^n$$

mit  $|c-a| \leq |x-a|$  für alle  $x$  aus einer Umgebung von  $a$ . Da  $f^{(n)}$  stetig und  $f^{(n)}(a) \neq 0$  ist, existiert eine Umgebung  $U$  von  $a$  mit  $f^{(n)}(c)f^{(n)}(a) > 0$  für alle  $c \in U$ . Jetzt liest man die Behauptung ab. ■

## Konvexe Funktionen

Wechselt die erste Ableitung einer Funktion auf einem Intervall nicht das Vorzeichen, so ist die Funktion dort monoton. Die Bedingung, dass die zweite Ableitung nicht das Vorzeichen ändert, führt auf die wichtige Klasse der konvexen bzw. konkaven Funktionen.

**Definition.** Sei  $D$  ein Intervall. Die Funktion  $f : D \rightarrow \mathbb{R}$  heißt *konvex*, wenn

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2)$$

für alle  $x_1, x_2 \in D$  und alle  $\lambda \in [0, 1]$  gilt.  $f$  heißt *konkav*, wenn  $-f$  konvex ist.

**3.6 Satz.** Sei  $D$  ein Intervall und  $f : D \rightarrow \mathbb{R}$  differenzierbar. Dann ist  $f$  konvex genau dann, wenn  $f'$  monoton wachsend ist.

*Beweis.* Die Konvexitätsbedingung ist äquivalent mit: Für alle  $x_1, x_2 \in D$  mit  $x_1 < x_2$  und alle  $x \in (x_1, x_2)$  gilt

$$f(x) \leq \frac{x_2 - x}{x_2 - x_1} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2). \quad (3.7)$$

Seien  $x_1, x_2$  wie oben. Ist  $f$  konvex, so folgt aus (3.7) durch Umrechnung

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_2) - f(x)}{x_2 - x}$$

und daraus durch Grenzübergang  $x \searrow x_1$  bzw.  $x \nearrow x_2$  die Ungleichung  $f'(x_1) \leq f'(x_2)$ .

Umgekehrt folgt aus dem Mittelwertsatz 2.2 die Existenz von Zahlen  $c_1 \in (x_1, x)$  und  $c_2 \in (x, x_2)$  mit

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(c_1), \quad \frac{f(x_2) - f(x)}{x_2 - x} = f'(c_2).$$

Ist nun  $f'$  monoton wachsend, so ist  $f'(c_1) \leq f'(c_2)$ , woraus die Ungleichung (3.7) sich durch Umrechnung ergibt. ■

Als Folgerung ergibt sich wegen Satz 2.6:

**3.8 Satz.** Sei  $D$  ein Intervall und  $f : D \rightarrow \mathbb{R}$  zweimal differenzierbar. Dann ist  $f$  genau dann konvex, wenn  $f''(x) \geq 0$  ist für alle  $x \in D$ .

Als Anwendung beweisen wir eine sehr nützliche Ungleichung. Dazu betrachten wir zunächst die Funktion  $f(x) = \ln x$  ( $x > 0$ ) und berechnen  $f''(x) = -\frac{1}{x^2} < 0$ , also ist  $\ln$  konkav. Für beliebige  $x, y > 0$  und  $p, q \in (1, \infty)$  mit  $\frac{1}{p} + \frac{1}{q} = 1$  gilt daher

$$\ln \left( \frac{1}{p}x + \frac{1}{q}y \right) \geq \frac{1}{p} \ln x + \frac{1}{q} \ln y,$$

folglich

$$\frac{1}{p}x + \frac{1}{q}y \geq e^{\frac{1}{p} \ln x} e^{\frac{1}{q} \ln y} = x^{\frac{1}{p}} y^{\frac{1}{q}},$$

also

$$x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{x}{p} + \frac{y}{q}. \quad (*)$$

Dies gilt trivialerweise auch für  $x = 0$  oder  $y = 0$ .

Hieraus können wir herleiten:

**3.9 Satz.** Für  $p, q \in (1, \infty)$  mit  $\frac{1}{p} + \frac{1}{q} = 1$  und beliebige  $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$  gilt

$$\sum_{i=1}^n |x_i y_i| \leq \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} \left( \sum_{i=1}^n |y_i|^q \right)^{\frac{1}{q}}$$

(Höldersche Ungleichung).

*Beweis.* O.B.d.A. sei die rechte Seite  $\neq 0$ . Setze

$$a_i := \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, \quad b_i = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}, \quad i = 1, \dots, n.$$

Aus der obigen Ungleichung (\*) folgt für alle  $i = 1, \dots, n$

$$\frac{|x_i y_i|}{\left(\sum |x_j|^p\right)^{\frac{1}{p}} \left(\sum |y_j|^q\right)^{\frac{1}{q}}} = a_i^{\frac{1}{p}} b_i^{\frac{1}{q}} \leq \frac{a_i}{p} + \frac{b_i}{q},$$

also

$$\frac{\sum_i |x_i y_i|}{\left(\sum_j |x_j|^p\right)^{\frac{1}{p}} \left(\sum_j |y_j|^q\right)^{\frac{1}{q}}} \leq \frac{\sum_i a_i}{p} + \frac{\sum_i b_i}{q} = \frac{1}{p} + \frac{1}{q} = 1. \quad \blacksquare$$



## 7 Integration

Notwendigkeit des Integralbegriffes und Hinweise zu seiner Präzisierung liegen auf der Hand. Betrachten wir etwa den physikalischen Begriff der Arbeit, die im einfachsten Fall, nämlich bei konstanter Kraft, als das Produkt „Kraft mal Weg“ definiert ist. Bei stückweise konstanter Kraft wird man sie entsprechend als Summe definieren. Stellen wir die Kraft als Funktion des Weges graphisch dar, so veranschaulicht sich die Arbeit als der „Flächeninhalt unter dem Graphen“. Ist nun die Kraft nicht mehr stückweise konstant, so wird man intuitiv immer noch den „Flächeninhalt unter dem Graphen“ als wohldefiniert ansehen und als Maß für die entsprechende Arbeit nehmen. Um aus dieser anschaulichen Vorstellung wirklich eine Definition zu machen, werden wir versuchen, eine gegebene Funktion durch stückweise konstante Funktionen zu approximieren. Für jede der approximierenden Funktionen ist dann die Fläche unter dem Graphen bzw. die Arbeit erklärt, und wir werden hoffen, dass diese Werte einem Grenzwert zustreben. Damit dies wirklich der Fall ist und der Grenzwert nicht von der Auswahl der approximierenden Folge abhängt, müssen wir aber eine besonders gute Art der Approximation wählen. Unser Integral wird daher nur für solche Funktionen erklärt sein, die sich in dieser Weise durch stückweise konstante Funktionen approximieren lassen. Das ist aber zunächst ausreichend.

### 7.1 Regelfunktionen

Im folgenden liegt stets ein festes kompaktes Intervall  $[a, b]$  zugrunde.

$B([a, b])$  sei die Menge aller beschränkten Funktionen  $f : [a, b] \rightarrow \mathbb{R}$ . Für reelle Funktionen  $f, g$  mit demselben Definitionsbereich  $D$  erklärt man, wie schon erwähnt,  $f + g$  und  $\lambda f$  ( $\lambda \in \mathbb{R}$ ) durch  $(f + g)(x) := f(x) + g(x)$  und  $(\lambda f)(x) := \lambda f(x)$  für  $x \in D$ . Mit  $f, g \in B([a, b])$  ist dann auch  $f + g \in B([a, b])$  und  $\lambda f \in B([a, b])$  für  $\lambda \in \mathbb{R}$ . Da die Vektorraumaxiome trivialerweise erfüllt sind, ist  $B([a, b])$  also ein reeller Vektorraum. Der Nullvektor ist die Funktion, die identisch gleich 0 ist. Wir wollen sie ebenfalls mit 0 bezeichnen.

**Definition.** Für  $f \in B([a, b])$  sei

$$\|f\| := \sup_{x \in [a, b]} |f(x)| := \sup\{|f(x)| \mid x \in [a, b]\}.$$

**1.1 Satz.** Für  $f, g \in B([a, b])$  gilt

(a)  $\|f\| \geq 0$ , und  $\|f\| = 0$  nur für  $f = 0$  (Nullfunktion),

(b)  $\|\lambda f\| = |\lambda| \|f\|$  für  $\lambda \in \mathbb{R}$ ,

(c)  $\|f + g\| \leq \|f\| + \|g\|$ .

*Beweis.* (a) und (b) sind trivial, und (c) folgt aus

$$|(f + g)(x)| = |f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\|. \quad \blacksquare$$

Eine auf einem reellen Vektorraum  $V$  erklärte Abbildung  $\|\cdot\| : V \rightarrow \mathbb{R}$  mit den Eigenschaften (a), (b), (c) aus Satz 1.1 nennt man eine *Norm* auf  $V$ . Die hier auf  $B([a, b])$  erklärte Norm heißt *Supremumsnorm*.

Diese Norm verwenden wir nun zur Erklärung eines für unsere Zwecke geeigneten Approximationsbegriffes.

**Definition.** Seien  $f_n, f \in B([a, b])$  ( $n \in \mathbb{N}$ ). Die Folge  $(f_n)_{n \in \mathbb{N}}$  konvergiert *gleichmäßig gegen  $f$* , wenn

$$\lim_{n \rightarrow \infty} \|f_n - f\| = 0$$

gilt.

Es gilt also:

$(f_n)_{n \in \mathbb{N}}$  konvergiert gleichmäßig gegen  $f \Leftrightarrow$

$$\forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \geq n_0 \forall x \in [a, b] : |f_n(x) - f(x)| \leq \varepsilon.$$

Dies kann man sich leicht veranschaulichen: Zu jedem  $\varepsilon > 0$  müssen schließlich alle Funktionen der Folge im „ $\varepsilon$ -Streifen um  $f$ “ liegen.

Die hier erklärte gleichmäßige Konvergenz einer Funktionenfolge ist zu unterscheiden von der Konvergenz schlechthin, unter der man punktweise Konvergenz versteht. Man sagt, dass die Folge  $(f_n)_{n \in \mathbb{N}}$  (punktweise) gegen  $f$  konvergiert, wenn für jedes  $x \in [a, b]$  die Folge  $(f_n(x))_{n \in \mathbb{N}}$  gegen  $f(x)$  konvergiert. Es gilt also:

$(f_n)_{n \in \mathbb{N}}$  konvergiert gegen  $f \Leftrightarrow$

$$\forall x \in [a, b] \quad \forall \varepsilon \in \mathbb{R}^+ \quad \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0 : |f_n(x) - f(x)| \leq \varepsilon.$$

Der wesentliche Unterschied liegt darin, dass das  $n_0$  (bei gegebenem  $\varepsilon$ ) hier von  $x$  abhängen kann, bei gleichmäßiger Konvergenz (daher der Name) dagegen einheitlich für alle  $x \in [a, b]$  wählbar ist.

**Definition.**  $f \in B([a, b])$  heißt *Treppenfunktion*, wenn es Punkte

$$a = x_0 < x_1 < \dots < x_n = b$$

gibt, so dass  $f$  auf jedem offenen Teilintervall  $(x_{k-1}, x_k)$  konstant ist ( $k = 1, \dots, n$ ), und  $(x_0, \dots, x_n)$  heißt dann eine *zu  $f$  gehörige Unterteilung* von  $[a, b]$ . Sei  $T([a, b])$  die Menge der Treppenfunktionen auf  $[a, b]$ .

$T([a, b])$  ist ein Untervektorraum von  $B([a, b])$ . Zum Beweis ist nur zu zeigen, dass mit  $f, g \in T([a, b])$ ,  $\lambda \in \mathbb{R}$ , auch  $f + g \in T([a, b])$  und  $\lambda f \in T([a, b])$  ist. Letzteres ist trivial. Zum Beweis des Ersteren sei  $(x_0, \dots, x_n)$  eine zu  $f$  und  $(y_0, \dots, y_m)$  eine zu  $g$  gehörige Unterteilung von  $[a, b]$ . Sei  $(z_0, \dots, z_k)$  die Unterteilung, die durch Vereinigung beider Teilpunktfolgen und Ordnen nach der Größe entsteht. Auf jedem Intervall  $(z_{j-1}, z_j)$  ist  $f + g$  konstant, also ist  $f + g \in T([a, b])$ .

**Definition.** Eine Funktion  $f : [a, b] \rightarrow \mathbb{R}$  heißt *Regelfunktion*, wenn es eine gleichmäßig gegen  $f$  konvergierende Folge von Treppenfunktionen auf  $[a, b]$  gibt.  $R([a, b])$  sei die Menge der Regelfunktionen auf  $[a, b]$ .

$R([a, b])$  ist ein Untervektorraum von  $B([a, b])$ . Zunächst ist klar, dass jede Regelfunktion beschränkt ist und dass mit  $f \in R([a, b])$  auch  $\lambda f \in R([a, b])$  für  $\lambda \in \mathbb{R}$  gilt. Seien jetzt  $f, g \in R([a, b])$ . Es gibt also zwei Folgen  $(f_n), (g_n)$  in  $T([a, b])$ , die gleichmäßig gegen  $f$  bzw.  $g$  konvergieren. Aus

$$\|(f + g) - (f_n + g_n)\| \leq \|f - f_n\| + \|g - g_n\|$$

folgt dann, dass die Folge  $(f_n + g_n)$  von Treppenfunktionen gleichmäßig gegen  $f + g$  konvergiert, also ist  $f + g \in R([a, b])$ .

Es fragt sich, wie man einer gegebenen Funktion „ansetzen“ kann, ob sie Regelfunktion ist. Dies wird durch folgenden Satz beantwortet:

**1.2 Satz.** Die Funktion  $f : [a, b] \rightarrow \mathbb{R}$  ist genau dann Regelfunktion, wenn für alle  $c \in [a, b]$  die einseitigen Grenzwerte

$$\lim_{x \nearrow c} f(x) \quad \text{und} \quad \lim_{x \searrow c} f(x)$$

existieren.

*Beweis.* „ $\Rightarrow$ “: Sei  $f$  Regelfunktion. Es gibt also eine Folge  $(t_n)_{n \in \mathbb{N}}$  in  $T([a, b])$  mit  $\lim_{n \rightarrow \infty} \|t_n - f\| = 0$ . Eine Treppenfunktion hat offenbar überall rechtsseitige und linksseitige Grenzwerte. Die Behauptung ergibt sich daher aus der folgenden allgemeineren:

**Behauptung.** Konvergiert die Folge  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$  und hat jedes  $f_n$  überall einseitige Grenzwerte, so gilt dies auch für  $f$ .

*Beweis.* (etwa für rechtsseitige Grenzwerte). Sei  $c \in [a, b)$ . Zu  $\varepsilon \in \mathbb{R}^+$  existiert ein  $n \in \mathbb{N}$  mit  $\|f - f_n\| < \varepsilon/3$ . Weiter existiert nach dem Cauchy-Kriterium 2.3 aus Kapitel 4 (das völlig analog für einseitige Grenzwerte gilt) ein  $\delta \in \mathbb{R}^+$  mit

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{3} \quad \text{für alle } x, y \in [a, b] \text{ mit } c < x, y < c + \delta.$$

Für diese  $x, y$  gilt also

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| < \varepsilon.$$

Aus Satz 2.3 aus Kapitel 4 (mit  $D := (c, b)$ ) folgt die Existenz von  $\lim_{x \searrow c} f(x)$ . ■

„ $\Leftarrow$ “: Es existiere  $\lim_{x \searrow c} f(x)$  für alle  $c \in [a, b)$  und  $\lim_{x \nearrow c} f(x)$  für alle  $c \in (a, b]$ .

Sei  $n \in \mathbb{N}$ . Zu jedem  $c \in [a, b]$  existiert nach Voraussetzung und nach Satz 2.3 aus Kapitel 4 ein  $\delta_c \in \mathbb{R}^+$  mit

$$|f(x) - f(y)| < \frac{1}{n}$$

für alle  $x, y \in [a, b]$  mit  $c < x, y < c + \delta_c$  oder  $c - \delta_c < x, y < c$ . Das System  $\{U_{\delta_c}(c) \mid c \in [a, b]\}$  ist eine offene Überdeckung von  $[a, b]$ , enthält also nach dem Überdeckungssatz von Heine-Borel 1.8 aus Kapitel 4 eine endliche Teilüberdeckung. Es gibt also endlich viele Punkte  $c_1, \dots, c_k \in [a, b]$  mit zugehörigen Zahlen  $\delta_j := \delta_{c_j}$ , so dass  $[a, b] \subset \bigcup_{j=1}^k U_{\delta_j}(c_j)$  ist. Wir ordnen die Zahlen  $c_1, \dots, c_k, c_1 \pm \delta_1, \dots, c_k \pm \delta_k$  (soweit sie in  $[a, b]$  liegen) nach der Größe und erhalten eine Unterteilung

$$a = x_0 < x_1 < \dots < x_m = b$$

von  $[a, b]$ . Wähle  $z_j \in (x_{j-1}, x_j)$  ( $j = 1, \dots, m$ ) und setze

$$t_n(x) := \begin{cases} f(z_j) & \text{falls } x \in (x_{j-1}, x_j) \text{ für ein } j \in \{1, \dots, m\}, \\ f(x_j) & \text{falls } x = x_j \text{ für ein } j \in \{0, \dots, m\}. \end{cases}$$

Dann ist  $t_n \in T([a, b])$ . Sei  $x \in [a, b]$ . Ist  $x = x_j$  für ein  $j$ , so gilt  $|f(x) - t_n(x)| = 0$ . Andernfalls gibt es genau ein  $j$  mit  $x \in (x_{j-1}, x_j)$ . Da  $(x_{j-1}, x_j)$  Teilmenge eines offenen Intervalls  $(c_i - \delta_i, c_i + \delta_i)$  ist und entweder links oder rechts von  $c_i$  liegt, folgt

$$|f(x) - t_n(x)| = |f(x) - f(z_j)| < \frac{1}{n}.$$

Da  $x \in [a, b]$  beliebig war, ist  $\|f - t_n\| \leq \frac{1}{n}$ . Da  $n \in \mathbb{N}$  beliebig war, ist damit eine Folge  $(t_n)_{n \in \mathbb{N}}$  in  $T([a, b])$  gefunden, die gleichmäßig gegen  $f$  konvergiert. ■

## 7.2 Das Integral einer Regelfunktion

Zuerst erklären wir das Integral einer Treppenfunktion:

**Definition.** Sei  $f \in T([a, b])$ . Sei durch

$$a = x_0 < x_1 < \dots < x_n = b$$

eine Unterteilung von  $[a, b]$  gegeben, so dass  $f$  auf  $(x_{j-1}, x_j)$  den Wert  $c_j$  annimmt. Dann ist die Zahl

$$I(f) := \sum_{j=1}^n c_j (x_j - x_{j-1})$$

unabhängig von der Wahl der Unterteilung; sie heißt *Integral* von  $f$  über  $[a, b]$  und wird mit

$$\int_a^b f \quad \text{oder} \quad \int_a^b f(x) dx$$

bezeichnet.

Die Behauptung, dass  $I(f)$  nur von  $f$  und nicht von der speziellen Unterteilung abhängt, ist leicht zu sehen: Ist noch eine andere Unterteilung zu  $f$  gegeben, so bilde man die gemeinsame Verfeinerung. Sie entsteht aus der ursprünglichen Unterteilung durch Einführung weiterer Teilpunkte. Es genügt zu zeigen, dass die Summe sich nicht ändert, wenn man einen weiteren Teilpunkt einfügt. Da dieser ein Intervall zerlegt, auf dem  $f$  konstant ist, ist das aber klar: ein Summand  $c_j(x_j - x_{j-1})$  wird ersetzt durch  $c_j(x - x_{j-1}) + c_j(x_j - x)$ , was dasselbe ist.

Der folgende Satz bringt die wichtigsten Eigenschaften des Integrals (d.h. der Abbildung  $I : T([a, b]) \rightarrow \mathbb{R}$ ) zum Ausdruck.

**2.1 Satz.** Die Abbildung  $I : T([a, b]) \rightarrow \mathbb{R} : f \mapsto \int_a^b f$ , ist ein lineares Funktional, d.h. es gilt

$$(a) \int_a^b (f + g) = \int_a^b f + \int_a^b g \text{ für alle } f, g \in T([a, b]),$$

$$(b) \int_a^b (\lambda f) = \lambda \int_a^b f \text{ für alle } f \in T([a, b]) \text{ und alle } \lambda \in \mathbb{R}.$$

Dieses Funktional ist monoton, d.h. es gilt

$$(c) \text{ Aus } f \leq g \text{ (d.h. } f(x) \leq g(x) \forall x \in [a, b]) \text{ folgt } \int_a^b f \leq \int_a^b g.$$

Ferner gilt

$$(d) \left| \int_a^b f \right| \leq (b - a) \|f\| \text{ für } f \in T([a, b]),$$

das Funktional ist also beschränkt.

Ein lineares Funktional  $\varphi$  auf einem normierten Vektorraum  $(V, \|\cdot\|)$  heißt *beschränkt*, wenn es eine Konstante  $K \in \mathbb{R}$  gibt mit  $|\varphi(x)| \leq K\|x\|$  für alle  $x \in V$ .

*Beweis.* Sind  $f, g \in T([a, b])$  gegeben, so wähle man zu jeder dieser Funktionen eine zugehörige Unterteilung von  $[a, b]$  und bilde dann die gemeinsame Verfeinerung. Sie ist eine zu  $f, g$  und  $f + g$  gehörige Unterteilung. Die Behauptungen (a) und (c) folgen jetzt sofort aus der Definition, (b) ist trivial. Schließlich gilt

$$\begin{aligned} \left| \int_a^b f \right| &= \left| \sum_{j=1}^n c_j (x_j - x_{j-1}) \right| \leq \sum_{j=1}^n |c_j| (x_j - x_{j-1}) \\ &\leq \max\{|c_1|, \dots, |c_n|\} \sum_{j=1}^n (x_j - x_{j-1}) \leq \|f\| (b - a). \quad \blacksquare \end{aligned}$$

Nun setzen wir die Abbildung  $I$  unter Beibehaltung ihrer Eigenschaften fort auf den Raum  $R([a, b])$  der Regelfunktionen.

**Definition.** Sei  $f \in R([a, b])$ . Sei  $(t_n)_{n \in \mathbb{N}}$  eine Folge in  $T([a, b])$ , die gleichmäßig gegen  $f$  konvergiert. Dann existiert

$$\int_a^b f := \lim_{n \rightarrow \infty} \int_a^b t_n$$

und ist unabhängig von der Wahl der Folge  $(t_n)$ .  $\int_a^b f$  heißt das *Integral* von  $f$  über  $[a, b]$  und wird auch mit

$$\int_a^b f(x) dx$$

bezeichnet.

*Beweis der Wohldefiniertheit.* Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $\|t_n - f\| < \varepsilon/2(b-a)$  für  $n \geq n_0$ . Für  $m, n \geq n_0$  gilt also

$$\|t_m - t_n\| \leq \|t_m - f\| + \|t_n - f\| < \frac{\varepsilon}{b-a}$$

und daher nach Satz 2.1

$$\left| \int_a^b t_m - \int_a^b t_n \right| = \left| \int_a^b (t_m - t_n) \right| \leq \|t_m - t_n\|(b-a) < \varepsilon.$$

Also ist  $(\int_a^b t_n)_{n \in \mathbb{N}}$  eine Cauchyfolge reeller Zahlen und daher konvergent.

Sei  $(u_n)_{n \in \mathbb{N}}$  eine weitere Folge in  $T([a, b])$ , die gleichmäßig gegen  $f$  konvergiert. Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit

$$\|f - t_n\| < \frac{\varepsilon}{2(b-a)} \quad \text{und} \quad \|f - u_n\| < \frac{\varepsilon}{2(b-a)} \quad \text{für } n \geq n_0.$$

Für  $n \geq n_0$  gilt also

$$\|t_n - u_n\| \leq \|t_n - f\| + \|f - u_n\| < \frac{\varepsilon}{b-a}$$

und daher nach Satz 2.1

$$\left| \int_a^b t_n - \int_a^b u_n \right| \leq \|t_n - u_n\|(b-a) < \varepsilon.$$

Also ist  $\left(\int_a^b t_n - \int_a^b u_n\right)_{n \in \mathbb{N}}$  eine Nullfolge und somit

$$\lim_{n \rightarrow \infty} \int_a^b t_n = \lim_{n \rightarrow \infty} \int_a^b u_n. \quad \blacksquare$$

Ist speziell  $f$  eine Treppenfunktion, so liefert die neue Definition denselben Integralwert wie die alte, wie sich aus der Unabhängigkeit des Grenzwertes von der approximierenden Folge ergibt.

**2.2 Satz.** Für alle  $f, g \in R([a, b])$  und  $\lambda \in \mathbb{R}$  gilt

$$(a) \int_a^b (f + g) = \int_a^b f + \int_a^b g,$$

$$(b) \int_a^b (\lambda f) = \lambda \int_a^b f,$$

$$(c) f \leq g \Rightarrow \int_a^b f \leq \int_a^b g,$$

$$(d) \left| \int_a^b f \right| \leq \int_a^b |f| \leq (b - a) \|f\|.$$

*Beweis.* Seien  $(t_n)_{n \in \mathbb{N}}, (u_n)_{n \in \mathbb{N}}$  Folgen in  $T([a, b])$ , die gleichmäßig gegen  $f$  bzw.  $g$  konvergieren. Dann konvergiert  $(t_n + u_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f + g$ , also gilt

$$\begin{aligned} \int_a^b (f + g) &= \lim_{n \rightarrow \infty} \int_a^b (t_n + u_n) = \lim_{n \rightarrow \infty} \left[ \int_a^b t_n + \int_a^b u_n \right] \\ &= \lim_{n \rightarrow \infty} \int_a^b t_n + \lim_{n \rightarrow \infty} \int_a^b u_n = \int_a^b f + \int_a^b g. \end{aligned}$$

Damit ist (a) bewiesen. (b) folgt analog.

Ist  $f \leq g$ , so gelten für jedes  $\varepsilon \in \mathbb{R}^+$  für fast alle  $n$  die Ungleichungen  $t_n \leq f + \varepsilon$  und  $u_n \geq g - \varepsilon$ , also  $t_n \leq u_n + 2\varepsilon$ , woraus

$$\begin{aligned} \int_a^b f &= \lim_{n \rightarrow \infty} \int_a^b t_n \leq \lim_{n \rightarrow \infty} \int_a^b (u_n + 2\varepsilon) \\ &= \lim_{n \rightarrow \infty} \left( \int_a^b u_n + 2\varepsilon(b-a) \right) = \int_a^b g + 2\varepsilon(b-a) \end{aligned}$$

folgt. Da  $\varepsilon > 0$  beliebig war, folgt (c).

Zu (d) ist zunächst zu bemerken, dass mit  $f$  auch  $|f|$  eine Regelfunktion ist. Für Treppenfunktionen ist das klar, und daraus folgt es allgemein: Wenn  $(t_n)$  gleichmäßig gegen  $f$  konvergiert, konvergiert  $(|t_n|)$  wegen

$$\left| |t_n| - |f| \right| \leq \|t_n - f\|$$

(klar wegen  $||x| - |y|| \leq |x - y|$ ) gleichmäßig gegen  $|f|$ .

Ferner folgt aus

$$\left| \|t_n\| - \|f\| \right| \leq \|t_n - f\|$$

(wegen  $\|t_n\| \leq \|t_n - f\| + \|f\|$  etc.) dass  $(\|t_n\|)$  gegen  $\|f\|$  konvergiert. Damit ergibt sich

$$\left| \int_a^b f \right| = \left| \lim_{n \rightarrow \infty} \int_a^b t_n \right| = \lim_{n \rightarrow \infty} \left| \int_a^b t_n \right| \leq \lim_{n \rightarrow \infty} \int_a^b |t_n| = \int_a^b |f|$$

und

$$\lim_{n \rightarrow \infty} \int_a^b |t_n| \leq \lim_{n \rightarrow \infty} (b-a) \|t_n\| = (b-a) \|f\|. \quad \blacksquare$$

Ist  $f \in R([a, b])$  und  $[c, d] \subset [a, b]$ , so ist die Einschränkung  $f|_{[c, d]}$  offenbar eine Regelfunktion. Ihr Integral ist also definiert; wir bezeichnen es mit

$$\int_c^d f.$$

Zur Ergänzung definiert man für  $c \leq d$

$$\int_d^c f := - \int_c^d f, \quad \int_c^c f := 0.$$

Mit diesen Festsetzungen gilt dann allgemein die im folgenden Satz ausgedrückte Additivitätseigenschaft des Integrals hinsichtlich der Integrationsbereiche.

**2.3 Satz.** Für  $f \in R([a, b])$  und  $u, v, w \in [a, b]$  gilt

$$\int_u^v f + \int_v^w f = \int_u^w f.$$

*Beweis.* Zunächst sei  $u < v < w$ . Die Behauptung folgt für Treppenfunktionen aus der Definition des Integrals, allgemein dann durch Approximation. Die übrigen Fälle ergeben sich hieraus und mit den obigen Festsetzungen. Ist z.B.  $u < w < v$ , so ist

$$\int_u^v f + \int_v^w f = \int_u^v f - \int_w^v f = \int_u^w f.$$

Analog schließt man in den anderen Fällen. ■

Wir schließen diesen Abschnitt mit zwei manchmal nützlichen weiteren Eigenschaften des Integrals.

**2.4 Satz.** Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge von Regelfunktionen, die gleichmäßig gegen eine Funktion  $f$  konvergiert. Dann ist  $f$  Regelfunktion, und es gilt

$$\int_a^b f = \lim_{n \rightarrow \infty} \int_a^b f_n.$$

*Beweis.* Es ist leicht zu sehen, dass  $f$  Regelfunktion ist. Aus Satz 2.2 folgt

$$\left| \int_a^b f - \int_a^b f_n \right| \leq (b-a) \|f - f_n\|$$

und daraus die Behauptung. ■

**Beispiel.** Die Aussage von Satz 2.4 wird i.a. falsch, wenn „gleichmäßig“ ersetzt wird durch „punktweise“: Setze

$$f_n(x) := \begin{cases} n & \text{für } 0 < x < \frac{1}{n}, \\ 0 & \text{für } x = 0 \text{ und für } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Dann ist  $f_n$  eine Treppenfunktion auf  $[0, 1]$ , und die Folge  $(f_n)_{n \in \mathbb{N}}$  konvergiert punktweise gegen 0. Es gilt

$$\lim_{n \rightarrow \infty} \int_0^1 f_n = 1 \neq 0 = \int_0^1 \lim_{n \rightarrow \infty} f_n.$$

**2.5 Satz** (Mittelwertsatz der Integralrechnung). *Seien  $f, g \in R([a, b])$ , sei  $f$  stetig und  $g \geq 0$ . Dann existiert ein  $c \in [a, b]$  mit*

$$\int_a^b fg = f(c) \int_a^b g.$$

(Speziell für  $g \equiv 1$  also  $\int_a^b f = f(c)(b - a)$ .)

*Beweis.* Die stetige Funktion  $f$  nimmt nach Satz 3.2 aus Kapitel 4 auf dem kompakten Intervall  $[a, b]$  ein Minimum  $m$  und ein Maximum  $M$  an. Aus  $m \leq f \leq M$  und  $g \geq 0$  folgt  $mg \leq fg \leq Mg$  und daraus nach Satz 2.2

$$m \int_a^b g \leq \int_a^b fg \leq M \int_a^b g.$$

Es ist also  $\int_a^b fg = \alpha \int_a^b g$  mit einer Zahl  $\alpha \in [m, M]$ . Nach dem Zwischenwertsatz 3.4 aus Kapitel 4 gibt es eine Zahl  $c \in [a, b]$  mit  $f(c) = \alpha$ . ■

## 7.3 Integration und Differentiation

Die zu Anfang dieses Kapitels erläuterte Interpretation des Integrals als verallgemeinerte Summe ist nicht die einzige Motivation für die Einführung des Integralbegriffs. Eine andere liegt darin, dass man Integration als Umkehrung von Differentiation ansehen kann. Dass ein solcher Zusammenhang bestehen muß, macht man sich leicht schon anschaulich klar, wenn man das Integral einer etwa stetigen positiven Funktion  $f$  als „Fläche unter dem Graphen“ interpretiert und dann für kleine positive  $h$  die Größen

$$\int_a^{x+h} f(t) dt - \int_a^x f(t) dt \quad \text{und} \quad f(x)h$$

miteinander vergleicht. Der Zusammenhang zwischen Integration und Differentiation ist Gegenstand des Hauptsatzes der Differential- und Integralrechnung.

**Definition.** Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig. Eine Funktion  $F : [a, b] \rightarrow \mathbb{R}$  heißt *Stammfunktion* von  $f$ , wenn  $F$  differenzierbar und  $F' = f$  ist.

Nicht jede Funktion besitzt eine Stammfunktion, wohl aber jede stetige Funktion.

**3.1 Satz** (Hauptsatz der Differential- und Integralrechnung). *Die Funktion  $f : [a, b] \rightarrow \mathbb{R}$  sei stetig. Die durch*

$$F(x) := \int_a^x f = \int_a^x f(t) dt, \quad x \in [a, b],$$

erklärte Funktion  $F$  ist Stammfunktion von  $f$ .

Jede Stammfunktion  $G$  von  $f$  ist von der Form  $G = F + c$  mit einer Konstanten  $c$ .

*Beweis.* Sei  $x \in [a, b]$ . Für  $h \in \mathbb{R}$  mit  $x + h \in [a, b]$  gilt

$$\begin{aligned} |F(x+h) - F(x) - f(x)h| &= \left| \int_a^{x+h} f(t) dt - \int_a^x f(t) dt - f(x)h \right| \\ &= \left| \int_x^{x+h} f(t) dt - \int_x^{x+h} f(x) dt \right| = \left| \int_x^{x+h} (f(t) - f(x)) dt \right| \\ &\leq \int_{\min\{x, x+h\}}^{\max\{x, x+h\}} |f(t) - f(x)| dt, \end{aligned}$$

wobei Satz 2.3 benutzt worden ist. Sei nun  $\varepsilon > 0$  gegeben. Da  $f$  auf  $[a, b]$  nach Satz 3.5 Kapitel 4 gleichmäßig stetig ist, existiert ein  $\delta > 0$ , so dass  $|f(t) - f(x)| < \varepsilon$  für alle  $x, t \in [a, b]$  mit  $|t - x| < \delta$ . Für  $|h| < \delta$  folgt also

$$|F(x+h) - F(x) - f(x)h| \leq |h|\varepsilon.$$

Damit ist nach Definition der Ableitung

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

gezeigt, also ist  $F$  Stammfunktion von  $f$ .

Ist auch  $G$  Stammfunktion von  $f$ , so ist  $F' = f = G'$ , also  $(F - G)' = 0$  auf  $[a, b]$ ; nach Satz 2.5 aus Kapitel 6 ist also  $F - G$  konstant. ■

Der Hauptsatz der Differential- und Integralrechnung enthält zwei wichtige Aussagen: Die erste Aussage ist, dass das Integral einer stetigen Funktion nach der oberen Grenze differenzierbar ist und dass die Ableitung den Integranden ergibt. Die zweite Aussage ist, dass die Stammfunktion eindeutig ist. Beides zusammen hat die folgende wichtige Konsequenz. Ist  $F$  eine Stammfunktion der stetigen Funktion  $f$ , so ist mit einer Konstanten  $c$

$$F(b) - F(a) = \left( c + \int_a^b f \right) - \left( c + \int_a^a f \right) = \int_a^b f.$$

Man schreibt auch  $F(b) - F(a) = [F(x)]_{x=a}^{x=b} = [F(x)]_a^b$ . Wir können für stetiges  $f$  also das Integral  $\int_a^b f$  berechnen, wenn wir eine Stammfunktion von  $f$  kennen. Wir kommen hierauf im nächsten Abschnitt zurück. Zunächst, ebenfalls im Hinblick auf die Berechnung von Integralen, noch eine Bemerkung und eine wichtige Folgerung.

**Bemerkung.** Die Funktion  $h : [a, b] \rightarrow \mathbb{R}$  sei stetig differenzierbar und es sei  $H(x) := \int_a^x h'(t) dt$ . Dann folgt aus dem Hauptsatz, dass  $H' = h'$  und somit gibt es eine Konstante  $c$  mit  $H = h + c$ . Dies und die Folgerung des Hauptsatzes implizieren

$$\int_a^b h'(t) dt = H(b) - H(a) = h(b) - h(a) = [h(x)]_a^b.$$

**3.2 Satz** (Partielle Integration oder Produktintegration). *Für stetig differenzierbare Funktionen  $f, g : [a, b] \rightarrow \mathbb{R}$  gilt*

$$\int_a^b f(x)g'(x) dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g(x) dx.$$

*Beweis.* Nach der Produktregel gilt

$$(fg)' = f'g + fg',$$

also gilt nach obiger Bemerkung

$$\int_a^b fg' + \int_a^b f'g = \int_a^b (fg)' = [f(x)g(x)]_a^b. \quad \blacksquare$$

Hiermit sind wir insbesondere in der Lage, einen neuen Beweis der Taylorformel mit einer anderen nützlichen Darstellung des Restgliedes anzugeben:

**3.3 Satz** (Taylorsche Formel mit Restglied in Integralform). *Sei  $D \subset \mathbb{R}$  ein Intervall und  $f : D \rightarrow \mathbb{R}$   $(n + 1)$ -mal stetig differenzierbar. Dann gilt für  $a, x \in D$*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + R_{n+1}(x)$$

mit

$$R_{n+1}(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt.$$

*Beweis.* Wir beweisen die Behauptung durch Induktion nach  $n$ . Für  $n = 0$  lautet die Behauptung

$$f(x) = f(a) + \int_a^x f'(t) dt,$$

sie ist richtig nach Satz 3.1. Sei die Behauptung bewiesen für  $n - 1$ , also

$$R_n(x) = \frac{1}{(n-1)!} \int_a^x f^{(n)}(t)(x-t)^{n-1} dt = \int_a^x f^{(n)}(t)g'(t) dt$$

mit

$$g(t) := -\frac{(x-t)^n}{n!}.$$

Partielle Integration ergibt

$$\begin{aligned} R_n(x) &= \left[ -f^{(n)}(t) \frac{(x-t)^n}{n!} \right]_{t=a}^{t=x} + \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt \\ &= \frac{f^{(n)}(a)}{n!} (x-a)^n + \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt, \end{aligned}$$

woraus die Behauptung folgt. \blacksquare

## 7.4 Berechnung von Integralen

Der Hauptsatz der Differential- und Integralrechnung bietet die Möglichkeit, die Integrale vieler spezieller Funktionen explizit anzugeben. In anderen Fällen wird man sich zur numerischen Berechnung von Integralen verschiedener Näherungsmethoden bedienen müssen. Wir wollen im folgenden einige Verfahren zur Integralberechnung und Rechenregeln für Integrale zusammenstellen.

Prinzipiell kann man natürlich zur näherungsweise Integralberechnung unmittelbar die Integraldefinition heranziehen. In manchen einfachen Fällen erhält man so sogar den genauen Wert. Eine ungleich bequemere Methode der Integralberechnung liefert aber der Hauptsatz der Differential- und Integralrechnung. Ist  $f : [a, b] \rightarrow \mathbb{R}$  stetig und kennen wir eine Stammfunktion  $F$  von  $f$ , so folgt aus der Bemerkung hinter Satz 3.1

$$\int_a^b f(x) dx = F(b) - F(a).$$

Um zum Beispiel  $\int_0^a \cos x dx$  zu berechnen, erinnern wir uns, dass  $\sin' = \cos$  ist und erhalten

$$\int_0^a \cos x dx = \sin a - \sin 0 = \sin a.$$

Allgemein können wir uns eine Liste machen von allen stetigen Funktionen, die uns als Ableitungen anderer Funktionen begegnet sind. Von diesen Funktionen kennen wir also Stammfunktionen und können daher die Integrale explizit angeben.

**Beispiele ( $F$  Stammfunktion von  $f$ ).**

$f(x)$	$F(x)$
$e^{ax} \ (a \neq 0)$	$\frac{1}{a}e^{ax}$
$\frac{1}{x}$	$\ln x$
$x^\alpha \ (\alpha \neq -1)$	$\frac{1}{\alpha+1}x^{\alpha+1}$
$a^x \ (a \neq 1)$	$\frac{1}{\ln a}a^x$
$\cos x$	$\sin x$
$\sin x$	$-\cos x$
$\frac{1}{\cos^2 x}$	$\tan x$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin x$
$\frac{1}{1+x^2}$	$\arctan x$

Viele weitere Stammfunktionen findet man (unter der Überschrift „Unbestimmte Integrale“) in den Formelsammlungen und Integraltafeln. Hat man durch eigene Berechnungen eine Stammfunktion explizit ermittelt, so sollte man stets die Probe durch Differenzieren machen. Es sei darauf hingewiesen, dass häufig für Stammfunktionen eine bequeme, aber nicht ganz korrekte Schreibweise üblich ist, wie zum Beispiel

$$\int \frac{1}{1+x^2} dx = \arctan x.$$

Korrekt wäre

$$\int_a^x \frac{1}{1+t^2} dt = \arctan x + c,$$

aber die obige Schreibweise ist so verbreitet, dass wir sie auch benutzen werden. Man bezeichnet Ausdrücke wie

$$\int \frac{1}{1+x^2} dx$$

auch als „unbestimmte Integrale“; das sind also nichts anderes als Stammfunktionen.

Häufig kann man unbekannte Integrale durch Umformungen auf bekannte zurückführen. Diese Umformungsregeln erhält man, wenn man Differentiationsregeln mit Hilfe des Hauptsatzes der Differential- und Integralrechnung in Integrationsregeln übersetzt. Die erste solche Regel ist in Satz 3.2 enthalten: Für stetig differenzierbare Funktionen  $f, g$  gilt

$$\int_a^b f g' = [fg]_a^b - \int_a^b f' g.$$

Vor der Behandlung von Beispielen eine Bemerkung: Wenn wir bei gegebenem  $f$  das Integral

$$\int_a^b f$$

für alle  $b$  (des Definitionsbereiches von  $f$ ) bestimmen können, haben wir damit natürlich eine Stammfunktion von  $f$  gefunden. Dies wird bei den Beispielen stets der Fall sein.

**Beispiele.** (1)

$$\begin{aligned} \int_a^b x e^x dx &= \int_a^b f(x) g'(x) dx && \begin{cases} f(x) = x, \\ g(x) = e^x \end{cases} \\ &= [x e^x]_a^b - \int_a^b 1 \cdot e^x dx = [x e^x]_a^b - [e^x]_a^b. \end{aligned}$$

Mit der üblichen Schreibweise für unbestimmte Integrale, die wir von nun an verwenden wollen, lautet das Ergebnis also

$$\int x e^x dx = x e^x - e^x.$$

Das bedeutet, wie gesagt, dass  $F(x) := x e^x - e^x$  eine Stammfunktion von  $f(x) := x e^x$  ist. Die Probe durch Differenzieren (die man immer machen sollte) bestätigt das.

Ganz ähnlich:

$$\begin{aligned} \int x \sin x dx & && \begin{cases} f(x) = x \\ g(x) = -\cos x \end{cases} \\ &= -x \cos x + \int \cos x dx \\ &= -x \cos x + \sin x. \end{aligned}$$

Zwei Kunstgriffe sind im Zusammenhang mit partieller Integration oft nützlich: Man kann immer den Faktor 1 einfügen und als  $g'$  ansehen. Zweitens kann man eventuell durch partielle Integration eine Gleichung für das fragliche Integral erhalten und hieraus das Integral berechnen.

$$\begin{aligned} (2) \quad \int \ln x dx &= \int f(x) g'(x) dx && \begin{cases} f(x) = \ln x \\ g(x) = x \end{cases} \\ &= x \ln x - \int \frac{1}{x} x dx = x \ln x - x. \end{aligned}$$

$$\begin{aligned} (3) \quad \int \frac{1}{x} \ln x dx &= \int f(x) g'(x) dx && \begin{cases} f(x) = \ln x \\ g(x) = \ln x \end{cases} \\ &= (\ln x)^2 - \int \frac{1}{x} \ln x dx. \end{aligned}$$

Also ist

$$\int \frac{1}{x} \ln x dx = \frac{1}{2} (\ln x)^2.$$

Manchmal muß man partielle Integration mehrfach anwenden, um zu einer Gleichung zu kommen, aus der das Integral bestimmbar ist:

$$(4) \int e^x \sin x \, dx = -e^x \cos x + \int e^x \cos x \, dx \\ = -e^x \cos x + e^x \sin x - \int e^x \sin x \, dx.$$

Also ist

$$\int e^x \sin x \, dx = \frac{1}{2} e^x (\sin x - \cos x).$$

Manchmal erfordert die „partielle Integration“, also die Bestimmung von  $g$  zu  $g'$  in  $\int f g'$ , ihrerseits partielle Integration:

$$(5) \int (\ln x)^2 \, dx = \int \ln x \ln x \, dx.$$

Wir wissen nach Beispiel (2), dass

$$\frac{d}{dx}(x \ln x - x) = \ln x$$

ist, also ergibt sich

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

Schließlich ermöglicht es partielle Integration bisweilen, zu einer Rekursionsformel zu kommen. Als Beispiel betrachten wir

$$(6) I_m := \int_a^b \sin^m x \, dx \quad \text{für } m \in \mathbb{N}.$$

Für  $m \geq 2$  ist

$$\begin{aligned}
 I_m &= - \int_a^b \sin^{m-1} x \cos' x \, dx \\
 &= [-\cos x \sin^{m-1} x]_a^b + (m-1) \int_a^b \sin^{m-2} x \cos^2 x \, dx \\
 &= [-\cos x \sin^{m-1} x]_a^b + (m-1) \int_a^b \sin^{m-2} x (1 - \sin^2 x) \, dx \\
 &= [-\cos x \sin^{m-1} x]_a^b + (m-1) I_{m-2} - (m-1) I_m,
 \end{aligned}$$

also

$$I_m = -\frac{1}{m} [\cos x \sin^{m-1} x]_a^b + \frac{m-1}{m} I_{m-2}.$$

Dies ist eine Rekursionsformel, mit deren Hilfe sich  $I_m$  berechnen läßt, denn  $I_0$  und  $I_1$  sind wohlbekannt.

Die zweite wichtige (und vielseitiger anwendbare) Integralumformung ist die Substitutionsregel. Man erhält sie, kurz gesagt, durch Integration der Kettenregel.

**4.1 Satz** (Substitutionsregel). *Sei  $f : [a, b] \rightarrow \mathbb{R}$  stetig und  $g : [c, d] \rightarrow [a, b]$  stetig differenzierbar. Dann gilt*

$$\int_c^d f(g(t))g'(t) \, dt = \int_{g(c)}^{g(d)} f(x) \, dx.$$

Ist  $g : [c, d] \rightarrow [a, b]$  außerdem bijektiv, so gilt

$$\int_a^b f(x) \, dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(t))g'(t) \, dt.$$

*Beweis.* Zu  $f$  gibt es nach Satz 3.1 eine Stammfunktion  $F$  auf  $[a, b]$ . Nach der Kettenregel 1.5 aus Kapitel 6 gilt

$$(F \circ g)'(t) = F'(g(t))g'(t) = f(g(t))g'(t)$$

für  $t \in [c, d]$ , also

$$\begin{aligned} \int_c^d f(g(t))g'(t) dt &= \int_c^d (F \circ g)'(t) dt = (F \circ g)(d) - (F \circ g)(c) \\ &= F(g(d)) - F(g(c)) = \int_{g(c)}^{g(d)} F'(x) dx = \int_{g(c)}^{g(d)} f(x) dx. \end{aligned}$$

Sei jetzt  $g : [c, d] \rightarrow [a, b]$  bijektiv. Aus den Eigenschaften von  $g$  folgert man leicht, dass entweder  $g(c) = a$ ,  $g(d) = b$  oder  $g(c) = b$ ,  $g(d) = a$  gelten muß. Im ersten Fall ist dann die Behauptung klar; im zweiten Fall ergibt sie sich aus

$$\int_b^a f(x) dx = \int_{g(c)}^{g(d)} f(x) dx = \int_{g^{-1}(b)}^{g^{-1}(a)} f(g(t))g'(t) dt$$

und Satz 2.3. ■

Die Anwendung dieser Regel wollen wir zunächst an einem sehr einfachen Beispiel studieren. Es sei das Integral

$$\int_c^d \sin^3 t \cos t dt$$

zu berechnen. Es fällt sofort ins Auge, dass  $\cos$  die Ableitung von  $\sin$  ist, also wird man versuchen, den Integranden in der Form  $f(g(t))g'(t)$  mit  $g(t) = \sin t$  zu schreiben. Man hat also  $f(x) = x^3$  zu wählen. Dann ergibt die Substitutionsregel

$$\begin{aligned} \int_c^d \sin^3 t \cos t dt &= \int_c^d f(g(t))g'(t) dt \\ &= \int_{\sin c}^{\sin d} f(x) dx = \int_{\sin c}^{\sin d} x^3 dx = \left[ \frac{1}{4} x^4 \right]_{\sin c}^{\sin d} \\ &= \frac{1}{4} \sin^4 d - \frac{1}{4} \sin^4 c. \end{aligned}$$

Dies war die korrekte und einwandfrei geschriebene Anwendung der Substitutionsregel. In der Praxis verwendet man eine nicht ganz einwandfreie, aber leicht zu merkende Schreibweise, die wir jetzt an demselben Beispiel erläutern wollen. Man würde  $x$  statt  $g$  schreiben und sagen, dass man „die Substitution

$$\sin t = x$$

macht“. Hierbei muß man im Kopf haben, dass  $x$  als Funktion von  $t$  aufzufassen ist. Man kann sie nach  $t$  differenzieren:

$$\frac{dx}{dt} = \cos t.$$

Schreibt man dies formal in der Weise

$$dx = \cos t \, dt$$

und setzt dies in das Integral ein, so erhält man

$$\int_c^d \sin^3 t \cos t \, dt = \int x^3 \, dx,$$

wobei noch die Grenzen einzusetzen sind nach der Regel: Ist  $t = c$ , so ist  $x = \sin c$  usw., also

$$= \int_{\sin c}^{\sin d} x^3 \, dx.$$

Man kommt bei diesen formalen Umformungen also zum richtigen Ergebnis. Bezeichnet nun  $F$  eine Stammfunktion des Integranden (nämlich  $F(x) = \frac{1}{4}x^4$ ), so ist dieses Integral

$$= [F(x)]_{x=\sin c}^{x=\sin d} = [F(\sin t)]_{t=c}^{t=d}$$

Man erhält also das richtige Ergebnis, wenn man am Ende wieder  $x = \sin t$  einsetzt. Da hier  $d$  beliebig ist, haben wir in Wirklichkeit nicht nur ein bestimmtes Integral gefunden, sondern eine Stammfunktion.

Die folgenden Beispiele zur Anwendung der Substitutionsregel behandeln wir immer in dieser sehr einprägsamen Form. Es sollte keine Mühe bereiten, die Umformungen auch in korrekter Weise darzustellen. Die Merkregel ist aber zur Auffindung passender Substitutionen (für die es kein Patentrezept gibt) sehr hilfreich.

Es versteht sich in den folgenden Beispielen von selbst, dass die Integranden i.a. nur auf passenden Intervallen erklärt sind; dies wird nicht jeweils im Einzelnen angegeben. Da wir in Wahrheit immer Stammfunktionen bestimmen, geben wir, der Konvention entsprechend, keine Integrationsgrenzen an.

$$(1) \int \tan t \, dt = \int \frac{\sin t}{\cos t} \, dt \quad \left[ \text{Substitution: } \begin{array}{l} x = \cos t \\ dx = -\sin t \, dt \end{array} \right]$$

$$\int \tan t \, dt = - \int \frac{1}{x} \, dx = -\ln |x| = -\ln |\cos t|.$$

Das läßt sich verallgemeinern. Zu bestimmen sei etwa

$$\int \frac{g'(t)}{g(t)} dt \quad \text{mit } g > 0. \quad \left[ \text{Substitution: } \begin{array}{l} x = g(t) \\ dx = g'(t) dt \end{array} \right]$$

$$\int \frac{g'(t)}{g(t)} dt = \int \frac{1}{x} dx = \ln x = \ln g(t).$$

$$(2) \int \frac{x}{(1+x^2)^2} dx \quad \left[ \text{Substitution } \begin{array}{l} 1+x^2 = t \\ 2x dx = dt \end{array} \right]$$

$$= \frac{1}{2} \int \frac{1}{t^2} dt = -\frac{1}{2} \frac{1}{t} = -\frac{1}{2} \frac{1}{1+x^2}.$$

- (3) Ein etwas schwierigeres Beispiel, bei dem das vorstehende nach algebraischer Umformung und partieller Integration benutzt wird:

$$\int \frac{1}{(1+x^2)^2} dx = \int \left( \frac{1}{1+x^2} - \frac{x^2}{(1+x^2)^2} \right) dx$$

$$= \arctan x - \int \frac{x^2}{(1+x^2)^2} dx,$$

$$\int \frac{x^2}{(1+x^2)^2} dx = \int x \cdot \frac{x}{(1+x^2)^2} dx \quad (\text{vgl. (2)})$$

$$= -\frac{1}{2} \frac{x}{1+x^2} + \frac{1}{2} \int \frac{1}{1+x^2} dx$$

$$= -\frac{1}{2} \frac{x}{1+x^2} + \frac{1}{2} \arctan x$$

$$\Rightarrow \int \frac{1}{(1+x^2)^2} dx = \frac{1}{2} \frac{x}{1+x^2} + \frac{1}{2} \arctan x.$$

$$(4) \int_a^b \sqrt{1-x^2} dx \quad \text{mit } [a, b] \subset [-1, 1]$$

Hier substituieren wir  $x = \sin t$ . Im Integranden fassen wir also  $x$  auf als  $x = \sin(\arcsin x)$  und substituieren dann  $\arcsin x = t$ . Hierbei ist wesentlich, dass die Abbildung

$$\sin : \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right] \rightarrow [-1, 1]$$

bijektiv ist (und somit eine Umkehrabbildung besitzt). Im Grunde benutzen wir also die zweite, speziellere Form der Substitutionsregel 4.1. Wir erhalten

$$\begin{aligned}
\int \sqrt{1-x^2} dx &= \int \sqrt{1-\sin^2 t} \cos t dt \\
&= \int \cos^2 t dt \quad [\text{partielle Integration}] \\
&= \frac{1}{2} \sin t \cos t + \frac{1}{2} t \\
&= \frac{1}{2} \sin t \sqrt{1-\sin^2 t} + \frac{1}{2} t \\
&= \frac{1}{2} x \sqrt{1-x^2} + \frac{1}{2} \arcsin x.
\end{aligned}$$

Speziell ergibt sich (mit einem Grenzübergang)

$$\int_{-1}^1 \sqrt{1-x^2} dx = \left[ \frac{1}{2} x \sqrt{1-x^2} + \frac{1}{2} \arcsin x \right]_{-1}^1 = \frac{\pi}{2}.$$

Damit ist die Deutung der Zahl  $\pi$  als Flächeninhalt des Kreises vom Radius 1 gewonnen.

- (5) Häufig muß man Substitution und partielle Integration kombinieren, wie im folgenden Beispiel.

$$\begin{aligned}
\int \arctan x dx &= x \arctan x - \int \frac{x}{1+x^2} dx \\
\int \frac{x}{1+x^2} dx &\quad \left[ \begin{array}{l} \text{Substitution } x^2 = t \\ 2x dx = dt \end{array} \right] \\
&= \frac{1}{2} \int \frac{1}{1+t} dt = \frac{1}{2} \ln(1+t) = \frac{1}{2} \ln(1+x^2),
\end{aligned}$$

also

$$\int \arctan x dx = x \arctan x - \frac{1}{2} \ln(1+x^2).$$

### Numerische Integration

Schon für relativ einfache Funktionen, die aus den speziellen, elementaren Funktionen zusammengesetzt sind, ist es oft prinzipiell unmöglich, explizit die Stammfunktion durch elementare Funktionen auszudrücken. Integrale über solche Funktionen können nur näherungsweise (mit beliebiger Genauigkeit) berechnet werden. Hierzu gibt es verschiedene Verfahren. Natürlich kann man grundsätzlich auf die Integraldefinition zurückgehen, also die gegebene Funktion durch Treppenfunktionen approximieren. Das wäre jedoch meist

mit unnötig viel Rechenaufwand verbunden. Es gibt bessere Verfahren, die sich vor allem dadurch auszeichnen, dass bei Kenntnis von Ableitungen der Funktion günstige Fehlerabschätzungen möglich sind. Wir wollen ein solches Verfahren betrachten.

Zur Berechnung von  $\int_a^b f(x) dx$  wird man im allgemeinen so vorgehen, dass man zuerst das Intervall  $[a, b]$  in genügend kleine Teilintervalle zerlegt und dann in jedem Teilintervall die Funktion  $f$  ersetzt durch eine Funktion, deren Integral man berechnen kann. Die naheliegendste Ersatzfunktion ist eine affine. Wir wollen zunächst nur ein Teilintervall betrachten, etwa das Intervall  $[-h, h]$ . Wir ersetzen  $f$  in  $[-h, h]$  durch die affine Funktion, die in den Endpunkten des Intervalls dieselben Werte annimmt wie  $f$ . Deren Integral ist offenbar

$$h(f(-h) + f(h)) =: A.$$

Das Problem ist jetzt, eine Fehlerabschätzung, d.h. eine obere Schranke für

$$\left| \int_{-h}^h f(x) dx - A \right|$$

zu finden. Ohne weitere Information über  $f$  können wir natürlich keine solche Schranke aufstellen. Nun ist die obige Differenz gleich Null, wenn  $f$  eine affine Funktion ist. Wir hoffen daher, dass die Differenz klein ist, wenn  $f$  nur wenig von einer affinen Funktion abweicht. Da affine Funktionen durch  $f'' = 0$  gekennzeichnet sind, können wir das Abweichen von einer affinen Funktion durch die Größe  $\|f''\|$  messen, vorausgesetzt, dass  $f''$  existiert und beschränkt ist.

**4.2 Satz** (Sehnen- oder Trapezregel). *Sei  $h > 0$  und  $f : [-h, h] \rightarrow \mathbb{R}$  zweimal stetig differenzierbar, sei*

$$A := h(f(-h) + f(h)).$$

*Dann gilt*

$$\int_{-h}^h f(x) dx = A - \frac{2}{3}h^3 f''(c)$$

*mit passendem  $c \in [-h, h]$ , also*

$$\left| \int_{-h}^h f(x) dx - A \right| \leq \frac{2}{3}h^3 \|f''\|.$$

*Beweis.* Durch zweimalige partielle Integration erhalten wir

$$\begin{aligned}\int_{-h}^h f(x) dx &= [xf(x)]_{-h}^h - \int_{-h}^h x f'(x) dx \\ &= A - \left[ \frac{1}{2}(x^2 - h^2)f'(x) \right]_{-h}^h + \int_{-h}^h \frac{1}{2}(x^2 - h^2)f''(x) dx.\end{aligned}$$

Hier ist

$$\left[ \frac{1}{2}(x^2 - h^2)f'(x) \right]_{-h}^h = 0$$

(deshalb wurde bei der zweiten partiellen Integration zu  $x$  die Stammfunktion  $\frac{1}{2}(x^2 - h^2)$  gewählt). Wegen  $h^2 - x^2 \geq 0$  für  $x \in [-h, h]$  folgt aus dem Mittelwertsatz der Integralrechnung 2.5 mit passendem  $c \in [-h, h]$

$$\begin{aligned}\int_{-h}^h (h^2 - x^2)f''(x) dx &= f''(c) \int_{-h}^h (h^2 - x^2) dx \\ &= f''(c) \left[ h^2x - \frac{x^3}{3} \right]_{-h}^h = f''(c) \frac{4}{3}h^3,\end{aligned}$$

also

$$\int_{-h}^h f(x) dx = A - \frac{2}{3}h^3 f''(c)$$

und daher die Behauptung. ■

Jetzt wollen wir, wie oben angedeutet, das Intervall  $[a, b]$  in Teilintervalle zerlegen und die Sehnenregel auf die Teilintervalle anwenden.

Sei  $n \in \mathbb{N}$ . Wir zerlegen  $[a, b]$  in  $n$  Intervalle gleicher Länge und setzen dazu

$$h := \frac{b-a}{n}, \quad x_j := a + jh \quad \text{für } j = 0, \dots, n.$$

Anwendung der Sehnenregel auf das Intervall  $[x_{j-1}, x_j]$  ( $\frac{h}{2}$  spielt also jetzt die Rolle des früheren  $h$ ) ergibt

$$\int_{x_{j-1}}^{x_j} f(x) dx = \frac{h}{2} [f(x_{j-1}) + f(x_j)] - \frac{h^3}{12} f''(c_j)$$

mit  $c_j \in [x_{j-1}, x_j]$  ( $j = 1, \dots, n$ ). Summieren wir über  $j$  und setzen

$$A_n := h \left[ \frac{1}{2}f(a) + f(a+h) + \dots + f(a+(n-1)h) + \frac{1}{2}f(b) \right],$$

so erhalten wir

$$\int_a^b f(x) dx = A_n - \frac{h^3}{12} \sum_{j=1}^n f''(c_j),$$

also

$$\left| \int_a^b f(x) dx - A_n \right| \leq \frac{h^3}{12} n \|f''\| = \frac{h^2}{12} (b-a) \|f''\|.$$

Wir sehen also, dass der Fehler, den wir bei der Approximation des Integrals  $\int_a^b f(x) dx$  durch die endliche Näherungssumme  $A_n$  machen, wie das Quadrat der Schrittweite  $h$  gegen Null geht.

## 7.5 Parameterabhängige Integrale

In der Taylorschen Formel 3.3 hatten wir das Restglied in der Form

$$R_{n+1}(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt$$

angegeben. Dies ist ein Beispiel für ein „parameterabhängiges Integral“. Im Integranden und in der oberen Grenze kommt ein „Parameter“ vor, d.h. eine Zahl, die aus einer gegebenen Teilmenge von  $\mathbb{R}$  gewählt werden kann, so dass also durch das Integral eine auf dieser Teilmenge definierte Funktion erklärt wird. Wir wollen Eigenschaften einer so erklärten Funktion betrachten; dabei beschränken wir uns auf den Fall, dass der Parameter im Integranden vorkommt. Zur exakten Formulierung benötigen wir den Begriff einer „Funktion von zwei Veränderlichen“.

Wir bezeichnen das kartesische Produkt  $\mathbb{R} \times \mathbb{R}$ , also die Menge aller geordneten Paare reeller Zahlen, mit  $\mathbb{R}^2$ .

**Definition.** Sei  $A \subset \mathbb{R}^2$ . Eine Funktion  $f : A \rightarrow \mathbb{R}$  heißt „Funktion von zwei (reellen) Veränderlichen“. Statt  $f((x, y))$  schreibt man  $f(x, y)$ .

Die Funktion  $f : A \rightarrow \mathbb{R}$  heißt *gleichmäßig stetig*  $\Leftrightarrow$

$$\forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall (x, t), (x', t') \in A : \\ |x - x'| < \delta \wedge |t - t'| < \delta \Rightarrow |f(x, t) - f(x', t')| < \varepsilon.$$

Ist  $D, [a, b] \subset \mathbb{R}$  und  $f : D \times [a, b] \rightarrow \mathbb{R}$  gleichmäßig stetig, so ist insbesondere für jedes  $x \in D$  die Funktion

$$f(x, \cdot) : [a, b] \rightarrow \mathbb{R} : t \mapsto f(x, t)$$

stetig und daher eine Regelfunktion. Wir können also

$$F(x) := \int_a^b f(x, t) dt \quad \text{für } x \in D$$

definieren. Wir fragen jetzt nach Eigenschaften dieser Funktion  $F$  wie Stetigkeit und Differenzierbarkeit. Es sei darauf hingewiesen, dass die im folgenden getroffenen Voraussetzungen wesentlich stärker sind als notwendig wäre. (Für allgemeinere Aussagen siehe z.B. Barner-Flohr.) Wir begnügen uns jedoch mit diesen einfach zu beweisenden Resultaten, die für viele Anwendungen ausreichen.

**5.1 Satz.** Sei  $D, [a, b] \subset \mathbb{R}$ , sei  $f : D \times [a, b] \rightarrow \mathbb{R}$  eine gleichmäßig stetige Funktion. Dann ist die durch

$$F(x) := \int_a^b f(x, t) dt \quad \text{für } x \in D$$

definierte Funktion  $F$  stetig.

*Beweis.* Sei  $x \in D$  und  $\varepsilon \in \mathbb{R}^+$  gegeben. Wegen der gleichmäßigen Stetigkeit von  $f$  gibt es ein  $\delta \in \mathbb{R}^+$  mit

$$|f(x, t) - f(x', t)| < \frac{\varepsilon}{b - a}$$

für alle  $t \in [a, b]$  und alle  $x, x' \in D$  mit  $|x - x'| < \delta$ . Für  $x, x' \in D$  mit  $|x - x'| < \delta$  folgt also

$$|F(x) - F(x')| \leq \int_a^b |f(x, t) - f(x', t)| dt \leq \varepsilon. \quad \blacksquare$$

Jetzt fragen wir nach der Differenzierbarkeit von  $F$ . Sie erfordert natürlich eine geeignete Differenzierbarkeitsvoraussetzung an  $f$ .

**Definition.** Sei  $f : D \times [a, b] \rightarrow \mathbb{R}$  eine Funktion. Sie heißt *partiell differenzierbar nach der ersten Veränderlichen*, wenn für alle  $t \in [a, b]$  die Funktion

$$f(\cdot, t) : D \rightarrow \mathbb{R} : x \mapsto f(x, t)$$

differenzierbar ist. Ihre Ableitung an der Stelle  $x$  wird dann mit  $\partial_1 f(x, t)$  bezeichnet. Die Funktion

$$\partial_1 f : D \times [a, b] \rightarrow \mathbb{R} : (x, t) \mapsto \partial_1 f(x, t)$$

heißt *partielle Ableitung von  $f$  nach der ersten Veränderlichen*.

**5.2 Satz.** Sei  $[c, d], [a, b] \subset \mathbb{R}$ , sei  $f : [c, d] \times [a, b] \rightarrow \mathbb{R}$  eine gleichmäßig stetige Funktion, deren partielle Ableitung  $\partial_1 f$  existiert und gleichmäßig stetig ist. Dann ist die durch

$$F(x) := \int_a^b f(x, t) dt \quad \text{für } x \in [c, d]$$

definierte Funktion  $F$  differenzierbar, und es gilt

$$F'(x) = \int_a^b \partial_1 f(x, t) dt$$

(„Vertauschbarkeit von Differentiation und Integration“).

*Beweis.* Sei  $x \in [c, d]$  und  $\varepsilon \in \mathbb{R}^+$  gegeben. Für  $h \neq 0$  mit  $x + h \in [c, d]$  gilt

$$\frac{F(x+h) - F(x)}{h} - \int_a^b \partial_1 f(x, t) dt = \int_a^b \left( \frac{f(x+h, t) - f(x, t)}{h} - \partial_1 f(x, t) \right) dt.$$

Für jedes  $t \in [a, b]$  wenden wir auf die Funktion  $f(\cdot, t)$  den Mittelwertsatz der Differentialrechnung 2.2 aus Kapitel 6 an. Danach existiert ein  $\vartheta_t \in (0, 1)$  mit

$$\frac{f(x+h, t) - f(x, t)}{h} = \partial_1 f(x + \vartheta_t h, t).$$

Nach Voraussetzung ist  $\partial_1 f$  gleichmäßig stetig, also existiert ein  $\delta \in \mathbb{R}^+$  mit

$$|\partial_1 f(x, t) - \partial_1 f(x', t)| < \frac{\varepsilon}{b-a}$$

für alle  $t \in [a, b]$  und alle  $x, x' \in [c, d]$  mit  $|x - x'| < \delta$ . Sei  $|h| < \delta$ . Dann gilt  $|x - (x + \vartheta_t h)| = \vartheta_t |h| < \delta$ , also

$$\left| \frac{f(x+h, t) - f(x, t)}{h} - \partial_1 f(x, t) \right| = |\partial_1 f(x + \vartheta_t h, t) - \partial_1 f(x, t)| < \frac{\varepsilon}{b-a}$$

und daher

$$\left| \frac{F(x+h) - F(x)}{h} - \int_a^b \partial_1 f(x, t) dt \right| < \varepsilon. \quad \blacksquare$$

Nun fragen wir analog, ob man „unter dem Integralzeichen integrieren“ darf.

**5.3 Satz.** Sei  $f : [c, d] \times [a, b] \rightarrow \mathbb{R}$  gleichmäßig stetig. Dann gilt

$$\int_c^d \left( \int_a^b f(x, t) dt \right) dx = \int_a^b \left( \int_c^d f(x, t) dx \right) dt$$

(„Vertauschbarkeit der Integrationsreihenfolge“).

*Beweis.* Setze

$$G(y) := \int_c^y \left( \int_a^b f(x, t) dt \right) dx \quad \text{für } y \in [c, d].$$

Nach Satz 5.1 ist der Integrand des äußeren Integrals stetig; nach dem Hauptsatz der Differential- und Integralrechnung 3.1 ist daher  $G$  differenzierbar und

$$G'(y) = \int_a^b f(y, t) dt.$$

Setze

$$H(y) := \int_a^b \left( \int_c^y f(x, t) dx \right) dt \quad \text{für } y \in [c, d].$$

Die durch

$$h(y, t) := \int_c^y f(x, t) dx \quad \text{für } (y, t) \in [c, d] \times [a, b]$$

definierte Funktion  $h$  ist (wieder nach Satz 3.1) partiell differenzierbar nach der ersten Veränderlichen, und es gilt

$$\partial_1 h = f.$$

Also ist  $\partial_1 h$  gleichmäßig stetig. Nach Satz 5.2 folgt die Differenzierbarkeit der Funktion  $H$  und die Gleichung

$$H'(y) = \int_a^b \partial_1 h(y, t) dt = \int_a^b f(y, t) dt.$$

Es ist also  $(G - H)' = 0$  und daher  $G - H = \text{const.}$  Wegen  $G(c) = 0 = H(c)$  ist  $G = H$ . Daraus folgt die Behauptung. ■

Durch ein Beispiel wollen wir noch zeigen, dass man die Integrationsreihenfolge keineswegs immer vertauschen darf.

**Beispiel.** Die Funktion  $f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  sei definiert durch

$$f(x, y) := \begin{cases} \frac{1}{y^2} & \text{für } 0 < x < y < 1, \\ -\frac{1}{x^2} & \text{für } 0 < y < x < 1, \\ 0 & \text{sonst.} \end{cases}$$

Ist  $y = 0$  oder  $y = 1$ , so ist  $f(x, y) = 0$  für alle  $x \in [0, 1]$ . Sei  $y \in (0, 1)$ . Dann ist

$$f(x, y) = \begin{cases} 0 & \text{für } x = 0, \\ \frac{1}{y^2} & \text{für } 0 < x < y, \\ 0 & \text{für } x = y, \\ -\frac{1}{x^2} & \text{für } y < x < 1, \\ 0 & \text{für } x = 1. \end{cases}$$

Also ist  $f(\cdot, y)$  auf dem Intervall  $(0, y)$  konstant und auf dem Intervall  $(y, 1)$  Einschränkung einer auf  $[y, 1]$  stetigen Funktion. Es folgt, dass  $f(\cdot, y)$  Regelfunktion ist. Daher ist das Integral

$$F(y) := \int_0^1 f(x, y) dx$$

definiert. Es ist  $F(0) = F(1) = 0$ . Sei  $y \in (0, 1)$ . Dann ist

$$F(y) = \int_0^y \frac{1}{y^2} dx + \int_y^1 \left(-\frac{1}{x^2}\right) dx = \left[\frac{1}{y^2}x\right]_{x=0}^{x=y} + \left[\frac{1}{x}\right]_{x=y}^{x=1} = 1.$$

Also ist  $F$  Regelfunktion und

$$\int_0^1 F(y) dy = 1.$$

Analog findet man, dass für jedes  $x \in [0, 1]$  die Funktion  $f(x, \cdot)$  Regelfunktion ist und dass

$$\begin{aligned} G(x) &:= \int_0^1 f(x, y) dy \\ &= \int_0^x \left(-\frac{1}{x^2}\right) dy + \int_x^1 \frac{1}{y^2} dy = -1 \end{aligned}$$

für  $x \in (0, 1)$  sowie  $G(0) = G(1) = 0$  gilt. Also ist  $G$  Regelfunktion und  $\int_0^1 G(x) dx = -1$ . Damit ist

$$\int_0^1 \int_0^1 f(x, y) dx dy = 1 \neq -1 = \int_0^1 \int_0^1 f(x, y) dy dx$$

gezeigt.

## 7.6 Uneigentliche Integrale

Bisher können wir nur Regelfunktionen integrieren. Eine solche Funktion ist stets auf einem kompakten Intervall definiert und beschränkt. Es kommt jedoch häufig vor, dass Funktionen über unendliche Intervalle oder unbeschränkte Funktionen zu integrieren sind. In naheliegender Weise können solche Integrale durch Grenzübergänge definiert werden. Wir betrachten zunächst den Fall eines unbeschränkten Integrationsintervalls.

**Definition.** Sei  $a \in \mathbb{R}$ , sei  $f : [a, \infty) \rightarrow \mathbb{R}$  eine Funktion derart, dass für jedes  $b \in [a, \infty)$  die Einschränkung  $f|_{[a, b]}$  eine Regelfunktion ist. Falls der Grenzwert

$$\lim_{b \rightarrow \infty} \int_a^b f(x) dx$$

existiert, bezeichnet man ihn mit  $\int_a^\infty f(x) dx$  und sagt, dass das (uneigentliche) Integral  $\int_a^\infty f(x) dx$  konvergiert.

Ganz analog definiert man  $\int_{-\infty}^a f(x) dx$  und

$$\int_{-\infty}^{\infty} f(x) dx := \int_{-\infty}^a f(x) dx + \int_a^{\infty} f(x) dx,$$

wobei  $a \in \mathbb{R}$  beliebig und die rechte Seite offenbar unabhängig von  $a$  ist.

Im Folgenden wird von allen auftretenden Funktionen stillschweigend vorausgesetzt, dass ihre Einschränkungen auf kompakte Intervalle Regelfunktionen sind.

Für die Konvergenz uneigentlicher Integrale gilt das folgende Cauchy-Kriterium.

**6.1 Satz** (Cauchy-Kriterium). *Das Integral  $\int_a^{\infty} f(x) dx$  konvergiert*

$$\Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \quad \exists u \in \mathbb{R} \quad \forall x_1, x_2 \geq u : \left| \int_{x_1}^{x_2} f(x) dx \right| < \varepsilon.$$

Der Satz ist analog zum Cauchy-Kriterium 2.3 aus Kapitel 3 für Reihen und zum Cauchy-Kriterium 2.3 aus Kapitel 4 für Grenzwerte von Funktionen; man beweist ihn analog wie letzteres.

**Beispiele.** (1)  $\int_1^{\infty} \frac{1}{x^s} dx$ . Für  $b \geq 1$  und  $s \neq 1$  ist

$$\int_1^b \frac{1}{x^s} dx = \left[ \frac{1}{1-s} \frac{1}{x^{s-1}} \right]_1^b = \frac{1}{s-1} \left( 1 - \frac{1}{b^{s-1}} \right).$$

Also ist das Integral  $\int_1^{\infty} \frac{1}{x^s} dx$  konvergent für  $s > 1$ , nicht konvergent für  $s < 1$ . Für  $s = 1$  ist es wegen  $\int_1^b \frac{1}{x} dx = \ln b$  ebenfalls nicht konvergent.

(2)  $\int_{-\infty}^{\infty} \frac{1}{1+x^2} dx$ . Für  $b \geq 0$  gilt

$$\int_{-b}^0 \frac{1}{1+x^2} dx = [\arctan x]_{-b}^0 = -\arctan(-b) \rightarrow \frac{\pi}{2} \quad \text{für } b \rightarrow \infty,$$

analog

$$\int_0^b \frac{1}{1+x^2} dx \rightarrow \frac{\pi}{2}, \quad \text{also} \quad \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = \pi.$$

- (3)  $\int_0^{\infty} \frac{\sin x}{x} dx$  ist konvergent (an der Stelle 0 ist der Integrand gleich 1 zu setzen; er ist dann stetig). In der Tat, für  $b > a > 0$  ergibt partielle Integration

$$\int_a^b \frac{\sin x}{x} dx = \left[ -\frac{1}{x} \cos x \right]_a^b - \int_a^b \frac{\cos x}{x^2} dx.$$

Sei  $\varepsilon \in \mathbb{R}^+$ . Für  $x_2 > x_1 > \frac{1}{\varepsilon}$  gilt

$$\left| \int_{x_1}^{x_2} \frac{\cos x}{x^2} dx \right| \leq \int_{x_1}^{x_2} \frac{1}{x^2} dx = \left[ -\frac{1}{x} \right]_{x_1}^{x_2} = \frac{1}{x_1} - \frac{1}{x_2} < \frac{1}{x_1} < \varepsilon.$$

Nach dem Cauchy-Kriterium 6.1 ist also  $\int_a^{\infty} \frac{\cos x}{x^2} dx$  konvergent und daher auch  $\int_a^{\infty} \frac{\sin x}{x} dx$  konvergent.

### Das Integralkriterium für Reihen

An dieser Stelle können wir ein Konvergenzkriterium für Reihen nachtragen, dessen Formulierung den Begriff des uneigentlichen Integrals erfordert.

**6.2 Satz** (Integralkriterium für Reihen). *Sei  $f : [1, \infty) \rightarrow \mathbb{R}^+$  eine monoton fallende Funktion. Dann ist die Reihe*

$$\sum_{n=1}^{\infty} f(n)$$

*genau dann konvergent, wenn das uneigentliche Integral*

$$\int_1^{\infty} f(x) dx$$

*konvergent ist.*

*Beweis.* Für  $k \in \mathbb{N}$  und  $k-1 \leq x \leq k$  gilt  $f(k) \leq f(x) \leq f(k-1)$  und daher

$$f(k) \leq \int_{k-1}^k f(x) dx \leq f(k-1).$$

Summation ergibt

$$\sum_{k=2}^n f(k) \leq \int_1^n f(x) dx \leq \sum_{k=1}^{n-1} f(k).$$

Wegen  $f \geq 0$  liest man hieran die Behauptung ab.  $\blacksquare$

**Beispiele.** (1)  $\sum_{n=1}^{\infty} \frac{1}{n^s}$  konvergiert für  $s > 1$  und divergiert für  $s \leq 1$  (vgl. Beispiel (1) nach Satz 6.1)

(2)  $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)^s}$  ist für  $s > 1$  konvergent, für  $s \leq 1$  divergent. In der Tat, erhalten wir für  $b \geq 2$  und  $s \neq 1$  [ Substitution  $x = e^t$  ]

$$\int_2^b \frac{1}{x(\ln x)^s} dx = \int_{\ln 2}^{\ln b} \frac{1}{t^s} dt = \left[ \frac{1}{1-s} \frac{1}{t^{s-1}} \right]_{\ln 2}^{\ln b} = \frac{1}{s-1} \left[ \frac{1}{(\ln 2)^{s-1}} - \frac{1}{(\ln b)^{s-1}} \right].$$

Für  $s > 1$  ist also  $\int_2^{\infty} \frac{1}{x(\ln x)^s} dx$  konvergent, für  $s < 1$  divergent. Für  $s = 1$  gilt

$$\int_2^b \frac{1}{x \ln x} dx = \int_{\ln 2}^{\ln b} \frac{1}{t} dt = [\ln t]_{\ln 2}^{\ln b} = \ln \ln b - \ln \ln 2,$$

also ist  $\int_2^{\infty} \frac{1}{x \ln x} dx$  ebenfalls divergent.

Wir kommen zum zweiten Typ uneigentlicher Integrale, bei denen der Integrand an einer Integrationsgrenze nicht erklärt ist und auch nicht so erklärt werden kann, dass eine Regelfunktion entsteht.

**Definition.** Sei  $f : (a, b] \rightarrow \mathbb{R}$  eine Funktion derart, dass für jedes  $\varepsilon \in \mathbb{R}^+$  (mit  $\varepsilon < b - a$ ) die Einschränkung  $f|_{[a+\varepsilon, b]}$  eine Regelfunktion ist. Man sagt, „das Integral  $\int_a^b f(x) dx$  konvergiert“, wenn

$$\lim_{\varepsilon \searrow 0} \int_{a+\varepsilon}^b f(x) dx$$

existiert und bezeichnet dann diesen Grenzwert mit  $\int_a^b f(x) dx$ .

Analog wird  $\int_a^b f(x) dx$  behandelt, wenn  $f$  nur auf  $[a, b)$  erklärt ist. Ist  $f$  nur auf  $(a, b)$  erklärt, so bezeichnet man  $\int_a^b f(x) dx$  als konvergent, wenn für ein (und dann für jedes  $a$ )  $c \in (a, b)$  die Integrale

$$\int_a^c f(x) dx \quad \text{und} \quad \int_c^b f(x) dx$$

konvergieren; in diesem Fall setzt man

$$\int_a^b f(x) dx := \int_a^c f(x) dx + \int_c^b f(x) dx.$$

**Beispiele.** (1) Wegen

$$\int_{\varepsilon}^1 \frac{1}{x^s} dx = \begin{cases} \frac{1}{1-s}(1 - \varepsilon^{1-s}) & \text{für } s \neq 1, \\ -\ln \varepsilon & \text{für } s = 1 \end{cases}$$

ist  $\int_0^1 \frac{1}{x^s} dx$  für  $s < 1$  konvergent, für  $s \geq 1$  jedoch nicht.

(2) Wegen  $\int_{\varepsilon}^1 \ln x dx = [x \ln x - x]_{\varepsilon}^1 = -1 - \varepsilon \ln \varepsilon + \varepsilon$  und  $\lim_{\varepsilon \searrow 0} \varepsilon \ln \varepsilon = 0$  (wie aus Behauptung 1.10 aus Kapitel 5 folgt) ist  $\int_0^1 \ln x dx = -1$ .

(3)  $\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} dx$  konvergiert: Es ist

$$\begin{aligned} \int_0^{1-\varepsilon} \frac{1}{\sqrt{1-x^2}} dx &= [\arcsin x]_0^{1-\varepsilon} = \arcsin(1-\varepsilon) \\ \Rightarrow \lim_{\varepsilon \rightarrow 0} \int_0^{1-\varepsilon} \frac{1}{\sqrt{1-x^2}} dx &= \frac{\pi}{2}. \end{aligned}$$

Analog  $\int_{-1}^0 \frac{1}{\sqrt{1-x^2}} dx = \frac{\pi}{2}$ .

(4) Im Beispiel  $\int_0^1 \sin \frac{1}{x} dx$  ist der Integrand beschränkt, aber  $\lim_{x \searrow 0} \sin \frac{1}{x}$  existiert nicht; daher ist der Integrand nicht Einschränkung einer auf  $[0, 1]$  erklärten Regelfunktion.

Die Substitution  $x = \frac{1}{t}$  ergibt

$$\int_{\varepsilon}^1 \sin \frac{1}{x} dx = \int_1^{\frac{1}{\varepsilon}} \frac{\sin t}{t^2} dt.$$

Da das Integral  $\int_1^{\infty} \frac{\sin t}{t^2} dt$  konvergiert (z.B. nach dem Cauchy-Kriterium), ist  $\int_0^1 \sin \frac{1}{x} dx$  konvergent.



## 8 Funktionenreihen

In diesem Kapitel wollen wir etwas ausführlicher die Möglichkeit studieren, Funktionen durch unendliche Reihen darzustellen. Wir hatten schon im Anschluß an die Taylorformel kurz die Möglichkeit erörtert, bei beliebig oft differenzierbaren Funktionen von den Taylorpolynomen zur Taylorreihe überzugehen. Schon wesentlich früher hatten wir unendliche Reihen benutzt, um gewisse spezielle Funktionen einzuführen, zum Beispiel die Exponentialfunktion durch

$$\exp x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Hierbei ergibt sich unter anderem die Frage, wie man von Eigenschaften der Reihenglieder auf Eigenschaften der Funktion schließen kann. Wir können zum Beispiel zur Berechnung der Ableitung versuchen, „gliedweise“ zu differenzieren. Formal vorgehend, erhält man

$$\exp' x = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{dx^k}{dx} = \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \exp x.$$

Das Ergebnis ist richtig; aber führt ein solches Vorgehen in jedem Fall zum richtigen Ergebnis? Diese und ähnliche Fragen werden im folgenden beantwortet. Wir betrachten zunächst allgemein konvergente Folgen von Funktionen und erst dann spezielle Reihen.

### 8.1 Konvergenz von Funktionenfolgen

Wir kennen bereits zwei verschiedene Konvergenzbegriffe für Funktionenfolgen, und an diese sei zunächst erinnert.

Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge von reellen Funktionen, die sämtlich denselben Definitionsbereich  $D$  haben. Für jedes  $x \in D$  ist dann  $(f_n(x))_{n \in \mathbb{N}}$  eine Folge

reeller Zahlen, und für diese ist ein Konvergenzbegriff wohldefiniert. Wenn für jedes  $x \in D$  die Folge  $(f_n(x))_{n \in \mathbb{N}}$  konvergiert, ist durch

$$f(x) := \lim_{n \rightarrow \infty} f_n(x) \quad (x \in D)$$

eine neue Funktion  $f$  auf  $D$  erklärt, die wir als *Grenzfunktion* der Folge bezeichnen können.

**Definition.** Seien  $f, f_n$  ( $n \in \mathbb{N}$ ) reelle Funktionen auf  $D$ . Die Folge  $(f_n)_{n \in \mathbb{N}}$  *konvergiert (punktweise)* gegen  $f$ , geschrieben

$$\lim_{n \rightarrow \infty} f_n = f \quad \text{oder} \quad f_n \rightarrow f \quad (n \rightarrow \infty),$$

wenn  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  für alle  $x \in D$  gilt.

**Beispiele.** (1) Sei  $D = \mathbb{R}$  und

$$f_n(x) := \sum_{k=0}^n \frac{x^k}{k!}.$$

Dann konvergiert  $(f_n)_{n \in \mathbb{N}}$  gegen  $\exp$ .

(2) Sei  $D = [0, 1]$  und

$$f_n(x) := x^n.$$

Dann konvergiert  $(f_n)_{n \in \mathbb{N}}$  gegen die durch

$$f(x) := \begin{cases} 1 & \text{für } x = 1, \\ 0 & \text{für } 0 \leq x < 1. \end{cases}$$

erklärte Funktion  $f$ .

Bei diesem Beispiel fällt auf, dass zwar jede Funktion  $f_n$  der Folge stetig ist, dass aber die Grenzfunktion unstetig ist. Stetigkeit überträgt sich also bei punktweiser Konvergenz i.a. nicht auf die Grenzfunktion. Um den oft wünschenswerten Schluß von der Stetigkeit der Folgenglieder auf die Stetigkeit der Grenzfunktion zu ermöglichen, braucht man einen Konvergenzbegriff, der schärfer ist als punktweise Konvergenz. Dies ist die bereits in Abschnitt 7.2 benutzte gleichmäßige Konvergenz, die wir jetzt etwas allgemeiner definieren wollen.

**Definition.** Seien  $f, f_n$  ( $n \in \mathbb{N}$ ) reelle Funktionen auf  $D$ , sei  $D' \subset D$ . Die Folge  $(f_n)_{n \in \mathbb{N}}$  *konvergiert gleichmäßig in  $D'$  gegen  $f$* , wenn gilt

$$\forall \varepsilon \in \mathbb{R}^+ \quad \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0 \quad \forall x \in D' : |f_n(x) - f(x)| \leq \varepsilon.$$

Statt „gleichmäßig in  $D'$ “ sagt man kurz „gleichmäßig“.

Ist ein fester Definitionsbereich  $D$  gegeben, so können wir wie früher für Intervalle die *Supremumsnorm* einer beschränkten Funktion  $f : D \rightarrow \mathbb{R}$  erklären durch

$$\|f\| := \sup_{x \in D} |f(x)|.$$

Dann gilt also (für Funktionen  $f, f_n$  auf  $D$ ):

$$\begin{aligned} (f_n)_{n \in \mathbb{N}} &\text{ konvergiert gleichmäßig gegen } f \\ \Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \|f_n - f\| &\leq \varepsilon \\ \Leftrightarrow \lim_{n \rightarrow \infty} \|f_n - f\| &= 0. \end{aligned}$$

Man beachte, dass  $\|f_n - f\| < \varepsilon$  impliziert, dass  $\|f_n - f\|$  definiert, also  $f_n - f$  beschränkt ist. Es wird aber nicht vorausgesetzt, dass  $f, f_n$  beschränkt sind.

Für die Konvergenz von Folgen reeller Zahlen kennen wir das Kriterium von Cauchy. Ein ganz analoges Kriterium gilt auch für die gleichmäßige Konvergenz von Funktionenfolgen. Im folgenden liege stets ein fester Definitionsbereich  $D$  zugrunde.

**Definition.** Die Folge  $(f_n)_{n \in \mathbb{N}}$  von Funktionen auf  $D$  heißt *Cauchy-Folge* genau dann, wenn

$$\forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n, m \geq n_0 : \|f_n - f_m\| < \varepsilon.$$

**1.1 Satz.** Die Folge  $(f_n)_{n \in \mathbb{N}}$  konvergiert genau dann gleichmäßig, wenn sie eine Cauchy-Folge ist.

*Beweis.* „ $\Rightarrow$ “: Sei  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig konvergent gegen  $f$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $\|f_n - f\| < \varepsilon/2$  für  $n \geq n_0$ . Für alle  $n, m \geq n_0$  gilt also

$$\|f_n - f_m\| \leq \|f_n - f\| + \|f - f_m\| < \varepsilon,$$

wobei die Dreiecksungleichung für die Supremumsnorm benutzt wurde.

„ $\Leftarrow$ “: Sei  $(f_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge. Sei  $\varepsilon \in \mathbb{R}^+$ . Nach Voraussetzung existiert ein  $n_0 \in \mathbb{N}$  mit  $\|f_n - f_m\| < \varepsilon$  für alle  $n, m \geq n_0$ . Insbesondere gilt also für jedes  $x \in D$

$$|f_n(x) - f_m(x)| < \varepsilon \quad \text{für } n, m \geq n_0.$$

Die Folge  $(f_n(x))_{n \in \mathbb{N}}$  ist also eine Cauchy-Folge reeller Zahlen und daher nach dem gewöhnlichen Cauchy-Kriterium konvergent gegen eine Zahl, die wir  $f(x)$  nennen. Da  $x \in D$  beliebig war, ist damit eine Funktion  $f : D \rightarrow \mathbb{R}$  erklärt. Wir behaupten, dass  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$  konvergiert.

Sei  $\varepsilon \in \mathbb{R}^+$  und dazu  $n_0$  wie oben. Für beliebiges  $x \in D$  gilt dann

$$|f_n(x) - f_m(x)| < \varepsilon \quad \text{für } n, m \geq n_0.$$

Der Grenzübergang  $m \rightarrow \infty$  liefert

$$|f_n(x) - f(x)| \leq \varepsilon \quad \text{für } n \geq n_0.$$

Da dies für alle  $x \in D$  gilt, folgt  $\|f_n - f\| \leq \varepsilon$  für alle  $n \geq n_0$ . ■

Der folgende Satz rückt die Bedeutung der gleichmäßigen Konvergenz ins rechte Licht.

**1.2 Satz.** *Seien  $f_n, f$  Funktionen auf  $D$  ( $n \in \mathbb{N}$ ), sei  $a \in D$ . Sind alle  $f_n$  stetig in  $a$  und konvergiert  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$ , so ist  $f$  stetig in  $a$ .*

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $m \in \mathbb{N}$  mit  $\|f_m - f\| < \varepsilon/3$ , also

$$|f_m(x) - f(x)| < \frac{\varepsilon}{3} \quad \text{für alle } x \in D.$$

Da  $f_m$  in  $a$  stetig ist, existiert ein  $\delta \in \mathbb{R}^+$  mit

$$|f_m(x) - f_m(a)| < \frac{\varepsilon}{3} \quad \text{für alle } x \in D \text{ mit } |x - a| < \delta.$$

Sei jetzt  $x \in D$  und  $|x - a| < \delta$ . Dann gilt

$$\begin{aligned} |f(x) - f(a)| &\leq |f(x) - f_m(x)| + |f_m(x) - f_m(a)| + |f_m(a) - f(a)| \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned} \quad \blacksquare$$

**Bemerkung.** Natürlich kann man nicht umgekehrt schließen, d.h. wenn  $(f_n)_{n \in \mathbb{N}}$  punktweise gegen  $f$  konvergiert und alle  $f_n$  sowie  $f$  stetig sind, braucht keineswegs die Konvergenz gleichmäßig zu sein.

Kann man in Satz 1.2 „stetig“ durch „differenzierbar“ ersetzen? Das ist nicht der Fall.

**Beispiel.** Sei  $D = \mathbb{R}$ ,  $f_n(x) := \sqrt{x^2 + \frac{1}{n}}$  ( $n \in \mathbb{N}$ ) und  $f(x) := |x|$  für  $x \in \mathbb{R}$ . Dann konvergiert  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$ , wie aus

$$|f_n(x) - f(x)| = \left| \sqrt{x^2 + \frac{1}{n}} - \sqrt{x^2} \right| = \frac{x^2 + \frac{1}{n} - x^2}{\sqrt{x^2 + \frac{1}{n}} + \sqrt{x^2}} \leq \frac{1}{\sqrt{n}}$$

folgt. Jede Funktion  $f_n$  ist differenzierbar, aber  $f$  ist in 0 nicht differenzierbar.

Es kann auch sein, dass zwar die Grenzfunktion  $f$  der Folge  $(f_n)_{n \in \mathbb{N}}$  differenzierbar ist, aber die Folge  $(f'_n)_{n \in \mathbb{N}}$  nicht gegen  $f'$  konvergiert.

**Beispiel.** Sei  $f_n(x) = \frac{1}{n} \sin nx$  und  $f(x) = 0$  für  $x \in \mathbb{R}$ . Dann konvergiert  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$ , alle  $f_n$  sowie  $f$  sind differenzierbar, aber  $(f'_n)_{n \in \mathbb{N}}$  konvergiert nicht punktweise gegen  $f'$ . Es ist nämlich  $f'_n(x) = \cos nx$  und z.B.  $\lim_{n \rightarrow \infty} f'_n(0) = 1$ ,  $f'(0) = 0$ .

Um wirklich Grenzübergang und Differentiation vertauschen zu können, braucht man stärkere Voraussetzungen, z.B. die folgenden:

**1.3 Satz.** Sei  $(f_n)_{n \in \mathbb{N}}$  eine Folge differenzierbarer Funktionen auf  $[a, b]$ . Die Folge  $(f'_n)_{n \in \mathbb{N}}$  konvergiere gleichmäßig, und  $(f_n)_{n \in \mathbb{N}}$  konvergiere an wenigstens einer Stelle  $x_0 \in [a, b]$ . Dann konvergiert  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen eine differenzierbare Funktion  $f$  und  $(f'_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f'$ .

*Beweis.* Wir zeigen zuerst, dass  $(f_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge ist. Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Nach Voraussetzung und Satz 1.1 existiert ein  $n_1 \in \mathbb{N}$  mit

$$\|f'_m - f'_n\| < \frac{\varepsilon}{2(b-a)} \quad \text{für alle } m, n \geq n_1.$$

Da  $(f_n(x_0))_{n \in \mathbb{N}}$  eine Cauchy-Folge ist, existiert ein  $n_2 \in \mathbb{N}$  mit

$$|f_m(x_0) - f_n(x_0)| < \frac{\varepsilon}{2} \quad \text{für } m, n \geq n_2.$$

Sei  $x \in [a, b]$ . Es ist

$$|f_m(x) - f_n(x)| \leq |(f_m - f_n)(x) - (f_m - f_n)(x_0)| + |f_m(x_0) - f_n(x_0)|.$$

Nach dem Mittelwertsatz 2.2 aus Kapitel 6 existiert ein  $z \in [a, b]$  mit

$$(f_m - f_n)(x) - (f_m - f_n)(x_0) = (f'_m(z) - f'_n(z))(x - x_0).$$

Für alle  $m, n \geq n_0 := \max\{n_1, n_2\}$  folgt

$$\begin{aligned} |f_m(x) - f_n(x)| &\leq |f'_m(z) - f'_n(z)| |x - x_0| + \frac{\varepsilon}{2} \\ &\leq \|f'_m - f'_n\| (b-a) + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

Da  $x \in [a, b]$  beliebig war, folgt  $\|f_m - f_n\| \leq \varepsilon$  für  $m, n \geq n_0$ . Also ist  $(f_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge und daher nach Satz 1.1 gleichmäßig konvergent gegen eine Funktion  $f$ .

Wir zeigen jetzt die Differenzierbarkeit von  $f$ . Sei  $c \in [a, b]$ . Setze

$$g_n(x) := \begin{cases} \frac{f_n(x) - f_n(c)}{x - c} - f'_n(c) & \text{für } x \in [a, b] \setminus \{c\}, \\ 0 & \text{für } x = c. \end{cases}$$

Wegen der Differenzierbarkeit von  $f_n$  in  $c$  ist  $g_n$  in  $c$  stetig. Wir zeigen zunächst, dass  $(g_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge ist. Sei  $\varepsilon \in \mathbb{R}^+$  vorgegeben. Nach Voraussetzung und Satz 1.1 existiert ein  $n_0 \in \mathbb{N}$  mit

$$\|f'_m - f'_n\| < \frac{\varepsilon}{2} \quad \text{für alle } m, n \geq n_0.$$

Sei  $x \in [a, b]$ . Im Fall  $x \neq c$  ist

$$g_m(x) - g_n(x) = \frac{(f_m - f_n)(x) - (f_m - f_n)(c)}{x - c} - (f_m - f_n)'(c).$$

Nach dem Mittelwertsatz existiert ein  $z \in [a, b]$  mit

$$\frac{(f_m - f_n)(x) - (f_m - f_n)(c)}{x - c} = (f_m - f_n)'(z).$$

Für  $m, n \geq n_0$  folgt

$$\begin{aligned} |g_m(x) - g_n(x)| &\leq |f'_m(z) - f'_n(z)| + |f'_m(c) - f'_n(c)| \\ &\leq 2\|f'_m - f'_n\| \leq \varepsilon. \end{aligned}$$

Dies gilt trivialerweise auch für  $x = c$ . Da  $x \in [a, b]$  beliebig war, ist also  $\|g_m - g_n\| \leq \varepsilon$  für  $m, n \geq n_0$ . Also ist  $(g_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge.

Nach Satz 1.1 ist  $(g_n)_{n \in \mathbb{N}}$  gleichmäßig konvergent gegen eine Funktion  $g$ , die  $g(c) = 0$  erfüllt und nach Satz 1.2 in  $c$  stetig ist. Nun gilt für alle  $x \in [a, b] \setminus \{c\}$

$$\frac{f_n(x) - f_n(c)}{x - c} - f'_n(c) = g_n(x),$$

woraus durch Grenzübergang

$$\frac{f(x) - f(c)}{x - c} - \lim_{n \rightarrow \infty} f'_n(c) = g(x)$$

folgt. Wegen  $\lim_{x \rightarrow c} g(x) = g(c) = 0$  folgt

$$\lim_{x \rightarrow c} \frac{f(x) - f(c)}{x - c} = \lim_{n \rightarrow \infty} f'_n(c).$$

Also ist  $f$  in  $c$  differenzierbar und  $f'(c) = \lim_{n \rightarrow \infty} f'_n(c)$ . Damit ist die Differenzierbarkeit von  $f$  gezeigt, ferner die punktweise Konvergenz von  $(f'_n)_{n \in \mathbb{N}}$  gegen  $f'$ . Diese Konvergenz ist nach Voraussetzung gleichmäßig. ■

## 8.2 Potenzreihen

Nachdem wir früher Reihen reeller Zahlen und jetzt Funktionenfolgen betrachtet haben, ist klar, was allgemein unter einer Funktionenreihe zu verstehen ist. Sei  $(f_k)_{k \in \mathbb{N}_0}$  eine Funktionenfolge auf  $D$ . Dann verstehen wir unter

$$\sum_{k=0}^{\infty} f_k$$

die Funktionenfolge der Partialsummen  $(\sum_{k=0}^n f_k)_{n \in \mathbb{N}}$ , und wir bezeichnen  $\sum_{k=0}^{\infty} f_k$  als Funktionenreihe. Ist die Folge punktweise konvergent, so bezeichnen wir mit  $\sum_{k=0}^{\infty} f_k$  auch die Grenzfunktion. Die Schreibweise

$$\sum_{k=0}^{\infty} f_k = f \quad \text{in } D$$

bedeutet also definitionsgemäß:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n f_k(x) = f(x) \quad \text{für alle } x \in D.$$

Konvergiert die Folge  $(\sum_{k=0}^n f_k)_{n \in \mathbb{N}}$  in  $D' \subset D$  gleichmäßig (gegen  $f$ ), so sagen wir, dass die Reihe  $\sum_{k=0}^{\infty} f_k$  in  $D'$  gleichmäßig (gegen  $f$ ) konvergiert.

Die Begriffe der punktweisen oder gleichmäßigen Konvergenz von Reihen sind also nichts Neues gegenüber den Funktionenfolgen. Auch die Sätze aus Abschnitt 8.1 gelten natürlich sinngemäß für Funktionenreihen. Neu gegenüber der Folgenkonvergenz ist lediglich - wie schon bei Zahlenreihen - der Begriff der absoluten Konvergenz. Definitionsgemäß konvergiert die Reihe  $\sum f_k$  absolut, wenn die Reihe  $\sum |f_k|$  konvergiert.

In Verallgemeinerung von Satz 2.8 aus Kapitel 3 haben wir das folgende wichtige Kriterium für gleichmäßige und absolute Konvergenz.

**2.1 Satz** (Majorantenkriterium). *Sei  $f_k : D \rightarrow \mathbb{R}$ ,  $c_k \in \mathbb{R}^+$  ( $k \in \mathbb{N}_0$ ) gegeben. Gilt*

$$\|f_k\| \leq c_k \quad \text{für alle } k \in \mathbb{N}_0$$

*und ist die Reihe  $\sum_{k=0}^{\infty} c_k$  konvergent, so konvergiert die Reihe  $\sum_{k=0}^{\infty} f_k$  absolut und gleichmäßig.*

*Beweis.* Sei  $\varepsilon \in \mathbb{R}^+$ . Da  $\sum c_k$  konvergiert, existiert ein  $n_0 \in \mathbb{N}$  mit

$$\sum_{k=m}^{m+p} c_k < \varepsilon$$

für alle  $m \geq n_0, p \in \mathbb{N}_0$ . Für diese  $m, p$  gilt also

$$\left\| \sum_{k=m}^{m+p} f_k \right\| \leq \left\| \sum_{k=m}^{m+p} |f_k| \right\| \leq \sum_{k=m}^{m+p} \|f_k\| \leq \sum_{k=m}^{m+p} c_k < \varepsilon.$$

Die Folge der Partialsummen von  $\sum f_k$  und die Folge der Partialsummen von  $\sum |f_k|$  sind also Cauchyfolgen und daher nach Satz 1.1 gleichmäßig konvergent. ■

Das Vorstehende und die Ergebnisse aus Abschnitt 8.1 wollen wir jetzt anwenden auf den besonders wichtigen Spezialfall der Potenzreihen.

Unter einer *Potenzreihe zur Stelle  $a$*  versteht man eine Funktionenreihe

$$\sum_{k=0}^{\infty} f_k \quad \text{mit } f_k(x) = a_k(x-a)^k \quad \text{für } x \in \mathbb{R},$$

wo  $a \in \mathbb{R}$  und  $(a_k)_{k \in \mathbb{N}_0}$  eine Folge reeller Zahlen ist. Hier ist die folgende ungenaue, aber bequeme Sprechweise üblich. Man sagt

$$\text{„die Potenzreihe } \sum_{k=0}^{\infty} a_k(x-a)^k \text{“}$$

statt

$$\text{„die Potenzreihe } \sum_{k=0}^{\infty} f_k \text{ mit } f_k(x) = a_k(x-a)^k \text{ für } x \in \mathbb{R} \text{“}.$$

Wir fragen jetzt nach der Menge aller  $x \in \mathbb{R}$ , für die eine gegebene Potenzreihe

$$\sum_{k=0}^{\infty} a_k(x-a)^k \tag{2.2}$$

konvergiert. Auf jeden Fall konvergiert sie trivialerweise für  $x = a$  (es gibt Potenzreihen, die für kein anderes  $x$  konvergieren). Allgemein gibt das Wurzelkriterium 2.10 aus Kapitel 3 Auskunft. Nach ihm ist (für gegebenes  $x \in \mathbb{R}$ ) die Reihe (2.2) absolut konvergent, wenn

$$\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n(x-a)^n|} = |x-a| \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} < 1$$

ist, und sie ist divergent, wenn  $|x-a| \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|} > 1$  ist.

Mit der Schreibweise  $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$  vermeiden wir lästige Fallunterscheidungen. Wir definieren

$$-\infty < y < \infty \quad \text{für } y \in \mathbb{R}.$$

Außerdem definieren wir  $\frac{1}{\infty} = 0$  und  $\frac{1}{0} = \infty$ , ferner  $y + \infty = \infty$ ,  $y - \infty = -\infty$  für  $y \in \mathbb{R}$ . Es sei daran erinnert, dass wir in Abschnitt 3.2  $\limsup \sqrt[n]{|a_n|} = \infty$  definiert hatten im Fall, dass die Folge  $(\sqrt[n]{|a_n|})_{n \in \mathbb{N}}$  nicht beschränkt ist.

Für unsere gegebene Potenzreihe setzen wir jetzt

$$r := \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}}.$$

Dann gilt also: Für alle  $x \in \mathbb{R}$  mit  $|x - a| < r$  ist (2.2) absolut konvergent, für alle  $x \in \mathbb{R}$  mit  $|x - a| > r$  ist (2.2) divergent. Das Intervall  $(a - r, a + r)$  heißt daher *Konvergenzintervall* der Potenzreihe.

Wie steht es mit gleichmäßiger Konvergenz? Sei (2.2) absolut konvergent für ein  $x_0$ . Für alle  $x \in \mathbb{R}$  mit  $|x - a| \leq |x_0 - a|$  gilt dann

$$|a_k(x - a)^k| \leq |a_k||x_0 - a|^k \quad \text{für } k \in \mathbb{N}_0,$$

und die Reihe  $\sum_{k=0}^{\infty} |a_k||x_0 - a|^k$  ist konvergent. Nach dem Majorantenkriterium 2.1 folgt, dass die Reihe (2.2) in  $[a - |x_0 - a|, a + |x_0 - a|]$  absolut und gleichmäßig konvergiert. Wir fassen zusammen:

**2.3 Satz.** *Zu der Potenzreihe*

$$\sum_{k=0}^{\infty} a_k(x - a)^k \tag{2.4}$$

*gibt es ein  $r \in \overline{\mathbb{R}}$ ,  $r \geq 0$ , so dass die Reihe (2.4) für  $|x - a| < r$  absolut konvergiert und für  $|x - a| > r$  divergiert. Die Zahl  $r$  heißt Konvergenzradius der Reihe (2.4) und ist gegeben durch*

$$r = \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}}.$$

*Das offene Intervall  $(a - r, a + r)$  heißt Konvergenzintervall der Reihe (2.4). In jedem kompakten Teilintervall des Konvergenzintervalls konvergiert die Reihe (2.4) gleichmäßig.*

**Bemerkung.** Achtung! Über die Konvergenz in den Endpunkten des Konvergenzintervalls wird hier nichts ausgesagt (und läßt sich auch allgemein nichts sagen). Es gibt Potenzreihen, die in keinem, einem oder beiden Endpunkten des Konvergenzintervalls konvergieren.

Ferner beachte man, dass i.A. nicht im ganzen Konvergenzintervall gleichmäßige Konvergenz vorliegt, sondern nur in kompakten Teilintervallen.

**Beispiel.**

$$\sum_{k=1}^{\infty} \frac{1}{k^m} x^k \quad \text{mit einem } m \in \mathbb{N}_0. \quad (2.5)$$

Wegen  $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$  (vgl. Behauptung 1.11 aus Kapitel 5) ist der Konvergenzradius = 1. Ferner gilt:

- für  $m = 0$  ist (2.5) divergent in 1 und  $-1$
- für  $m = 1$  ist (2.5) konvergent in  $-1$ , divergent in 1
- für  $m = 2$  ist (2.5) konvergent in  $-1$  und 1.

Nehmen wir an, die Potenzreihe

$$\sum_{k=0}^{\infty} a_k (x - a)^k \quad (2.6)$$

habe den Konvergenzradius  $0 < r < \infty$  und konvergiere etwa auch noch im Endpunkt  $a + r$  des Konvergenzintervalls. Wir wollen zeigen, dass sie dann im Intervall  $[a, a + r]$  gleichmäßig konvergiert. Der Unterschied zur früheren Argumentation ist, dass jetzt nicht notwendig absolute Konvergenz vorliegt, daher ist das Majorantenkriterium nicht anwendbar. O.B.d.A. können wir uns auf den Fall  $a = 0$ ,  $r = 1$  beschränken, der durch eine einfache Transformation erreichbar ist.

**2.7 Satz.** *Ist  $\sum_{k=0}^{\infty} a_k$  konvergent, so ist  $\sum_{k=0}^{\infty} a_k x^k$  gleichmäßig konvergent in  $[0, 1]$ .*

*Beweis.* Betrachte die Restglieder  $b_k := \sum_{j=k+1}^{\infty} a_j$  für  $k \in \mathbb{N}_0$ . Dann ist  $(b_k)_{k \in \mathbb{N}}$  eine Nullfolge, und es ist  $b_{k-1} - b_k = a_k$ . Also ergibt sich für  $n > m$

$$\sum_{k=m+1}^n a_k x^k = b_m x^{m+1} - b_n x^n + \sum_{k=m+1}^{n-1} b_k (x^{k+1} - x^k).$$

Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $n_0 \in \mathbb{N}$  mit  $|b_k| < \varepsilon/3$  für  $k \geq n_0$ . Sei jetzt  $m \geq n_0$ ,  $p \in \mathbb{N}$ . Dann gilt für beliebiges  $x \in [0, 1]$

$$\begin{aligned} \left| \sum_{k=m+1}^{m+p} a_k x^k \right| &\leq |b_m| + |b_{m+p}| + \sum_{k=m+1}^{m+p-1} |b_k| |x^{k+1} - x^k| \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \underbrace{\sum_{k=m+1}^{m+p-1} (x^k - x^{k+1})}_{= x^{m+1} - x^{m+p}} \leq \varepsilon. \end{aligned}$$

Aus dem Cauchy-Kriterium 1.1 folgt jetzt die behauptete gleichmäßige Konvergenz. ■

Aus Satz 2.7 und Satz 1.2 folgt insbesondere:

**2.8 Folgerung.** *Ist*

$$f(x) = \sum_{k=0}^{\infty} a_k x^k \quad \text{für } x \in [0, 1],$$

so ist  $f$  in  $[0, 1]$  stetig.

Eine durch eine konvergente Potenzreihe dargestellte Funktion ist also stetig. Allgemein sagen wir, falls

$$f(x) = \sum_{k=0}^{\infty} a_k (x-a)^k \quad \text{für } x \in (a-r, a+r) \quad (2.9)$$

mit  $r > 0$  gilt, die Funktion  $f$  sei durch die obige Potenzreihe *dargestellt*, oder sie sei in eine Potenzreihe um  $a$  *entwickelt*.

Wir untersuchen jetzt die Differenzierbarkeit einer derart dargestellten Funktion. Dazu betrachten wir die durch gliedweise Differentiation von (2.9) entstehende Potenzreihe

$$\sum_{k=1}^{\infty} k a_k (x-a)^{k-1} \quad (2.10)$$

Wegen  $\limsup_{n \rightarrow \infty} \sqrt[n]{(n+1)|a_{n+1}|} = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}$  hat sie denselben Konvergenzradius wie (2.9). In jedem kompakten Teilintervall des Konvergenzintervalls  $(a-r, a+r)$  ist also (2.10) nach Satz 2.3 gleichmäßig konvergent; aus Satz 1.3 folgt daher die Differenzierbarkeit von  $f$  und die Konvergenz von (2.10) gegen  $f'$ . Auf  $f'$  kann man natürlich wieder denselben Schluß anwenden, usw. Auf diese Weise erhalten wir den folgenden Satz:

**2.11 Satz.** *Sei*

$$f(x) = \sum_{k=0}^{\infty} a_k (x-a)^k \quad \text{für } x \in (a-r, a+r)$$

mit positivem Konvergenzradius  $r$ . Dann ist  $f$  beliebig oft differenzierbar, und für  $n \in \mathbb{N}$  gilt

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1) \cdots (k-n+1) a_k (x-a)^{k-n},$$

wobei die rechts stehende Reihe denselben Konvergenzradius  $r$  hat. Speziell ist

$$a_n = \frac{f^{(n)}(a)}{n!}.$$

Die letzte Aussage zeigt insbesondere, dass zwei verschiedene Potenzreihen (zur selben Stelle  $a$ ) nicht dieselbe Funktion darstellen können. Mit anderen Worten: Aus

$$\sum_{k=0}^{\infty} a_k(x-a)^k = \sum_{k=0}^{\infty} b_k(x-a)^k$$

mit Konvergenz in einem Intervall um  $a$  folgt  $a_k = b_k$  für  $k \in \mathbb{N}_0$  (Eindeutigkeitssatz für Potenzreihen).

## Taylorreihen

Wir haben eben gezeigt: Wenn die Funktion  $f$  durch eine (in einer Umgebung von  $a$  konvergente) Potenzreihe um  $a$  dargestellt wird, so ist diese gegeben durch

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

Hier sei kurz an Abschnitt 6.3 erinnert: Ist  $f$  in einer Umgebung von  $a$  beliebig oft differenzierbar, so hatten wir die Reihe

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

als die *Taylorreihe* von  $f$  zur Stelle  $a$  bezeichnet. Wenn also  $f$  überhaupt durch eine Potenzreihe um  $a$  dargestellt werden kann, dann nur durch die Taylorreihe. Im konkreten Fall kommt es also darauf an, den Konvergenzradius der Taylorreihe zu ermitteln. Ist er positiv, so folgt aber allein daraus noch nicht, dass die Taylorreihe gegen die Funktion konvergiert, wie ein Beispiel in Abschnitt 6.3 zeigte. Um zu zeigen, dass eine gegebene Funktion im Konvergenzintervall wirklich durch ihre Taylorreihe dargestellt wird, muß man also entweder zeigen, dass das Restglied gegen Null konvergiert, oder, falls dies nicht gelingt, auf andere Weise schließen. Hierfür im folgenden einige Beispiele. Für die Abschätzung des Restgliedes haben wir die durch die Taylorformel gegebenen Darstellungen zur Verfügung: Wird

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + R_{n+1}(x)$$

gesetzt (und ist  $f$  auf einem Intervall definiert), so gilt nach Satz 3.2 aus Kapitel 6

$$R_{n+1}(x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1}$$

mit einem (von  $n$  und  $x$  abhängenden)  $c$  zwischen  $a$  und  $x$ , und nach Satz 3.3 aus Kapitel 7 gilt

$$R_{n+1}(x) = \frac{1}{n!} \int_a^x f^{(n+1)}(t)(x-t)^n dt.$$

In manchen, aber nicht in allen Fällen kann man hiermit die gewünschte Konvergenz der Taylorreihe gegen  $f$  zeigen.

Wir wollen als Beispiele die Taylorreihen für einige der im Kapitel über spezielle Funktionen betrachteten Funktionen untersuchen. Für die Funktionen  $\exp$ ,  $\sin$ ,  $\cos$  wurde (für  $a = 0$ ) bereits früher gezeigt, dass das Restglied gegen Null geht.

Wir betrachten die Logarithmus-Funktion. Als Entwicklungsstelle kommt nur ein Punkt des Definitionsbereiches  $\mathbb{R}^+$  in Frage. Wir wählen  $a = 1$ . Für  $f = \ln$  beweist man leicht durch vollständige Induktion

$$f^{(k)}(x) = (-1)^{k-1}(k-1)!x^{-k}, \quad (k \in \mathbb{N})$$

speziell  $f^{(k)}(1) = (-1)^{k-1}(k-1)!$ . Die Taylorreihe zur Stelle 1 lautet also

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

Der Konvergenzradius dieser Reihe ist offenbar 1; sie stellt also in (0,2) eine Funktion  $g$  dar. Eine Abschätzung des Restgliedes stößt auf Schwierigkeiten: Da man nichts über  $c_n$  weiß, kann man nicht ausschließen, dass  $c_n = 1/2$  für alle  $n$  ist. Wegen

$$R_{n+1}(x) = \frac{(-1)^n (x-1)^{n+1}}{n+1 c_n^{n+1}}$$

würde dann aber für  $0 < x < 1/2$  gelten:

$$\left| \frac{x-1}{c_n} \right| > 1,$$

also  $R_{n+1}(x) \not\rightarrow 0$ . Man kann aber auf andere Weise leicht zeigen, dass  $g$  in (0,2) mit der Funktion  $\ln$  übereinstimmt. Dazu differenzieren wir die Funktion  $\ln - g$ : Es ist

$$\ln' x - g'(x) = \frac{1}{x} - \sum_{k=1}^{\infty} (-1)^{k-1} (x-1)^{k-1} = \frac{1}{x} - \frac{1}{1+(x-1)} = 0.$$

Die Funktion  $\ln -g$  ist also in  $(0,2)$  konstant; da sie an der Stelle  $x = 1$  gleich Null ist, ist sie überall Null. Damit ist

$$\ln x = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (x-1)^k \quad \text{für } x \in (0, 2)$$

gezeigt. Die rechts stehende Reihe konvergiert nach dem Leibnizkriterium auch für  $x = 2$ . Da die dargestellte Funktion nach Folgerung 2.8 (passend transformiert) in 2 noch stetig ist, stimmt sie dort mit  $\ln 2$  überein. Wir können das Ergebnis auch in der Form

$$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} x^k \quad \text{für } -1 < x \leq 1$$

schreiben.

Als nächstes Beispiel betrachten wir die Funktion  $f = \arctan$ , die wir in eine Potenzreihe um 0 entwickeln wollen. Zur Berechnung der  $n$ -ten Ableitung an der Stelle 0 kann man den Ansatz

$$f^{(n)}(x) = \frac{P_n(x)}{(1+x^2)^n}$$

machen und findet für  $P_n$  eine Rekursionsformel, aus der sich herleiten läßt, dass

$$f^{(2n)}(0) = 0, \quad f^{(2n+1)}(0) = (-1)^n (2n)!$$

ist. Die Taylorreihe der Funktion  $\arctan$  lautet also

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1}.$$

Der Konvergenzradius ist offenbar 1. Sei  $g$  die in  $(-1, 1)$  dargestellte Funktion. Es gilt

$$g'(x) = \sum_{k=0}^{\infty} (-1)^k x^{2k} = \frac{1}{1+x^2} = \arctan' x,$$

ferner  $g(0) = 0 = \arctan 0$ . Also ist  $g(x) = \arctan x$ . Wir haben also

$$\begin{aligned} \arctan x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1} \quad \text{in } (-1, 1). \end{aligned}$$

Nach dem Leibniz-Kriterium konvergiert die Reihe auch für  $x = 1$  und  $x = -1$ . Nach Folgerung 2.8 ist die dargestellte Funktion dort stetig, stimmt also mit der Funktion  $\arctan$  überein.

Als letztes Beispiel wollen wir die allgemeine Potenz  $x \mapsto x^\alpha$  betrachten. Als Entwicklungsstelle wählen wir ebenfalls 1, aber es schreibt sich bequemer, die Funktion  $x \mapsto (1+x)^\alpha$  zu nehmen und um 0 zu entwickeln. Für  $f(x) = (1+x)^\alpha$  ( $x > -1$ ) berechnet man

$$f^{(k)}(x) = \alpha(\alpha-1) \cdots (\alpha-k+1)(1+x)^{\alpha-k}.$$

Zur übersichtlicheren Schreibweise wollen wir wie früher den Binomialkoeffizienten  $\binom{\alpha}{k}$  erklären durch

$$\binom{\alpha}{k} := \frac{\alpha(\alpha-1) \cdots (\alpha-k+1)}{k!}.$$

Dann ist die Taylorreihe der Funktion  $f$  zur Stelle 0 gegeben durch

$$\sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

Im Fall  $\alpha \in \mathbb{N}_0$  sind fast alle Koeffizienten 0, es liegt also ein Polynom vor. Wir setzen jetzt  $\alpha \notin \mathbb{N}_0$  voraus.

Aus dem Quotientenkriterium 2.9 aus Kapitel 3 folgt sofort, dass der Konvergenzradius gleich 1 ist. Die Reihe stellt also in  $(-1, 1)$  eine Funktion  $g$  dar. Setzen wir

$$h(x) := \frac{g(x)}{(1+x)^\alpha} \quad \text{für } -1 < x < 1,$$

so gilt

$$h'(x) = \frac{g'(x)(1+x)^\alpha - g(x)\alpha(1+x)^{\alpha-1}}{(1+x)^{2\alpha}} = \frac{g'(x)(1+x) - \alpha g(x)}{(1+x)^{\alpha+1}}.$$

Nun ist

$$\begin{aligned} (1+x)g'(x) &= (1+x) \sum_{k=1}^{\infty} \binom{\alpha}{k} kx^{k-1} \\ &= \sum_{k=0}^{\infty} \binom{\alpha}{k+1} (k+1)x^k + \sum_{k=1}^{\infty} \binom{\alpha}{k} kx^k \\ &= \sum_{k=0}^{\infty} \left\{ \binom{\alpha}{k+1} (k+1) + \binom{\alpha}{k} k \right\} x^k \\ &= \alpha \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k = \alpha g(x), \end{aligned}$$

also  $h' = 0$  und daher  $h = \text{const.}$  Wegen  $h(0) = 1$  folgt

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \text{für } -1 < x < 1.$$

Man nennt diese Reihe auch die „binomische Reihe“. Das Konvergenzverhalten an den Stellen  $\pm 1$  ist etwas schwieriger zu beurteilen:

**Bemerkung.** Sei  $\alpha \in \mathbb{R} \setminus \mathbb{N}_0$ . Die binomische Reihe  $\sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$  ist für

	in $x = 1$	in $x = -1$
$\alpha > 0$	konvergent	konvergent
$-1 < \alpha < 0$	konvergent	divergent
$\alpha \leq -1$	divergent	divergent

### 8.3 Fourierreihen

Ein wichtiger Typ von Reihenentwicklungen wird gegeben durch die Fourierreihen. Fourier war Physiker, und wir wollen zunächst kurz die physikalische Fragestellung, bei deren Behandlung Fourier (1807) erstmals die heute nach ihm benannten Reihen verwendete, heuristisch erläutern.

Fourier befaßte sich mit Problemen der Wärmeleitung in homogenen Medien. Als besonders einfaches (idealisiertes) Wärmeleitungsproblem betrachten wir einen geraden Stab der Länge  $\ell$ , den wir als „unendlich dünn“ annehmen. Seine Punkte können also durch eine Koordinate  $x$  beschrieben werden, die das Intervall  $[0, \ell]$  durchläuft. In diesem Stab herrsche an der Stelle  $x$  zur Zeit  $t \geq 0$  die Temperatur  $T(x, t)$ . Wir nehmen an, die beiden Enden des Stabes würden von außen ständig auf der Temperatur Null gehalten, und im übrigen sei der Stab vollkommen isoliert. Zur Zeit  $t = 0$  herrsche eine bekannte Temperaturverteilung, gegeben durch eine Funktion  $f : [0, \ell] \rightarrow \mathbb{R}$ . Von der Temperaturverteilung  $T(x, t)$  ist also bekannt, dass

$$T(x, 0) = f(x) \quad \text{für } x \in [0, \ell]$$

und

$$T(0, t) = T(\ell, t) = 0 \quad \text{für } t \geq 0$$

ist. Im Lauf der Zeit werden sich nun auf Grund der Wärmeleitung die Temperaturunterschiede ausgleichen. Die Frage ist, welche Temperaturverteilung zu einer Zeit  $t > 0$  herrscht, also wie man die Funktion  $T(x, t)$  aus der Anfangsverteilung  $T(x, 0)$  berechnen kann.

Physikalische Grundannahmen und Überlegungen führen zu der Folgerung, dass die Funktion  $T$  (unter, wie üblich, geeigneten Differenzierbarkeitsannahmen) der partiellen Differentialgleichung

$$\frac{\partial T}{\partial t} = \mu \frac{\partial^2 T}{\partial x^2}$$

genügen muß, wobei  $\mu > 0$  eine Materialkonstante ist. Wir stellen nun zunächst die Frage der Anfangsverteilung zurück und suchen allgemeiner eine Lösung  $\varphi : [0, \ell] \times [0, \infty) \rightarrow \mathbb{R}$  des Problems

$$\frac{\partial \varphi}{\partial t} = \mu \frac{\partial^2 \varphi}{\partial x^2}, \quad \varphi(0, t) = \varphi(\ell, t) = 0. \quad (3.1)$$

Um spezielle Lösungen zu finden, kann man untersuchen, ob es Lösungen der Form

$$\varphi(x, t) = h(t)g(x)$$

gibt („Separationsansatz“, „Trennung der Veränderlichen“). Hierzu muß wegen

$$\frac{\partial \varphi}{\partial t} = h'(t)g(x), \quad \frac{\partial^2 \varphi}{\partial x^2} = h(t)g''(x)$$

also

$$\frac{h'(t)}{h(t)} = \mu \frac{g''(x)}{g(x)}$$

sein. Da die linke Seite nicht von  $x$  und die rechte nicht von  $t$  abhängt, handelt es sich um eine Konstante  $c$ . Somit werden wir auf die Gleichungen

$$\begin{aligned} h'(t) &= ch(t), \\ \mu g''(x) &= cg(x) \end{aligned}$$

geführt. Die erste Differentialgleichung läßt sich sofort lösen:

$$h(t) = be^{ct}$$

mit einer Konstanten  $b$ . Bei der zweiten ist die Randbedingung

$$g(0) = g(\ell) = 0$$

zu berücksichtigen. Es liegt daher der Ansatz  $g(x) = \sin ax$  nahe. Die Differentialgleichung ist erfüllt, wenn  $a\ell = k\pi$  mit ganzzahligem  $k$  ist. Für die Konstanten ergibt sich also

$$a = \frac{k\pi}{\ell}, \quad c = -\frac{\mu k^2 \pi^2}{\ell^2}.$$

Insgesamt erhalten wir, dass bei beliebigem  $k \in \mathbb{N}$  durch

$$\varphi_k(x, t) = b_k e^{-\frac{\mu k^2 \pi^2}{\ell^2} t} \sin \frac{k\pi x}{\ell}$$

mit  $b_k = \text{const.}$  eine Lösung des Problems (3.1) gegeben ist.

Natürlich kann hierdurch i.A. noch nicht die gesuchte Temperaturverteilung  $T(x, t)$  gegeben sein, denn es sollte ja  $T(x, 0)$  eine vorgeschriebene Funktion  $f(x)$  sein. Wir erhalten aber

$$\varphi_k(x, 0) = b_k \sin \frac{k\pi x}{\ell},$$

also für jedes  $k$  eine sehr spezielle Funktion.

Nun beachte man aber, dass unser Problem „linear“ ist. Dies hat zur Folge, dass die Summe von Lösungen auch eine Lösung ist, also ist durch

$$\varphi(x, t) := \sum_{k=1}^n \varphi_k(x, t) = \sum_{k=1}^n b_k e^{-\frac{\mu k^2 \pi^2}{\ell^2} t} \sin \frac{k\pi x}{\ell}$$

( $n \in \mathbb{N}$ ) ebenfalls eine Lösung gegeben. Für diese Lösung ist

$$\varphi(x, 0) = \sum_{k=1}^n b_k \sin \frac{k\pi x}{\ell},$$

und im Allgemeinen wird es immer noch nicht möglich sein, die Zahl  $n$  und die Koeffizienten  $b_k$  so zu wählen, dass  $\varphi(x, 0) = f(x)$  ist. Hier setzt nun Fouriers entscheidende (und damals kühne) Überlegung ein. Er geht davon aus, dass man die weitgehend willkürliche Funktion  $f$  durch die *unendliche* Reihe

$$\sum_{k=1}^{\infty} b_k \sin \frac{k\pi x}{\ell}$$

darstellen kann, wenn man die Koeffizienten  $b_k$  geeignet bestimmt. Alsdann setze man

$$T(x, t) := \sum_{k=1}^{\infty} b_k e^{-\frac{\mu k^2 \pi^2}{\ell^2} t} \sin \frac{k\pi x}{\ell}.$$

Wenn die obere Reihe für jedes  $x \in [0, \ell]$  absolut konvergiert, dann nach dem Majorantenkriterium auch die untere. Die „Anfangsbedingung“  $T(x, 0) = f(x)$  (für  $x \in [0, \ell]$ ) und die „Randbedingung“  $T(0, t) = T(\ell, t) = 0$  (für  $t \geq 0$ ) sind dann erfüllt. Falls es erlaubt ist, die Reihe gliedweise zu differenzieren (einmal nach  $t$ , zweimal nach  $x$ ), erhält man, dass  $T$  auch der

vorgeschriebenen partiellen Differentialgleichung genügt. Das Problem der Temperaturverteilung ist damit als gelöst anzusehen.

Der springende Punkt bei dieser Argumentation ist natürlich die Frage, ob es möglich ist, eine gegebene Funktion  $f : [0, \ell] \rightarrow \mathbb{R}$  mit  $f(0) = f(\ell) = 0$  darzustellen als Reihe der Form

$$f(x) = \sum_{k=1}^{\infty} b_k \sin \frac{k\pi x}{\ell}.$$

Mit einer Variante dieser Fragestellung befassen wir uns im Folgenden. Es bedeutet dabei keine Einschränkung der Allgemeinheit,  $l = \pi$  anzunehmen. Andererseits soll die Voraussetzung  $f(0) = f(\ell) = 0$  fallengelassen werden, und wir fragen allgemeiner nach der Entwicklungsmöglichkeit in Reihen der Form

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

Es ist bequem, Funktionen zu betrachten, die auf ganz  $\mathbb{R}$  definiert sind. Wenn eine solche Funktion durch die obige Reihe dargestellt wird, ist  $f(x + 2\pi) = f(x)$  für  $x \in \mathbb{R}$ . Funktionen mit dieser Eigenschaft nennen wir  $2\pi$ -periodisch.

Wir nehmen nun an, für eine gegebene Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  existiere in der Tat eine Darstellung der Form

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \quad x \in \mathbb{R},$$

wobei die Reihe sogar gleichmäßig konvergent sei. Dann lassen sich die Koeffizienten  $a_k, b_k$  folgendermaßen berechnen. Multiplikation mit  $\cos mx$  und Integration über  $[0, 2\pi]$  (beachte, dass  $f$  wegen der gleichmäßigen Konvergenz stetig ist) ergibt

$$\begin{aligned} & \int_0^{2\pi} f(x) \cos mx \, dx \\ &= \frac{a_0}{2} \int_0^{2\pi} \cos mx \, dx + \int_0^{2\pi} \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \cos mx \, dx. \end{aligned}$$

Die rechts stehende Reihe ist ebenfalls gleichmäßig konvergent (nämlich gegen  $f(x) \cos mx$ ; beachte  $\|fg\| \leq \|f\| \|g\|$ ), darf also nach Satz 2.4 aus Kapitel 7 gliedweise integriert werden. Es ist also

$$\int_0^{2\pi} f(x) \cos mx \, dx$$

$$= \frac{a_0}{2} \int_0^{2\pi} \cos mx \, dx + \sum_{k=1}^{\infty} \left( a_k \int_0^{2\pi} \cos kx \cos mx \, dx + b_k \int_0^{2\pi} \sin kx \cos mx \, dx \right).$$

Nun findet man leicht durch partielle Integration:

$$\int_0^{2\pi} \cos kx \cos mx \, dx = \int_0^{2\pi} \sin kx \sin mx \, dx = 0 \quad \text{für } k \neq m$$

$$\int_0^{2\pi} \sin kx \cos mx \, dx = 0,$$

$$\int_0^{2\pi} \cos^2 kx \, dx = \int_0^{2\pi} \sin^2 kx \, dx = \pi \quad \text{für } k \geq 1.$$

Damit erhält man

$$a_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos mx \, dx \quad \text{für } m \in \mathbb{N}_0.$$

Ganz analog beweist man

$$b_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin mx \, dx \quad \text{für } m \in \mathbb{N}.$$

Man bezeichnet nun ganz allgemein für jede  $2\pi$ -periodische Funktion  $f$  und für  $k \in \mathbb{N}_0$  die Zahlen

$$a_k := \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx,$$

$$b_k := \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx,$$

( $k \in \mathbb{N}_0$ ) falls diese Integrale existieren, als die *Fourierkoeffizienten* von  $f$  und nennt die Funktionenreihe

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

die *Fourierreihe* von  $f$ . Dabei ist (analog wie bei Taylorreihen) nichts darüber ausgesagt, ob diese Reihe überhaupt (und für welche  $x$  bzw. in welchem Sinne) konvergiert und ob sie im Konvergenzfall gegen  $f(x)$  konvergiert. Die Beantwortung dieser Fragen erfordert eingehendere Untersuchungen.

Bevor wir hierauf eingehen, wollen wir aber die hier sehr zweckmäßige komplexe Schreibweise einführen. Es ist ja

$$e^{ix} = \cos x + i \sin x, \quad \cos x = \frac{1}{2}(e^{ix} + e^{-ix}), \quad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix}).$$

Damit ergibt sich für  $k \in \mathbb{N}_0$

$$\begin{aligned} a_k \cos kx + b_k \sin kx &= a_k \cdot \frac{1}{2}(e^{ikx} + e^{-ikx}) - b_k \cdot \frac{i}{2}(e^{ikx} - e^{-ikx}) \\ &= \frac{1}{2}(a_k - ib_k)e^{ikx} + \frac{1}{2}(a_k + ib_k)e^{-ikx} \\ &= c_k e^{ikx} + c_{-k} e^{-ikx} \end{aligned}$$

mit

$$c_k := \frac{1}{2}(a_k - ib_k), \quad c_{-k} := \frac{1}{2}(a_k + ib_k),$$

also für  $n \in \mathbb{N}$

$$\begin{aligned} \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx) \\ &= c_0 + \sum_{k=1}^n (c_k e^{ikx} + c_{-k} e^{-ikx}) \\ &= \sum_{k=-n}^n c_k e^{ikx} \end{aligned}$$

Die Folge  $(\sum_{k=-n}^n c_k e^{ikx})_{n \in \mathbb{N}}$  kürzt man ab mit

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx}.$$

Dies ist also ebenfalls die Fourierreihe von  $f$ , nur in komplexer Schreibweise.

Es ist weiter zweckmäßig, auch komplexwertige Funktionen  $f : \mathbb{R} \rightarrow \mathbb{C}$  zuzulassen. Jede solche Funktion läßt sich eindeutig darstellen in der Form  $f = u + iv$  mit reellen Funktionen  $u, v$ . Man definiert dann (falls  $u$  und  $v$ , eingeschränkt auf  $[a, b]$ , Regelfunktionen sind)

$$\int_a^b f(x) dx := \int_a^b u(x) dx + i \int_a^b v(x) dx.$$

Mit dieser Konvention ist für  $k \geq 0$

$$\begin{aligned} c_k &= \frac{1}{2}(a_k - ib_k) \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(x)(\cos kx - i \sin kx) dx \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{-ikx} dx \end{aligned}$$

und

$$c_{-k} = \frac{1}{2}(a_k + ib_k) = \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{ikx} dx,$$

somit

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x)e^{-ikx} dx \quad \text{für } k \in \mathbb{Z}.$$

Wir wollen hier auch komplexwertige Funktionen  $f$  zulassen, erklären dann die Fourierkoeffizienten  $c_k$  von  $f$  durch diese Gleichung (falls die Integrale existieren) und nennen die Reihe

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx}$$

die *Fourierreihe* von  $f$ .

Die Konvergenztheorie der Fourierreihen wird dann besonders einfach und übersichtlich, wenn man nicht nach der (i.a. selbst bei stetigen reellen Funktionen nicht vorliegenden) punktwisen Konvergenz fragt, sondern nach Konvergenz in einem schwächeren Sinne. Mit dieser auch für Anwendungen wichtigen Konvergenz wollen wir uns hier befassen.

### Konvergenz im quadratischen Mittel

Wir betrachten die Menge  $V$  der Funktionen  $f : \mathbb{R} \rightarrow \mathbb{C}$  mit  $f(x+2\pi) = f(x)$  für  $x \in \mathbb{R}$  und der Eigenschaft, dass Real- und Imaginärteil von  $f$ , eingeschränkt auf  $[0, 2\pi]$ , Regelfunktionen sind. Offenbar ist  $V$  (mit den üblichen Verknüpfungen) ein Vektorraum über  $\mathbb{C}$ .

**Definition.** Für  $f, g \in V$  sei

$$\langle f, g \rangle := \frac{1}{2\pi} \int_0^{2\pi} f(x) \overline{g(x)} dx.$$

Die komplexe Zahl  $\langle f, g \rangle$  heißt das *Skalarprodukt* von  $f$  und  $g$ .

**3.2 Satz.** Die Abbildung  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$  hat die folgenden Eigenschaften (für  $f, g, h \in V$  und  $\lambda \in \mathbb{C}$ ):

$$(1) \langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$$

$$(2) \langle f, g + h \rangle = \langle f, g \rangle + \langle f, h \rangle$$

$$(3) \langle \lambda f, g \rangle = \lambda \langle f, g \rangle$$

$$(4) \langle f, \lambda g \rangle = \overline{\lambda} \langle f, g \rangle$$

$$(5) \langle g, f \rangle = \overline{\langle f, g \rangle}$$

$$(6) \langle f, f \rangle \geq 0 \text{ (d.h. } \langle f, f \rangle \text{ ist reellwertig und nichtnegativ)}$$

$$(7) |\langle f, g \rangle|^2 \leq \langle f, f \rangle \langle g, g \rangle$$

$$(8) \sqrt{\langle f + g, f + g \rangle} \leq \sqrt{\langle f, f \rangle} + \sqrt{\langle g, g \rangle}.$$

*Beweis.* (1) – (6) folgen unmittelbar aus der Definition; bei (6) beachte man  $f\overline{f} = |f|^2 \geq 0$ . Wegen (6) sind die Wurzeln in (8) im Reellen erklärt.

(7) Nach (6) gilt für  $f, g \in V$  und  $\lambda \in \mathbb{C}$

$$\langle f + \lambda g, f + \lambda g \rangle \geq 0,$$

also unter Verwendung von (1) – (5)

$$\langle f, f \rangle + \overline{\lambda} \langle f, g \rangle + \lambda \overline{\langle f, g \rangle} + \lambda \overline{\lambda} \langle g, g \rangle \geq 0.$$

Ist  $\langle g, g \rangle \neq 0$ , so kann man

$$\lambda = -\frac{\langle f, g \rangle}{\langle g, g \rangle}$$

einsetzen und erhält die behauptete Ungleichung. Ist  $\langle g, g \rangle = 0$ , so setze man  $\lambda = -n \langle f, g \rangle$ ; es folgt  $|\langle f, g \rangle|^2 \leq 1/2n \langle f, f \rangle$ , also (da  $n \in \mathbb{N}$  beliebig ist)  $|\langle f, g \rangle| \leq 0 = \langle f, f \rangle \langle g, g \rangle$ .

(8) Aus (7) folgt

$$\operatorname{Re}\langle f, g \rangle \leq |\langle f, g \rangle| \leq \sqrt{\langle f, f \rangle \langle g, g \rangle},$$

also

$$\begin{aligned} \langle f + g, f + g \rangle &= \langle f, f \rangle + \langle f, g \rangle + \overline{\langle f, g \rangle} + \langle g, g \rangle \\ &= \langle f, f \rangle + \langle g, g \rangle + 2\operatorname{Re}\langle f, g \rangle \\ &\leq \langle f, f \rangle + \langle g, g \rangle + 2\sqrt{\langle f, f \rangle \langle g, g \rangle} \\ &= \left( \sqrt{\langle f, f \rangle} + \sqrt{\langle g, g \rangle} \right)^2, \end{aligned}$$

woraus (8) folgt. ■

**Definition.** Für  $f \in V$  ist

$$\|f\|_2 := \sqrt{\langle f, f \rangle}$$

die  $L_2$ -Norm oder *Hilbert-Norm* von  $f$ .

Die Ungleichungen (7) und (8) aus Satz 3.2 schreiben sich damit in der Form

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$$

(*Cauchy-Schwarzsche Ungleichung*) und

$$\|f + g\|_2 \leq \|f\|_2 + \|g\|_2;$$

letzteres ist die Dreiecksungleichung für die  $L_2$ -Norm. Für die  $L_2$ -Norm gilt auch  $\|\lambda f\|_2 = |\lambda| \|f\|_2$  und  $\|f\|_2 = 0$  für  $f = 0$ . Allerdings folgt aus  $\|f\|_2 = 0$  nicht, dass  $f$  die Nullfunktion ist.

Wir erinnern uns daran, dass wir schon früher Konvergenz im Sinne einer Norm erklärt hatten, und definieren:

**Definition.** Seien  $f, f_n \in V$  ( $n \in \mathbb{N}$ ). Die Folge  $(f_n)_{n \in \mathbb{N}}$  heißt *konvergent im quadratischen Mittel* (oder konvergent in der  $L_2$ -Norm) gegen  $f$ , wenn

$$\lim_{n \rightarrow \infty} \|f - f_n\|_2 = 0$$

ist.

Die Bezeichnung „im quadratischen Mittel“ ist naheliegend wegen

$$\|f - f_n\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x) - f_n(x)|^2 dx;$$

der rechts stehende Ausdruck läßt sich als Mittelwert der quadratischen Abweichung von  $f$  und  $f_n$  ansehen. Man beachte, dass eine im quadratischen Mittel konvergierende Funktionenfolge nicht punktweise konvergent zu sein braucht.

Wir betrachten jetzt speziell die durch

$$e_k(x) := e^{ikx}$$

definierten Elemente  $e_k \in V$  ( $k \in \mathbb{Z}$ ) und berechnen  $\langle e_k, e_m \rangle$ . Nach Definition ist

$$\begin{aligned} \langle e_k, e_m \rangle &= \frac{1}{2\pi} \int_0^{2\pi} e^{ikx} \overline{e^{imx}} dx = \frac{1}{2\pi} \int_0^{2\pi} e^{i(k-m)x} dx \\ &= \frac{1}{2\pi} \left( \int_0^{2\pi} \cos(k-m)x dx + i \int_0^{2\pi} \sin(k-m)x dx \right) \\ &= \begin{cases} 1 & \text{für } k = m, \\ 0 & \text{für } k \neq m. \end{cases} \end{aligned}$$

Allgemein nennt man zwei Elemente  $f, g \in V$  mit  $\langle f, g \rangle = 0$  *orthogonal*, und eine Folge  $(b_k)_{k \in \mathbb{Z}}$  in  $V$  heißt *Orthonormalsystem*, wenn

$$\langle b_k, b_m \rangle = \begin{cases} 1 & \text{für } k = m \\ 0 & \text{für } k \neq m \end{cases}$$

ist.

Im folgenden sei jetzt zunächst ein beliebiges Orthonormalsystem  $(b_k)_{k \in \mathbb{Z}}$  gegeben. Ist  $f \in V$ , so nennt man die durch

$$c_k := \langle f, b_k \rangle$$

definierten komplexen Zahlen die *Fourierkoeffizienten* von  $f$  bezüglich des gegebenen Orthonormalsystems. Die früher definierten speziellen Fourierkoeffizienten sind also genau diejenigen bezüglich des speziellen Orthonormalsystems  $(e_k)_{k \in \mathbb{Z}}$ . Die Reihe

$$\sum_{k \in \mathbb{Z}} c_k b_k$$

nennt man die *Fourierreihe* von  $f$  bezüglich des Orthonormalsystems  $(b_k)_{k \in \mathbb{Z}}$ .

Sei jetzt  $f \in V$  gegeben, und seien  $\alpha_{-n}, \dots, \alpha_n$  beliebige komplexe Zahlen. Wir wollen

$$\left\| f - \sum_{k=-n}^n \alpha_k b_k \right\|_2^2$$

berechnen und aus dem Ergebnis wichtige Folgerungen ziehen. Es ist (Summation jeweils von  $-n$  bis  $n$ )

$$\begin{aligned} \left\| f - \sum \alpha_k b_k \right\|_2^2 &= \left\langle f - \sum \alpha_k b_k, f - \sum \alpha_j b_j \right\rangle \\ &= \langle f, f \rangle - \sum \alpha_k \underbrace{\langle b_k, f \rangle}_{c_k} - \sum \bar{\alpha}_j \underbrace{\langle f, b_j \rangle}_{c_j} + \left\langle \sum \alpha_k b_k, \sum \alpha_j b_j \right\rangle \end{aligned}$$

und

$$\left\langle \sum \alpha_k b_k, \sum \alpha_j b_j \right\rangle = \sum_{k,j} \alpha_k \bar{\alpha}_j \langle b_k, b_j \rangle = \sum_k \alpha_k \bar{\alpha}_k.$$

Ferner ist

$$\begin{aligned} \sum |c_k - \alpha_k|^2 &= \sum (c_k - \alpha_k)(\bar{c}_k - \bar{\alpha}_k) \\ &= \sum c_k \bar{c}_k - \sum \alpha_k \bar{c}_k - \sum \bar{\alpha}_k c_k + \sum \alpha_k \bar{\alpha}_k. \end{aligned}$$

Damit ergibt sich

$$\left\| f - \sum \alpha_k b_k \right\|_2^2 = \|f\|_2^2 - \sum |c_k|^2 + \sum |c_k - \alpha_k|^2.$$

Hieran lesen wir folgendes ab:

- (1)  $\|f - \sum_{k=-n}^n \alpha_k b_k\|_2^2$  wird genau dann minimal, wenn  $\alpha_k = c_k$  für  $k = -n, \dots, n$  ist. Dies ist also eine Charakterisierung der Fourierkoeffizienten durch eine Extremaleigenschaft:  $f$  wird durch eine Linearkombination der Vektoren  $b_{-n}, \dots, b_n$  im Sinne der  $L_2$ -Norm genau dann am besten approximiert, wenn man als Koeffizienten die Fourierkoeffizienten wählt.

- (2) Wählen wir jetzt  $\alpha_k = c_k$ , so lautet die obige Gleichung

$$\|f\|_2^2 - \sum_{k=-n}^n |c_k|^2 = \left\| f - \sum_{k=-n}^n c_k b_k \right\|_2^2,$$

und dies ist  $\geq 0$ . Es folgt die Konvergenz der Reihe  $\sum_{k=-\infty}^{\infty} |c_k|^2$  und die Ungleichung

$$\sum_{k=-\infty}^{\infty} |c_k|^2 \leq \|f\|_2^2.$$

Genau dann gilt hier das Gleichheitszeichen, wenn

$$\lim_{n \rightarrow \infty} \left\| f - \sum_{k=-n}^n c_k b_k \right\|_2 = 0$$

ist.

Wir fassen die erhaltenen Ergebnisse zusammen und ergänzen sie durch einige neue Bezeichnungen.

**3.3 Satz.** Sei  $(b_k)_{k \in \mathbb{Z}}$  ein Orthonormalsystem in  $V$ , sei  $f \in V$ , und seien  $c_k = \langle f, b_k \rangle$ ,  $k \in \mathbb{Z}$ , die Fourierkoeffizienten von  $f$  bezüglich des Orthonormalsystems  $(b_k)_{k \in \mathbb{Z}}$ . Dann gilt

$$\sum_{k=-\infty}^{\infty} |c_k|^2 \leq \|f\|_2^2 \quad (\text{Besselsche Ungleichung}).$$

Die Fourierreihe von  $f$  konvergiert genau dann im quadratischen Mittel gegen  $f$ , d.h. es gilt

$$\lim_{n \rightarrow \infty} \left\| f - \sum_{k=-n}^n c_k b_k \right\|_2 = 0,$$

wenn

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = \|f\|_2^2 \quad (\text{Parsevalsche Vollständigkeitsrelation}) \quad (3.4)$$

gilt. Das Orthonormalsystem  $(b_k)_{k \in \mathbb{Z}}$  heißt vollständig, wenn (3.4) für jedes  $f \in V$  gilt.

Nach diesen allgemeinen Betrachtungen kehren wir zum speziellen Orthonormalsystem  $(e_k)_{k \in \mathbb{Z}}$  zurück und zeigen:

**3.5 Satz.** Das Orthonormalsystem  $(e_k)_{k \in \mathbb{Z}}$  ist vollständig.

Für jede Funktion  $f \in V$  konvergiert also die Fourierreihe

$$\sum_{k=-\infty}^{\infty} c_k e^{ikx}$$

mit

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx$$

im quadratischen Mittel gegen  $f(x)$ , d.h. es ist

$$\lim_{n \rightarrow \infty} \int_0^{2\pi} \left| f(x) - \sum_{k=-n}^n c_k e^{ikx} \right|^2 dx = 0$$

oder, anders geschrieben,

$$\lim_{n \rightarrow \infty} \|f - S_n[f]\|_2 = 0,$$

wenn wir (wie auch im folgenden) mit

$$S_n[f] := \sum_{k=-n}^n c_k e^{ikx}$$

die  $n$ -te Partialsumme der Fourierreihe von  $f$  bezeichnen. Nach Satz 3.3 ist die Konvergenz der Fourierreihe gleichwertig mit (3.4), oder anders geschrieben mit

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 dx. \quad (3.6)$$

Zum Beweis von (3.6) nehmen wir zunächst  $f$  von einer sehr speziellen Gestalt an. Sei  $0 \leq a \leq b \leq 2\pi$  und

$$f(x) = \chi_{[a,b]}(x) := \begin{cases} 1 & \text{für } a \leq x \leq b, \\ 0 & \text{für } 0 \leq x < a \text{ und } b < x \leq 2\pi; \end{cases}$$

im übrigen sei  $f(x+2\pi) = f(x)$ . Für die komplexen Fourierkoeffizienten von  $f$  erhalten wir

$$c_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx = \frac{b-a}{2\pi}$$

und für  $k \neq 0$

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx = \frac{1}{2\pi} \left[ \frac{-1}{ik} e^{-ikx} \right]_a^b = \frac{i}{2\pi k} (e^{-ikb} - e^{-ika}),$$

also

$$|c_k|^2 = \frac{1 - \cos k(b-a)}{2\pi^2 k^2}.$$

Somit ist

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{(b-a)^2}{4\pi^2} + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} - \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{\cos k(b-a)}{k^2}.$$

Zur Berechnung der letzten Summe folgt eine längere Zwischenrechnung.

**Behauptung.**

$$\sum_{k=1}^{\infty} \frac{\sin kx}{k} = \frac{\pi - x}{2} \quad \text{für } 0 < x < 2\pi.$$

*Beweis.*

$$\begin{aligned} 1 + 2 \sum_{k=1}^n \cos kt &= \sum_{k=-n}^n e^{ikt} = e^{-int} \sum_{k=0}^{2n} e^{ikt} \\ &= e^{-int} \frac{e^{i(2n+1)t} - 1}{e^{it} - 1} = \frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{e^{\frac{i}{2}t} - e^{-\frac{i}{2}t}} \\ &= \frac{\sin(n + \frac{1}{2})t}{\sin \frac{1}{2}t}, \end{aligned}$$

also

$$\sum_{k=1}^n \cos kt = \frac{\sin(n + \frac{1}{2})t}{2 \sin \frac{1}{2}t} - \frac{1}{2}.$$

Integration von  $\pi$  bis  $x$  ergibt

$$\sum_{k=1}^n \frac{\sin kx}{k} = \int_{\pi}^x \frac{\sin(n + \frac{1}{2})t}{2 \sin \frac{1}{2}t} dt - \frac{x - \pi}{2}.$$

Für  $n \rightarrow \infty$  konvergiert das Integral gegen 0, da die reellen Fourierkoeffizienten einer Regelfunktion wegen der Besselschen Ungleichung gegen 0 konvergieren. ■

**Behauptung.** Für  $\delta \in (0, \pi)$  ist die Reihe  $\sum_{k=1}^{\infty} \sin kx/k$  gleichmäßig konvergent in  $[\delta, 2\pi - \delta]$ .

*Beweis.* Setze

$$s_n(x) := \sum_{k=1}^n \sin kx = \operatorname{Im} \left( \sum_{k=1}^n e^{ikx} \right).$$

Für  $\delta \leq x \leq 2\pi - \delta$  gilt

$$\begin{aligned} |s_n(x)| &\leq \left| \sum_{k=1}^n e^{ikx} \right| = |e^{ix}| \left| \frac{e^{inx} - 1}{e^{ix} - 1} \right| \\ &\leq \frac{2}{\left| e^{\frac{ix}{2}} - e^{-\frac{ix}{2}} \right|} = \frac{1}{\sin \frac{x}{2}} \leq \frac{1}{\sin \frac{\delta}{2}}. \end{aligned}$$

Für  $m > n > 0$  folgt

$$\begin{aligned} \left| \sum_{k=n}^m \frac{\sin kx}{k} \right| &= \left| \sum_{k=n}^m \frac{s_k(x) - s_{k-1}(x)}{k} \right| \\ &= \left| \sum_{k=n}^m s_k(x) \left( \frac{1}{k} - \frac{1}{k+1} \right) + \frac{s_m(x)}{m+1} - \frac{s_{n-1}(x)}{n} \right| \\ &\leq \frac{1}{\sin \frac{\delta}{2}} \left( \frac{1}{n} - \frac{1}{m+1} + \frac{1}{m+1} + \frac{1}{n} \right) = \frac{2}{n \sin \frac{\delta}{2}}. \end{aligned}$$

Nach dem Cauchy-Kriterium 1.1 folgt die gleichmäßige Konvergenz. ■

**Behauptung.**

$$\sum_{k=1}^{\infty} \frac{\cos kx}{k^2} = \left( \frac{x - \pi}{2} \right)^2 - \frac{\pi^2}{12} \quad \text{für } x \in [0, 2\pi].$$

*Beweis.* Setze  $F(x) := \sum_{k=1}^{\infty} \cos kx/k^2$ . Diese Reihe ist nach dem Majorantenkriterium gleichmäßig konvergent. Für beliebiges  $\delta > 0$  ist die Reihe  $-\sum_{k=1}^{\infty} \sin kx/k$  auf  $[\delta, 2\pi - \delta]$  nach der 2. Behauptung gleichmäßig konvergent. Nach Satz 1.3 und der 1. Behauptung folgt

$$F'(x) = \frac{x - \pi}{2}, \quad \text{also } F(x) = \left( \frac{x - \pi}{2} \right)^2 + c.$$

Nun ist

$$\int_0^{2\pi} F(x) dx = \frac{\pi^3}{6} + 2\pi c,$$

und dies ist andererseits nach Satz 2.4 aus Kapitel 7

$$= \int_0^{2\pi} \sum_{k=1}^{\infty} \frac{\cos kx}{k^2} dx = \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^{2\pi} \cos kx dx = 0.$$

Also ist  $c = -\frac{\pi^2}{12}$ , womit die 3. Behauptung bewiesen ist. ■

Jetzt können wir unsere oben begonnene Rechnung fortsetzen und erhalten

$$\begin{aligned} \sum_{k=-\infty}^{\infty} |c_k|^2 &= \frac{(b-a)^2}{4\pi^2} + \frac{1}{\pi^2} \cdot \frac{\pi^2}{6} - \frac{1}{\pi^2} \left( \frac{b-a-\pi}{2} \right)^2 + \frac{1}{\pi^2} \cdot \frac{\pi^2}{12} \\ &= \frac{b-a}{2\pi} = \|f\|_2^2. \end{aligned}$$

Die Gleichung (3.6) ist für diese spezielle Funktion  $f$  also bewiesen.

Nun sei  $f : [0, 2\pi] \rightarrow \mathbb{R}$  eine Treppenfunktion. Dann gibt es endlich viele Funktionen  $f_1, \dots, f_r$  der oben betrachteten Art und reelle Zahlen  $\alpha_1, \dots, \alpha_r$  mit  $f = \sum_{j=1}^r \alpha_j f_j$ . Allgemein sei mit  $S_n[g]$  die  $n$ -te Partialsumme der Fourierreihe von  $g$  bezeichnet, also

$$S_n[g](x) := \sum_{k=-n}^n c_k[g] e^{ikx}, \quad c_k[g] := \frac{1}{2\pi} \int_0^{2\pi} g(x) e^{-ikx} dx.$$

Dann gilt

$$\begin{aligned} \|f - S_n[f]\|_2 &= \left\| \sum_{j=1}^r \alpha_j f_j - \sum_{j=1}^r \alpha_j S_n[f_j] \right\|_2 \\ &\leq \sum_{j=1}^r |\alpha_j| \|f_j - S_n[f_j]\|_2 \rightarrow 0 \quad \text{für } n \rightarrow \infty \end{aligned}$$

nach dem bereits Bewiesenen und Satz 3.3.

Schließlich sei  $f : [0, 2\pi] \rightarrow \mathbb{R}$  eine Regelfunktion. Zu  $\varepsilon \in \mathbb{R}^+$  existiert eine Treppenfunktion  $t$  auf  $[0, 2\pi]$  mit  $|f(x) - t(x)| \leq \varepsilon/2$  für  $x \in [0, 2\pi]$ . Für  $g := f - t$  folgt

$$\|g - S_n[g]\|_2^2 = \|g\|_2^2 - \sum_{k=-n}^n |c_k[g]|^2 \leq \|g\|_2^2 \leq \left(\frac{\varepsilon}{2}\right)^2.$$

Nach dem bereits Bewiesenen existiert ein  $n_0 \in \mathbb{N}$  mit  $\|t - S_n[t]\|_2 < \varepsilon/2$  für  $n \geq n_0$ . Für  $n \geq n_0$  gilt also

$$\|f - S_n[f]\|_2 \leq \|g - S_n[g]\|_2 + \|t - S_n[t]\|_2 < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Somit ist

$$\lim_{n \rightarrow \infty} \|f - S_n[f]\|_2 = 0.$$

Die Ausdehnung auf  $f \in V$  ist jetzt trivial und wir haben (3.6) für alle  $f \in V$  bewiesen.

Gegenüber der Konvergenz im quadratischen Mittel ist die Beurteilung der punktweisen Konvergenz einer Fourierreihe wesentlich schwieriger. Selbst für stetiges  $f$  braucht die Folge  $(S_n[f])_{n \in \mathbb{N}}$  der Partialsummen der Fourierreihe von  $f$  nicht punktweise gegen  $f$  zu konvergieren. Erstaunlicherweise gilt aber, dass die arithmetischen Mittel dieser Partialsummen sogar gleichmäßig gegen  $f$  konvergieren, wie der folgende Satz zeigt.

**3.7 Satz** (Fejér). *Ist  $f : \mathbb{R} \rightarrow \mathbb{R}$   $2\pi$ -periodisch und stetig und wird*

$$\sigma_n[f] := \frac{1}{n}(S_0[f] + \cdots + S_{n-1}[f])$$

*gesetzt, so konvergiert die Folge  $(\sigma_n[f])_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$ .*

*Beweis.* Es ist

$$\begin{aligned} S_n[f](x) &= \sum_{k=-n}^n \left( \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-ikt} dt \right) e^{ikx} \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) \sum_{k=-n}^n e^{ik(x-t)} dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(t) D_n(x-t) dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_n(t) dt \quad [\text{Substitution und Periodizität}] \end{aligned}$$

mit

$$D_n(x) := \sum_{k=-n}^n e^{ikx}.$$

Wir bemerken

$$\frac{1}{2\pi} \int_0^{2\pi} D_n(x) dx = 1$$

und

$$(e^{ix} - 1)D_n(x) = e^{i(n+1)x} - e^{-inx}.$$

Multiplikation mit  $e^{-\frac{ix}{2}}$  ergibt

$$D_n(x) = \frac{\sin(n + \frac{1}{2})x}{\sin \frac{x}{2}}.$$

Setze

$$F_n := \frac{1}{n}(D_0 + \dots + D_{n-1}),$$

dann ist

$$\frac{1}{2\pi} \int_0^{2\pi} F_n(x) dx = 1$$

und

$$F_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} \frac{\sin(k + \frac{1}{2})x}{\sin \frac{x}{2}}.$$

Aus

$$\begin{aligned} \sum_{k=0}^{n-1} e^{i(k+\frac{1}{2})x} &= e^{\frac{ix}{2}} \sum_{k=0}^{n-1} e^{ikx} = e^{\frac{ix}{2}} \frac{e^{inx} - 1}{e^{ix} - 1} \\ &= \frac{e^{inx} - 1}{e^{\frac{ix}{2}} - e^{-\frac{ix}{2}}} = \frac{\cos nx - 1 + i \sin nx}{2i \sin \frac{x}{2}} \end{aligned}$$

folgt

$$\sum_{k=0}^{n-1} \sin(k + \frac{1}{2})x = \frac{1 - \cos nx}{2 \sin \frac{x}{2}} = \frac{\sin^2 \frac{nx}{2}}{\sin \frac{x}{2}},$$

also

$$F_n(x) = \frac{1}{n} \left( \frac{\sin \frac{nx}{2}}{\sin \frac{x}{2}} \right)^2,$$

was insbesondere  $F_n(x) \geq 0$  impliziert. Nun ist

$$\begin{aligned} \sigma_n[f](x) &= \sum_{k=0}^{n-1} \frac{1}{n} S_k[f](x) = \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_k(t) dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) F_n(t) dt, \end{aligned}$$

also

$$\begin{aligned} & |f(x) - \sigma_n[f](x)| \\ &= \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} F_n(t) dt \cdot f(x) - \frac{1}{2\pi} \int_0^{2\pi} f(x-t) F_n(t) dt \right| \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x-t) - f(x)| F_n(t) dt. \end{aligned}$$

Zu gegebenem  $\varepsilon > 0$  existiert ein  $\delta > 0$  mit  $\delta < \pi$  und

$$|f(x-t) - f(x)| \leq \varepsilon \quad \text{für } |t| < \delta,$$

da  $f$  nach Satz 3.5 aus Kapitel 4 gleichmäßig stetig ist; ferner ist

$$|f(x-t) - f(x)| \leq M \quad \text{für alle } x \text{ und } t$$

mit passendem  $M$ . Es folgt

$$\begin{aligned} & \int_{-\pi}^{\pi} |f(x-t) - f(x)| F_n(t) dt \\ &= \int_{|t| \leq \delta} |f(x-t) - f(x)| F_n(t) dt + \int_{|t| \geq \delta} |f(x-t) - f(x)| F_n(t) dt \\ &\leq \varepsilon 2\pi + M \int_{|t| \geq \delta} \frac{1}{n} \left[ \frac{\sin \frac{nt}{2}}{\sin \frac{t}{2}} \right]^2 dt \leq \varepsilon + \frac{M}{n} \frac{2\pi}{(\sin \frac{\delta}{2})^2}. \end{aligned}$$

Für  $n \geq n_0$  mit passendem  $n_0$  ist der letzte Summand kleiner als  $\varepsilon 2\pi$ , also ist  $|f(x) - \sigma_n[f](x)| < 2\varepsilon$  für  $n \geq n_0$ . ■

Wir wollen an den Satz von Fejér noch zwei Bemerkungen anfügen. Zunächst soll noch einmal kurz das Wesen der dort verwendeten Summierung herausgestellt werden. Für eine Reihe

$$\sum_{k=0}^{\infty} a_k$$

bildet man die Mittel der Partialsummen, also

$$\begin{aligned}
\sigma_n &= \frac{s_0 + s_1 + \cdots + s_n}{n+1} \\
&= \frac{a_0 + (a_0 + a_1) + (a_0 + a_1 + a_2) + \cdots + (a_0 + a_1 + \cdots + a_n)}{n+1} \\
&= a_0 + \frac{n}{n+1}a_1 + \frac{n-1}{n+1}a_2 + \cdots + \frac{1}{n+1}a_n \\
&= \sum_{k=0}^n \left(1 - \frac{k}{n+1}\right) a_k.
\end{aligned}$$

Gegenüber der Bildung der Partialsumme  $\sum_{k=0}^n a_k$  werden hierbei also die Summanden  $a_k$  mit Gewichten versehen, die von  $n$  abhängen und mit  $k$  abnehmen.

Sodann wollen wir darauf hinweisen, dass man aus dem Satz von Fejér leicht einen wichtigen Satz über die gleichmäßige Approximierbarkeit stetiger reeller Funktionen durch Polynome herleiten kann.

**3.8 Satz** (Approximationssatz von Weierstraß). *Sei  $g : [a, b] \rightarrow \mathbb{R}$  stetig und  $\varepsilon > 0$ . Dann existiert ein Polynom  $P$  mit*

$$|g(x) - P(x)| \leq \varepsilon \quad \text{für alle } x \in [a, b].$$

*Beweis.* O.B.d.A. sei  $a = -1$  und  $b = 1$ . Setze

$$f(\varphi) := g(\cos \varphi) \quad \text{für } -\pi \leq \varphi \leq \pi$$

und ergänze  $f$  zu einer  $2\pi$ -periodischen Funktion. Wegen  $f(-\pi) = f(\pi)$  ist  $f$  stetig. Nach dem Satz von Fejér existiert ein  $n \in \mathbb{N}$  mit

$$|f(\varphi) - \sigma_n[f](\varphi)| \leq \varepsilon \quad \text{für alle } \varphi \in \mathbb{R}.$$

Die reelle Fourierreihe von  $f$  ist von der Form

$$f(\varphi) \sim \frac{a_0}{2} + \sum_{k=0}^{\infty} a_k \cos k\varphi,$$

denn wegen  $f(\varphi) = f(-\varphi)$  verschwinden alle  $b_k$ . Da  $\cos k\varphi$  sich als Polynom in  $\cos \varphi$  ausdrücken läßt, ist

$$\sigma_n[f](\varphi) = P(\cos \varphi)$$

mit einem Polynom  $P$ . Es ist also

$$|g(\cos \varphi) - P(\cos \varphi)| \leq \varepsilon \quad \text{für alle } \varphi$$

und daher

$$|g(x) - P(x)| \leq \varepsilon \quad \text{für alle } x \in [-1, 1]. \quad \blacksquare$$



## Analysis II



## 9 Metrische Räume

In diesem Kapitel sollen einige Aussagen über Konvergenz, Stetigkeit und damit zusammenhängende Begriffsbildungen in einem allgemeineren Rahmen betrachtet werden. Diese Verallgemeinerung geschieht nicht um ihrer selbst willen, sondern ist zweckmäßig, damit man in verschiedenartigen konkreten Fällen nicht gleichartige Sachverhalte stets neu beweisen muß.

Zur Motivation soll zunächst an einige Definitionen aus Analysis I erinnert werden. In Analysis I, Abschnitt 3.1, wurde definiert:

**Definition.** Sei  $(a_n)_{n \in \mathbb{N}}$  eine Folge in  $\mathbb{R}$ , sei  $a \in \mathbb{R}$ . Die Folge  $(a_n)_{n \in \mathbb{N}}$  heißt *konvergent gegen  $a$* , wenn gilt:

$$\forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \geq n_0 : |a_n - a| < \varepsilon.$$

In Abschnitt 5.2 haben wir in völlig gleichlautender Weise die Konvergenz einer Folge komplexer Zahlen erklärt. Dabei bezeichnete  $|z|$  den Betrag einer komplexen Zahl  $z$ .

In Abschnitt 8.1 tauchte eine ähnliche Definition auf. Sei  $D \subseteq \mathbb{R}$ . Für jede beschränkte Funktion  $f : D \rightarrow \mathbb{R}$  ist durch

$$\|f\| := \sup_{x \in D} |f(x)|$$

die *Supremumsnorm* von  $f$  erklärt. Seien jetzt  $f, f_n$  ( $n \in \mathbb{N}$ ) beschränkte reelle Funktionen auf  $D$ .

**Definition.** Die Folge  $(f_n)_{n \in \mathbb{N}}$  *konvergiert gleichmäßig* (oder *im Sinne der Supremumsnorm*) gegen  $f$ , wenn gilt:

$$\forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \|f_n - f\| < \varepsilon.$$

Die Definition sieht formal wieder genau gleich aus wie oben, obwohl die Folgenglieder keine Zahlen mehr sind, sondern Funktionen, und die Norm

$\|\cdot\|$  daher eine andere Bedeutung hat. Trotzdem erinnern wir uns, dass einige einfache Aussagen über diese Konvergenz sich völlig analog wie im Bereich der reellen Zahlen beweisen ließen. Wie eine genauere Analyse dieser Beweise zeigt, werden dabei nur wenige formale Eigenschaften des „Abstandes“  $\|f_n - f\|$  benutzt. Man kann daher eine abstrakte Konvergenztheorie entwickeln, wenn man nur auf einer Grundmenge einen Abstandsbegriff mit entsprechenden formalen Eigenschaften zur Verfügung hat. Einen solchen Abstand nennt man eine „Metrik“ und eine Menge mit einer Metrik dann einen „metrischen Raum“.

Nicht nur Konvergenz läßt sich in diesem allgemeinen Rahmen behandeln. In Analysis I, Abschnitt 5.2, haben wir erklärt:

**Definition.** Sei  $f : D \rightarrow \mathbb{R}$  eine Funktion (mit  $D \subseteq \mathbb{R}$ ) und  $a \in D$ . Die Funktion  $f$  heißt *stetig in  $a$* , wenn gilt:

$$\forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall x \in D : (|x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon).$$

Auch hier kommen nur die „Abstände“  $|x - a|$  und  $|f(x) - f(a)|$  vor. Die Beweise einiger einfacher Aussagen über stetige Funktionen benutzen nur die formalen Eigenschaften dieser Abstände. Dementsprechend läßt sich auch Stetigkeit allgemein in metrischen Räumen behandeln.

Im Zusammenhang mit der Untersuchung stetiger Funktionen hatten wir ferner in Analysis I sogenannte topologische Begriffe benutzt wie *Umgebung*, *offen*, *abgeschlossen*, *kompakt*. Auch diese Begriffe erfordern zu ihrer Definition nur den Abstandsbegriff (oder noch weniger) und lassen sich daher in wesentlich allgemeinerem Rahmen erfolgreich einsetzen.

## 9.1 Metrische und topologische Grundbegriffe

**1.1 Definition.** Sei  $M$  eine nichtleere Menge. Eine Metrik auf  $M$  ist eine Abbildung  $d : M \times M \rightarrow \mathbb{R}$  mit folgenden Eigenschaften. Für alle  $x, y, z \in M$  gilt

- (a)  $d(x, y) \geq 0$ ;  $d(x, y) = 0 \Leftrightarrow x = y$ ,
- (b)  $d(x, y) = d(y, x)$ ,
- (c)  $d(x, z) \leq d(x, y) + d(y, z)$  („Dreiecksungleichung“).

Ist  $d$  eine Metrik auf  $M$ , so heißt die Zahl  $d(x, y)$  der Abstand von  $x$  und  $y$ , und das Paar  $(M, d)$  heißt metrischer Raum.

**Bemerkung.** Sei  $(M, d)$  ein metrischer Raum. Die Elemente von  $M$  werden häufig als „Punkte“ bezeichnet. Oft wird auch einfach  $M$  als metrischer Raum bezeichnet, wenn aus dem Zusammenhang klar ist, welche Metrik auf  $M$  vorgegeben ist.

**Beispiele.** (1)  $M = \mathbb{R}, d(x, y) := |x - y|$ ,

(2)  $M = \mathbb{C}, d(x, y) := |x - y|$ ,

(3) Sei  $M$  die Menge aller endlichen 0-1-Folgen der Länge  $n$ . Für  $x = (\xi_1, \dots, \xi_n), y = (\eta_1, \dots, \eta_n) \in M, \xi_i, \eta_i \in \{0, 1\}$ , sei  $d(x, y)$  die Anzahl der Indizes  $i \in \{1, \dots, n\}$  mit  $\xi_i \neq \eta_i$ . Dann ist  $d$  eine Metrik auf  $M$ , und  $d(x, y)$  heißt der *Hamming-Abstand* von  $x$  und  $y$ ,

(4)  $M = \mathbb{R}^n$ ,

$$d((x_1, \dots, x_n), (y_1, \dots, y_n)) := \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

(„euklidische Metrik“, „ $\ell^2$ -Metrik“),

(5)  $M = \mathbb{R}^2$ ,

$$d((x_1, x_2), (y_1, y_2)) := |x_1 - y_1| + |x_2 - y_2|$$

(„Taxi-Metrik“, „ $\ell^1$ -Metrik“),

(6)  $M \neq \emptyset$ ,

$$d(x, y) := \begin{cases} 1, & \text{wenn } x \neq y, \\ 0, & \text{wenn } x = y. \end{cases}$$

(„diskrete Metrik“),

(7) Sei  $V$  ein Vektorraum über  $\mathbb{R}$  oder  $\mathbb{C}$ . Eine *Norm* auf  $V$  ist eine Abbildung  $\|\cdot\| : V \rightarrow \mathbb{R}$  mit folgenden Eigenschaften. Für alle  $x, y \in V$  gilt

(a)  $\|x\| \geq 0; \|x\| = 0 \Leftrightarrow x = 0$  (Nullvektor),

(b)  $\|\lambda x\| = |\lambda| \|x\|$  für  $\lambda \in \mathbb{R}$  bzw.  $\lambda \in \mathbb{C}$ ,

(c)  $\|x + y\| \leq \|x\| + \|y\|$ .

Ist  $\|\cdot\|$  eine Norm auf  $V$ , so wird durch

$$d(x, y) := \|x - y\| \quad \text{für } x, y \in V$$

eine Metrik auf  $V$  induziert.

- (8) Sei  $M$  die Menge aller Folgen reeller Zahlen. Auf  $M$  wird eine Metrik erklärt durch

$$d(x, y) := \sum_{j=1}^{\infty} \frac{1}{2^j} \frac{|x_j - y_j|}{1 + |x_j - y_j|}$$

für  $x = (x_j)_{j \in \mathbb{N}}, y = (y_j)_{j \in \mathbb{N}} \in M$ .

- (9) Seien  $(M, d), (M', d')$  metrische Räume. Dann werden auf dem Produkt  $M \times M'$  Metriken erklärt durch

$$d_1((x, x'), (y, y')) := d(x, y) + d'(x', y')$$

und durch

$$d_2((x, x'), (y, y')) := \sqrt{d(x, y)^2 + d'(x', y')^2}.$$

Unmittelbar einzusehen ist auch die Richtigkeit der folgenden Definition und Behauptung.

**1.2 Definition.** Sei  $(M, d)$  ein metrischer Raum und  $\emptyset \neq M' \subseteq M$  eine Teilmenge. Dann wird durch die Einschränkung

$$d' := d|_{M' \times M'}$$

auf  $M'$  eine Metrik  $d'$  gegeben.  $(M', d')$  heißt Unterraum oder Teilraum von  $(M, d)$ .

Im folgenden sei  $(M, d)$  ein gegebener metrischer Raum.

**1.3 Lemma.** Für alle  $x, \bar{x}, y, \bar{y} \in M$  gilt die „Vierecksungleichung“

$$|d(x, \bar{x}) - d(y, \bar{y})| \leq d(x, y) + d(\bar{x}, \bar{y}).$$

*Beweis.* Aus der Dreiecksungleichung folgt

$$\begin{aligned} d(x, \bar{x}) &\leq d(x, y) + d(y, \bar{x}) \\ &\leq d(x, y) + d(y, \bar{y}) + d(\bar{y}, \bar{x}) \\ &= d(x, y) + d(y, \bar{y}) + d(\bar{x}, \bar{y}), \end{aligned}$$

also

$$d(x, \bar{x}) - d(y, \bar{y}) \leq d(x, y) + d(\bar{x}, \bar{y}).$$

Analog ergibt sich

$$d(y, \bar{y}) - d(x, \bar{x}) \leq d(x, y) + d(\bar{x}, \bar{y})$$

und damit die Behauptung. ■

Wir übertragen nun einige aus Analysis I bekannte Begriffsbildungen metrischer und topologischer Art in den allgemeineren Rahmen metrischer Räume und beweisen darüber eine Reihe einfacher Aussagen, die im Spezialfall bereits vertraut sind.

**1.4 Definition.** Eine Teilmenge  $A \subseteq M$  heißt beschränkt, wenn es eine Zahl  $c \in \mathbb{R}^+$  gibt mit  $d(x, y) < c$  für alle  $x, y \in A$ . Ist  $A$  beschränkt, so heißt die reelle Zahl

$$\delta(A) := \sup_{x, y \in A} d(x, y)$$

der Durchmesser von  $A$ .

**1.5 Satz.** Die Vereinigung von endlich vielen beschränkten Mengen ist beschränkt.

Der Beweis ist eine einfache Übungsaufgabe.

**1.6 Definition.** Sei  $x \in M$ ,  $\varepsilon \in \mathbb{R}^+$ . Die Menge

$$U(x, \varepsilon) := \{y \in M \mid d(x, y) < \varepsilon\}$$

heißt offene Kugel um  $x$  mit Radius  $\varepsilon$ .

**1.7 Definition.** Eine Teilmenge  $A \subseteq M$  heißt Umgebung des Punktes  $x \in M$ , wenn ein  $\varepsilon \in \mathbb{R}^+$  existiert mit  $U(x, \varepsilon) \subseteq A$ . Die Menge  $A$  heißt offen, wenn zu jedem  $x \in A$  ein  $\varepsilon \in \mathbb{R}^+$  existiert mit  $U(x, \varepsilon) \subseteq A$ .

Eine Menge ist also genau dann offen, wenn sie Umgebung für jeden ihrer Punkte ist.

**1.8 Satz.** Jede offene Kugel ist offen.

*Beweis.* Die offene Kugel  $U(x, \varepsilon)$  in  $M$  sei gegeben. Sei  $y \in U(x, \varepsilon)$ . Dann ist  $d(x, y) < \varepsilon$ , also  $\varepsilon' := \varepsilon - d(x, y) > 0$ . Ist  $z \in U(y, \varepsilon')$ , so ist  $d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \varepsilon' = \varepsilon$ , also  $z \in U(x, \varepsilon)$ . Da  $z \in U(y, \varepsilon')$  beliebig war, ist  $U(y, \varepsilon') \subseteq U(x, \varepsilon)$ . Da  $y \in U(x, \varepsilon)$  beliebig war, ist  $U(x, \varepsilon)$  also offen. ■

Das System aller offenen Teilmengen von  $M$  hat die folgenden Eigenschaften.

**1.9 Satz.** Es gilt

- (a) Die Vereinigung von beliebig vielen offenen Mengen ist offen.
- (b) Der Durchschnitt von endlich vielen offenen Mengen ist offen.
- (c)  $\emptyset$  und  $M$  sind offen.

*Beweis.* (a) Sei  $(A_i)_{i \in I}$  ( $I$  eine Indexmenge) eine Familie offener Teilmengen von  $M$ . Setze  $A := \bigcup_{i \in I} A_i$ . Sei  $x \in A$ . Dann gibt es ein  $i \in I$  mit  $x \in A_i$ . Da  $A_i$  offen ist, existiert ein  $\varepsilon \in \mathbb{R}^+$  mit  $U(x, \varepsilon) \subseteq A_i \subseteq A$ . Da  $x \in A$  beliebig war, ist  $A$  offen.

(b) Sei  $(A_i)_{i \in E}$  ( $E$  eine endliche Indexmenge) eine endliche Familie offener Teilmengen von  $M$ . Setze  $B := \bigcap_{i \in E} A_i$ . Sei  $x \in B$ . Für jedes  $i \in E$  gilt  $x \in A_i$ ; da  $A_i$  offen ist, existiert ein  $\varepsilon_i \in \mathbb{R}^+$  mit  $U(x, \varepsilon_i) \subseteq A_i$ . Setze  $\varepsilon := \min\{\varepsilon_i \mid i \in E\}$ . Da  $E$  endlich ist, existiert das Minimum und ist positiv. Es gilt  $U(x, \varepsilon) \subseteq U(x, \varepsilon_i) \subseteq A_i$  für alle  $i \in E$ , also  $U(x, \varepsilon) \subseteq B$ . Da  $x \in B$  beliebig war, ist  $B$  offen.

(c) ist trivial. ■

**1.10 Definition.** Sei  $A \subseteq M$  und  $x \in M$ . Der Punkt  $x$  heißt Berührungspunkt von  $A$ , wenn jede Umgebung von  $x$  einen Punkt von  $A$  enthält, und  $x$  heißt Häufungspunkt von  $A$ , wenn jede Umgebung von  $x$  einen Punkt aus  $A \setminus \{x\}$  enthält. Die Menge  $A$  heißt abgeschlossen, wenn sie alle ihre Berührungspunkte enthält.

Es sei daran erinnert, dass man (unter Bezugnahme auf eine bestimmte vorliegende Grundmenge  $M$ ) unter dem Komplement einer Teilmenge  $A \subseteq M$  die Menge

$$A^c := M \setminus A$$

versteht und dass  $(A^c)^c = A$  ist.

**1.11 Satz.** Die Teilmenge  $A \subseteq M$  ist genau dann abgeschlossen, wenn ihr Komplement  $A^c$  offen ist.

*Beweis.* Sei  $A$  abgeschlossen. Sei  $x \in A^c$ . Dann ist  $x$  nicht Berührungspunkt von  $A$ , also existiert ein  $\varepsilon \in \mathbb{R}^+$  mit  $U(x, \varepsilon) \cap A = \emptyset$ , das heißt mit  $U(x, \varepsilon) \subseteq A^c$ . Da  $x \in A^c$  beliebig war, ist  $A^c$  offen.

Sei  $A^c$  offen. Sei  $x$  Berührungspunkt von  $A$ . Angenommen, es wäre  $x \notin A$ , also  $x \in A^c$ . Da  $A^c$  offen ist, existiert ein  $\varepsilon \in \mathbb{R}^+$  mit  $U(x, \varepsilon) \subseteq A^c$ , also mit  $U(x, \varepsilon) \cap A = \emptyset$ , folglich ist  $x$  kein Berührungspunkt von  $A$ . Aus dem Widerspruch folgt  $x \in A$ . Da  $x$  ein beliebiger Berührungspunkt von  $A$  war, ist  $A$  abgeschlossen. ■

**Beispiel.** In einem Raum mit diskreter Metrik  $d$  ist jede Teilmenge offen, denn mit  $x \in A$  ist  $U(x, 1) = \{x\} \subseteq A$ . Nach Satz 1.11 ist daher auch jede Teilmenge abgeschlossen.

Im allgemeinen ist eine Teilmenge eines metrischen Raumes weder offen noch abgeschlossen.

Unter Verwendung von Satz 1.11 und den de Morganschen Regeln  $(\bigcup A_i)^c = \bigcap A_i^c$  und  $(\bigcap A_i)^c = \bigcup A_i^c$  ergibt sich aus Satz 1.9 sofort die folgende duale Aussage.

**1.12 Satz.** *Es gilt*

- (a) *Der Durchschnitt von beliebig vielen abgeschlossenen Mengen ist abgeschlossen.*
- (b) *Die Vereinigung von endlich vielen abgeschlossenen Mengen ist abgeschlossen.*
- (c)  *$M$  und  $\emptyset$  sind abgeschlossen.*

**1.13 Definition.** *Sei  $A \subseteq M$  und  $x \in M$ . Der Punkt  $x$  heißt innerer Punkt von  $A$ , wenn ein  $\varepsilon \in \mathbb{R}^+$  existiert mit  $U(x, \varepsilon) \subseteq A$  (also wenn  $A$  Umgebung von  $x$  ist). Der Punkt  $x$  heißt Randpunkt von  $A$ , wenn jede Umgebung von  $x$  Punkte aus  $A$  und aus  $A^c$  enthält. Die Menge  $A^\circ$  aller inneren Punkte von  $A$  heißt das Innere oder der offene Kern von  $A$ ; die Menge  $\bar{A}$  aller Berührungspunkte von  $A$  heißt die abgeschlossene Hülle von  $A$ , und die Menge  $\partial A$  aller Randpunkte von  $A$  heißt der Rand von  $A$ .*

**Bemerkung.** Unmittelbar aus den Definitionen folgt  $\bar{A} = A \cup \partial A$  und  $\partial A = \bar{A} \setminus A^\circ$ .

**1.14 Satz.** *Sei  $A \subseteq M$ . Das Innere  $A^\circ$  ist die Vereinigung aller offenen Teilmengen von  $A$ ;  $A^\circ$  ist offen.  $\bar{A}$  ist abgeschlossen und ist der Durchschnitt aller abgeschlossenen Teilmengen von  $M$ , die  $A$  enthalten.*

Der Beweis kann als Übungsaufgabe dienen.

**1.15 Definition.** *Eine Teilmenge  $A \subseteq M$  heißt dicht, wenn  $\bar{A} = M$  ist.*

**Beispiel.**  $\mathbb{Q}$  (vgl. Satz 4.8, Kapitel 1) und  $\mathbb{R} \setminus \mathbb{Q}$  sind dicht in  $\mathbb{R}$ .

## 9.2 Konvergenz und Vollständigkeit

Im folgenden liegt wieder ein fester metrischer Raum  $(M, d)$  zugrunde.

Die Konvergenz einer Punktfolge in einem metrischen Raum kann wörtlich so wie für Folgen reeller Zahlen erklärt werden, wenn der Abstand reeller Zahlen durch den allgemeinen Abstand ersetzt wird.

**2.1 Definition.** Sei  $(x_n)_{n \in \mathbb{N}}$  eine Folge in  $M$ , sei  $x \in M$ . Die Folge  $(x_n)_{n \in \mathbb{N}}$  konvergiert gegen  $x$ , und  $x$  heißt Grenzwert (Grenzpunkt, Limes) der Folge, geschrieben

$$\lim_{n \rightarrow \infty} x_n = x \quad \text{oder} \quad x_n \rightarrow x \quad (n \rightarrow \infty),$$

wenn zu jeder Umgebung  $U$  von  $x$  ein  $n_0 \in \mathbb{N}$  existiert mit

$$x_n \in U \quad \text{für alle } n \geq n_0.$$

Die Folge  $(x_n)_{n \in \mathbb{N}}$  heißt konvergent, wenn sie konvergent gegen ein  $x \in M$  ist.

Man beachte, dass bei Verwendung der Schreibweisen  $\lim x_n = x$  etc. stets klar sein muß, auf welche Metrik sich diese Konvergenz bezieht.

Wir geben noch zwei offensichtlich äquivalente Umformulierungen der Konvergenzdefinition an:

$$\begin{aligned} \lim_{n \rightarrow \infty} x_n = x &\Leftrightarrow \forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n \geq n_0 : d(x_n, x) < \varepsilon \\ &\Leftrightarrow \lim_{n \rightarrow \infty} d(x_n, x) = 0. \end{aligned}$$

Einige elementare Aussagen über Konvergenz lassen sich analog wie in Analysis I beweisen.

**2.2 Satz.** Der Limes einer konvergenten Folge ist eindeutig bestimmt.

*Beweis.* Sei  $(x_n)_{n \in \mathbb{N}}$  eine Folge in  $M$ , die sowohl gegen  $x$  als auch gegen  $y$  konvergiert. Angenommen, es wäre  $x \neq y$ . Dann ist  $d(x, y) > 0$ . Zu  $\varepsilon := \frac{1}{2}d(x, y)$  gibt es nach Voraussetzung ein  $n_0 \in \mathbb{N}$  mit  $d(x, x_n) < \varepsilon$  für  $n \geq n_0$  und ein  $n_1 \in \mathbb{N}$  mit  $d(y, x_n) < \varepsilon$  für  $n \geq n_1$ . Für  $n \geq \max\{n_0, n_1\}$  gilt also

$$d(x, y) \leq d(x, x_n) + d(x_n, y) < \varepsilon + \varepsilon = d(x, y).$$

Aus diesem Widerspruch folgt  $x = y$ . ■

Die Folge  $(x_n)_{n \in \mathbb{N}}$  heißt beschränkt, wenn die Menge  $\{x_n : n \in \mathbb{N}\}$  beschränkt ist.

**2.3 Satz.** Jede konvergente Folge ist beschränkt.

*Beweis.* Sei  $(x_n)_{n \in \mathbb{N}}$  konvergent gegen  $x$ . Dann gibt es ein  $n_0 \in \mathbb{N}$  mit  $d(x, x_n) < 1$  für  $n \geq n_0$ . Mit

$$c := \max\{1, d(x, x_1), \dots, d(x, x_{n_0-1})\}$$

gilt also für beliebige  $n, m \in \mathbb{N}$

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) \leq 2c. \quad \blacksquare$$

**Bemerkung.** Der wichtige Satz von Bolzano-Weierstraß (Satz 1.3, Kapitel 4) aus Analysis I besagt, dass in  $\mathbb{R}$  jede beschränkte Folge reeller Zahlen eine konvergente Teilfolge besitzt. Diese Aussage läßt sich nicht auf allgemeine metrische Räume übertragen. Zum Beispiel besitzt eine injektive Folge in einem Raum mit diskreter Metrik keine konvergente Teilfolge, ist aber beschränkt.

**2.4 Satz.** *Jeder Berührungspunkt von  $A \subseteq M$  ist Limes einer Folge in  $A$ .*

*Beweis.* Sei  $x$  Berührungspunkt von  $A$ . Für jedes  $n \in \mathbb{N}$  ist dann  $U(x, \frac{1}{n}) \cap A \neq \emptyset$ ; wir können also einen Punkt  $x_n \in U(x, \frac{1}{n}) \cap A$  auswählen. Damit ist eine Folge  $(x_n)_{n \in \mathbb{N}}$  definiert, die wegen  $d(x, x_n) < \frac{1}{n}$  gegen  $x$  konvergiert. ■

Hieraus folgt insbesondere: Die Menge  $A \subseteq M$  ist genau dann abgeschlossen, wenn der Grenzwert jeder konvergenten Folge in  $A$  ebenfalls in  $A$  liegt.

Wenn zwei Folgen konvergieren, so konvergieren auch die Abstände entsprechender Punkte:

**2.5 Lemma.** *Aus  $\lim x_n = x$  und  $\lim y_n = y$  folgt  $\lim d(x_n, y_n) = d(x, y)$ .*

*Beweis.* Nach der Vierecksungleichung (Lemma 1.3) ist

$$|d(x_n, y_n) - d(x, y)| \leq d(x_n, x) + d(y_n, y),$$

woran man die Behauptung abliest. ■

**2.6 Definition.** *Die Folge  $(x_n)_{n \in \mathbb{N}}$  in  $M$  heißt Cauchy-Folge, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $n_0 \in \mathbb{N}$  existiert mit*

$$d(x_m, x_n) < \varepsilon \quad \text{für alle } m, n \geq n_0.$$

**2.7 Satz.** *Jede konvergente Folge ist eine Cauchy-Folge.*

*Beweis.* Sei  $\lim x_n = x$  und  $\varepsilon \in \mathbb{R}^+$ . Dann existiert ein  $n_0 \in \mathbb{N}$  mit  $d(x, x_n) < \varepsilon/2$  für  $n \geq n_0$ . Für  $n, m \geq n_0$  gilt also

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \varepsilon. \quad \blacksquare$$

Dass umgekehrt nicht jede Cauchy-Folge konvergiert, zeigt schon das Beispiel des Raumes der rationalen Zahlen. Andererseits gilt in  $\mathbb{R}$  das Cauchysche Konvergenzkriterium, und hierauf (bzw. auf äquivalenten Aussagen) beruhen letzten Endes die meisten wesentlichen Sätze aus Analysis I. Die metrischen Räume, in denen jede Cauchy-Folge konvergiert, sind daher besonders wichtig.

**2.8 Definition.** Ein metrischer Raum heißt vollständig, wenn in ihm jede Cauchy-Folge konvergiert. Eine Teilmenge  $A$  eines metrischen Raumes heißt vollständig, wenn sie als Unterraum (mit der induzierten Metrik) vollständig ist, wenn also jede Cauchy-Folge in  $A$  gegen einen Punkt von  $A$  konvergiert.

**2.9 Satz.** Jede vollständige Teilmenge eines metrischen Raumes ist abgeschlossen. In einem vollständigen metrischen Raum ist jede abgeschlossene Teilmenge vollständig.

*Beweis.* Sei  $M$  ein metrischer Raum und  $A \subseteq M$  eine vollständige Teilmenge. Sei  $x$  ein Berührungspunkt von  $A$ . Nach Satz 2.4 gibt es eine Folge in  $A$ , die gegen  $x$  konvergiert. Nach Satz 2.7 ist diese Folge eine Cauchy-Folge, wegen der Vollständigkeit von  $A$  ist sie also konvergent gegen einen Punkt  $y \in A$ . Nach Satz 2.2 ist  $x = y$ , also  $x \in A$ . Somit ist  $A$  abgeschlossen.

Sei  $M$  ein vollständiger metrischer Raum und  $A \subseteq M$  eine abgeschlossene Teilmenge. Sei  $(x_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge in  $A$ . Da  $M$  vollständig ist, konvergiert sie gegen einen Punkt  $x \in M$ . Dieser ist Berührungspunkt von  $A$  und daher Element von  $A$ . Somit ist  $A$  vollständig. ■

**2.10 Definition.** Ein normierter Vektorraum, der (mit der durch die Norm induzierten Metrik) vollständig ist, heißt Banachraum.

Das folgende Beispiel wird im nächsten Abschnitt von Bedeutung sein. Wir betrachten eine Menge  $\emptyset \neq D \subseteq \mathbb{R}$  und die Menge  $B(D)$  aller beschränkten reellen Funktionen  $f: D \rightarrow \mathbb{R}$ . Mit den wie üblich (d.h. punktweise) für Funktionen erklärten Operationen der Addition und der Multiplikation mit Skalaren ist  $B(D)$  ein reeller Vektorraum, und die Supremumsnorm  $\|\cdot\|$  ist darauf eine Norm. Eine Folge  $(f_n)_{n \in \mathbb{N}}$  in  $B(D)$  ist nach der obigen Festsetzung eine Cauchy-Folge, wenn gilt:

$$\forall \varepsilon \in \mathbb{R}^+ \exists n_0 \in \mathbb{N} \forall n, m \geq n_0 : \|f_n - f_m\| < \varepsilon.$$

Diese Definition einer Cauchy-Folge von Funktionen auf  $D$  haben wir auch schon in Analysis I, Abschnitt 8.1, verwendet. Dort haben wir gezeigt: Die Folge  $(f_n)_{n \in \mathbb{N}}$  konvergiert genau dann gleichmäßig, wenn sie eine Cauchy-Folge ist. Gleichmäßige Konvergenz ist aber dasselbe wie Konvergenz im Sinne der Supremumsnorm. Wir haben also:

**2.11 Satz.**  $B(D)$  ist ein Banachraum.

Mit  $C(D)$  bezeichnen wir nun die Teilmenge der stetigen Funktionen in  $B(D)$ . Natürlich ist  $C(D)$  ein Untervektorraum. Wir behaupten, dass er abgeschlossen ist. Sei also  $f$  ein Berührungspunkt von  $C(D)$ . Nach Satz 2.4 ist  $f$  Limes einer Folge  $(f_n)_{n \in \mathbb{N}}$  in  $C(D)$ . Die Konvergenz bezieht sich hier auf die Supremumsnorm, die Folge  $(f_n)_{n \in \mathbb{N}}$  konvergiert also gleichmäßig gegen  $f$ .

Da alle  $f_n$  stetig sind, ist  $f$  nach Satz 1.2 aus Kapitel 8, Analysis I ebenfalls stetig, also Element von  $C(D)$ . Somit ist  $C(D)$  abgeschlossen. Aus Satz 2.11 und Satz 2.9 folgt, dass  $C(D)$  vollständig ist. Wir haben also gezeigt:

**2.12 Satz.**  $C(D)$  ist ein Banachraum.

### 9.3 Der Banachsche Fixpunktsatz

Eine wesentliche Aufgabe der Analysis, vor allem in den Anwendungen, besteht im „Lösen von Gleichungen“. Hierbei kann es sich um die Bestimmung von Nullstellen komplizierter Funktionen, Auflösung linearer Gleichungssysteme, Lösen von gewöhnlichen oder partiellen Differentialgleichungen, Berechnen von implizit definierten Funktionen, Integralgleichungen und vieles andere handeln. In manchen Fällen kann man hier das Verfahren der „sukzessiven Approximationen“ verwenden, das sowohl die Existenz einer Lösung zu beweisen gestattet als auch eine schrittweise Berechnung von Näherungslösungen mit vorgeschriebener Genauigkeit ermöglicht. Das zugrundeliegende Prinzip läßt sich allgemein formulieren und beweisen als ein Fixpunktsatz für gewisse Abbildungen eines vollständigen metrischen Raumes in sich. In diesem Satz, den wir jetzt beweisen wollen, haben wir ein besonders eindrucksvolles Beispiel für die Bedeutung und Auswirkung der Vollständigkeit von metrischen Räumen.

**3.1 Definition.** Sei  $(M, d)$  ein metrischer Raum. Eine Abbildung  $f : M \rightarrow M$  heißt kontrahierend, wenn es eine reelle Zahl  $c < 1$  gibt mit

$$d(f(x), f(y)) \leq cd(x, y) \quad \text{für alle } x, y \in M.$$

Jede Zahl  $c$  mit dieser Eigenschaft heißt eine Lipschitzkonstante der Abbildung  $f$ . Ein Punkt  $x \in M$  heißt Fixpunkt von  $f$ , wenn  $f(x) = x$  ist.

**3.2 Satz** (Banachscher Fixpunktsatz). Sei  $(M, d)$  ein vollständiger metrischer Raum und  $f : M \rightarrow M$  eine kontrahierende Abbildung. Dann hat  $f$  genau einen Fixpunkt.

Ist  $x_0 \in M$  beliebig und wird rekursiv

$$x_{n+1} := f(x_n) \quad \text{für } n \in \mathbb{N}_0$$

definiert, so konvergiert die Folge  $(x_n)_{n \in \mathbb{N}}$  gegen den Fixpunkt  $x$ , und es gilt die Fehlerabschätzung

$$d(x_n, x) \leq \frac{c}{1-c} d(x_{n-1}, x_n) \leq \frac{c^n}{1-c} d(x_0, x_1),$$

wenn  $c < 1$  eine Lipschitzkonstante der Abbildung  $f$  ist.

*Beweis.* Sei  $M$  vollständig und  $f : M \rightarrow M$  eine Abbildung mit einer Lipschitzkonstanten  $c < 1$ . Dass  $f$  höchstens einen Fixpunkt hat, ist klar: Gilt  $f(x) = x$  und  $f(y) = y$ , so ist

$$d(x, y) = d(f(x), f(y)) \leq cd(x, y),$$

wegen  $c < 1$  also  $d(x, y) = 0$  und daher  $x = y$ .

Wir wählen nun  $x_0 \in M$  beliebig und definieren rekursiv  $x_{n+1} := f(x_n)$  für  $n \in \mathbb{N}_0$ . Wir wollen zeigen, dass  $(x_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge ist. Für  $n \geq 1$  gilt

$$d(x_n, x_{n+1}) = d(f(x_{n-1}), f(x_n)) \leq cd(x_{n-1}, x_n).$$

Durch vollständige Induktion bekommt man daraus

$$d(x_n, x_{n+1}) \leq c^n d(x_0, x_1)$$

und hieraus für  $n \geq 1$  und  $p \geq 1$  unter mehrfacher Verwendung der Dreiecksungleichung

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq (c^n + c^{n+1} + \cdots + c^{n+p-1})d(x_0, x_1) \\ &= c^n \frac{1 - c^p}{1 - c} d(x_0, x_1) \\ &\leq \frac{c^n}{1 - c} d(x_0, x_1). \end{aligned}$$

Zu gegebenem  $\varepsilon \in \mathbb{R}^+$  gibt es also ein  $n_0 \in \mathbb{N}$  mit  $d(x_n, x_m) < \varepsilon$  für alle  $n, m \geq n_0$ ; somit ist  $(x_n)_{n \in \mathbb{N}}$  eine Cauchy-Folge. Da  $M$  als vollständig vorausgesetzt ist, existiert ein Punkt  $x \in M$  mit  $\lim x_n = x$ . Wegen  $d(f(x), f(x_n)) \leq cd(x, x_n)$  gilt

$$f(x) = \lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} x_{n+1} = x.$$

Also ist  $x$  ein Fixpunkt von  $f$ .

Die Fehlerabschätzung ergibt sich aus

$$\begin{aligned} d(x_n, x_{n+p}) &\leq d(x_n, x_{n+1}) + \cdots + d(x_{n+p-1}, x_{n+p}) \\ &\leq (c + c^2 + \cdots + c^p) d(x_{n-1}, x_n) \end{aligned}$$

durch den Grenzübergang  $p \rightarrow \infty$ , wobei Lemma 2.5 zu beachten ist. Es folgt

$$d(x_n, x) \leq \frac{c}{1 - c} d(x_{n-1}, x_n) \leq \frac{c^n}{1 - c} d(x_0, x_1).$$

Damit ist alles bewiesen. ■

Bei konkreter Anwendung wird man, um einen Fixpunkt der kontrahierenden Abbildung  $f$  näherungsweise zu berechnen, so vorgehen: Man wählt  $x_0$  beliebig (aber möglichst nahe beim Fixpunkt, wenn man eine Vermutung über dessen ungefähre Lage hat), berechnet  $x_1 = f(x_0)$  (wir unterstellen, dass dies möglich ist – meist wird es auch nur näherungsweise möglich sein), berechnet  $x_2 = f(x_1)$ , und so weiter. Wegen der schrittweisen Annäherung an den Fixpunkt spricht man von einem „Iterationsverfahren“ oder von „sukzessiven Approximationen“. Nach dem  $n$ -ten Schritt liefert die Ungleichung

$$d(x_n, x) \leq \frac{c}{1-c} d(x_{n-1}, x_n)$$

eine Information darüber, wie weit man höchstens vom Fixpunkt entfernt ist. Wenn man zu Beginn des Verfahrens schon abschätzen will, wieviele Schritte man höchstens braucht, um den Faktor unter eine vorgegebene Schranke zu drücken, so erhält man eine solche Information (nach Berechnung von  $x_1$ ) aus der Ungleichung

$$d(x_n, x) \leq \frac{c^n}{1-c} d(x_0, x_1).$$

**Bemerkung.** In Satz 3.2 ist die Voraussetzung wesentlich, dass die Abbildung  $f$  eine Lipschitzkonstante  $c < 1$  besitzt. Die schwächere Voraussetzung

$$d(f(x), f(y)) < d(x, y) \quad \text{für alle } x, y \in M \text{ mit } x \neq y$$

reicht nicht aus, um die Existenz eines Fixpunktes zu zeigen. Zum Beispiel sei  $M = \mathbb{R}$  (mit der üblichen Metrik) und  $f : \mathbb{R} \rightarrow \mathbb{R}$  definiert durch

$$f(x) := x - \frac{\pi}{2} - \arctan x \quad \text{für } x \in \mathbb{R}.$$

Es ist

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

also  $0 \leq f'(x) < 1$  für alle  $x \in \mathbb{R}$ . Nach dem Mittelwertsatz der Differentialrechnung folgt daraus

$$|f(x) - f(y)| < |x - y| \quad \text{für alle } x, y \in \mathbb{R} \text{ mit } x \neq y.$$

Da aber  $f(x) < x$  für alle  $x \in \mathbb{R}$  gilt, hat  $f$  keinen Fixpunkt.

Wir behandeln zwei Anwendungsbeispiele für den Banachschen Fixpunktsatz.

**Beispiele.** Zunächst betrachten wir eine reelle Funktion  $f : [a, b] \rightarrow [a, b]$ . Es sei eine Lösung der Gleichung  $f(x) = x$  zu finden. Ist  $f$  stetig, so existiert

nach dem Zwischenwertsatz jedenfalls eine Lösung, denn es ist  $f(x) - x \geq 0$  für  $x = a$  und  $\leq 0$  für  $x = b$ . Für die Eindeutigkeit der Lösung und die näherungsweise Berechenbarkeit durch sukzessive Approximationen ist zum Beispiel hinreichend, dass  $f$  differenzierbar ist und eine Konstante  $c < 1$  existiert mit

$$|f'(x)| \leq c \quad \text{für alle } x \in [a, b].$$

Hieraus folgt nämlich nach dem Mittelwertsatz der Differentialrechnung

$$|f(x) - f(y)| \leq c|x - y| \quad \text{für alle } x, y \in [a, b],$$

so dass Satz 3.2 anwendbar ist mit  $M = [a, b]$ .

Das nächste Beispiel ist etwas schwieriger, aber interessanter und besonders wichtig. Es handelt sich um das sogenannte *Anfangswertproblem für eine gewöhnliche Differentialgleichung erster Ordnung*. In der üblichen Schreibweise lautet es

$$y' = f(x, y), \quad y(x_0) = y_0.$$

Wir präzisieren und erläutern das folgendermaßen. Es seien  $I, J \subseteq \mathbb{R}$  zwei offene Intervalle und  $f : I \times J \rightarrow \mathbb{R}$  eine gegebene reelle Funktion. Ferner sei ein Punkt  $(x_0, y_0) \in I \times J$  gegeben. Gesucht ist eine in einer Umgebung  $U \subseteq I$  von  $x_0$  definierte differenzierbare reelle Funktion  $y$  (mit  $y(U) \subseteq J$ ), für die

$$\begin{aligned} y'(x) &= f(x, y(x)) & \text{für alle } x \in U, \\ y(x_0) &= y_0 \end{aligned}$$

gilt (man veranschauliche sich die Aufgabenstellung durch ein „Richtungsfeld“).

Wir wollen zeigen, dass man die Existenz einer solchen Lösung aus dem Banachschen Fixpunktsatz erhält, wenn die Funktion  $f$  die folgende Voraussetzung erfüllt.

**Voraussetzung.**  $f$  sei stetig in  $I \times J$  (die Erklärung erfolgt erst in Abschnitt 9.4) und beschränkt und genüge einer Lipschitzbedingung in der zweiten Veränderlichen, das heißt es gebe eine Zahl  $L \in \mathbb{R}$  mit

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|$$

für alle  $x \in I$  und alle  $y_1, y_2 \in J$ .

Wenn wir eine (in einem Intervall um  $x_0$  erklärte) stetige Funktion  $y$  haben mit

$$y(x) = y_0 + \int_{x_0}^x f(t, y(t)) dt, \quad (3.3)$$

so ist  $y$  nach dem Hauptsatz der Differential- und Integralrechnung (Analysis I, Kapitel 7, Satz 3.2) differenzierbar, und es gilt

$$\begin{aligned} y'(x) &= f(x, y(x)), \\ y(x_0) &= y_0, \end{aligned} \quad (3.4)$$

das heißt  $y$  löst unser Anfangswertproblem. Diese Tatsache werden wir im folgenden ausnutzen.

Wir müssen nun eine geeignete Situation herstellen, in der wir den Banachschen Fixpunktsatz anwenden können. Nach Voraussetzung existiert eine Konstante  $a \in \mathbb{R}$  mit

$$|f(x, y)| < a \quad \text{für } (x, y) \in I \times J.$$

Wir können  $\delta > 0$  wählen, so dass  $\delta L < 1$  und

$$[x_0 - \delta, x_0 + \delta] \times [y_0 - \delta a, y_0 + \delta a] \subseteq I \times J$$

ist. Sodann sei  $C([x_0 - \delta, x_0 + \delta])$  der reelle Vektorraum aller stetigen reellen Funktionen auf dem Intervall  $[x_0 - \delta, x_0 + \delta]$  mit der Supremumsnorm  $\|\cdot\|$  und

$$M := \{g \in C([x_0 - \delta, x_0 + \delta]) \mid g(x_0) = y_0 \text{ und } |g(x) - y_0| \leq \delta a \text{ für } |x - x_0| \leq \delta\}.$$

Der Raum  $C([x_0 - \delta, x_0 + \delta])$  ist nach Satz 2.12 vollständig. Die Teilmenge  $M$  ist offenbar abgeschlossen und daher nach Satz 2.9 ebenfalls vollständig. Wir definieren nun eine Abbildung  $F : M \rightarrow M$  in der folgenden Weise. Für  $g \in M$  sei die Funktion  $F(g)$  erklärt durch

$$F(g)(x) := y_0 + \int_{x_0}^x f(t, g(t)) dt \quad \text{für } x \in [x_0 - \delta, x_0 + \delta].$$

Die Funktion  $F(g)$  ist stetig, sie erfüllt  $F(g)(x_0) = y_0$ , und für jedes  $x \in [x_0 - \delta, x_0 + \delta]$  gilt

$$|F(g)(x) - y_0| = \left| \int_{x_0}^x f(t, g(t)) dt \right| \leq \int_{x_0}^x |f(t, g(t))| dt \leq \delta a.$$

Also ist  $F(g) \in M$ , so dass in der Tat eine Abbildung  $F : M \rightarrow M$  erklärt worden ist. Wir zeigen, dass  $F$  kontrahierend ist. Sei  $g, h \in M$ . Für  $x \in [x_0 - \delta, x_0 + \delta]$  gilt

$$\begin{aligned}
|F(g)(x) - F(h)(x)| &= \left| \int_{x_0}^x [f(t, g(t)) - f(t, h(t))] dt \right| \\
&\leq \int_{x_0}^x |f(t, g(t)) - f(t, h(t))| dt \\
&\leq L \int_{x_0}^x |g(t) - h(t)| dt \\
&\leq L\delta \|g - h\|.
\end{aligned}$$

Da dies für alle  $x \in [x_0 - \delta, x_0 + \delta]$  gilt, ist

$$\|F(g) - F(h)\| \leq L\delta \|g - h\|.$$

Nach Wahl von  $\delta$  ist  $L\delta < 1$ , also  $F$  in der Tat kontrahierend. Jetzt folgt aus Satz 3.2, dass  $F$  einen Fixpunkt besitzt. Es gibt also ein Element  $y \in M$  mit  $F(y) = y$ . Die Funktion  $y$  ist auf  $[x_0 - \delta, x_0 + \delta]$  erklärt, dort stetig, und sie erfüllt

$$y(x) = y_0 + \int_{x_0}^x f(t, y(t)) dt \quad \text{für } x \in [x_0 - \delta, x_0 + \delta].$$

Wie oben bemerkt, folgt hieraus

$$\begin{aligned}
y'(x) &= f(x, y(x)), \\
y(x_0) &= y_0,
\end{aligned}$$

woraus insbesondere folgt, dass  $y$  stetig differenzierbar ist. Aus der eindeutigen Bestimmtheit des Fixpunktes folgt auch, dass  $y$  als Lösung des Anfangswertproblems auf  $[x_0 - \delta, x_0 + \delta]$  eindeutig bestimmt ist. Wir fassen das gerade Bewiesene zusammen.

**3.5 Satz** (Picard, Lindelöf). *Seien  $I, J \subset \mathbb{R}$  zwei offene Intervalle und  $f : I \times J \rightarrow \mathbb{R}$  eine gegebene stetige reelle Funktion. Diese sei beschränkt und genüge einer Lipschitzbedingung in der zweiten Veränderlichen, das heißt es gibt Zahlen  $a, L \in \mathbb{R}^+$  mit*

$$\begin{aligned}
|f(x, y_1)| &\leq a, \\
|f(x, y_1) - f(x, y_2)| &\leq L|y_1 - y_2|
\end{aligned}$$

für alle  $x \in I$  und alle  $y_1, y_2 \in J$ . Dann existiert für alle  $(x_0, y_0) \in I \times J$  ein  $\delta > 0$  und genau eine stetig differenzierbare Funktion  $y : [x_0 - \delta, x_0 + \delta] \rightarrow \mathbb{R}$ , die das Anfangswertproblem (3.4) im Intervall  $[x_0 - \delta, x_0 + \delta]$  löst.

## 9.4 Stetigkeit und Zusammenhang

Im folgenden seien  $(M, d)$ ,  $(M', d')$  metrische Räume. Wir wollen Abbildungen  $f : M \rightarrow M'$  betrachten. Im Spezialfall  $M' \subseteq \mathbb{R}$  sprechen wir auch von reellen Funktionen auf  $M$ . Die Definition der Stetigkeit einer Abbildung läßt sich wörtlich aus dem in Analysis I betrachteten Spezialfall übernehmen.

**4.1 Definition.** Sei  $f : M \rightarrow M'$  eine Abbildung, sei  $x \in M$ . Die Abbildung  $f$  heißt stetig in  $x$ , wenn zu jeder Umgebung  $V$  von  $f(x)$  eine Umgebung  $U$  von  $x$  existiert mit  $f(U) \subseteq V$ . Die Abbildung  $f$  heißt stetig, wenn sie stetig in  $x$  ist für alle  $x \in M$ .

Gemäß der Definition der Umgebungen können wir die Stetigkeit in äquivalenter Weise auch folgendermaßen erklären:

$f$  stetig in  $x \Leftrightarrow$

$$\forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall y \in M : (d(x, y) < \delta \Rightarrow d'(f(x), f(y)) < \varepsilon).$$

**4.2 Satz.** Die Abbildung  $f : M \rightarrow M'$  ist genau dann stetig, wenn für jede offene Menge  $A \subseteq M'$  das Urbild  $f^{-1}(A)$  offen ist.

*Beweis.* Sei  $f$  stetig. Sei  $A \subseteq M'$  offen. Sei  $x \in f^{-1}(A)$ . Dann ist  $f(x) \in A$ . Da  $A$  offen ist, ist  $A$  Umgebung von  $f(x)$ . Da  $f$  in  $x$  stetig ist, existiert eine Umgebung  $U \subseteq M$  von  $x$  mit  $f(U) \subseteq A$ , also mit  $U \subseteq f^{-1}(A)$ . Somit ist  $f^{-1}(A)$  offen.

Für jede offene Menge  $A \subseteq M'$  sei  $f^{-1}(A)$  offen. Sei  $x \in M$ . Sei  $V$  eine Umgebung von  $f(x)$ ; sie enthält eine offene Umgebung  $V_0$  von  $f(x)$ . Dann ist  $U := f^{-1}(V_0)$  offen, also Umgebung von  $x$ , und es gilt  $f(U) \subseteq V_0 \subseteq V$ . Also ist  $f$  stetig in  $x$ . ■

Einige Aussagen über stetige Abbildungen zwischen metrischen Räumen lassen sich ganz analog wie im Spezialfall  $M = M' = \mathbb{R}$  beweisen.

**4.3 Satz.** Die Abbildung  $f : M \rightarrow M'$  ist genau dann stetig in  $x$ , wenn für jede Folge  $(x_n)_{n \in \mathbb{N}}$  in  $M$  aus  $\lim_{n \rightarrow \infty} x_n = x$  stets  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$  folgt.

*Beweis.* Sei  $f : M \rightarrow M'$  stetig in  $x$ . Sei  $(x_n)_{n \in \mathbb{N}}$  eine Folge in  $M$  mit  $\lim_{n \rightarrow \infty} x_n = x$ . Sei  $V$  eine Umgebung von  $f(x)$ . Wegen der Stetigkeit von  $f$  in  $x$  existiert eine Umgebung  $U$  von  $x$  mit  $f(U) \subseteq V$ . Wegen  $\lim_{n \rightarrow \infty} x_n = x$  existiert ein  $n_0 \in \mathbb{N}$  mit  $x_n \in U$  für  $n \geq n_0$ . Für alle  $n \geq n_0$  gilt also  $f(x_n) \in f(U) \subseteq V$ . Da  $V$  eine beliebige Umgebung von  $f(x)$  war, ist damit  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$  gezeigt.

Umgekehrt folge aus  $\lim_{n \rightarrow \infty} x_n = x$  stets  $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ . Angenommen,  $f$  wäre nicht stetig in  $x$ . Dann gibt es eine Umgebung  $V$  von  $f(x)$  derart, dass für alle Umgebungen  $U$  von  $x$  gilt  $f(U) \not\subseteq V$ . Insbesondere können wir zu jedem  $n \in \mathbb{N}$  einen Punkt  $x_n \in U(x, 1/n)$  auswählen mit  $f(x_n) \notin V$ . Damit ist eine Folge  $(x_n)_{n \in \mathbb{N}}$  in  $M$  definiert, die wegen  $d(x, x_n) < 1/n$  gegen  $x$  konvergiert, für die aber die Folge  $(f(x_n))_{n \in \mathbb{N}}$  wegen  $f(x_n) \notin V$  nicht gegen  $f(x)$  konvergiert. Das ist ein Widerspruch zur Voraussetzung; also ist  $f$  stetig in  $x$ . ■

**4.4 Satz.** *Seien  $M, M', M''$  metrische Räume. Sei  $f : M \rightarrow M'$  stetig in  $x$  und  $g : M' \rightarrow M''$  stetig in  $f(x)$ . Dann ist  $g \circ f$  stetig in  $x$ .*

*Beweis.* Zu einer vorgegebenen Umgebung  $V$  von  $(g \circ f)(x) = g(f(x))$  gibt es eine Umgebung  $W$  von  $f(x)$  mit  $g(W) \subseteq V$ . Zu  $W$  gibt es eine Umgebung  $U$  von  $x$  mit  $f(U) \subseteq W$ . Es folgt  $(g \circ f)(U) = g(f(U)) \subseteq g(W) \subseteq V$ . ■

Ebenfalls in Analogie zu Analysis I definieren wir noch:

**4.5 Definition.** *Die Abbildung  $f : M \rightarrow M'$  heißt gleichmäßig stetig, wenn zu jedem  $\varepsilon \in \mathbb{R}^+$  ein  $\delta \in \mathbb{R}^+$  existiert mit*

$$d'(f(x), f(y)) < \varepsilon \quad \text{für alle } x, y \in M \text{ mit } d(x, y) < \delta.$$

Offensichtlich ist jede gleichmäßig stetige Abbildung auch stetig. Insbesondere (aber nicht nur in diesem Fall) ist eine Abbildung  $f : M \rightarrow M'$  gleichmäßig stetig, wenn es eine Konstante  $c$  gibt mit

$$d'(f(x), f(y)) \leq cd(x, y) \quad \text{für alle } x, y \in M.$$

Abbildungen mit dieser Eigenschaft nennt man *Lipschitzabbildungen* (für  $(M', d') = (M, d)$  und  $c < 1$  kamen sie bereits im Banachschen Fixpunktsatz vor).

Wir betrachten jetzt speziell stetige Abbildungen  $f : M \rightarrow \mathbb{R}$ , also reellwertige stetige Funktionen, und wir wollen Satz 1.2 aus Analysis I, Kapitel 8 verallgemeinern. Wir können ganz allgemein (d.h. für beliebige Mengen  $M$ ) für beschränkte Funktionen  $f : M \rightarrow \mathbb{R}$  die Supremumsnorm erklären durch

$$\|f\| := \sup_{x \in M} |f(x)|.$$

Wir sagen dann, völlig analog zu dem früher betrachteten Spezialfall  $M \subseteq \mathbb{R}$ , dass die Folge  $(f_n)_{n \in \mathbb{N}}$  von Funktionen auf  $M$  *gleichmäßig gegen  $f$  konvergiert*, wenn  $\lim_{n \rightarrow \infty} \|f_n - f\| = 0$  ist. Offenbar impliziert gleichmäßige Konvergenz punktweise Konvergenz.

**4.6 Satz.** Die Folge  $(f_n)_{n \in \mathbb{N}}$  von stetigen reellen Funktionen auf  $M$  konvergiere gleichmäßig gegen die Funktion  $f$ . Dann ist  $f$  stetig.

*Beweis.* Sei  $x \in M$  und  $\varepsilon \in \mathbb{R}^+$ . Da  $(f_n)_{n \in \mathbb{N}}$  gleichmäßig gegen  $f$  konvergiert, existiert ein  $n_0 \in \mathbb{N}$  mit

$$|f_n(y) - f(y)| < \frac{\varepsilon}{3} \quad \text{für } n \geq n_0 \text{ und alle } y \in M.$$

Da  $f_{n_0}$  stetig ist, existiert eine Umgebung  $U$  von  $x$  mit

$$|f_{n_0}(x) - f_{n_0}(y)| < \frac{\varepsilon}{3} \quad \text{für alle } y \in U.$$

Für  $y \in U$  gilt also

$$|f(x) - f(y)| \leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x) - f_{n_0}(y)| + |f_{n_0}(y) - f(y)| < \varepsilon.$$

Da  $\varepsilon \in \mathbb{R}^+$  beliebig war, ist  $f$  stetig in  $x$ . Da  $x \in M$  beliebig war, ist  $f$  stetig. ■

Ein besonders wichtiger Satz aus Analysis I über stetige Funktionen ist der Zwischenwertsatz: Eine auf einem Intervall definierte stetige reelle Funktion nimmt jeden Wert zwischen zwei Funktionswerten an. Wenn man versuchen will, diesen Satz auf stetige Funktionen auf einem metrischen Raum zu verallgemeinern, steht man vor dem folgenden Problem. Wie triviale Gegenbeispiele zeigen, ist für die Gültigkeit des Zwischenwertsatzes die Voraussetzung unentbehrlich, dass der Definitionsbereich ein Intervall ist. Aber ein Intervall ist unter Verwendung der Anordnungsrelation für reelle Zahlen definiert worden; diese Definition läßt sich also nicht auf Teilmengen beliebiger metrischer Räume übertragen. Es zeigt sich aber, dass die für die Gültigkeit des Zwischenwertsatzes wesentliche Eigenschaft der Intervalle auch ohne Verwendung der Anordnungsrelation und damit in verallgemeinerungsfähiger Weise formuliert werden kann. Die hier gemeinte wesentliche Eigenschaft besteht darin, dass ein Intervall sozusagen „aus einem Stück besteht“. Man bezeichnet diese Eigenschaft als Zusammenhang und definiert sie allgemein folgendermaßen.

**4.7 Definition.** Der metrische Raum  $M$  heißt zusammenhängend, wenn es keine offenen Teilmengen  $A_1, A_2 \subseteq M$  gibt mit

$$A_1 \neq \emptyset, \quad A_2 \neq \emptyset, \quad A_1 \cap A_2 = \emptyset, \quad A_1 \cup A_2 = M.$$

Eine Teilmenge  $A \subseteq M$  heißt zusammenhängend, wenn sie als Teilraum (mit der induzierten Metrik) zusammenhängend ist.

Wenn  $M = A_1 \cup A_2$  und  $A_1 \cap A_2 = \emptyset$  ist (man sagt dann,  $M$  sei in die Mengen  $A_1, A_2$  zerlegt), so sind  $A_1$  und  $A_2$  nach Satz 1.11 genau dann beide offen, wenn sie abgeschlossen sind. Man kann also die Definition des Zusammenhangs auch äquivalent folgendermaßen fassen: Der metrische Raum  $M$  ist genau dann zusammenhängend, wenn  $\emptyset$  und  $M$  die einzigen zugleich offenen und abgeschlossenen Teilmengen von  $M$  sind.

dass die Definition 4.7 das Gewünschte leistet, zeigt der folgende Satz.

**4.8 Satz.** *Eine stetige reelle Funktion auf einem zusammenhängenden metrischen Raum nimmt jeden Wert zwischen zwei Funktionswerten an.*

*Beweis.* Sei  $M$  zusammenhängend und  $f : M \rightarrow \mathbb{R}$  stetig. Sei  $c \in \mathbb{R}$  eine Zahl, die zwischen zwei von  $f$  angenommenen Funktionswerten liegt. Dann sind die Mengen  $A_1 := f^{-1}((-\infty, c))$  und  $A_2 := f^{-1}((c, \infty))$  nicht leer und nach Satz 4.2 offen, ferner ist  $A_1 \cap A_2 = \emptyset$ . Würde  $c$  nicht als Funktionswert angenommen, so wäre  $A_1 \cup A_2 = M$ , also wäre  $M$  nicht zusammenhängend, ein Widerspruch. ■

Der Zwischenwertsatz aus Analysis I (Satz 3.4, Kapitel 4) ergibt sich allerdings nicht unmittelbar als Spezialfall von Satz 4.8, da wir noch nicht wissen, dass ein Intervall zusammenhängend ist. Dies kann man umgekehrt gerade mit dem Zwischenwertsatz beweisen.

**4.9 Satz.** *Jedes Intervall in  $\mathbb{R}$  ist zusammenhängend.*

*Beweis.* Allgemeiner sei  $M$  ein metrischer Raum mit der Eigenschaft, dass jede stetige Funktion  $f : M \rightarrow \mathbb{R}$  jeden Wert zwischen zwei Funktionswerten annimmt. Seien  $A_1, A_2 \subseteq M$  nichtleere offene Teilmengen mit  $A_1 \cap A_2 = \emptyset$ . Angenommen, es wäre  $A_1 \cup A_2 = M$ . Setze

$$f(x) := \begin{cases} 1 & \text{für } x \in A_1, \\ 0 & \text{für } x \in A_2. \end{cases}$$

Die damit erklärte Funktion  $f : M \rightarrow \mathbb{R}$  ist stetig, weil Urbilder offener Mengen offen sind. Sie nimmt, da  $A_1$  und  $A_2$  nicht leer sind, die Werte 0 und 1 an, aber keinen Wert, der echt zwischen 0 und 1 liegt. Aus dem Widerspruch folgt, dass  $M$  zusammenhängend ist. ■

## 9.5 Kompaktheit

Unter den möglichen Eigenschaften metrischer Räume sind zwei von herausragender Bedeutung in der Analysis, da sie direkt oder indirekt in die Vor-

aussetzungen vieler wichtiger Sätze eingehen. Dies sind die Vollständigkeit und die jetzt zu behandelnde Kompaktheit.

In Analysis I (Abschnitt 4.1) hatten wir eine Teilmenge von  $\mathbb{R}$  als *kompakt* bezeichnet, wenn sie beschränkt und abgeschlossen ist. Die Voraussetzung der Kompaktheit der Menge  $K \subseteq \mathbb{R}$  spielte in den folgenden wichtigen Sätzen aus Analysis I eine wesentliche Rolle.

- Analysis I, Kapitel 4, Satz 1.6: Jede Folge in  $K$  besitzt eine Teilfolge, die gegen ein Element von  $K$  konvergiert.
- Analysis I, Kapitel 4, Satz 3.2: Sei  $f : K \rightarrow \mathbb{R}$  stetig. Dann nimmt  $f$  ein Maximum an.
- Analysis I, Kapitel 4, Satz 3.5: Sei  $f : K \rightarrow \mathbb{R}$  stetig. Dann ist  $f$  gleichmäßig stetig.

Wir möchten diese Sätze ausdehnen auf allgemeine metrische Räume. Wenn wir auch dort unter kompakten Mengen solche verstehen, die beschränkt und abgeschlossen sind, so sind aber die Sätze im allgemeinen falsch! Zum Beispiel ist der metrische Raum  $(\mathbb{N}, d)$ , wo  $d$  die diskrete Metrik ist, beschränkt und abgeschlossen. Die Folge  $(n)_{n \in \mathbb{N}}$  besitzt aber keine konvergente Teilfolge, da keine Teilfolge eine Cauchy-Folge ist. Und die durch  $f(n) = n$  erklärte Funktion  $f : (\mathbb{N}, d) \rightarrow (\mathbb{R}, |\cdot|)$  ist stetig, da in  $(\mathbb{N}, d)$  jede Teilmenge offen ist, sie nimmt aber kein Maximum an. Diese Beispiele zeigen, dass die Verallgemeinerung der obigen Sätze höchstens dann gelten kann, wenn im allgemeinen Fall die Kompaktheit einschneidender definiert wird. Wie dies zweckmäßig zu geschehen hat, wird durch den Überdeckungssatz von Heine-Borel nahegelegt. Nach diesem Satz (Analysis I, Kapitel 4, Satz 1.8) und seiner Umkehrung (Analysis I, Kapitel 4, Satz 1.9) ist eine Teilmenge  $A \subseteq \mathbb{R}$  genau dann kompakt, wenn jede offene Überdeckung von  $A$  eine endliche Teilüberdeckung enthält. Diese Überdeckungseigenschaft verwenden wir nun im allgemeinen Fall als Definition der Kompaktheit.

**5.1 Definition.** Sei  $(M, d)$  ein metrischer Raum,  $A \subseteq M$  eine Teilmenge und  $(U_i)_{i \in I}$  eine Familie von Teilmengen von  $M$ . Die Familie  $(U_i)_{i \in I}$  heißt Überdeckung der Menge  $A$ , wenn  $A \subseteq \bigcup_{i \in I} U_i$  gilt, und offene Überdeckung von  $A$ , wenn außerdem alle  $U_i$  offene Mengen sind.

**5.2 Definition.** Die Teilmenge  $A \subseteq M$  heißt kompakt (überdeckungskompakt), wenn jede offene Überdeckung von  $A$  eine endliche Teilüberdeckung von  $A$  enthält.

**5.3 Satz.** Jede kompakte Teilmenge eines metrischen Raumes ist abgeschlossen und beschränkt. In einem kompakten metrischen Raum ist jede abgeschlossene Teilmenge kompakt.

*Beweis.* Sei  $(M, d)$  ein metrischer Raum. Sei  $A \subseteq M$  kompakt. Sei  $x$  ein Berührungspunkt von  $A$ . Angenommen,  $x \notin A$ . Dann ist  $(U(y, \frac{1}{2}d(x, y)))_{y \in A}$  eine offene Überdeckung von  $A$ , enthält also eine endliche Teilüberdeckung  $(U(y_i, \frac{1}{2}d(x, y_i)))_{i=1, \dots, n}$ . Setze  $\varepsilon := \min\{d(x, y_i) \mid i = 1, \dots, n\}$ . Dann gilt

$$U(x, \frac{1}{2}\varepsilon) \cap U(y_i, \frac{1}{2}d(x, y_i)) = \emptyset \quad \text{für } i = 1, \dots, n,$$

also  $U(x, \frac{1}{2}\varepsilon) \cap A = \emptyset$ . Somit ist  $x$  nicht Berührungspunkt von  $A$ , ein Widerspruch. Damit ist die Abgeschlossenheit von  $A$  gezeigt.

Da die Überdeckung  $(U(y, 1))_{y \in A}$  von  $A$  eine endliche Überdeckung von  $A$  enthält, ist  $A$  nach Satz 1.5 beschränkt.

Sei jetzt  $M$  kompakt und  $A \subseteq M$  abgeschlossen. Sei  $(U_i)_{i \in I}$  eine offene Überdeckung von  $A$ . Dann ist  $(U_i \cup A^c)_{i \in I}$  eine offene Überdeckung von  $M$ , enthält also eine endliche Teilüberdeckung  $(U_i \cup A^c)_{i \in E}$  ( $E \subseteq I$  endlich).  $(U_i)_{i \in E}$  ist eine endliche Überdeckung von  $A$ . Damit ist die Kompaktheit von  $A$  gezeigt. ■

Wir müssen nun zeigen, dass unser allgemeiner Kompaktheitsbegriff wirklich das Gewünschte leistet, also eine Ausdehnung der am Anfang dieses Abschnitts zitierten Sätze aus Analysis I ermöglicht.

**5.4 Satz.** *Ist  $f : M \rightarrow M'$  eine stetige Abbildung eines kompakten metrischen Raumes  $M$  in einen metrischen Raum  $M'$ , so ist das Bild  $f(M)$  kompakt.*

*Beweis.* Sei  $(U_i)_{i \in I}$  eine offene Überdeckung von  $f(M)$ . Nach Satz 4.2 ist  $(f^{-1}(U_i))_{i \in I}$  eine offene Überdeckung von  $M$ , enthält also eine endliche Teilüberdeckung  $(f^{-1}(U_i))_{i \in E}$  ( $E \subseteq I$  endlich). Dann ist  $(U_i)_{i \in E}$  eine endliche Überdeckung von  $f(M)$ . ■

**5.5 Folgerung.** *Eine stetige Funktion  $f : M \rightarrow \mathbb{R}$  auf einem kompakten metrischen Raum  $M$  nimmt ein Maximum an.*

*Beweis.* Nach Satz 5.4 ist  $f(M)$  kompakt, also abgeschlossen und beschränkt. Jede abgeschlossene, beschränkte und nichtleere Menge reeller Zahlen enthält ein größtes Element. ■

**5.6 Satz.** *Jede stetige Abbildung  $f : M \rightarrow M'$  eines kompakten metrischen Raumes  $(M, d)$  in einen metrischen Raum  $(M', d')$  ist gleichmäßig stetig.*

*Beweis.* Sei  $M$  kompakt und  $f : M \rightarrow M'$  stetig. Sei  $\varepsilon \in \mathbb{R}^+$  gegeben. Zu jedem  $z \in M$  existiert, da  $f$  in  $z$  stetig ist, ein  $\delta(z) \in \mathbb{R}^+$  mit

$$d'(f(x), f(z)) < \frac{\varepsilon}{2} \quad \text{für alle } x \in M \text{ mit } d(x, z) < \delta(z).$$

Das System  $(U(z, \frac{1}{2}\delta(z)))_{z \in M}$  ist eine offene Überdeckung von  $M$ , enthält also eine endliche Teilüberdeckung  $(U(z_i, \frac{1}{2}\delta(z_i)))_{i=1, \dots, n}$ . Setze  $\delta := \min\{\frac{1}{2}\delta(z_i) \mid i = 1, \dots, n\}$ . Seien jetzt  $x, y \in M$  beliebige Punkte mit  $d(x, y) < \delta$ . Es gibt ein  $i \in \{1, \dots, n\}$  mit  $x \in U(z_i, \frac{1}{2}\delta(z_i))$ . Dann ist  $d(x, z_i) < \frac{1}{2}\delta(z_i) < \delta(z_i)$ , ferner  $d(y, z_i) \leq d(y, x) + d(x, z_i) < \delta + \frac{1}{2}\delta(z_i) \leq \delta(z_i)$ . Es folgt

$$d'(f(x), f(y)) \leq d'(f(x), f(z_i)) + d'(f(z_i), f(y)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Also ist  $f$  gleichmäßig stetig. ■

**5.7 Satz.** *Sei  $A \subseteq M$  kompakt. Dann besitzt jede Folge in  $A$  eine Teilfolge, die gegen einen Punkt von  $A$  konvergiert.*

*Beweis.* Sei  $A$  kompakt und  $(x_i)_{i \in \mathbb{N}}$  eine Folge in  $A$ . Für  $n \in \mathbb{N}$  setzen wir  $A_n := \{x_i \mid i \geq n\}$ . Wir zeigen zuerst, dass es ein Element  $z \in A$  gibt mit  $z \in \bigcap_{n \in \mathbb{N}} \overline{A_n}$ . Anderenfalls gäbe es für alle  $y \in A$  einen Index  $m \in \mathbb{N}$  mit  $y \notin \overline{A_m}$ , d.h.  $A \subseteq \bigcup_{n \in \mathbb{N}} (M \setminus \overline{A_n})$ . Also ist die Familie  $(M \setminus \overline{A_n})_{n \in \mathbb{N}}$  eine offene Überdeckung von  $A$ , enthält also eine endliche Teilüberdeckung von  $A$ . Es gibt daher ein  $m \in \mathbb{N}$  mit  $A \subseteq \bigcup_{n=1}^m (M \setminus \overline{A_n}) \subseteq \bigcup_{n=1}^m (M \setminus A_n)$ . Wegen  $x_m \in A$  und  $x_m \in A_n$  für  $n = 1, \dots, m$  ist das ein Widerspruch. Somit existiert ein Punkt  $z \in A \cap \bigcap_{n \in \mathbb{N}} \overline{A_n}$ . Wir definieren nun rekursiv eine Teilfolge von  $(x_i)_{i \in \mathbb{N}}$ . Da  $z$  Berührungspunkt von  $A_1 = \{x_i \mid i \geq 1\}$  ist, existiert ein  $i_1 \in \mathbb{N}$  mit  $x_{i_1} \in U(z, 1)$ . Seien  $i_1, \dots, i_{k-1}$  schon definiert. Da  $z$  Berührungspunkt von  $A_{i_{k-1}+1} = \{x_i \mid i \geq i_{k-1} + 1\}$  ist, existiert ein  $i_k > i_{k-1}$  mit  $x_{i_k} \in U(z, \frac{1}{k})$ . Damit ist eine Teilfolge  $(x_{i_k})_{k \in \mathbb{N}}$  von  $(x_i)_{i \in \mathbb{N}}$  definiert, die gegen den Punkt  $z \in A$  konvergiert. ■

Mengen eines metrischen Raumes mit der Eigenschaft aus Satz 5.7 nennt man *folgenkompakt*, d.h. wir haben gezeigt, dass in metrischen Räumen Überdeckungskompaktheit Folgenkompaktheit impliziert.

**5.8 Folgerung.** *Jeder kompakte metrische Raum ist vollständig.*

*Beweis.* Jede Cauchy-Folge in einem kompakten metrischen Raum besitzt nach Satz 5.7 eine Teilfolge, die gegen einen Punkt  $z$  konvergiert. Aus der Cauchy-Eigenschaft der Folge ergibt sich mit der Dreiecksungleichung, dass sie selbst gegen  $z$  konvergieren muß. ■



# 10 Der euklidische Raum

## 10.1 Der $n$ -dimensionale euklidische Vektorraum

Unser Arbeitsgebiet in den folgenden Kapiteln wird der  $n$ -dimensionale euklidische Raum sein. Diese Begriffsbildung ist aus der Linearen Algebra bekannt, und überhaupt werden Methoden der Linearen Algebra in diesem zweiten Teil der Analysis viel stärker benutzt als im ersten Teil. Wir wollen zunächst an einige Definitionen aus der Vorlesung Lineare Algebra I erinnern und einige für unsere Zwecke praktische Bezeichnungen festlegen. Sodann werden wir sehen, dass der euklidische Raum gegenüber den in Kapitel 9 betrachteten allgemeinen metrischen Räumen noch einige sehr spezielle metrische Eigenschaften hat. Da er zugleich ein endlichdimensionaler Vektorraum ist, ergeben sich aus der Wechselwirkung zwischen Metrik und Vektorraumstruktur zusätzliche Aussagen, die später ständig benutzt werden.

Im folgenden ist  $n$  eine natürliche Zahl. Unter dem  $n$ -dimensionalen euklidischen Raum werden wir einen reellen Vektorraum der Dimension  $n$  mit einem (positiv definiten) Skalarprodukt verstehen. Bekanntlich sind je zwei  $n$ -dimensionale reelle Vektorräume isomorph; wir können daher von *dem*  $n$ -dimensionalen reellen Vektorraum sprechen und den weiteren Betrachtungen einen bestimmten zugrundelegen. Hierfür wählen wir das  $n$ -fache kartesische Produkt

$$\mathbb{R}^n := \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{-mal}}$$

also die Menge aller geordneten  $n$ -Tupel reeller Zahlen, zusammen mit den durch

$$\begin{aligned}(x_1, \dots, x_n) + (y_1, \dots, y_n) &:= (x_1 + y_1, \dots, x_n + y_n), \\ \lambda(x_1, \dots, x_n) &:= (\lambda x_1, \dots, \lambda x_n) \quad \text{für } \lambda \in \mathbb{R}\end{aligned}$$

definierten Vektorraumoperationen. Den Nullvektor  $(0, \dots, 0)$  von  $\mathbb{R}^n$  bezeichnen wir auch mit  $0$  (also – wie üblich – mit demselben Symbol wie die reelle Zahl  $0$ ). Durch

$$E_1 := (1, 0, \dots, 0), \quad E_2 := (0, 1, 0, \dots, 0), \dots, \quad E_n := (0, \dots, 0, 1)$$

ist eine Basis  $(E_1, \dots, E_n)$  von  $\mathbb{R}^n$  gegeben; wir nennen sie die *Standardbasis* von  $\mathbb{R}^n$ . Für  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$  ist

$$X = x_1 E_1 + \dots + x_n E_n,$$

also  $x_i$  die  $i$ -te Koordinate des Vektors  $X$  bezüglich der Standardbasis; wir nennen  $x_i$  kurz die  $i$ -te *Koordinate* von  $X$ . Die Abbildung

$$p_i : \mathbb{R}^n \rightarrow \mathbb{R} : (x_1, \dots, x_n) \mapsto x_i,$$

die also jedem Vektor seine  $i$ -te Koordinate zuordnet, heißt  $i$ -te *Projektion* ( $i = 1, \dots, n$ ).

**Bemerkung.** In der Sprechweise machen wir keinen Unterschied zwischen „Punkten“ und „Vektoren“. Beides bedeutet hier also dasselbe, nämlich Elemente von  $\mathbb{R}^n$ . Wir verwenden beide Ausdrücke und sprechen von „Punkten“ meist dann, wenn die Vektorraumstruktur keine Rolle spielt.

Um auf dem Vektorraum  $\mathbb{R}^n$  Analysis betreiben zu können, brauchen wir (zum Beispiel) eine Metrik. Sie sollte mit der Vektorraumstruktur gekoppelt sein. Wir führen sie daher ein durch eine spezielle Norm, die durch ein Skalarprodukt induziert ist. An diese Begriffsbildungen soll zunächst in allgemeinerem Rahmen erinnert werden.

**1.1 Definition.** *Ein Skalarprodukt auf dem reellen Vektorraum  $V$  ist eine Abbildung*

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R} : (u, v) \mapsto \langle u, v \rangle$$

mit folgenden Eigenschaften:

$$(a) \quad \langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle, \quad \langle u, \lambda v + \mu w \rangle = \lambda \langle u, v \rangle + \mu \langle u, w \rangle,$$

$$(b) \quad \langle u, v \rangle = \langle v, u \rangle,$$

$$(c) \quad \langle u, u \rangle > 0, \text{ falls } u \neq 0$$

für alle  $u, v, w \in V$  und alle  $\lambda, \mu \in \mathbb{R}$ .

Ist  $\langle \cdot, \cdot \rangle$  ein Skalarprodukt auf  $V$ , so wird das Paar  $(V, \langle \cdot, \cdot \rangle)$  als euklidischer Vektorraum bezeichnet.

Sind  $(V, \langle \cdot, \cdot \rangle), (V', \langle \cdot, \cdot \rangle')$  zwei  $n$ -dimensionale reelle Vektorräume mit Skalarprodukt, so gibt es, wie man in der Linearen Algebra zeigt, einen Vektorraumisomorphismus  $f : V \rightarrow V'$  mit  $\langle f(u), f(v) \rangle' = \langle u, v \rangle$  für alle  $u, v \in V$ .

In diesem Sinne sind also je zwei euklidische Vektorräume gleicher (endlicher) Dimension isomorph. Alle Skalarprodukte auf einem  $n$ -dimensionalen reellen Vektorraum sind also im wesentlichen gleichwertig, und wir können insbesondere bei der Behandlung von  $\mathbb{R}^n$  ein spezielles wählen.

**1.2 Definition.** Für  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ ,  $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$  sei

$$\langle X, Y \rangle := x_1 y_1 + \dots + x_n y_n.$$

Dann ist  $\langle \cdot, \cdot \rangle$  ein Skalarprodukt auf  $\mathbb{R}^n$ ; es heißt das  $\cdot$ -Skalarprodukt.

dass in der Tat ein Skalarprodukt definiert wird, ist klar. Wir werden es den späteren Betrachtungen zugrundelegen. Unter dem  $n$ -dimensionalen euklidischen Vektorraum wird dann immer  $\mathbb{R}^n$  mit dem Standard-Skalarprodukt verstanden. Zunächst schließen sich aber an die Definition des Skalarprodukts noch einige Begriffsbildungen und Hilfssätze an, die wir ohne Mehraufwand im allgemeineren Rahmen behandeln können.

**1.3 Satz.** Ist  $\langle \cdot, \cdot \rangle$  ein Skalarprodukt auf dem reellen Vektorraum  $V$ , so gilt für alle  $u, v \in V$

$$\langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle$$

(Cauchy-Schwarzsche Ungleichung) und

$$\sqrt{\langle u+v, u+v \rangle} \leq \sqrt{\langle u, u \rangle} + \sqrt{\langle v, v \rangle}$$

(Minkowskische Ungleichung).

*Beweis.* Für  $\lambda \in \mathbb{R}$  und  $u, v \in V$  gilt nach Definition 1.1

$$\langle u, u \rangle + 2\lambda \langle u, v \rangle + \lambda^2 \langle v, v \rangle = \langle u + \lambda v, u + \lambda v \rangle \geq 0.$$

Ist  $\langle v, v \rangle \neq 0$ , so kann man

$$\lambda := -\frac{\langle u, v \rangle}{\langle v, v \rangle}$$

einsetzen und erhält die erste Ungleichung. Ist  $\langle v, v \rangle = 0$ , so ist  $v = 0$  und daher  $\langle u, v \rangle = 0$ ; die Cauchy-Schwarzsche Ungleichung gilt also trivialerweise. Die Minkowskische Ungleichung folgt aus der Cauchy-Schwarzschen:

$$\begin{aligned} \langle u+v, u+v \rangle &= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle \\ &\leq \langle u, u \rangle + 2\sqrt{\langle u, u \rangle \langle v, v \rangle} + \langle v, v \rangle \\ &= \left( \sqrt{\langle u, u \rangle} + \sqrt{\langle v, v \rangle} \right)^2. \quad \blacksquare \end{aligned}$$

**1.4 Lemma.** Sei  $(V, \langle \cdot, \cdot \rangle)$  ein euklidischer Vektorraum. Dann wird durch

$$\|u\| := \sqrt{\langle u, u \rangle} \quad \text{für } u \in V$$

auf  $V$  eine Norm erklärt. Sie heißt die durch das Skalarprodukt induzierte Norm.

*Beweis.* (Zum Begriff der Norm siehe Beispiele in Abschnitt 9.1.) Nach Definition 1.1 ist  $\langle u, u \rangle \geq 0$ , also ist  $\|u\| = \sqrt{\langle u, u \rangle}$  als reelle Zahl definiert. dass die Axiome für eine Norm erfüllt sind, folgt aus Definition 1.1 und der Minkowskischen Ungleichung, die gerade die Dreiecksungleichung für die induzierte Norm ist. ■

Gemäß der Beispiele in Abschnitt 9.1 wird damit jetzt durch

$$d(u, v) := \|u - v\| \quad \text{für } u, v \in V$$

auf  $V$  auch eine Metrik  $d$  gegeben. Im Falle des  $n$ -dimensionalen euklidischen Vektorraums ist

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

für  $X = (x_1, \dots, x_n)$ ,  $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$ . Wir bezeichnen diese Zahl als den *euklidischen Abstand* der Punkte  $X, Y$  und  $d$  auch als die *euklidische Metrik* auf  $\mathbb{R}^n$ .

Mit Hilfe des Skalarprodukts lassen sich weitere, anschaulich vertraute geometrische Grundbegriffe erklären:

**1.5 Definition.** Sei  $(V, \langle \cdot, \cdot \rangle)$  ein euklidischer Vektorraum. Die Vektoren  $u, v \in V$  heißen orthogonal (oder senkrecht), wenn  $\langle u, v \rangle = 0$  ist. Für  $u, v \in V \setminus \{0\}$  wird die durch

$$\frac{\langle u, v \rangle}{\|u\| \|v\|} = \cos \varphi, \quad 0 \leq \varphi \leq \pi,$$

definierte reelle Zahl  $\varphi$  als der Winkel zwischen  $u$  und  $v$  bezeichnet.

dass  $\varphi$  existiert, ist klar wegen

$$|\langle u, v \rangle| \leq \|u\| \|v\|,$$

was gerade die Cauchy-Schwarzsche Ungleichung in neuer Schreibweise ist. Die Eindeutigkeit von  $\varphi$  folgt aus der strengen Monotonie von  $\cos$  auf  $[0, \pi]$ .

**Bemerkung.** Für orthogonale  $u, v \in V$  gilt der „Satz des Pythagoras“:

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Wegen  $\langle u, v \rangle = 0$  folgt das aus

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle = \|u\|^2 + \|v\|^2.$$

Eine Basis eines euklidischen Vektorraumes, deren Elemente paarweise orthogonal und von der Norm 1 sind, heißt *orthonormiert*. Die Standardbasis des  $\mathbb{R}^n$  ist also orthonormiert (bezüglich des Standard-Skalarprodukts).

Von nun an legen wir speziell den  $n$ -dimensionalen euklidischen Raum zugrunde, also  $\mathbb{R}^n$  mit dem Standard-Skalarprodukt  $\langle \cdot, \cdot \rangle$ , der daraus abgeleiteten Norm  $\|\cdot\|$  und der hierdurch induzierten Metrik. Der euklidische Abstand zweier Punkte  $X, Y \in \mathbb{R}^n$  ist durch  $\|X - Y\|$  gegeben. Auf diesen metrischen Raum können wir nun alles anwenden, was wir in Kapitel 9 definiert und bewiesen haben. Durch die Koppelung der Metrik mit der Struktur eines endlichdimensionalen Vektorraumes ergeben sich einige Besonderheiten und Vereinfachungen, die wir jetzt zusammenstellen wollen.

**1.6 Lemma.** Für  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$  sei

$$\|X\|_{\max} := \max\{|x_1|, \dots, |x_n|\}.$$

Dann ist  $\|\cdot\|_{\max}$  eine Norm auf  $\mathbb{R}^n$ , und es gilt

$$\|X\|_{\max} \leq \|X\| \leq \sqrt{n}\|X\|_{\max}.$$

Der Beweis ist trivial.

Diese einfachen Ungleichungen sind ein bequemes Hilfsmittel, um Aussagen über  $\mathbb{R}^n$  auf dem Weg über die Koordinaten auf Aussagen über  $\mathbb{R}$  zurückzuführen. Die folgenden Sätze werden dies demonstrieren.

**1.7 Satz.** Die Menge  $\mathbb{Q}^n$  der Punkte mit rationalen Koordinaten ist dicht in  $\mathbb{R}^n$ .

*Beweis.* Sei  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$  gegeben. Sei  $\varepsilon \in \mathbb{R}^+$ . Da  $\mathbb{Q}$  dicht in  $\mathbb{R}$  ist, gibt es zu jedem  $i \in \{1, \dots, n\}$  eine Zahl  $y_i \in \mathbb{Q}$  mit

$$|x_i - y_i| < \frac{\varepsilon}{\sqrt{n}}.$$

Dann ist  $Y := (y_1, \dots, y_n) \in \mathbb{Q}^n$ , und nach Lemma 1.6 ist

$$\|X - Y\| \leq \sqrt{n}\|X - Y\|_{\max} < \varepsilon.$$

In jeder Umgebung eines beliebigen Punktes von  $\mathbb{R}^n$  liegen also Punkte aus  $\mathbb{Q}^n$ . ■

Der folgende Satz führt die Konvergenz von Punktfolgen im  $\mathbb{R}^n$  auf die Konvergenz der Koordinatenfolgen zurück; wir werden ihn häufig zu benutzen haben.

**1.8 Satz.** Sei  $(X_k)_{k \in \mathbb{N}}$  eine Folge in  $\mathbb{R}^n$ , sei  $X_k = (x_1^{(k)}, \dots, x_n^{(k)})$  für  $k \in \mathbb{N}$ , sei  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ . Dann gilt

$$\lim_{k \rightarrow \infty} X_k = X \Leftrightarrow \lim_{k \rightarrow \infty} x_i^{(k)} = x_i \quad \text{für } i = 1, \dots, n.$$

Ferner gilt

$$(X_k)_{k \in \mathbb{N}} \text{ ist Cauchy-Folge} \Leftrightarrow (x_i^{(k)})_{k \in \mathbb{N}} \text{ ist Cauchy-Folge für } i = 1, \dots, n.$$

*Beweis.* Gelte  $\lim_{k \rightarrow \infty} X_k = X$ . Sei  $i \in \{1, \dots, n\}$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Es gibt ein  $k_0 \in \mathbb{N}$  mit  $\|X - X_k\| < \varepsilon$  für  $k \geq k_0$ . Für alle  $k \geq k_0$  gilt also

$$|x_i - x_i^{(k)}| \leq \|X - X_k\|_{\max} \leq \|X - X_k\| < \varepsilon.$$

Damit ist  $\lim_{k \rightarrow \infty} x_i^{(k)} = x_i$  gezeigt.

Gelte  $\lim_{k \rightarrow \infty} x_i^{(k)} = x_i$  für  $i = 1, \dots, n$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Für  $i \in \{1, \dots, n\}$  gibt es ein  $k_i \in \mathbb{N}$  mit  $|x_i - x_i^{(k)}| < \varepsilon/\sqrt{n}$  für  $k \geq k_i$ . Für  $k \geq k_0 := \max\{k_1, \dots, k_n\}$  gilt dann

$$\|X - X_k\| \leq \sqrt{n} \|X - X_k\|_{\max} < \varepsilon.$$

Damit ist  $\lim_{k \rightarrow \infty} X_k = X$  gezeigt.

Die zweite Äquivalenz zeigt man völlig analog. ■

Als Anwendung kann man zeigen, dass die Vektorraumoperationen und das Skalarprodukt stetig sind. Wegen Satz 4.3 aus Kapitel 9 ist das äquivalent mit den folgenden Aussagen.

**1.9 Satz.** Für konvergente Folgen  $(X_k)_{k \in \mathbb{N}}$ ,  $(Y_k)_{k \in \mathbb{N}}$  in  $\mathbb{R}^n$  und  $(\lambda_k)_{k \in \mathbb{N}}$  in  $\mathbb{R}$  gilt

$$\begin{aligned} \lim_{k \rightarrow \infty} (X_k + Y_k) &= \lim_{k \rightarrow \infty} X_k + \lim_{k \rightarrow \infty} Y_k, \\ \lim_{k \rightarrow \infty} \lambda_k X_k &= \left( \lim_{k \rightarrow \infty} \lambda_k \right) \left( \lim_{k \rightarrow \infty} X_k \right), \\ \lim_{k \rightarrow \infty} \langle X_k, Y_k \rangle &= \left\langle \lim_{k \rightarrow \infty} X_k, \lim_{k \rightarrow \infty} Y_k \right\rangle. \end{aligned}$$

Der Beweis ergibt sich sofort aus Satz 1.8 und bekannten Aussagen über konvergente Folgen reeller Zahlen. Aus dem zweiten Teil von Satz 1.8 erhält man die folgende Aussage.

**1.10 Satz.**  $\mathbb{R}^n$  ist vollständig.

*Beweis.* Sei  $(X_k)_{k \in \mathbb{N}}$  mit  $X_k = (x_1^{(k)}, \dots, x_n^{(k)})$  eine Cauchy-Folge in  $\mathbb{R}^n$ . Für  $i \in \{1, \dots, n\}$  ist  $(x_i^{(k)})_{k \in \mathbb{N}}$  nach Satz 1.8 eine Cauchy-Folge in  $\mathbb{R}$ , wegen der Vollständigkeit von  $\mathbb{R}$  also konvergent gegen eine Zahl  $x_i \in \mathbb{R}$ . Mit  $X := (x_1, \dots, x_n)$  gilt  $\lim_{k \rightarrow \infty} X_k = X$  nach Satz 1.8. ■

Ebenso wichtig wie die durch Satz 1.10 ausgedrückte Tatsache, dass in  $\mathbb{R}^n$  das Cauchysche Konvergenzkriterium gilt, ist die Tatsache, dass der Satz von Bolzano-Weierstraß sich auf den  $\mathbb{R}^n$  übertragen läßt.

**1.11 Satz (Bolzano-Weierstraß).** Jede beschränkte Folge in  $\mathbb{R}^n$  besitzt eine konvergente Teilfolge.

*Beweis.* Sei  $(X_k)_{k \in \mathbb{N}}$  mit  $X_k = (x_1^{(k)}, \dots, x_n^{(k)})$  eine beschränkte Folge in  $\mathbb{R}^n$ . Wegen Lemma 1.6 ist die Folge  $(x_1^{(k)})_{k \in \mathbb{N}}$  beschränkt; sie besitzt also nach dem in  $\mathbb{R}$  gültigen Satz von Bolzano-Weierstraß eine gegen eine Zahl  $x_1 \in \mathbb{R}$  konvergierende Teilfolge  $(x_1^{(k_j)})_{j \in \mathbb{N}}$ . Die Folge  $(x_2^{(k_j)})_{j \in \mathbb{N}}$  ist ebenfalls beschränkt, sie besitzt also eine gegen ein  $x_2 \in \mathbb{R}$  konvergierende Teilfolge. So weiter schließend, erhält man nach endlich vielen Schritten eine streng monotone Folge  $(m_j)_{j \in \mathbb{N}}$  in  $\mathbb{N}$  und Zahlen  $x_1, \dots, x_n \in \mathbb{R}$  mit

$$\lim_{j \rightarrow \infty} x_i^{(m_j)} = x_i \quad \text{für } i = 1, \dots, n.$$

Mit  $X := (x_1, \dots, x_n)$  gilt also nach Satz 1.8

$$\lim_{j \rightarrow \infty} X_{m_j} = X. \quad \blacksquare$$

Hieraus können wir zum Beispiel eine Verallgemeinerung des aus Analysis I bekannten Intervallschachtelungsprinzips (Kapitel 2, Satz 2.3) herleiten.

**1.12 Satz.** Sei  $(A_i)_{i \in \mathbb{N}}$  eine Folge abgeschlossener, beschränkter, nichtleerer Teilmengen von  $\mathbb{R}^n$  mit  $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ . Dann ist  $\bigcap_{i \in \mathbb{N}} A_i \neq \emptyset$ .

*Beweis.* Da  $A_i \neq \emptyset$  ist, können wir ein  $X_i \in A_i$  auswählen ( $i \in \mathbb{N}$ ). Die Folge  $(X_i)_{i \in \mathbb{N}}$  ist (wegen  $X_i \in A_1$  und der Voraussetzung) beschränkt, besitzt also nach Satz 1.11 eine konvergente Teilfolge  $(X_{i_k})_{k \in \mathbb{N}}$ . Sei  $X$  ihr Limes. Sei  $i \in \mathbb{N}$ . Wegen  $X_{i_k} \in A_{i_k} \subseteq A_i$  für  $i_k \geq i$  ist  $X$  Berührungspunkt von  $A_i$ ; also ist  $X \in A_i$ . ■

Nun können wir auch zeigen, dass in  $\mathbb{R}^n$  der Überdeckungssatz von Heine-Borel gilt:

**1.13 Satz.** *In  $\mathbb{R}^n$  ist jede abgeschlossene, beschränkte Menge kompakt.*

*Beweis.* Der Beweis verläuft völlig analog zum Beweis von Satz 1.8, Kapitel 4 in Analysis I; dabei sind lediglich Intervalle  $[a_i, b_i]$  zu ersetzen durch Würfel  $[a_i^{(1)}, b_i^{(1)}] \times \cdots \times [a_i^{(n)}, b_i^{(n)}]$ . ■

Nach diesen wichtigen Beispielen für die Übertragbarkeit von Sätzen von  $\mathbb{R}$  auf  $\mathbb{R}^n$  kann man sich fragen, welche Rolle dabei die spezielle Wahl der euklidischen Norm  $\|\cdot\|$  spielt. Die Antwort lautet, dass man ebenso gut eine beliebige andere Norm hätte wählen können. Dies liegt daran, dass zwei beliebige Normen auf dem Vektorraum  $\mathbb{R}^n$  äquivalent sind in folgendem Sinne.

**1.14 Satz.** *Sind  $\|\cdot\|_1$  und  $\|\cdot\|_2$  zwei Normen auf  $\mathbb{R}^n$ , so gibt es Zahlen  $k, K \in \mathbb{R}^+$  mit*

$$k\|X\|_1 \leq \|X\|_2 \leq K\|X\|_1 \quad \text{für alle } X \in \mathbb{R}^n.$$

*Beweis.* Es genügt, die Ungleichungen für eine spezielle Norm  $\|\cdot\|_1$  zu beweisen; das allgemeine Resultat folgt dann nämlich durch mehrmalige Anwendung des spezielleren. Wir wählen hierzu die Norm  $\|\cdot\|_{\max}$  aus Lemma 1.6.

Sei  $(E_1, \dots, E_n)$  die Standardbasis des  $\mathbb{R}^n$  und

$$K := \|E_1\|_2 + \cdots + \|E_n\|_2.$$

Für  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$  gilt dann

$$\begin{aligned} \|X\|_2 &= \|x_1 E_1 + \cdots + x_n E_n\|_2 \\ &\leq |x_1| \|E_1\|_2 + \cdots + |x_n| \|E_n\|_2 \\ &\leq K \|X\|_{\max}. \end{aligned}$$

Damit ist schon die Existenz von  $K$  gezeigt.

Wir setzen nun  $f(X) := \|X\|_2$  für  $X \in \mathbb{R}^n$  und zeigen die Stetigkeit der Funktion  $f$ . Für  $X, Y \in \mathbb{R}^n$  gilt

$$\begin{aligned} \left| \|X\|_2 - \|Y\|_2 \right| &\leq \|X - Y\|_2 && \text{(Dreiecksungleichung)} \\ &\leq K \|X - Y\|_{\max} \\ &\leq K \|X - Y\| && \text{(nach Lemma 1.6)}. \end{aligned}$$

Also ist  $f$  stetig. Nun ist die Menge

$$A := \{X \in \mathbb{R}^n : \|X\|_{\max} = 1\}$$

beschränkt (denn für  $X \in A$  ist  $d(X, 0) = \|X\| \leq \sqrt{n}\|X\|_{\max} = \sqrt{n}$ ) und abgeschlossen (denn aus  $X_i \in A$  und  $X_i \rightarrow X$  folgt  $1 = \|X_i\|_{\max} \rightarrow \|X\|_{\max}$  nach Lemma 2.5 aus Kapitel 9, also  $\|X\|_{\max} = 1$ ). Nach Satz 1.13 ist  $A$  kompakt. Nach Folgerung 5.5 aus Kapitel 9 nimmt  $f$  auf  $A$  ein Minimum  $k \geq 0$  an. Wäre  $k = 0$ , so gäbe es ein  $X \in A$  mit  $\|X\|_2 = 0$ , also  $X = 0$ , ein Widerspruch. Also ist  $k > 0$ . Wir haben also  $\|X\|_2 \geq k > 0$  für alle  $X \in \mathbb{R}^n$  mit  $\|X\|_{\max} = 1$ . Für beliebiges  $X \in \mathbb{R}^n \setminus \{0\}$  gilt dann mit  $\lambda := 1/\|X\|_{\max}$  die Gleichung  $\|\lambda X\|_{\max} = 1$ , also  $|\lambda|\|X\|_2 = \|\lambda X\|_2 \geq k$ , folglich  $\|X\|_2 \geq k\|X\|_{\max}$ . Für  $X = 0$  gilt diese Ungleichung trivialerweise. Damit ist auch die Existenz von  $k$  gezeigt. ■

## 10.2 Abbildungen und Koordinatenfunktionen

Abbildungen zwischen euklidischen Räumen treten in der höherdimensionalen Analysis in verschiedenster Form auf. In diesem Abschnitt sollen einige grundlegende Erläuterungen zu allgemeinen Abbildungen  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  gegeben werden. Wir betrachten zunächst lineare Abbildungen und wiederholen einige einschlägige Begriffe aus der Linearen Algebra.

Eine Abbildung  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  heißt bekanntlich *linear*, wenn

$$F(\lambda X + \mu Y) = \lambda F(X) + \mu F(Y)$$

für alle  $X, Y \in \mathbb{R}^n$  und alle  $\lambda, \mu \in \mathbb{R}$  gilt. Lineare Abbildungen lassen sich durch Matrizen beschreiben. Wie vereinbart, wollen wir dabei immer die Standardbasen zugrundelegen. Sei also  $(E_1, \dots, E_n)$  die Standardbasis von  $\mathbb{R}^n$  und  $(E'_1, \dots, E'_k)$  diejenige von  $\mathbb{R}^k$ . Für jedes  $j \in \{1, \dots, n\}$  ist der Vektor  $F(E_j)$  eindeutig linear kombinierbar aus den Vektoren  $E'_1, \dots, E'_k$ , also gilt

$$F(E_j) = \sum_{i=1}^k a_{ij} E'_i \quad \text{für } j = 1, \dots, n$$

mit reellen Zahlen  $a_{ij}$ . Durch die  $k \times n$ -Matrix

$$(a_{ij})_{\substack{i=1, \dots, k \\ j=1, \dots, n}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kn} \end{pmatrix}$$

ist umgekehrt die Abbildung  $F$  festgelegt, denn für  $X = \sum_{j=1}^n x_j E_j \in \mathbb{R}^n$  ist

$$F(X) = \sum_{j=1}^n x_j F(E_j) = \sum_{j=1}^n x_j \sum_{i=1}^k a_{ij} E'_i = \sum_{i=1}^k \left( \sum_{j=1}^n a_{ij} x_j \right) E'_i.$$

Die  $i$ -te Koordinate des Bildvektors ist also gegeben durch

$$\sum_{j=1}^n a_{ij} x_j.$$

Die linearen Abbildungen  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  bilden in bekannter Weise einen reellen Vektorraum. Es ist zweckmäßig, hierauf eine Norm einzuführen. Wir nehmen dazu die euklidische Norm des (beliebig geordneten)  $(kn)$ -Tupels der Matrixeinträge, durch die  $F$  bezüglich der Standardbasen beschrieben wird.

**2.1 Lemma.** Sei  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  eine lineare Abbildung. Setze

$$\|F\| := \sqrt{\sum_{i=1}^k \sum_{j=1}^n a_{ij}^2},$$

wobei  $(a_{ij})_{\substack{i=1,\dots,k \\ j=1,\dots,n}}$  die der Abbildung  $F$  wie oben zugeordnete Matrix ist. Dann gilt

$$\|F(X)\| \leq \|F\| \|X\| \quad \text{für alle } X \in \mathbb{R}^n.$$

**Bemerkung.** Man beachte, dass hier ungenauerweise dasselbe Symbol  $\|\cdot\|$  für drei verschiedene Funktionen benutzt wird:  $\|F(X)\|$  ist die Norm von  $F(X)$  in  $\mathbb{R}^k$ ,  $\|X\|$  ist die Norm von  $X$  in  $\mathbb{R}^n$ , und  $\|F\|$  ist eine Norm auf dem Vektorraum  $\text{Hom}(\mathbb{R}^n, \mathbb{R}^k)$  der linearen Abbildungen von  $\mathbb{R}^n$  in  $\mathbb{R}^k$ . Da aber keine Gefahr von Mißverständnissen besteht, ist diese unpräzise Bezeichnungsweise durchaus üblich.

*Beweis (Lemma 2.1).* Für  $X = (x_1, \dots, x_n) \in \mathbb{R}^n$  gilt

$$F(X) = \sum_{i=1}^k \left( \sum_{j=1}^n a_{ij} x_j \right) E'_i,$$

also

$$\|F(X)\|^2 = \sum_{i=1}^k \left( \sum_{j=1}^n a_{ij} x_j \right)^2.$$

Nach der Cauchy-Schwarzschen Ungleichung (in einem beliebigen  $\mathbb{R}^m$ )  $\langle A, B \rangle^2 \leq \|A\|^2 \|B\|^2$  ist

$$\left(\sum a_j b_j\right)^2 \leq \left(\sum a_j^2\right) \left(\sum b_j^2\right),$$

also

$$\|F(X)\|^2 \leq \sum_{i=1}^k \left(\sum_{j=1}^n a_{ij}^2\right) \left(\sum_{j=1}^n x_j^2\right) = \|F\|^2 \|X\|^2. \quad \blacksquare$$

**2.2 Folgerung.** Ist  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  eine lineare Abbildung, so ist

$$\|F(X) - F(Y)\| \leq \|F\| \|X - Y\| \quad \text{für } X, Y \in \mathbb{R}^n,$$

also ist  $F$  eine Lipschitzabbildung (mit Lipschitz-Konstante  $\|F\|$ ) und somit insbesondere gleichmäßig stetig.

Nun betrachten wir Abbildungen in euklidische Räume, die nicht notwendig linear sind. Der Definitionsbereich darf dann zunächst auch von allgemeinerer Natur sein. Sei zunächst  $M$  eine beliebige nichtleere Menge und  $F : M \rightarrow \mathbb{R}^n$  eine Abbildung. Für jedes  $x \in M$  können wir dann den Bildvektor  $F(x)$  als Linearkombination der Basisvektoren  $E_1, \dots, E_n$  des  $\mathbb{R}^n$  darstellen. Bezeichnen wir die  $i$ -te Koordinate von  $F(x)$  mit  $f_i(x)$ , so gilt also

$$F(x) = \sum_{i=1}^n f_i(x) E_i \quad \text{für } x \in M.$$

Dadurch sind reellwertige Funktionen  $f_i : M \rightarrow \mathbb{R}$ ,  $i = 1, \dots, n$ , definiert. Wir bezeichnen  $f_i$  als die  $i$ -te *Koordinatenfunktion* der Abbildung  $F$ . Sie kann auch als Komposition

$$f_i = p_i \circ F$$

dargestellt werden; dabei ist

$$p_i : \quad \mathbb{R}^n \quad \rightarrow \mathbb{R} \\ (x_1, \dots, x_n) \mapsto x_i$$

die  $i$ -te Projektion. Allgemein läßt sich also die Untersuchung einer Abbildung in den  $\mathbb{R}^n$  zurückführen auf die Untersuchung von  $n$  reellwertigen Funktionen, was oft bequem ist. Hierfür ein erstes Beispiel:

**2.3 Satz.** Sei  $M$  ein metrischer Raum,  $F : M \rightarrow \mathbb{R}^n$  eine Abbildung und  $x \in M$ . Dann gilt:

$$F \text{ ist stetig in } x \Leftrightarrow p_i \circ F \text{ ist stetig in } x \text{ für } i = 1, \dots, n.$$

*Beweis.* Ist  $F$  stetig in  $x$ , so ist wegen der Stetigkeit der Projektion  $p_i$  auch  $p_i \circ F$  stetig in  $x$  ( $i = 1, \dots, n$ ). Seien umgekehrt alle Koordinatenfunktionen

$p_i \circ F =: f_i$  stetig in  $x$ . Sei  $\varepsilon \in \mathbb{R}^+$ . Für  $i \in \{1, \dots, n\}$  gibt es eine Umgebung  $U_i$  von  $x$  in  $M$  mit

$$|f_i(x) - f_i(y)| < \frac{\varepsilon}{\sqrt{n}} \quad \text{für alle } y \in U_i.$$

Dann ist  $U := U_1 \cap \dots \cap U_n$  eine Umgebung von  $x$ , und für alle  $y \in U$  gilt  $|f_i(x) - f_i(y)| < \varepsilon/\sqrt{n}$  für  $i = 1, \dots, n$ , also

$$\|F(y) - F(x)\| \leq \sqrt{n} \|F(y) - F(x)\|_{\max} < \varepsilon. \quad \blacksquare$$

**Bemerkung.** Analoge Aussagen gelten offenbar, wenn „stetig in  $x$ “ ersetzt wird durch „stetig“ oder „gleichmäßig stetig“.

Die Stetigkeit einer Abbildung in einem Punkt kann man natürlich ganz analog wie in Analysis I auch unter Verwendung eines Grenzwertbegriffs für Funktionen formulieren. Der folgende Konvergenzbegriff für Abbildungen zwischen euklidischen Räumen wird später häufig benutzt.

**2.4 Definition.** Sei  $M \subseteq \mathbb{R}^n$ ,  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung,  $X_0$  ein Häufungspunkt von  $M$  und  $Y \in \mathbb{R}^k$ . Dann wird definiert

$$\begin{aligned} \lim_{X \rightarrow X_0} F(X) = Y &: \Leftrightarrow \\ \forall \varepsilon \in \mathbb{R}^+ \exists \delta \in \mathbb{R}^+ \forall X \in M \setminus \{X_0\} : (\|X - X_0\| < \delta \Rightarrow \|F(X) - Y\| < \varepsilon). \end{aligned}$$

Man beachte, dass die Aussage

$$\lim_{X \rightarrow X_0} F(X) = Y$$

nicht erfordert, dass  $F$  auch an der Stelle  $X_0$  definiert ist.

Die Stetigkeitsdefinition kann jetzt also auch folgendermaßen umformuliert werden:

$$F \text{ stetig in } X_0 \Leftrightarrow \lim_{X \rightarrow X_0} F(X) = F(X_0).$$

Ebenso wie Konvergenz einer Folge in  $\mathbb{R}^n$  gleichbedeutend ist mit Konvergenz der entsprechenden Koordinatenfolgen, kann man Limesbeziehungen für Abbildungen zurückführen auf Limesbeziehungen für die Koordinatenfunktionen:

**2.5 Satz.** Sei  $M \subseteq \mathbb{R}^n$ ,  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung,  $X_0$  ein Häufungspunkt von  $M$ ,  $Y = (y_1, \dots, y_k) \in \mathbb{R}^k$  und  $f_i := p_i \circ F$ , also  $F(X) = (f_1(X), \dots, f_k(X))$  für  $X \in M$ . Dann gilt:

$$\lim_{X \rightarrow X_0} F(X) = Y \Leftrightarrow \lim_{X \rightarrow X_0} f_i(X) = y_i \quad \text{für } i = 1, \dots, k.$$

*Beweis.* In Analogie zum Beweis von Satz 1.8 (leichte Übung). ■

Die formalen Eigenschaften des Konvergenzbegriffes aus Definition 2.4 sind sinngemäß dieselben wie früher im eindimensionalen Fall. Wir nennen nur ein später häufig benutztes Beispiel: Aus

$$\lim_{X \rightarrow X_0} F(X) = Y \quad \text{und} \quad \lim_{X \rightarrow X_0} G(X) = Z$$

folgt

$$\lim_{X \rightarrow X_0} (F + G)(X) = Y + Z.$$



## 11 Differentiation

Wir beginnen mit einigen Vorbetrachtungen. In diesem Kapitel soll die Differentialrechnung für Funktionen von  $n$  reellen Veränderlichen behandelt werden. Unter einer reellen Funktion von  $n$  reellen Veränderlichen verstehen wir eine Abbildung  $f : M \rightarrow \mathbb{R}$  mit  $M \subseteq \mathbb{R}^n$ . Für  $X \in \mathbb{R}^n$ ,  $X = (x_1, \dots, x_n)$  schreibt man üblicherweise

$$f(X) = f(x_1, \dots, x_n).$$

Man muß sich dabei stets darüber klar sein, dass diese Schreibweise die Festlegung einer Basis des  $\mathbb{R}^n$  voraussetzt, auch wenn dies nicht immer explizit gesagt wird. In der Schreibweise  $f(X)$  ist dagegen nicht auf eine spezielle Basis Bezug genommen; daher nennt man diese Schreibweise auch „koordinatenunabhängig“ oder „invariant“. Wenn wir in  $\mathbb{R}^n$  eine andere Basis  $(\tilde{E}_1, \dots, \tilde{E}_n)$  einführen, so können wir natürlich auch

$$f(X) = f\left(\sum_{i=1}^n \tilde{x}_i \tilde{E}_i\right) =: g(\tilde{x}_1, \dots, \tilde{x}_n)$$

schreiben, aber die Zuordnung  $(\tilde{x}_1, \dots, \tilde{x}_n) \mapsto g(\tilde{x}_1, \dots, \tilde{x}_n)$  ist natürlich eine andere als die Zuordnung  $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$ . Aus diesen und anderen Gründen ist es zweckmäßig, so weit wie möglich koordinatenunabhängige Definitionen, Schreibweisen und Schlußweisen zu benutzen. Bei der analytischen Behandlung von Funktionen von mehreren Veränderlichen empfiehlt es sich also, in  $f(X)$  immer das Argument  $X$ , also einen Punkt des Raumes, als das Wesentliche anzusehen und seine Beschreibung durch Koordinaten nur als Hilfsmittel zu betrachten.

Diese Betrachtungsweise führt uns auch dazu, an den Anfang der Differentialrechnung in höheren Dimensionen nicht die partiellen Ableitungen zu stellen. Was hierunter zu verstehen ist, ist schnell gesagt. Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  eine Funktion und  $\bar{X} = (\bar{x}_1, \dots, \bar{x}_n) \in \mathbb{R}^n$ . Als  $i$ -te partielle Ableitung von  $f$  an der Stelle  $\bar{X}$  wird die Ableitung der Funktion

$$t \mapsto f(\bar{x}_1, \dots, \bar{x}_{i-1}, t, \bar{x}_{i+1}, \dots, \bar{x}_n)$$

an der Stelle  $t = \bar{x}_i$  bezeichnet, falls sie existiert. Die  $i$ -te partielle Ableitung oder partielle Ableitung nach der  $i$ -ten Veränderlichen ist also die gewöhnliche Ableitung der Funktion, die sich ergibt, wenn alle Veränderlichen außer der  $i$ -ten festgehalten werden. Partielle Ableitungen beziehen sich also immer auf eine Basis. Abgesehen von diesem Schönheitsfehler ist es auch aus anderen Gründen nicht zweckmäßig, den Begriff der Differenzierbarkeit auf die Existenz der partiellen Ableitungen zu gründen. Dies soll jetzt durch Beispiele erläutert werden.

Wir betrachten Funktionen auf  $\mathbb{R}^2$ . In diesem Fall werden wir die Koordinaten (bezüglich der Standardbasis) eines Vektors  $X$  meist mit  $x, y$  statt  $x_1, x_2$  bezeichnen. Sei nun  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definiert durch

$$f(x, y) := \begin{cases} \frac{xy}{x^2+y^2} & \text{für } (x, y) \neq (0, 0), \\ 0 & \text{für } (x, y) = (0, 0). \end{cases}$$

Die partiellen Ableitungen an der Stelle  $(0, 0)$ , bezeichnet mit

$$\frac{\partial f}{\partial x}(0, 0) \quad \text{und} \quad \frac{\partial f}{\partial y}(0, 0),$$

existieren, denn die Funktionen  $x \mapsto f(x, 0) = 0$  und  $y \mapsto f(0, y) = 0$  sind differenzierbar. Die Funktion  $f$  ist also an der Stelle  $(0, 0)$  (und auch an jeder anderen Stelle) partiell differenzierbar. Sie ist aber an der Stelle  $(0, 0)$  nicht stetig! In der Tat liegen in jeder Umgebung von  $(0, 0)$  Punkte  $(a, a)$  mit  $a \neq 0$ , und es ist  $f(a, a) = 1/2$ , während doch  $f(0, 0) = 0$  ist. Nun sind wir aus Analysis I gewöhnt, dass jede differenzierbare Funktion auch stetig ist. Hier haben wir aber eine partiell differenzierbare Funktion von zwei Veränderlichen, die nicht stetig ist. Dies weist darauf hin, dass man unter der Differenzierbarkeit einer Funktion von mehreren Veränderlichen etwas anderes verstehen sollte als partielle Differenzierbarkeit.

Nun bezieht sich partielle Differenzierbarkeit auf eine spezielle Basis. Man könnte daher von einer Funktion schärfer fordern, dass sie bezüglich jeder Basis partiell differenzierbar ist. Aber auch daraus würde nicht die Stetigkeit folgen. Dies wird belegt durch das Beispiel

$$f(x, y) := \begin{cases} \frac{x^2y}{x^4+y^2} & \text{für } (x, y) \neq (0, 0) \\ 0 & \text{für } (x, y) = (0, 0). \end{cases}$$

In jeder Umgebung von  $(0, 0)$  liegen Punkte  $(a, a^2)$  mit  $a \neq 0$ , und hierfür ist  $f(a, a^2) = \frac{1}{2}$ . Also ist  $f$  nicht stetig in  $(0, 0)$ . Andererseits sind die Funktionen

$$y \mapsto f(0, y) = 0$$

und

$$t \mapsto f(t, \alpha t) = \frac{\alpha t}{t^2 + \alpha^2}, \quad \alpha \in \mathbb{R}$$

überall differenzierbar. Mit anderen Worten: Für jede Gerade  $G$  in  $\mathbb{R}^2$  durch den Nullpunkt ist die Einschränkung von  $f$  auf  $G$  differenzierbar (genauer: für jeden Vektor  $E \in \mathbb{R}^2$  ist die Funktion  $t \mapsto f(tE)$  differenzierbar). Später werden wir hierfür sagen, dass alle Richtungsableitungen dieser Funktion existieren, insbesondere alle partiellen Ableitungen bei beliebiger Basis. Trotzdem ist die Funktion nicht stetig.

## 11.1 Differenzierbarkeit

Die Vorbetrachtungen machen klar, dass wir uns gut überlegen müssen, was eigentlich eine sinnvolle Definition der Differenzierbarkeit einer Funktion von mehreren Veränderlichen ist. Wir können uns hier durch den eindimensionalen Fall leiten lassen, müssen ihn aber inhaltlich richtig verstehen und dazu etwas uminterpretieren.

Der Grundgedanke der Differentialrechnung besteht darin, eine gegebene Abbildung von einem euklidischen Raum in einen anderen dadurch zu untersuchen, dass man die Abbildung in der Nähe eines betrachteten Punktes möglichst gut approximiert durch „möglichst einfache“ Abbildungen. Besonders einfach sind konstante Abbildungen und, nach diesem Trivialfall, lineare Abbildungen. Zusammensetzungen aus beiden nennt man affine Abbildungen. Eine Abbildung  $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$  heißt also *affin*, wenn

$$F(X) = L(X) + Z \quad \text{für } X \in \mathbb{R}^n$$

ist, wobei  $L : \mathbb{R}^n \rightarrow \mathbb{R}^k$  eine lineare Abbildung und  $Z \in \mathbb{R}^k$  ein fester Vektor ist.

Betrachten wir nun zunächst den Fall einer Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$ . Eine affine Abbildung  $g : \mathbb{R} \rightarrow \mathbb{R}$  ist von der Form

$$g(x) = cx + t \quad \text{für } x \in \mathbb{R}$$

mit Konstanten  $c$  und  $t$ . Wir wollen nun zu gegebenem  $x_0 \in \mathbb{R}$  eine affine Funktion  $g$  finden, die  $f$  bei  $x_0$  „möglichst gut“ approximiert. Die erste Forderung an  $g$  ist natürlich, dass  $g(x_0) = f(x_0)$  sein soll; dies wird erfüllt durch

$$g(x) = c(x - x_0) + f(x_0).$$

Schreiben wir wie üblich  $x = x_0 + h$ , so verschwindet also die Funktion

$$h \mapsto f(x_0 + h) - g(x_0 + h) = f(x_0 + h) - f(x_0) - ch$$

für  $h = 0$ . „Gute Approximation“ interpretieren wir nun so, dass die Differenz  $f(x_0 + h) - f(x_0) - ch$  für  $h \rightarrow 0$  schneller klein werden soll als  $h$ . Damit ist gemeint, dass sogar noch

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - ch}{h} = 0 \quad (1.1)$$

sein soll. Hiermit sind wir genau bei der Definition der Differenzierbarkeit aus Analysis I angelangt: (1.1) ist äquivalent damit, dass  $f$  bei  $x_0$  differenzierbar und dass  $f'(x_0) = c$  ist. Die obigen Überlegungen legen es aber jetzt nahe, nicht die Zahl  $c$ , sondern die durch sie definierte lineare Abbildung  $L : h \mapsto ch$  in den Mittelpunkt zu stellen, also zu sagen: Die Funktion  $f$  heißt differenzierbar in  $x_0$ , wenn es eine lineare Funktion  $L : \mathbb{R} \rightarrow \mathbb{R}$  gibt mit

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - L(h)}{h} = 0.$$

In dieser Form läßt sich die Definition sofort auf Abbildungen zwischen höherdimensionalen Räumen übertragen.

**1.2 Definition.** Sei  $M \subseteq \mathbb{R}^n$  offen, sei  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung, sei  $X_0 \in M$ . Die Abbildung  $F$  heißt differenzierbar in  $X_0$ , wenn es eine lineare Abbildung  $L : \mathbb{R}^n \rightarrow \mathbb{R}^k$  gibt mit

$$\lim_{H \rightarrow 0} \frac{F(X_0 + H) - F(X_0) - L(H)}{\|H\|} = 0.$$

Gibt es eine solche lineare Abbildung, so ist sie eindeutig bestimmt; sie heißt dann das Differential von  $F$  in  $X_0$  und wird mit  $DF_{X_0}$  bezeichnet.  $F$  heißt differenzierbar, wenn  $F$  differenzierbar in  $X$  ist für alle  $X \in M$ .

*Wohldefiniertheit von Definition 1.2.* Zu zeigen ist, dass es höchstens eine lineare Abbildung mit der genannten Eigenschaft gibt. Seien also  $L_1, L_2 : \mathbb{R}^n \rightarrow \mathbb{R}^k$  zwei lineare Abbildungen dieser Art. Dann gilt

$$\lim_{H \rightarrow 0} \frac{L_1(H) - L_2(H)}{\|H\|} = 0.$$

Für jeden Vektor  $X \in \mathbb{R}^n \setminus \{0\}$  folgt insbesondere

$$0 = \lim_{t \searrow 0} \frac{L_1(tX) - L_2(tX)}{\|tX\|} = \frac{L_1(X) - L_2(X)}{\|X\|},$$

also  $L_1(X) = L_2(X)$ ; somit ist  $L_1 = L_2$ . ■

**Bemerkung.** In der älteren Literatur wird Differenzierbarkeit im Sinne von Definition 1.2 oft als „vollständige“ oder „totale“ Differenzierbarkeit bezeichnet.

In Definition 1.2 dürfen  $n$  und  $k$  beliebige natürliche Zahlen sein. Ist eine dieser Zahlen gleich 1, so kann man sich die Bedeutung des Differentials gut veranschaulichen.

Sei zunächst  $k = 1$  und, der Einfachheit halber,  $n = 2$ . In diesem Fall empfiehlt es sich, von der Funktion  $f : M \rightarrow \mathbb{R}$  (mit  $M \subseteq \mathbb{R}^2$ ) den Graphen

$$\text{Graph } f := \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in M, z = f(x, y)\}$$

zu betrachten. Sei  $X_0 \in M$  und  $Df_{X_0}$  das Differential von  $f$  in  $X_0$ . Dann ist also  $Df_{X_0} : \mathbb{R}^2 \rightarrow \mathbb{R}$  eine lineare Funktion. Differenzierbarkeit von  $f$  in  $X_0$  kann auch so formuliert werden, dass

$$f(X) = f(X_0) + Df_{X_0}(X - X_0) + R(X_0, X)$$

mit

$$\lim_{X \rightarrow X_0} \frac{R(X_0, X)}{\|X - X_0\|} = 0 \quad (1.3)$$

gilt. Die durch

$$g(X) := f(X_0) + Df_{X_0}(X - X_0), \quad X \in \mathbb{R}^2,$$

definierte Funktion  $g$  ist affin, und ihr Graph

$$\text{Graph } g = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \mathbb{R}^2, z = g(x, y)\}$$

ist eine Ebene durch den Punkt  $(X_0, f(X_0))$ . Die Beziehung (1.3), also

$$\lim_{X \rightarrow X_0} \frac{f(X) - g(X)}{\|X - X_0\|} = 0$$

besagt, dass der Graph von  $f$  bei Annäherung von  $X$  an  $X_0$  sich dem Graphen von  $g$  besonders gut anschmiegt. Der Graph von  $g$  wird daher auch als *Tangentialebene* des Graphen von  $f$  im Punkt  $(X_0, f(X_0))$  bezeichnet. Differenzierbarkeit einer reellen Funktion kann also auch als Existenz einer Tangentialebene an den Graphen interpretiert werden. Das Differential beschreibt die Stellung dieser Tangentialebene.

Betrachten wir jetzt den Fall  $n = 1$  und  $k \geq 2$ . Die Menge  $M \subseteq \mathbb{R}$  wollen wir etwa als Intervall annehmen. Die Abbildung  $F : M \rightarrow \mathbb{R}^k$  bezeichnet man als eine *parametrisierte Kurve*. (Eine physikalische Interpretation im Fall  $k = 3$  wäre etwa, dass  $F$  die Bahn eines Massepunktes in Abhängigkeit von der Zeit beschreibt.) Sei  $t_0 \in M$  und  $F$  differenzierbar in  $t_0$ . Das Differential  $DF_{t_0} : \mathbb{R} \rightarrow \mathbb{R}^k$  ist eine lineare Abbildung von  $\mathbb{R}$  in  $\mathbb{R}^k$ , also gilt  $DF_{t_0}(h) = hDF_{t_0}(1)$ . Wir setzen

$$DF_{t_0}(1) =: F'(t_0);$$

das ist also ein Vektor in  $\mathbb{R}^k$ . Die Limesgleichung in der Definition 1.2 der Differenzierbarkeit lautet dann

$$\lim_{h \rightarrow 0} \frac{F(t_0 + h) - F(t_0) - F'(t_0)h}{|h|} = 0,$$

was äquivalent ist mit

$$\lim_{h \rightarrow 0} \frac{F(t_0 + h) - F(t_0)}{h} = F'(t_0).$$

Damit ist eine anschauliche Deutung des Vektors  $F'(t_0)$  gefunden. Wir bezeichnen ihn als *Ableitung* der Abbildung  $F$  oder als *Tangentenvektor* der parametrisierten Kurve  $F$  in  $t_0$ . Ist  $f_i$  die  $i$ -te Koordinatenfunktion von  $F$ , also

$$F(t) = (f_1(t), \dots, f_k(t)) \quad \text{für } t \in M,$$

so ist wegen Satz 2.5 aus Kapitel 10

$$F'(t_0) = (f'_1(t_0), \dots, f'_k(t_0)).$$

Man erhält den Ableitungs- oder Tangentenvektor also durch koordinatenweises Differenzieren.

Wir kehren zum allgemeinen Fall zurück und wollen zunächst einige grundlegende Aussagen über differenzierbare Abbildungen zusammenstellen. Danach untersuchen wir insbesondere reellwertige Funktionen, die wir anschliessend in Gestalt von Koordinatenfunktionen wieder für die Behandlung allgemeiner differenzierbarer Abbildungen nutzbar machen.

Die folgenden Sätze zeigen, dass der durch Definition 1.2 eingeführte Differenzierbarkeitsbegriff, anders als die Forderung der partiellen Differenzierbarkeit, die aus Analysis I bekannten Implikationen von Differenzierbarkeit auszudehnen gestattet.

**1.4 Satz.** *Ist  $F : M \rightarrow \mathbb{R}^k$  in  $X_0$  differenzierbar, so ist  $F$  in  $X_0$  stetig.*

*Beweis.* Sei  $F$  in  $X_0$  differenzierbar. Zu gegebenem  $\varepsilon \in \mathbb{R}^+$  existiert dann ein  $\delta_1 \in \mathbb{R}^+$  mit

$$\frac{\|F(X_0 + H) - F(X_0) - DF_{X_0}(H)\|}{\|H\|} < \frac{\varepsilon}{2}$$

für alle  $H$  mit  $X_0 + H \in M$  und  $0 < \|H\| < \delta_1$ . Da lineare Abbildungen gleichmäßig stetig sind, existiert ein  $\delta_2 \in \mathbb{R}^+$  mit  $\|DF_{X_0}(H)\| < \varepsilon/2$  für alle  $H \in \mathbb{R}^n$  mit  $\|H\| < \delta_2$ . Für alle  $X_0 + H \in M$  mit  $\|H\| < \delta := \min\{\delta_1, \delta_2, 1\}$  gilt dann

$$\begin{aligned}
& \|F(X_0 + H) - F(X_0)\| \\
& \leq \|F(X_0 + H) - F(X_0) - DF_{X_0}(H)\| + \|DF_{X_0}(H)\| \\
& \leq \frac{\varepsilon}{2} \|H\| + \frac{\varepsilon}{2} \\
& \leq \varepsilon. \quad \blacksquare
\end{aligned}$$

**1.5 Satz.** Sind die Abbildungen  $F : M \rightarrow \mathbb{R}^k$  und  $G : M \rightarrow \mathbb{R}^k$  differenzierbar in  $X_0$ , so auch  $F + G$ , und es gilt

$$D(F + G)_{X_0} = DF_{X_0} + DG_{X_0}.$$

*Beweis.* Die Behauptung ergibt sich aus

$$\begin{aligned}
& (F + G)(X_0 + H) - (F + G)(X_0) - (DF_{X_0} + DG_{X_0})(H) \\
& = [F(X_0 + H) - F(X_0) - DF_{X_0}(H)] + [G(X_0 + H) - G(X_0) - DG_{X_0}(H)]
\end{aligned}$$

nach Division durch  $\|H\|$  und Limesbildung  $H \rightarrow 0$ .  $\blacksquare$

Die wichtige Kettenregel besagt, dass die Komposition differenzierbarer Abbildungen differenzierbar ist und dass das Differential der Komposition die Komposition der Differentiale ist.

**1.6 Satz (Kettenregel).** Seien  $M \subseteq \mathbb{R}^n$  und  $N \subseteq \mathbb{R}^k$  offen, seien  $F : M \rightarrow N$  und  $G : N \rightarrow \mathbb{R}^m$  Abbildungen. Sei  $X_0 \in M$ , sei  $F$  differenzierbar in  $X_0$  und  $G$  differenzierbar in  $F(X_0)$ . Dann ist  $G \circ F$  differenzierbar in  $X_0$ , und es gilt

$$D(G \circ F)_{X_0} = DG_{F(X_0)} \circ DF_{X_0}.$$

*Beweis.* Wir setzen  $F(X_0) =: Y_0$  und  $F(X_0 + H) - F(X_0) = Z_H$  für  $H \in \mathbb{R}^n \setminus \{0\}$  und  $X_0 + H \in M$ . Für diese  $H$  gilt

$$\begin{aligned}
& \frac{1}{\|H\|} [(G \circ F)(X_0 + H) - (G \circ F)(X_0) - DG_{F(X_0)} \circ DF_{X_0}(H)] \\
& = \frac{1}{\|H\|} [G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)] \\
& \quad + \frac{1}{\|H\|} [DG_{Y_0}(F(X_0 + H) - F(X_0)) - DG_{Y_0}(DF_{X_0}(H))] \\
& = \frac{1}{\|H\|} [G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)] \\
& \quad + DG_{Y_0} \left( \frac{1}{\|H\|} (F(X_0 + H) - F(X_0) - DF_{X_0}(H)) \right),
\end{aligned}$$

wobei zuletzt benutzt wurde, dass das Differential eine lineare Abbildung ist. Wegen der Differenzierbarkeit von  $F$  in  $X_0$  und der Stetigkeit der linearen Abbildung  $DG_{Y_0}$  gilt

$$\lim_{H \rightarrow 0} DG_{Y_0} \left( \frac{1}{\|H\|} (F(X_0 + H) - F(X_0) - DF_{X_0}(H)) \right) = 0.$$

Für den ersten Summanden haben wir

$$\begin{aligned} & \frac{1}{\|H\|} [G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)] \\ &= \begin{cases} 0, & \text{falls } Z_H = 0, \\ \frac{\|Z_H\|}{\|H\|} \frac{1}{\|Z_H\|} [G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)] & \text{sonst.} \end{cases} \end{aligned}$$

Aus der Differenzierbarkeit von  $G$  in  $Y_0$ , der Stetigkeit von  $F$  in  $X_0$  und der Beschränktheit von  $\|Z_H\|/\|H\|$  in einer Umgebung von 0 (die aus der Differenzierbarkeit von  $F$  in  $X_0$  folgt) ergibt sich jetzt

$$\lim_{H \rightarrow 0} \frac{1}{\|H\|} [G(Y_0 + Z_H) - G(Y_0) - DG_{Y_0}(Z_H)] = 0.$$

Daraus folgt die Behauptung.  $\blacksquare$

Bei konkreten Berechnungen wird man im allgemeinen die Kettenregel in Koordinatenschreibweise benutzen; hierauf kommen wir später zurück.

Den Fall  $n = 1$  der Kettenregel (Satz 1.6), der häufig vorkommen wird, schreiben wir noch mit Verwendung des Ableitungsvektors.

**1.7 Folgerung.** *Seien  $I \subseteq \mathbb{R}$  und  $N \subseteq \mathbb{R}^k$  offen, seien  $F : I \rightarrow N$  und  $G : N \rightarrow \mathbb{R}^m$  Abbildungen. Ist  $F$  differenzierbar in  $t_0 \in I$  und  $G$  differenzierbar in  $F(t_0)$ , so gilt*

$$(G \circ F)'(t_0) = DG_{F(t_0)}(F'(t_0)).$$

*Beweis.* Nach Satz 1.6 ist

$$D(G \circ F)_{t_0} = DG_{F(t_0)} \circ DF_{t_0};$$

für  $h \in \mathbb{R}$  folgt also

$$\begin{aligned} h(G \circ F)'(t_0) &= D(G \circ F)_{t_0}(h) = DG_{F(t_0)}(DF_{t_0}(h)) \\ &= DG_{F(t_0)}(hF'(t_0)) = hDG_{F(t_0)}(F'(t_0)). \end{aligned} \quad \blacksquare$$

Einer der nützlichsten Sätze aus Analysis I über differenzierbare Funktionen ist der Mittelwertsatz (Kapitel 6, Satz 2.2). Er besagt, dass man für eine differenzierbare Funktion  $f : [x, y] \rightarrow \mathbb{R}$  die Differenz  $f(y) - f(x)$  darstellen kann durch

$$f(y) - f(x) = f'(z)(y - x),$$

wobei  $z \in (x, y)$  eine passende Zwischenstelle ist. Ausgenutzt wird dies meist in folgender Weise. Weiß man, dass für alle  $z \in (x, y)$  die Abschätzung  $|f'(z)| \leq c$  mit einer Konstanten  $c$  gilt, so folgt

$$|f(y) - f(x)| \leq c|y - x|.$$

Mit anderen Worten, die Differenz der Funktionswerte bei  $x$  und  $y$  ist klein, wenn im ganzen Intervall der Betrag der Ableitung klein ist. In dieser Form läßt sich der Mittelwertsatz auf differenzierbare Abbildungen  $F$  von  $\mathbb{R}^n$  in den  $\mathbb{R}^k$  übertragen, nicht jedoch in der vorherigen Form einer Gleichung. Qualitativ besagt diese Verallgemeinerung des Mittelwertsatzes, dass der Abstand  $\|F(Y) - F(X)\|$  nicht groß ist, wenn das Differential von  $F$  längs der Verbindungsstrecke von  $X$  und  $Y$  nicht groß ist. Die Größe des Differentials wird dabei im Sinne der in Lemma 2.1 aus Kapitel 10 eingeführten Norm linearer Abbildungen gemessen. Mit

$$[X, Y] := \{(1 - \lambda)X + \lambda Y \mid 0 \leq \lambda \leq 1\}$$

bezeichnen wir die Verbindungsstrecke der Punkte  $X, Y \in \mathbb{R}^n$ .

**1.8 Satz.** *Sei  $M \subseteq \mathbb{R}^n$  offen, seien  $X, Y \in M$  Punkte mit  $[X, Y] \subseteq M$ . Die Abbildung  $F : M \rightarrow \mathbb{R}^k$  sei stetig in  $M$  und differenzierbar in  $(1 - \lambda)X + \lambda Y$  für alle  $\lambda \in (0, 1)$ . Gilt*

$$\|DF_Z\| \leq c \quad \text{für } Z = (1 - \lambda)X + \lambda Y \text{ mit } \lambda \in (0, 1),$$

mit einer Konstanten  $c$ , so gilt

$$\|F(Y) - F(X)\| \leq c\|Y - X\|.$$

*Beweis.* Zunächst behandeln wir einen Spezialfall. Sei  $G : [0, 1] \rightarrow \mathbb{R}^k$  eine stetige Abbildung derart, dass die Einschränkung  $G|_{(0,1)}$  differenzierbar ist und dass

$$\|G'(t)\| \leq c \quad \text{für alle } t \in (0, 1)$$

gilt. Wähle  $\varepsilon \in \mathbb{R}^+$  und setze

$$A := \{t \in [0, 1] \mid \|G(t) - G(0)\| \leq (c + \varepsilon)t + \varepsilon\}.$$

Da  $G$  in 0 stetig ist, enthält  $A$  jedenfalls ein Intervall  $[0, \tau]$  mit  $\tau > 0$ . Setze  $s := \sup A$ ; dann ist also  $0 < s \leq 1$ . Da  $G$  in  $s$  stetig ist, gilt

$$\|G(s) - G(0)\| \leq (c + \varepsilon)s + \varepsilon,$$

also  $s \in A$ . Angenommen, es wäre  $s < 1$ . Da  $G$  in  $s$  differenzierbar ist, gibt es ein  $h > 0$  mit  $s + h \leq 1$  und

$$\left\| \frac{G(s+h) - G(s)}{h} - G'(s) \right\| \leq \varepsilon,$$

also mit

$$\left\| \frac{G(s+h) - G(s)}{h} \right\| \leq \|G'(s)\| + \varepsilon \leq c + \varepsilon.$$

Es folgt

$$\begin{aligned} \|G(s+h) - G(0)\| &\leq \|G(s+h) - G(s)\| + \|G(s) - G(0)\| \\ &\leq (c + \varepsilon)h + (c + \varepsilon)s + \varepsilon \\ &= (c + \varepsilon)(s+h) + \varepsilon, \end{aligned}$$

also  $s+h \in A$ , im Widerspruch zur Definition von  $s$ . Damit ist  $s = 1$  bewiesen; es gilt also

$$\|G(1) - G(0)\| \leq c + 2\varepsilon.$$

Da  $\varepsilon \in \mathbb{R}^+$  beliebig war, folgt

$$\|G(1) - G(0)\| \leq c.$$

Nun sei  $F$  wie in Satz 1.8 vorausgesetzt. Wir definieren  $K : [0, 1] \rightarrow \mathbb{R}^n$  durch

$$K(t) := (1-t)X + tY \quad \text{für } t \in [0, 1].$$

Trivialerweise ist  $K$  differenzierbar und  $K'(t) = Y - X$ . Nach der Kettenregel (Satz 1.6) ist  $F \circ K$  differenzierbar in  $t \in (0, 1)$ , und es gilt

$$D(F \circ K)_t = DF_{K(t)} \circ DK_t,$$

also

$$(F \circ K)'(t) = DF_{K(t)}(K'(t)) = DF_{K(t)}(Y - X).$$

Nach Lemma 2.1 aus Kapitel 10 folgt

$$\|(F \circ K)'(t)\| = \|DF_{K(t)}(Y - X)\| \leq \|DF_{K(t)}\| \|Y - X\| \leq c \|Y - X\|.$$

Nach dem bereits Bewiesenen (mit  $G := F \circ K$  und  $c\|Y - X\|$  statt  $c$ ) folgt

$$\|F(Y) - F(X)\| = \|(F \circ K)(1) - (F \circ K)(0)\| \leq c\|Y - X\|. \quad \blacksquare$$

**1.9 Folgerung.** Sei  $M \subseteq \mathbb{R}^n$  offen und zusammenhängend, sei  $F : M \rightarrow \mathbb{R}^k$  eine differenzierbare Abbildung mit  $DF_X = 0$  für alle  $X \in M$ . Dann ist  $F$  konstant.

*Beweis.* Seien  $X, Y \in M$ . Da  $M$  zusammenhängend und offen ist, gibt es Punkte  $X_1, \dots, X_k$  mit  $X_1 = X$ ,  $X_k = Y$  und  $[X_i, X_{i+1}] \subseteq M$  für  $i = 1, \dots, k-1$ . (Denn für festes  $X \in M$  ist die Menge  $A_1$  aller  $Y \in M$ , die derart mit  $X$  durch einen Streckenzug in  $M$  verbindbar sind, offen und nicht leer. Die Menge  $A_2 := M \setminus A_1$  ist ebenfalls offen. Da  $A_1 \cup A_2 = M$ ,  $A_1 \cap A_2 = \emptyset$  und  $M$  zusammenhängend ist, muß  $A_2 = \emptyset$  sein, also  $M = A_1$ .) Für  $i \in \{1, \dots, k-1\}$  gilt nach Satz 1.8 die Gleichung  $F(X_{i+1}) - F(X_i) = 0$ . Hieraus folgt  $F(Y) = F(X)$ . Da  $X, Y \in M$  beliebig waren, folgt die Behauptung. ■

## 11.2 Partielle Ableitungen und Koordinatenfunktionen

Bei der Behandlung konkreter differenzierbarer Abbildungen wird man zweckmäßigerweise ihre Koordinatenfunktionen heranziehen. Differentiale und Kettenregel treten dann in Matrizen Schreibweise auf. Diese Umformulierungen wollen wir in diesem Abschnitt vornehmen. Dazu müssen wir zunächst partielle Ableitungen reellwertiger Funktionen betrachten.

Im ersten Teil des Folgenden sei stets  $M \subset \mathbb{R}^n$  eine offene Menge,  $X_0 \in M$  und  $f : M \rightarrow \mathbb{R}$  eine reelle Funktion.

Für reellwertige Funktionen hängt der Begriff des Differentials eng zusammen mit dem der Richtungsableitung.

**2.1 Definition.** Sei  $E \in \mathbb{R}^n \setminus \{0\}$  ein Vektor. Wenn die Funktion

$$t \mapsto f(X_0 + tE)$$

in 0 differenzierbar ist, so wird ihre Ableitung mit  $f'(X_0; E)$  bezeichnet, also

$$f'(X_0; E) := \lim_{h \rightarrow 0} \frac{f(X_0 + hE) - f(X_0)}{h}.$$

$f'(X_0; E)$  heißt Richtungsableitung von  $f$  bezüglich  $E$  im Punkt  $X_0$ .

**2.2 Satz.** Ist  $f$  in  $X_0$  differenzierbar, so existieren alle Richtungsableitungen von  $f$  in  $X_0$ , und es gilt

$$f'(X_0; E) = Df_{X_0}(E) \quad \text{für } E \in \mathbb{R}^n.$$

*Beweis.* Definieren wir  $g(t) := f(X_0 + tE)$  für alle  $t \in \mathbb{R}$  mit  $X_0 + tE \in M$ , so ist  $g$  nach der Kettenregel (Folgerung 1.7) in 0 differenzierbar, und es gilt  $g'(0) = Df_{X_0}(E)$ . ■

Die Richtungsableitungen in den Richtungen, die durch die Vektoren  $E_1, \dots, E_n$  der Standardbasis des  $\mathbb{R}^n$  gegeben sind, treten besonders häufig auf. Hierfür hat man daher besondere Bezeichnungen:

**2.3 Definition.** Für  $i \in \{1, \dots, n\}$  schreibt man

$$f'(X_0; E_i) =: \partial_i f(X_0),$$

falls diese Ableitung existiert.  $\partial_i f(X_0)$  heißt  $i$ -te partielle Ableitung von  $f$  in  $X_0$ .

Eine häufig verwendete Schreibweise ist auch

$$\partial_i f = \frac{\partial f}{\partial x_i}.$$

Wie partielle Ableitungen zu berechnen sind, ist klar: Für  $X_0 = (x_1^0, \dots, x_n^0)$  gilt

$$\begin{aligned} \partial_i f(X_0) &= f'(X_0; E_i) = \lim_{h \rightarrow 0} \frac{f(X_0 + hE_i) - f(X_0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f(x_1^0, \dots, x_{i-1}^0, x_i^0 + h, x_{i+1}^0, \dots, x_n^0) - f(x_1^0, \dots, x_n^0)}{h}. \end{aligned}$$

Das ist die gewöhnliche Ableitung der Funktion

$$t \mapsto f(x_1^0, \dots, x_{i-1}^0, t, x_{i+1}^0, \dots, x_n^0)$$

an der Stelle  $x_i^0$ . Hieraus erklärt sich die Bezeichnung „partielle Ableitung“, und es ergibt sich eine einfache Berechnungsvorschrift.

Durch die partiellen Ableitungen läßt sich nun auch die Richtungsableitung allgemein und damit das Differential einer differenzierbaren Funktion leicht ausdrücken. Für  $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$  gilt wegen der Linearität des Differentials

$$\begin{aligned} f'(X_0; Y) &= Df_{X_0}(y_1 E_1 + \dots + y_n E_n) \\ &= y_1 Df_{X_0}(E_1) + \dots + y_n Df_{X_0}(E_n) \\ &= y_1 \partial_1 f(X_0) + \dots + y_n \partial_n f(X_0) \\ &= \langle Y, (\partial_1 f(X_0), \dots, \partial_n f(X_0)) \rangle. \end{aligned}$$

dass man, wie hier, ein lineares Funktional auf dem  $\mathbb{R}^n$  als Skalarprodukt mit einem festen Vektor darstellen kann, ist aus der Linearen Algebra wohlbekannt. Für den in diesem Fall auftretenden Vektor hat man eine besondere Bezeichnung.

**2.4 Definition.** Existieren für  $f : M \rightarrow \mathbb{R}$  (mit  $M \subseteq \mathbb{R}^n$ ) die partiellen Ableitungen in  $X_0$ , so heißt der Vektor

$$\nabla f(X_0) := (\partial_1 f(X_0), \dots, \partial_n f(X_0))$$

der Gradient von  $f$  in  $X_0$ . Ist  $f$  in  $X_0$  differenzierbar, so gilt

$$f'(X_0; E) = Df_{X_0}(E) = \langle \nabla f(X_0), E \rangle \quad \text{für alle } E \in \mathbb{R}^n.$$

Man beachte, dass die Darstellung

$$\nabla f(X_0) = (\partial_1 f(X_0), \dots, \partial_n f(X_0))$$

von einer speziellen Basis des  $\mathbb{R}^n$  Gebrauch macht, dass der Gradient aber, wenn  $f$  in  $X_0$  differenzierbar ist, nicht von der gewählten Basis abhängt. Der Gradient  $\nabla f(X_0)$  ist ja durch die basisunabhängige Gleichung

$$\langle \nabla f(X_0), E \rangle = f'(X_0; E) \quad \text{für } E \in \mathbb{R}^n$$

festgelegt.

Der Gradient hat eine einfache anschauliche Bedeutung: Nach der Cauchy-Schwarzschen Ungleichung (Satz 1.3 aus Kapitel 10) gilt für  $\|E\| = 1$

$$f'(X_0; E) = \langle \nabla f(X_0), E \rangle \leq \|\nabla f(X_0)\|,$$

und das Gleichheitszeichen gilt (wie am Beweis von Satz 1.3 aus Kapitel 10 abzulesen ist) genau dann, wenn

$$\nabla f(X_0) = \lambda E \quad \text{mit einem } \lambda \geq 0$$

ist. Wir halten dies als Satz fest.

**2.5 Satz.** *Sei  $f$  in  $X_0$  differenzierbar und  $\nabla f(X_0) \neq 0$ . Dann nimmt die Richtungsableitung von  $f$  in  $X_0$  bezüglich Einheitsvektoren  $E$  genau für*

$$E = \frac{\nabla f(X_0)}{\|\nabla f(X_0)\|}$$

*ihr Maximum an; dieses Maximum ist gleich  $\|\nabla f(X_0)\|$ .*

Mit anderen Worten: Der Gradient weist in die Richtung stärksten Anstiegs der Funktion; seine Länge gibt die Größe dieses Anstiegs an.

Für differenzierbare reellwertige Funktionen gilt der Mittelwertsatz in Form einer Gleichung; er läßt sich bequem mit Hilfe des Gradienten formulieren.

**2.6 Satz** (Mittelwertsatz). *Seien  $X, Y \in M$  Punkte mit  $[X, Y] \subseteq M$ . Sei  $f : M \rightarrow \mathbb{R}$  stetig in  $M$  und differenzierbar in  $(1-t)X + tY$  für  $t \in (0, 1)$ . Dann gibt es eine Zahl  $\vartheta \in (0, 1)$  mit*

$$f(Y) - f(X) = \langle \nabla f((1-\vartheta)X + \vartheta Y), Y - X \rangle.$$

*Beweis.* Setze  $g(t) := f((1-t)X + tY)$  für  $0 \leq t \leq 1$ . Dann ist  $g$  stetig und nach der Kettenregel differenzierbar in  $(0, 1)$ . Nach dem Mittelwertsatz aus Analysis I (Kapitel 6, Satz 2.2) existiert daher eine Zahl  $\vartheta \in (0, 1)$  mit

$$\begin{aligned} f(Y) - f(X) &= g(1) - g(0) = g'(\vartheta) \\ &= \langle \nabla f((1-\vartheta)X + \vartheta Y), Y - X \rangle. \quad \blacksquare \end{aligned}$$

Wir wollen nun auf den grundsätzlichen Zusammenhang zwischen Differenzierbarkeit und partieller Differenzierbarkeit eingehen. Dabei heißt eine reellwertige Funktion *partiell differenzierbar*, wenn ihre partiellen Ableitungen existieren. Für eine partiell differenzierbare Funktion  $f : M \rightarrow \mathbb{R}$  ist also für jedes  $i \in \{1, \dots, n\}$  die Funktion

$$\partial_i f : M \rightarrow \mathbb{R} : X \mapsto \partial_i f(X)$$

erklärt, die man als *i-te partielle Ableitung* von  $f$  bezeichnet. Aus Differenzierbarkeit folgt partielle Differenzierbarkeit, nach Satz 2.2 sogar die Existenz aller Richtungsableitungen. Umgekehrt folgt aus der Existenz aller Richtungsableitungen nach einem Beispiel vom Anfang dieses Kapitels nicht einmal die Stetigkeit, also gewiß nicht die Differenzierbarkeit. Andererseits ist die partielle Differenzierbarkeit oft leicht nachprüfbar, und partielle Ableitungen sind in konkreten Fällen gut zu berechnen. Es sind daher Aussagen von Nutzen, die aus partieller Differenzierbarkeit und zusätzlichen Aussagen auf Differenzierbarkeit schließen lassen.

**2.7 Satz.** *Ist  $f : M \rightarrow \mathbb{R}$  in einer Umgebung von  $X_0$  partiell differenzierbar und sind die partiellen Ableitungen in  $X_0$  stetig, so ist  $f$  in  $X_0$  differenzierbar.*

*Beweis.* Sei  $U$  eine offene Kugel mit Mittelpunkt  $X_0 = (x_1, \dots, x_n)$ , in der  $f$  partiell differenzierbar ist. Für  $H = (h_1, \dots, h_n) \in \mathbb{R}^n$  mit  $X_0 + H \in U$  gilt nach dem Mittelwertsatz der Differentialrechnung

$$\begin{aligned} f(X_0 + H) - f(X_0) &= \sum_{i=1}^n [f(x_1, \dots, x_{i-1}, x_i + h_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) \\ &\quad - f(x_1, \dots, x_{i-1}, x_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n)] \\ &= \sum_{i=1}^n h_i \partial_i f(x_1, \dots, x_{i-1}, x_i + c_i h_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) \end{aligned}$$

mit geeigneten  $c_i \in (0, 1)$ , also

$$\begin{aligned} &\frac{1}{\|H\|} |f(X_0 + H) - f(X_0) - \langle \nabla f(X_0), H \rangle| \\ &= \frac{1}{\|H\|} \left| \sum_{i=1}^n h_i [\partial_i f(x_1, \dots, x_{i-1}, x_i + c_i h_i, \dots, x_n + h_n) - \partial_i f(X_0)] \right| \\ &\leq \sum_{i=1}^n |\partial_i f(x_1, \dots, x_{i-1}, x_i + c_i h_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) - \partial_i f(X_0)|. \end{aligned}$$

Da nach Voraussetzung jede partielle Ableitung  $\partial_i f$  in  $X_0$  stetig ist, ist der Limes der rechten Seite für  $\|H\| \rightarrow 0$  gleich 0. Daraus folgt die Behauptung. ■

Nach dieser Behandlung reellwertiger Funktionen kehren wir nun zu allgemeinen differenzierbaren Abbildungen in den  $\mathbb{R}^k$  zurück. Wir verwenden jetzt Koordinatenfunktionen und deren partielle Ableitungen.

Gegeben seien im folgenden, wenn nichts anderes gesagt ist, eine offene Menge  $M \subseteq \mathbb{R}^n$ , ein Punkt  $X_0 \in M$  und eine Abbildung  $F : M \rightarrow \mathbb{R}^k$ . Für  $i = 1, \dots, k$  sei  $f_i : M \rightarrow \mathbb{R}$  die  $i$ -te Koordinatenfunktion von  $F$  bezüglich der Standardbasis  $E'_1, \dots, E'_k$  des  $\mathbb{R}^k$ , also

$$F(X) = (f_1(X), \dots, f_k(X)) = \sum_{i=1}^k f_i(X) E'_i \quad \text{für } X \in M.$$

Mit  $X = (x_1, \dots, x_n)$ ,  $Y = (y_1, \dots, y_k)$  lautet also die Gleichung  $Y = F(X)$  in Koordinatenschreibweise

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_n), \\ &\vdots \\ y_k &= f_k(x_1, \dots, x_n). \end{aligned}$$

Differenzierbarkeit einer Abbildung ist, analog wie bei der Stetigkeit, gleichwertig mit Differenzierbarkeit der Koordinatenfunktionen.

**2.8 Satz.** *Die Abbildung  $F$  ist genau dann differenzierbar in  $X_0$ , wenn alle Koordinatenfunktionen  $f_i$  in  $X_0$  differenzierbar sind ( $i = 1, \dots, k$ ). Ist das der Fall, so gilt*

$$DF_{X_0}(H) = \sum_{i=1}^k (Df_i)_{X_0}(H) E'_i \quad \text{für } H \in \mathbb{R}^n.$$

*Beweis.* Sei  $L : \mathbb{R}^n \rightarrow \mathbb{R}^k$  eine lineare Abbildung. Nach Satz 2.5 aus Kapitel 10 gilt

$$\begin{aligned} \lim_{H \rightarrow 0} \frac{F(X_0 + H) - F(X_0) - L(H)}{\|H\|} &= 0 \\ \Leftrightarrow \lim_{H \rightarrow 0} \frac{f_i(X_0 + H) - f_i(X_0) - (p_i \circ L)(H)}{\|H\|} &= 0 \quad \text{für } i = 1, \dots, k. \end{aligned}$$

Daraus folgen die Behauptungen unmittelbar. ■

An der Gleichung in Satz 2.8 können wir nun sofort ablesen, durch welche Matrix das Differential von  $F$  in  $X_0$  bezüglich der Standardbasen beschrieben wird. Sei  $E_1, \dots, E_n$  die Standardbasis von  $\mathbb{R}^n$ . Ist  $L : \mathbb{R}^n \rightarrow \mathbb{R}^k$  eine lineare

Abbildung, so ist die ihr bezüglich der Standardbasen zugeordnete Matrix  $(a_{ij})_{\substack{i=1,\dots,k \\ j=1,\dots,n}}$  definiert durch

$$L(E_j) = \sum_{i=1}^k a_{ij} E'_i, \quad j = 1, \dots, n.$$

Setzen wir nun in der Gleichung in Satz 2.8 speziell  $H = E_j$  ein, so erhalten wir wegen  $DF_{X_0}(E_j) = \partial_j f(X_0)$  die Gleichungen

$$DF_{X_0}(E_j) = \sum_{i=1}^k \partial_j f_i(X_0) E'_i,$$

also  $a_{ij} = \partial_j f_i(X_0)$ . Wir halten das fest und definieren:

**2.9 Definition.** Sei  $F : M \rightarrow \mathbb{R}^k$  mit  $M \subseteq \mathbb{R}^n$  differenzierbar in  $X_0 \in M$ . Dann wird das Differential  $DF_{X_0}$  bezüglich der Standardbasen in  $\mathbb{R}^n$  und  $\mathbb{R}^k$  beschrieben durch die  $k \times n$ -Matrix

$$JF(X_0) := \begin{pmatrix} \partial_1 f_1(X_0) & \cdots & \partial_n f_1(X_0) \\ \vdots & & \vdots \\ \partial_1 f_k(X_0) & \cdots & \partial_n f_k(X_0) \end{pmatrix}.$$

Sie heißt die Funktionalmatrix oder Jacobische Matrix der Abbildung  $F$  an der Stelle  $X_0$ . Im Fall  $k = n$  wird die Determinante  $\det JF(X_0)$  der Funktionalmatrix als Funktionaldeterminante von  $F$  in  $X_0$  bezeichnet.

**Bemerkung.** Für die Funktionaldeterminante von  $F$  (im Fall  $n = k$ ) findet man auch oft die Bezeichnung

$$\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}.$$

Beschreibt man lineare Abbildungen (nach Einführung von Basen in den beteiligten Räumen) durch Matrizen, so wird einer Komposition von Abbildungen als Matrix bekanntlich das Matrizenprodukt der Matrizen der einzelnen Abbildungen zugeordnet. Mit der Bezeichnung aus Definition 2.9 lautet die Kettenregel aus Satz 1.6 daher jetzt folgendermaßen: Die Funktionalmatrix einer Komposition ist gleich dem Matrizenprodukt (in der richtigen Reihenfolge) der Funktionalmatrizen der einzelnen Abbildungen.

**2.10 Satz** (Kettenregel in Koordinatenschreibweise). Seien  $M \subseteq \mathbb{R}^n$  und  $N \subseteq \mathbb{R}^k$  offen, seien  $F : M \rightarrow N$  und  $G : N \rightarrow \mathbb{R}^m$  Abbildungen, sei  $F$  differenzierbar in  $X_0$  und  $G$  differenzierbar in  $F(X_0)$ . Dann gilt

$$J(G \circ F)(X_0) = JG(F(X_0)) \cdot JF(X_0)$$

(Matrizenprodukt), also mit  $G \circ F =: H$

$$\partial_j h_i(X_0) = \sum_{r=1}^k \partial_r g_i(F(X_0)) \partial_j f_r(X_0)$$

für  $i = 1, \dots, m$  und  $j = 1, \dots, n$ .

Etwas übersichtlicher ist vielleicht die Schreibweise

$$\frac{\partial h_i}{\partial x_j}(X_0) = \sum_{r=1}^k \frac{\partial g_i}{\partial y_r}(Y_0) \frac{\partial f_r}{\partial x_j}(X_0), \quad Y_0 = F(X_0).$$

Im Fall  $m = n = 1$  reduziert sich die Kettenregel in Koordinatenschreibweise auf eine einzige Gleichung, die wir für

$$h(t) = g(f_1(t), \dots, f_k(t))$$

in der Form

$$\frac{dh}{dt} = \frac{\partial g}{\partial x_1} \frac{df_1}{dt} + \dots + \frac{\partial g}{\partial x_k} \frac{df_k}{dt}$$

schreiben können. Hat hier  $df_i/dt$  (und damit auch  $dh/dt$ ) das Argument  $t$ , so muß  $\partial g/\partial x_i$  das Argument  $(f_1(t), \dots, f_k(t))$  haben.

### 11.3 Höhere Ableitungen, Taylorformel, lokale Extrema

Im folgenden seien stets eine offene Menge  $M \subseteq \mathbb{R}^n$ , ein Punkt  $X_0 \in M$  und eine reellwertige Funktion  $f : M \rightarrow \mathbb{R}$  gegeben.

In diesem Abschnitt betrachten wir höhere partielle Ableitungen reellwertiger Funktionen. Es liegt auf der Hand, wie sie zu definieren sind. Existiert die  $i$ -te partielle Ableitung  $\partial_i f$  von  $f$  in einer Umgebung von  $X_0$  und existiert die  $k$ -te partielle Ableitung von  $\partial_i f$  in  $X_0$ , so wird sie mit

$$\partial_k \partial_i f(X_0) \quad \text{oder} \quad \frac{\partial^2 f}{\partial x_k \partial x_i}(X_0)$$

bezeichnet. Entsprechend wird allgemein die partielle Ableitung

$$\partial_{i_r} \dots \partial_{i_1} f = \frac{\partial^r f}{\partial x_{i_r} \dots \partial x_{i_1}}$$

rekursiv definiert. Man bezeichnet sie als *partielle Ableitung  $r$ -ter Ordnung*. Existieren alle partiellen Ableitungen  $r$ -ter Ordnung in  $X_0$ , so heißt  $f$   *$r$ -mal partiell differenzierbar* in  $X_0$ , und wenn dies für alle  $X_0 \in M$  gilt, heißt  $f$   *$r$ -mal partiell differenzierbar*. Ist  $f$   $r$ -mal partiell differenzierbar und sind alle partiellen Ableitungen  $r$ -ter Ordnung der Funktion  $f$  stetig, so heißt  $f$   *$r$ -mal stetig differenzierbar* (im Fall  $r = 1$  kurz *stetig differenzierbar*). Die Menge der auf  $M$   $r$ -mal stetig differenzierbaren reellen Funktionen wird mit  $C^r(M)$  bezeichnet. Unter  $C^0(M)$  wird die Menge der stetigen reellen Funktionen auf  $M$  verstanden. Ein Element von  $C^r(M)$  heißt auch *Funktion der Klasse  $C^r$*  auf  $M$ . Es gilt also

$$C^0(M) \supset C^1(M) \supset C^2(M) \supset \dots,$$

und zwar jeweils mit strikter Inklusion.

**Bemerkung.** Wird mehrmals nach derselben Veränderlichen differenziert, so verwendet man abkürzende Schreibweisen wie z.B.

$$\frac{\partial^2 f}{\partial x \partial x} =: \frac{\partial^2 f}{\partial x^2},$$

$$\underbrace{\partial_i \cdots \partial_i}_{s\text{-mal}} \underbrace{\partial_k \cdots \partial_k}_{r\text{-mal}} f = \frac{\partial^{r+s} f}{\partial x_i^s \partial x_k^r}.$$

Das folgende Beispiel zeigt, dass es ohne zusätzliche Voraussetzungen nicht gleichgültig ist, in welcher Reihenfolge die partiellen Differentiationen ausgeführt werden.

**Beispiel.** Sei  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  erklärt durch

$$f(x, y) := \begin{cases} xy \frac{x^2 - y^2}{x^2 + y^2} & \text{für } (x, y) \neq (0, 0), \\ 0 & \text{für } (x, y) = (0, 0). \end{cases}$$

Die partiellen Ableitungen zweiter Ordnung existieren, und man erhält

$$\frac{\partial^2 f}{\partial y \partial x}(0, 0) = -1, \quad \frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1.$$

Derartiges kann jedoch nicht passieren, wenn die partiellen Ableitungen zweiter Ordnung noch stetig sind, wie der folgende Satz zeigt.

**3.1 Satz** (Vertauschbarkeit der Differentiationsreihenfolge). *Ist  $r \geq 2$  und  $f \in C^r(M)$ , so sind die partiellen Ableitungen von  $f$  bis zur  $r$ -ten Ordnung unabhängig von der Reihenfolge der Differentiationen, d.h. es gilt*

$$\partial_{i_1} \cdots \partial_{i_r} f = \partial_{i_{\sigma(1)}} \cdots \partial_{i_{\sigma(r)}} f$$

für jede Permutation  $\sigma$  der Zahlen  $1, \dots, r$ .

*Beweis.* Da jede Permutation der Zahlen  $1, \dots, r$  Produkt von solchen Permutationen ist, bei denen nur zwei benachbarte Zahlen vertauscht werden, genügt es, den Beweis für den Fall  $r = 2$  zu führen. Außerdem bedeutet es keine Einschränkung der Allgemeinheit,  $n = 2$  anzunehmen, weil bei den in Betracht kommenden Differentiationen alle bis auf zwei Veränderliche festgehalten werden.

Sei  $X = (x_1, x_2) \in M$  und  $U \subseteq M$  eine offene Kugel mit Mittelpunkt  $X$ . Im folgenden sei  $H = (h_1, h_2)$  so klein, dass  $X + H \in U$  ist. Dann gilt

$$\begin{aligned} & f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2) \\ &= \varphi(x_1 + h_1) - \varphi(x_1) \quad \text{mit } \varphi(t) := f(t, x_2 + h_2) - f(t, x_2) \\ &= \varphi'(x_1 + c_1 h_1) h_1 \quad \text{mit } c_1 \in (0, 1) \text{ (Mittelwertsatz)} \\ &= [\partial_1 f(x_1 + c_1 h_1, x_2 + h_2) - \partial_1 f(x_1 + c_1 h_1, x_2)] h_1 \\ &= [\psi(x_2 + h_2) - \psi(x_2)] h_1 \quad \text{mit } \psi(t) := \partial_1 f(x_1 + c_1 h_1, t) \\ &= \psi'(x_2 + c_2 h_2) h_1 h_2 \quad \text{mit } c_2 \in (0, 1) \text{ (Mittelwertsatz)} \\ &= \partial_2 \partial_1 f(x_1 + c_1 h_1, x_2 + c_2 h_2) h_1 h_2. \end{aligned}$$

Analog findet man

$$\begin{aligned} & f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2) \\ &= \partial_1 \partial_2 f(x_1 + d_1 h_1, x_2 + d_2 h_2) h_1 h_2 \end{aligned}$$

mit passenden  $d_1, d_2 \in (0, 1)$ . Für  $h_1 h_2 \neq 0$  folgt also

$$\partial_2 \partial_1 f(x_1 + c_1 h_1, x_2 + c_2 h_2) = \partial_1 \partial_2 f(x_1 + d_1 h_1, x_2 + d_2 h_2).$$

Man beachte, dass  $c_i$  und  $d_i$  hier von  $h_i$  abhängen ( $i = 1, 2$ ), da der Mittelwertsatz auf dem Intervall  $(x_1, x_1 + h_1)$  bzw.  $(x_2, x_2 + h_2)$  angewandt wurde. Es gilt jedoch  $c_i, d_i \in (0, 1)$  und daher mit  $h_i \rightarrow 0$  auch  $c_i h_i \rightarrow 0$  und  $d_i h_i \rightarrow 0$ . Wegen der vorausgesetzten Stetigkeit der partiellen Ableitungen zweiter Ordnung ergibt sich mit  $(h_1, h_2) \rightarrow (0, 0)$  die Behauptung

$$\partial_2 \partial_1 f(x_1, x_2) = \partial_1 \partial_2 f(x_1, x_2). \quad \blacksquare$$

Einer der wichtigsten Sätze aus Analysis I über mehrfach differenzierbare Funktionen war die Taylorformel (Kapitel 6, Satz 3.3). Wir wollen nun untersuchen, was sie für Funktionen von  $n$  Veränderlichen liefert, wenn man diese auf Geraden durch einen betrachteten Punkt einschränkt. Zuerst erinnern wir uns an die Taylorformel aus Analysis I; dabei können wir uns auf einen Spezialfall beschränken. Sei  $g : [-\varepsilon, h] \rightarrow \mathbb{R}$  mit  $\varepsilon, h > 0$  eine  $(k + 1)$ -mal differenzierbare Funktion ( $k \in \mathbb{N}$ ). Dann gilt

$$g(h) = \sum_{j=0}^k \frac{1}{j!} g^{(j)}(0) h^j + \frac{1}{(k+1)!} g^{(k+1)}(c) h^{k+1}$$

mit einer passenden Zwischenstelle  $c \in (0, h)$ .

Wir betrachten nun eine Funktion  $f : M \rightarrow \mathbb{R}$  mit offenem Definitionsbereich  $M \subseteq \mathbb{R}^n$  und einen Punkt  $X_0 \in M$  und machen die folgende Voraussetzung.

**Voraussetzung.** Sei  $k \in \mathbb{N}$ ,  $f \in C^k(M)$ , die partiellen Ableitungen  $k$ -ter Ordnung von  $f$  sind in  $M$  differenzierbar,  $H \in \mathbb{R}^n$  ist ein Vektor mit  $[X_0, X_0 + H] \subseteq M$ .

Nun setzen wir

$$g(t) := f(X_0 + tH) \quad \text{für } t \in [-\varepsilon, 1]$$

wobei  $\varepsilon > 0$  so gewählt sei, dass  $[X_0, X_0 - \varepsilon H] \subseteq M$  ist. Da  $f$  differenzierbar ist (denn die partiellen Ableitungen sind noch differenzierbar, also stetig, somit ist  $f$  nach Satz 2.7 differenzierbar), ist  $g$  nach der Kettenregel differenzierbar, und nach Satz 2.10 gilt

$$g'(t) = \sum_{i=1}^n \partial_i f(X_0 + tH) h_i,$$

wenn  $H = (h_1, \dots, h_n)$  ist. Nach Voraussetzung sind die partiellen Ableitungen  $\partial_i f$  ( $i = 1, \dots, n$ ) noch differenzierbare Funktionen, daher können wir abermals die Kettenregel anwenden und erhalten

$$g''(t) = \sum_{i=1}^n \sum_{j=1}^n \partial_j \partial_i f(X_0 + tH) h_i h_j.$$

So fortfahrend, erhalten wir schließlich

$$g^{(r)}(t) = \sum_{i_1, \dots, i_r=1}^n \partial_{i_1} \cdots \partial_{i_r} f(X_0 + tH) h_{i_1} \cdots h_{i_r}$$

für  $r = 1, \dots, k+1$ . Auf  $g$  können wir die oben zitierte eindimensionale Taylorformel anwenden (mit  $h = 1$ ). Es gibt also eine Zahl  $c \in (0, 1)$  mit

$$g(1) = \sum_{r=0}^k \frac{1}{r!} g^{(r)}(0) + \frac{1}{(k+1)!} g^{(k+1)}(c).$$

Setzen wir hier die oben berechneten Ableitungen von  $g$  ein, so erhalten wir den folgenden Satz.

**3.2 Satz (Taylorformel).** Sei  $M \subseteq \mathbb{R}^n$  offen,  $X_0 \in M$ ,  $H \in \mathbb{R}^n$  mit  $[X_0, X_0 + H] \subseteq M$ ,  $k \in \mathbb{N}$ ,  $f \in C^k(M)$ , und die partiellen Ableitungen  $k$ -ter Ordnung von  $f$  seien in  $M$  differenzierbar. Dann gibt es eine Zahl  $c \in (0, 1)$  mit

$$\begin{aligned}
 f(X_0 + H) &= f(X_0) + \sum_{r=1}^k \frac{1}{r!} \sum_{i_1, \dots, i_r=1}^n \partial_{i_1} \cdots \partial_{i_r} f(X_0) h_{i_1} \cdots h_{i_r} \\
 &\quad + \frac{1}{(k+1)!} \sum_{i_1, \dots, i_{k+1}=1}^n \partial_{i_1} \cdots \partial_{i_{k+1}} f(X_0 + cH) h_{i_1} \cdots h_{i_{k+1}}.
 \end{aligned}$$

Speziell für  $k = 2$  ist also

$$f(X_0 + H) = f(X_0) + \langle \nabla f(X_0), H \rangle + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(X_0) h_i h_j + R(X_0; H)$$

mit

$$R(X_0; H) = \frac{1}{6} \sum_{i,j,m=1}^n \partial_i \partial_j \partial_m f(Y) h_i h_j h_m$$

und passendem  $Y \in [X_0, X_0 + H]$ . Wegen  $|h_i| \leq \|H\|$  folgt

$$\lim_{H \rightarrow 0} \frac{R(X_0; H)}{\|H\|^2} = 0,$$

falls  $\partial_i \partial_j \partial_m f$  auf  $M$  beschränkt ist. Analog zu Analysis I, Kapitel 6, Satz 3.4 (und daraus herleitbar) gilt diese Schlußfolgerung auch schon unter der Voraussetzung, dass  $f \in C^1(M)$  ist und die partiellen Ableitungen von  $f$  in  $X_0$  differenzierbar sind.

Man kann den durch

$$\begin{aligned}
 f(X_0 + H) &= f(X_0) + \langle \nabla f(X_0), H \rangle + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(X_0) h_i h_j + R(X_0; H), \\
 \lim_{H \rightarrow 0} \frac{R(X_0; H)}{\|H\|^2} &= 0
 \end{aligned}$$

ausgedrückten Sachverhalt auch folgendermaßen beschreiben: In zweiter Näherung verhält sich die Funktion  $f$  an der Stelle  $X_0$  wie ein Polynom zweiten Grades. Anschaulich läßt sich folgendes sagen. Betrachten wir den Graphen von  $f$ , so gibt der erste Term  $f(X_0)$  der Taylorformel die Höhe des Graphen über  $X_0$  an, der zweite Term, der durch den Gradienten von  $f$  in  $X_0$  bestimmt ist, legt die Tangentialebene an den Graphen über  $X_0$  fest, und der nächste Term,

$$\frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(X_0) h_i h_j,$$

gibt eine erste Information über die Gestalt des Graphen bei  $X_0$ . Dies zeigt sich insbesondere bei der Untersuchung lokaler Extremwerte. Zuvor wollen wir für diesen Term eine besondere Bezeichnung einführen.

**3.3 Definition.** Sei  $f : M \rightarrow \mathbb{R}$  zweimal partiell differenzierbar in  $X_0$ . Dann heißt die durch

$$Q(f, X_0; H) := \sum_{i,j=1}^n \partial_i \partial_j f(X_0) h_i h_j \quad \text{für } H = (h_1, \dots, h_n) \in \mathbb{R}^n$$

definierte Funktion  $Q(f, X_0; \cdot)$  die Hesse-Form von  $f$  in  $X_0$ , und die Matrix

$$\text{Hess}(f)_{X_0} := \begin{pmatrix} \partial_1 \partial_1 f(X_0) & \cdots & \partial_1 \partial_n f(X_0) \\ \vdots & & \vdots \\ \partial_n \partial_1 f(X_0) & \cdots & \partial_n \partial_n f(X_0) \end{pmatrix}$$

heißt die Hessesche Matrix von  $f$  in  $X_0$ .

Falls  $f$  in einer Umgebung von  $X_0$  zweimal stetig differenzierbar ist, dann ist nach Satz 3.1 die Hessesche Matrix von  $f$  symmetrisch. Die Hesse-Form ist also eine quadratische Form, die Hessesche Matrix ist die ihr (bezüglich der Standardbasis) zugeordnete Matrix. Über quadratische Formen, deren Behandlung in der Linearen Algebra erfolgt, wollen wir hier nur das für die folgenden Betrachtungen Notwendige bereitstellen.

**3.4 Definition.** Eine Funktion  $q : \mathbb{R}^n \rightarrow \mathbb{R}$  mit

$$q(X) = \sum_{i,j=1}^n a_{ij}x_i x_j \quad \text{für } X = (x_1, \dots, x_n) \in \mathbb{R}^n$$

( $a_{ij} \in \mathbb{R}$  für  $i, j = 1, \dots, n$ ) heißt quadratische Form über  $\mathbb{R}^n$ . Die quadratische Form  $q$  heißt positiv definit, wenn  $q(X) > 0$  für  $X \in \mathbb{R}^n \setminus \{0\}$  gilt, positiv semidefinit, wenn  $q(X) \geq 0$  für  $X \in \mathbb{R}^n$  gilt, negativ definit (negativ semidefinit), wenn  $-q$  positiv definit (bzw. positiv semidefinit) ist, und indefinit, wenn  $q$  weder positiv noch negativ semidefinit ist.

Nach diesen Vorbereitungen können wir nun lokale Extremwerte differenzierbarer Funktionen untersuchen.

**3.5 Definition.** Die Funktion  $f : M \rightarrow \mathbb{R}$  hat in  $X_0$  ein lokales Maximum (lokales Minimum), wenn es eine Umgebung  $U \subseteq M$  von  $X_0$  gibt mit  $f(X_0) \geq f(Y)$  (bzw.  $f(X_0) \leq f(Y)$ ) für alle  $Y \in U$ . Gilt  $f(X_0) > f(Y)$  (bzw.  $f(X_0) < f(Y)$ ) für  $Y \in U \setminus \{X_0\}$ , so hat  $f$  in  $X_0$  ein striktes lokales Maximum (striktes lokales Minimum). Die Funktion  $f$  hat in  $X_0$  ein lokales Extremum, wenn sie dort ein lokales Maximum oder ein lokales Minimum hat.

Zunächst stellen wir, analog wie in Analysis I, für das Vorliegen eines lokalen Extremums eine notwendige Bedingung auf.

**3.6 Satz.** Sei  $M \subseteq \mathbb{R}^n$  offen und  $f : M \rightarrow \mathbb{R}$  differenzierbar in  $X_0$ . Hat  $f$  in  $X_0$  ein lokales Extremum, so ist  $\nabla f(X_0) = 0$ .

*Beweis.* Hat  $f$  in  $X_0$  ein lokales Extremum, so hat für jeden Vektor  $H \in \mathbb{R}^n$  die durch

$$g(t) := f(X_0 + tH) \quad \text{für } X_0 + tH \in M$$

erklärte Funktion  $g$  in 0 ein lokales Extremum. Nach dem Kriterium aus Analysis I (Kapitel 6, Satz 3.5) gilt daher

$$\langle \nabla f(X_0), H \rangle = g'(0) = 0.$$

Da  $H \in \mathbb{R}^n$  beliebig war, ist  $\nabla f(X_0) = 0$ . ■

**3.7 Definition.** Ist  $f$  differenzierbar in einer Umgebung von  $X_0$  und ist  $\nabla f(X_0) = 0$ , so heißt  $X_0$  ein kritischer Punkt von  $f$ .

Wir wollen nun hinreichende Bedingungen angeben für das Vorliegen eines strikten lokalen Extremums.

**3.8 Satz.** Sei  $M \subseteq \mathbb{R}^n$  offen und  $f \in C^2(M)$ . Sei  $X_0$  kritischer Punkt von  $f$ . Ist die Hesse-Form  $Q(f, X_0; \cdot)$  von  $f$  in  $X_0$  positiv definit (negativ definit), so hat  $f$  in  $X_0$  ein striktes lokales Minimum (striktes lokales Maximum).

*Beweis.* Sei etwa  $Q(f, X_0; \cdot)$  positiv definit. Wir zeigen zunächst die Existenz einer Zahl  $\delta \in \mathbb{R}^+$  mit  $U(X_0, \delta) \subseteq M$  und der Eigenschaft, dass für alle  $H \in \mathbb{R}^n$  mit  $\|H\| < \delta$  auch die Hesse-Form  $Q(f, X_0 + H; \cdot)$  positiv definit ist.

Da die Funktion  $Q(f, X_0; \cdot)$  stetig ist, nimmt ihre Einschränkung auf die wegen Satz 1.13 aus Kapitel 10 kompakte Menge  $\{E \in \mathbb{R}^n \mid \|E\| = 1\}$  nach Folgerung 5.5 aus Kapitel 9 ein Minimum an. Da  $Q(f, X_0; \cdot)$  positiv definit ist, ist dieses Minimum positiv. Es gibt also eine Zahl  $a \in \mathbb{R}^+$  mit

$$Q(f, X_0; E) > 2a \quad \text{für alle } E \in \mathbb{R}^n \text{ mit } \|E\| = 1.$$

Da nach Voraussetzung die partiellen Ableitungen zweiter Ordnung der Funktion  $f$  in  $X_0$  stetig sind (und  $M$  offen ist), gibt es ein  $\delta \in \mathbb{R}^+$  mit  $U(X_0, \delta) \subseteq M$  und

$$|\partial_i \partial_j f(X_0 + H) - \partial_i \partial_j f(X_0)| < \frac{a}{n^2} \quad \text{für alle } H \in \mathbb{R}^n \text{ mit } \|H\| < \delta$$

( $i, j = 1, \dots, n$ ). Für beliebiges  $E = (e_1, \dots, e_n) \in \mathbb{R}^n$  mit  $\|E\| = 1$  und für  $\|H\| < \delta$  folgt

$$\begin{aligned} & |Q(f, X_0 + H; E) - Q(f, X_0; E)| \\ & \leq \sum_{i,j=1}^n |\partial_i \partial_j f(X_0 + H) - \partial_i \partial_j f(X_0)| |e_i| |e_j| \\ & \leq \frac{a}{n^2} \sum_{i,j=1}^n 1 = a. \end{aligned}$$

Also ist  $Q(f, X_0 + H; E) > a > 0$  für  $\|E\| = 1$  und  $\|H\| < \delta$ ; daraus folgt  $Q(f, X_0 + H; Y) > 0$  für alle  $Y \in \mathbb{R}^n \setminus \{0\}$ . Für alle  $H \in \mathbb{R}^n$  mit  $\|H\| < \delta$  ist also  $Q(f, X_0 + H; \cdot)$  positiv definit.

Nun gibt es nach Satz 3.2 (für  $k = 1$ ) und wegen der Voraussetzung  $\nabla(X_0) = 0$  ein  $c \in (0, 1)$  mit

$$f(X_0 + H) = f(X_0) + \frac{1}{2} \sum_{i,j=1}^n \partial_i \partial_j f(X_0 + cH) h_i h_j.$$

Für  $0 < \|H\| < \delta$  folgt

$$f(X_0 + H) > f(X_0).$$

Die Funktion  $f$  hat also in  $X_0$  ein striktes lokales Minimum. ■

Umgekehrt läßt sich aus dem Vorliegen eines lokalen Extremums lediglich folgern, dass die Hesse-Form in diesem Punkt semidefinit ist:

**3.9 Satz.** *Sei  $M$  offen und  $f \in C^2(M)$ . In  $X_0$  habe  $f$  ein lokales Minimum (lokales Maximum). Dann ist die Hesse-Form  $Q(f, X_0, \cdot)$  positiv (negativ) semidefinit.*

*Beweis.* Die Funktion  $f$  habe etwa in  $X_0$  ein lokales Minimum. Angenommen,  $Q(f, X_0; \cdot)$  wäre nicht positiv semidefinit. Dann gibt es ein  $E = (e_1, \dots, e_n) \in \mathbb{R}^n$  mit  $Q(f, X_0; E) < 0$ . Analog wie im Beweis von Satz 3.8 zeigt man die Existenz eines  $\delta \in \mathbb{R}^+$  mit  $U(X_0, \delta) \subseteq M$  und  $Q(f, X_0 + H; E) < 0$  für alle  $H \in \mathbb{R}^n$  mit  $\|H\| < \delta$ . Wählen wir  $H := \lambda E$  mit hinreichend kleinem  $\lambda \in \mathbb{R}^+$ , so folgt  $Q(f, X_0 + cH; H) < 0$  für  $c \in (0, 1)$ . Aus Satz 3.2 ergibt sich dann  $f(X_0 + H) < f(X_0)$ . Da hier  $\|H\| > 0$  beliebig klein gewählt werden kann, hat  $f$  in  $X_0$  kein lokales Minimum, ein Widerspruch. ■

**3.10 Folgerung.** *Sei  $M$  offen und  $f \in C^2(M)$ . Ist die Hesse-Form  $Q(f, X_0; \cdot)$  von  $f$  in  $X_0$  indefinit, so hat  $f$  in  $X_0$  kein lokales Extremum.*

Aus den vorstehenden Sätzen ergibt sich, dass man zur Untersuchung von lokalen Extrema Kriterien haben muß für den Definitheitscharakter quadratischer Formen. Wir geben ein solches Kriterium an für den Fall  $n = 2$ .

**3.11 Satz.** *Die durch*

$$q(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 \quad \text{für } (x, y) \in \mathbb{R}^2$$

*definierte quadratische Form  $q$  ist genau dann positiv definit, wenn*

$$a_{11} > 0 \quad \text{und} \quad a_{11}a_{22} - a_{12}^2 > 0$$

*ist. Im Fall  $a_{11}a_{22} - a_{12}^2 < 0$  ist sie indefinit.*

*Beweis.* Sei  $q$  positiv definit. Dann ist  $a_{11} = q(1, 0) > 0$ . Durch Ausrechnen bestätigt man die Identität

$$q(x, y) = \frac{1}{a_{11}} [(a_{11}x + a_{12}y)^2 + (a_{11}a_{22} - a_{12}^2)y^2].$$

Daraus folgt

$$a_{11}a_{22} - a_{12}^2 = a_{11} q\left(-\frac{a_{12}}{a_{11}}, 1\right) > 0.$$

Umgekehrt folgt aus  $a_{11} > 0$  und  $a_{11}a_{22} - a_{12}^2 > 0$  wegen der obigen Identität sofort  $q(x, y) > 0$  für  $(x, y) \neq (0, 0)$ . ■

## 11.4 Differenzierbare Abbildungen

Im ersten Teil dieses Abschnitts befassen wir uns mit differenzierbaren Abbildungen zwischen Mengen gleicher Dimension. Solche Abbildungen treten in den Anwendungen unter anderem als sogenannte Koordinatentransformationen auf. Wir erläutern dies zunächst an einem speziellen Beispiel.

Gegeben sei eine offene Menge  $M \subseteq \mathbb{R}^2$  und eine Funktion  $f : M \rightarrow \mathbb{R}$ . Bei konkreten Fragestellungen ist es manchmal nicht zweckmäßig, einen Punkt aus  $M$  durch seine Koordinaten bezüglich der Standardbasis zu beschreiben, sondern man kann oft mit Vorteil andere Zahlenpaare wählen, durch die die Punkte ebenso festgelegt werden, die aber der Problemstellung besser angepaßt sind („krummlinige Koordinaten“). Nehmen wir etwa an, die Funktion  $f$  sei rotationssymmetrisch, das heißt  $f(X)$  hänge nur vom „Radius“  $\|X\|$  ab. Dann empfiehlt es sich, in  $M \setminus \{0\}$  Polarkoordinaten einzuführen: Jeder Punkt  $X$  ist eindeutig festgelegt durch seinen Abstand  $r$  vom Nullpunkt und (im Fall  $X \neq 0$ ) durch den Winkel  $\varphi$ , den der Vektor  $X$  mit  $E_1$  bildet. Für  $X = (x, y)$  gilt dann

$$\begin{aligned}x &= r \cos \varphi, \\y &= r \sin \varphi,\end{aligned}$$

und man schreibt etwa

$$f(x, y) = f(r \cos \varphi, r \sin \varphi) = \tilde{f}(r, \varphi)$$

und sagt, man habe damit  $f$  auf Polarkoordinaten bezogen. Dies ist folgendermaßen zu präzisieren. Wir setzen

$$U := \{(r, \varphi) \in \mathbb{R}^2 \mid r > 0, 0 \leq \varphi < 2\pi\}$$

und definieren eine Abbildung  $\tau : U \rightarrow \mathbb{R}^2$  durch

$$\tau(r, \varphi) := (r \cos \varphi, r \sin \varphi).$$

Offenbar ist  $\tau(U) = \mathbb{R}^2 \setminus \{0\}$ , und  $\tau$  ist injektiv. Die Abbildung  $\tau$  ist differenzierbar, ihre Umkehrabbildung  $\tau^{-1}$  ebenfalls, denn sie ist explizit gegeben durch

$$\begin{aligned}r &= \sqrt{x^2 + y^2}, \\ \varphi &= \operatorname{arccot} \left( \frac{x}{y} + k\pi \right) \quad \text{für } y \neq 0, \\ \varphi &= \operatorname{arctan} \left( \frac{y}{x} \right) + k\pi \quad \text{für } x \neq 0,\end{aligned}$$

wo  $k \in \{0, 1, 2\}$  jeweils so zu wählen ist, dass  $\varphi(1, 0) = 0$  und  $\varphi$  für  $(x, y) \neq (a, 0)$ ,  $a > 0$ , stetig ist. Die oben mit  $\tilde{f}$  bezeichnete Funktion ist

die Komposition  $\tilde{f} = f \circ \tau$ . Aus ihr gewinnt man  $f$  in der Form  $f = \tilde{f} \circ \tau^{-1}$ . Dabei sind jeweils noch die Definitionsbereiche zu beachten sowie die Tatsache, dass  $(0, 0)$  nicht im Bild von  $\tau$  liegt.

Allgemein bezeichnet man als *Koordinatentransformation* eine Abbildung  $F : M \rightarrow \mathbb{R}^n$  mit  $M \subseteq \mathbb{R}^n$  ( $M$  offen) mit den Eigenschaften:  $F$  ist injektiv,  $F$  und die Umkehrabbildung  $F^{-1}$  (definiert auf  $F(M)$ ) sind differenzierbar. Sei  $F$  eine Abbildung mit diesen Eigenschaften. Setzen wir  $X = (x_1, \dots, x_n)$ ,  $F(X) = Y = (y_1, \dots, y_n)$ , so ist also

$$\left. \begin{aligned} y_1 &= f_1(x_1, \dots, x_n), \\ &\vdots \\ y_n &= f_n(x_1, \dots, x_n), \end{aligned} \right\} \quad (4.1)$$

wobei  $f_1, \dots, f_n$  die Koordinatenfunktionen von  $F$  sind. Sind  $g_1, \dots, g_n$  die Koordinatenfunktionen der Umkehrabbildung  $F^{-1} : F(M) \rightarrow M$ , so ist also (4.1) äquivalent mit

$$\left. \begin{aligned} x_1 &= g_1(y_1, \dots, y_n), \\ &\vdots \\ x_n &= g_n(y_1, \dots, y_n). \end{aligned} \right\} \quad (4.2)$$

Im allgemeinen wird es nicht möglich sein, bei explizit gegebenen Funktionen  $f_1, \dots, f_n$  die Funktionen  $g_1, \dots, g_n$  explizit zu berechnen (also das Gleichungssystem (4.1) „aufzulösen nach  $x_1, \dots, x_n$ “). Häufig benötigt man jedoch nur die partiellen Ableitungen der Funktionen  $g_1, \dots, g_n$ , und diese kann man stets explizit berechnen aus den partiellen Ableitungen der Funktionen  $f_1, \dots, f_n$ . Dies ergibt sich in folgender Weise aus bekannten Resultaten.

Wir setzen voraus, dass die Abbildung  $F : M \rightarrow \mathbb{R}^n$  differenzierbar ist in einem Punkt  $X \in M$ , dass die Umkehrabbildung  $F^{-1}$  existiert und differenzierbar ist im Punkt  $Y := F(X)$ . Da die Abbildung  $F^{-1} \circ F$  auf  $M$  die Identität ist, gilt nach der Kettenregel

$$D(F^{-1})_Y \circ DF_X = \text{Identität},$$

oder in Koordinatenschreibweise

$$JF^{-1}(Y)JF(X) = \text{Einheitsmatrix}.$$

Es folgt, dass das Differential  $DF_X$  und damit die Funktionalmatrix  $JF(X)$  regulär sind und dass

$$JF^{-1}(Y) = JF(X)^{-1}$$

ist. Ausgeschrieben lautet diese Gleichung

$$\begin{pmatrix} \partial_1 g_1(Y) & \cdots & \partial_n g_1(Y) \\ \vdots & & \vdots \\ \partial_1 g_n(Y) & \cdots & \partial_n g_n(Y) \end{pmatrix} = \begin{pmatrix} \partial_1 f_1(X) & \cdots & \partial_n f_1(X) \\ \vdots & & \vdots \\ \partial_1 f_n(X) & \cdots & \partial_n f_n(X) \end{pmatrix}^{-1}.$$

Nach bekannten Regeln für die Berechnung einer inversen Matrix ergibt sich hieraus für jede Funktion  $\partial_j g_i$  eine Darstellung als Quotient, wo Zähler und Nenner Polynome in den Funktionen  $\partial_k f_m \circ F^{-1}$  ( $k, m = 1, \dots, n$ ) sind und der Nenner nicht verschwindet. Durch weitere Differentiationen (nach der Kettenregel) ergibt sich dann: Sind  $f_1, \dots, f_n$  Funktionen der Klasse  $C^r$ , so sind  $g_1, \dots, g_n$  von der Klasse  $C^r$ .

Von besonderer Wichtigkeit ist nun naheliegenderweise die Frage, ob und gegebenenfalls wie man einer gegebenen differenzierbaren Abbildung  $F : M \rightarrow \mathbb{R}^n$  (mit  $M \subseteq \mathbb{R}^n$ ) ansehen kann, ob sie eine differenzierbare Umkehrabbildung besitzt, also eine Koordinatentransformation definiert. Notwendig ist, wie wir eben gesehen haben, jedenfalls die Regularität (Umkehrbarkeit) des Differentials, also das Nichtverschwinden der Funktionaldeterminante. Für eine differenzierbare Funktion  $f : (a, b) \rightarrow \mathbb{R}$  ist die Bedingung  $f'(t) \neq 0$  für  $t \in (a, b)$  bekanntlich auch hinreichend für die Existenz und Differenzierbarkeit der Umkehrfunktion, aber im Fall  $n > 1$  ist die Sachlage nicht so einfach.

**Beispiel.** Sei  $M = \mathbb{R}^2 \setminus \{0\}$  und  $F : M \rightarrow \mathbb{R}^2$  definiert durch

$$F(x, y) := (x^2 - y^2, 2xy) \quad \text{für } (x, y) \in \mathbb{R}^2.$$

Die Funktionalmatrix

$$JF(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$$

hat die Determinante  $4(x^2 + y^2) \neq 0$ , ist also regulär. Wegen  $F(X) = F(-X)$  ist die Abbildung  $F$  aber nicht umkehrbar.

Die Regularität des Differentials an einer gegebenen Stelle reicht jedoch aus, um zumindest in einer hinreichend kleinen Umgebung dieser Stelle die Existenz und Differenzierbarkeit der Umkehrabbildung zu sichern. Diesen wichtigen Satz wollen wir nun beweisen.

Die Abbildung  $F : M \rightarrow \mathbb{R}^n$  heißt *von der Klasse  $C^r$* , wenn alle Koordinatenfunktionen von  $F$   $r$ -mal stetig differenzierbar sind. Wir stellen zunächst einen Hilfssatz bereit.

**4.3 Lemma.** Sei  $M \subseteq \mathbb{R}^n$  offen, sei  $F : M \rightarrow \mathbb{R}^n$  eine Abbildung der Klasse  $C^1$ , sei  $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$  eine lineare Abbildung. Sind  $X, Y \in M$  Punkte mit  $[X, Y] \subseteq M$ , so gilt

$$\|F(X) - F(Y) - L(X - Y)\| \leq \|X - Y\| \max_{Z \in [X, Y]} \|DF_Z - L\|.$$

*Beweis.* Setze  $G(X) := F(X) - L(X)$  für  $X \in M$ ; dann ist  $DG_Z = DF_Z - L$  für  $Z \in M$ . Da  $F$  von der Klasse  $C^1$  ist, ist die Funktion  $Z \mapsto \|DF_Z - L\|$  stetig, daher existiert

$$\max_{Z \in [X, Y]} \|DF_Z - L\| =: c.$$

Aus dem Mittelwertsatz (Satz 1.8) folgt jetzt

$$\|G(X) - G(Y)\| \leq c\|X - Y\|,$$

also die Behauptung. ■

**4.4 Satz** (über die Umkehrabbildung). Sei  $M \subseteq \mathbb{R}^n$  offen,  $X_0 \in M$ , sei  $F : M \rightarrow \mathbb{R}^n$  eine Abbildung der Klasse  $C^r$  (für ein  $r \in \mathbb{N}$ ). Das Differential  $DF_{X_0}$  von  $F$  in  $X_0$  sei regulär (d.h. die Funktionaldeterminante von  $F$  in  $X_0$  sei  $\neq 0$ ). Dann gibt es eine offene Umgebung  $U \subseteq M$  von  $X_0$  mit folgenden Eigenschaften:

- (a) die Einschränkung  $F|_U$  ist injektiv,
- (b) die Bildmenge  $V := F(U)$  ist offen,
- (c) die Umkehrabbildung  $(F|_U)^{-1} : V \rightarrow U$  ist von der Klasse  $C^r$ .

*Beweis.* Im folgenden bezeichnet  $I$  die identische Abbildung von  $\mathbb{R}^n$  auf sich. Zur Abkürzung wird für  $\alpha > 0$

$$U_\alpha := U(0, \alpha) = \{X \in \mathbb{R}^n \mid \|X\| < \alpha\}$$

gesetzt. Wir machen zunächst eine spezielle Voraussetzung.

**Voraussetzung.**  $X_0 = 0$ ,  $F(0) = 0$ ,  $DF_0 = I$ .

Nach Voraussetzung ist die Funktion  $X \mapsto \|DF_X - I\|$  in  $M$  stetig, und es ist  $\|DF_0 - I\| = 0$ . Daher existiert zu vorgegebenem  $\varepsilon > 0$  ein  $\alpha > 0$  mit  $U_\alpha \subseteq M$  und

$$\|DF_X - I\| \leq \varepsilon \quad \text{für alle } X \in U_\alpha.$$

Aus Lemma 4.3 folgt dann für alle  $X, Y \in U_\alpha$

$$\|F(X) - F(Y) - (X - Y)\| \leq \varepsilon \|X - Y\| \quad (4.5)$$

und hieraus durch Anwendung der Dreiecksungleichung

$$(1 - \varepsilon)\|X - Y\| \leq \|F(X) - F(Y)\|. \quad (4.6)$$

Wir wählen nun ein positives  $\varepsilon < 1$  und dazu  $\alpha > 0$  wie oben und so, dass  $\overline{U_\alpha} \subseteq M$  ist.

**Behauptung (1).** Es gilt  $U_{(1-\varepsilon)\alpha} \subseteq F(U_\alpha)$ .

*Beweis.* Sei  $Y \in U_{(1-\varepsilon)\alpha}$ . Es ist ein  $X \in U_\alpha$  zu finden mit  $Y = F(X)$ . Hierzu benutzen wir den Banachschen Fixpunktsatz. Definiere  $\Phi: \overline{U_\alpha} \rightarrow \mathbb{R}^n$  durch

$$\Phi(X) := Y - F(X) + X \quad \text{für } X \in \overline{U_\alpha}.$$

Für  $X \in \overline{U_\alpha}$  gilt wegen  $F(0) = 0$  nach (4.5)

$$\|\Phi(X)\| \leq \|Y\| + \|F(X) - X\| \leq \|Y\| + \varepsilon \|X\| < (1 - \varepsilon)\alpha + \varepsilon\alpha = \alpha,$$

also  $\Phi(X) \in U_\alpha$ . Somit bildet  $\Phi$  die Menge  $\overline{U_\alpha}$  in sich ab. Für  $X, Z \in \overline{U_\alpha}$  gilt nach (4.5)

$$\|\Phi(X) - \Phi(Z)\| = \|F(X) - F(Z) - (X - Z)\| \leq \varepsilon \|X - Z\|,$$

wegen  $\varepsilon < 1$  ist also  $\Phi$  kontrahierend. Da  $\overline{U_\alpha}$  nach Satz 1.10 aus Kapitel 10 und Satz 2.9 aus Kapitel 9 vollständig ist, gibt es nach dem Banachschen Fixpunktsatz (Satz 3.2 aus Kapitel 9) einen Punkt  $X \in \overline{U_\alpha}$  mit  $\Phi(X) = X$ . Wie oben gezeigt, ist  $X = \Phi(X) \in U_\alpha$ . Es ist also ein Punkt  $X \in U_\alpha$  gefunden mit  $Y = F(X)$ . Damit ist die 1. Behauptung bewiesen. ■

Nun setzen wir  $V := U_{(1-\varepsilon)\alpha}$  und  $U := F^{-1}(V)$ . Dann ist  $U$  eine offene Umgebung von 0 und  $F(U) = V$ . Aus (4.6) folgt, dass  $F|_U$  injektiv ist. Für den Moment bezeichne  $G: V \rightarrow U$  die Umkehrabbildung von  $F|_U$ .

**Behauptung (2).**  $G$  ist in 0 differenzierbar.

*Beweis.* Sei  $\varepsilon' \in \mathbb{R}^+$  gegeben. Wie anfangs gezeigt ((4.5) mit  $Y = 0$ ) existiert ein  $\alpha' \in \mathbb{R}^+$  mit  $U_{\alpha'} \subseteq M$  und

$$\|F(X) - X\| \leq \frac{\varepsilon'}{1 + \varepsilon'} \|X\| \quad \text{für } \|X\| < \alpha'$$

und daher

$$\|X\| \leq (1 + \varepsilon')\|F(X)\| \quad \text{für } \|X\| < \alpha'.$$

Sei  $H \in V$  ein Vektor mit  $\|H\| < \alpha'(1 - \varepsilon)$ . Für  $X := G(H)$  ist dann  $X \in U$ , also  $\|X\| < \alpha$  und daher nach (4.6)

$$\|X\| \leq \frac{1}{1 - \varepsilon}\|F(X)\| = \frac{1}{1 - \varepsilon}\|H\| < \alpha',$$

also

$$\|G(H) - H\| = \|X - F(X)\| \leq \frac{\varepsilon'}{1 + \varepsilon'}\|X\| \leq \varepsilon'\|F(X)\| = \varepsilon'\|H\|,$$

somit

$$\frac{\|G(H) - H\|}{\|H\|} \leq \varepsilon' \quad \text{für } 0 < \|H\| < \alpha'(1 - \varepsilon).$$

Damit ist

$$\lim_{H \rightarrow 0} \frac{G(H) - H}{\|H\|} = 0$$

gezeigt, also (wegen  $G(0) = 0$ ) die Differenzierbarkeit von  $G$  in 0 sowie  $DG_0 = I$ . ■

Wir lassen nun die spezielle Voraussetzung fallen, dass  $X_0 = 0$ ,  $F(0) = 0$  und  $DF_0 = I$  sein soll.

Es ist also jetzt vorausgesetzt, dass  $X_0 \in M$  ein Punkt ist derart, dass  $DF_{X_0}$  regulär ist. Durch passende Transformationen gewinnen wir eine Abbildung  $\tilde{F}$ , die die spezielle Voraussetzung erfüllt. Hierzu bezeichne  $T_Z : \mathbb{R}^n \rightarrow \mathbb{R}^n$  die Translation um den Vektor  $Z$ , also  $T_Z(X) := X + Z$  für  $X \in \mathbb{R}^n$ . Wir setzen

$$\begin{aligned} L &:= DF_{X_0}, \\ \tilde{M} &:= (L \circ T_{-X_0})(M) \\ \tilde{F}(X) &:= T_{-F(X_0)} \circ F \circ T_{X_0} \circ L^{-1}(X) \quad \text{für } X \in \tilde{M}. \end{aligned}$$

Wie man leicht nachprüft, gilt  $0 \in \tilde{M}$ ,  $\tilde{F}(0) = 0$  und  $D\tilde{F}_0 = I$ . Anwendung des bereits Bewiesenen auf  $\tilde{F}$  ergibt dann die lokale Umkehrbarkeit und Differenzierbarkeit von  $F$ . Hierzu beachte man, dass

$$F = T_{F(X_0)} \circ \tilde{F} \circ L \circ T_{-X_0}$$

ist. Die Differenzierbarkeit der Umkehrabbildung ergibt sich zunächst nur im Punkt  $F(X_0)$ . Nun ist aber in  $X_0$  die Funktionaldeterminante von  $F$  verschieden von Null, und das gilt, da sie stetig ist, noch in einer ganzen Umgebung von  $X_0$ . Auf jeden Punkt dieser Umgebung kann man das bereits Bewiesene anwenden, erhält also die Differenzierbarkeit der Umkehrabbildung. Dass mit  $F$  auch die lokale Umkehrung von der Klasse  $C^r$  ist, haben wir bereits vor Lemma 4.3 begründet. ■

Eine injektive Abbildung  $F : M \rightarrow \mathbb{R}^n$  (mit offenem  $M \subseteq \mathbb{R}^n$ ) der Klasse  $C^r$  (für ein  $r \in \mathbb{N}$ ) mit überall regulärem Differential nennt man auch einen *Diffeomorphismus* der Klasse  $C^r$  von  $M$  auf  $F(M)$ . Ist nun  $F : M \rightarrow \mathbb{R}^n$  eine Abbildung der Klasse  $C^r$  und  $DF_{X_0}$  regulär, so gibt es nach Satz 4.4 eine offene Umgebung  $U$  von  $X_0$  derart, dass  $F|_U$  ein Diffeomorphismus der Klasse  $C^r$  von  $U$  auf  $F(U)$  ist. Für diesen Sachverhalt sagt man auch,  $F$  sei bei  $X_0$  lokal ein  $C^r$ -Diffeomorphismus.

Als eine Folgerung aus dem Satz über die Umkehrabbildung wollen wir nun eine Aussage über sogenannte „implizite Funktionen“ gewinnen. Dieses für manche Anwendungen wichtige Thema wollen wir zunächst anschaulich erläutern.

Verschiedenartige Aufgabenstellungen erfordern die Untersuchung von Gebilden im euklidischen Raum, wie „Kurven“ oder „Flächen“, die beschrieben werden durch Gleichungen, die zwischen den Koordinaten ihrer Punkte bestehen. Als einfaches Beispiel betrachten wir im  $\mathbb{R}^2$  die „Einheitssphäre“

$$S^2 := \{X \in \mathbb{R}^2 \mid x_1^2 + x_2^2 - 1 = 0\}.$$

Die genauere Untersuchung dieser Punktmenge kann dadurch erleichtert werden, dass man die definierende Gleichung  $x_1^2 + x_2^2 = 1$  nach einer Variablen auflöst, indem man etwa

$$x_2 = \sqrt{1 - x_1^2}$$

schreibt. Man betrachtet dann die Punktmenge

$$S_+^1 := \{X \in \mathbb{R}^2 \mid x_2 = \sqrt{1 - x_1^2}, x_1^2 < 1\}.$$

Allerdings ist  $S_+^1$  nur ein echter Teil von  $S^1$ , die „obere Halbsphäre“. Offenbar kann die ganze Menge  $S^1$  nicht einheitlich dadurch dargestellt werden, dass man eine Koordinate als Funktion der beiden anderen schreibt. Aber zur lokalen Untersuchung von  $S^1$  ist eine derartige Darstellung gut geeignet. Es stört nicht, dass man zur Untersuchung von  $S^1$  nicht mit nur einer solchen Darstellung auskommt. Will man  $S^1$  in einer Umgebung des Punktes  $(0, -1)$  untersuchen, so benutzt man die „Auflösung“  $x_2 = -\sqrt{1 - x_1^2}$ , in einer Umgebung von  $(1, 0)$  benutzt man die Auflösung  $x_1 = \sqrt{1 - x_2^2}$ , usw.

Analog kann man Punktmenge untersuchen, die durch mehrere Gleichungen zwischen den Koordinaten definiert sind. Als Beispiel betrachten wir die Punktmenge

$$K := \left\{ X \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3^2 - 1 = 0 \text{ und } \left(x_1 - \frac{1}{2}\right)^2 + x_2^2 - \frac{1}{4} = 0 \right\},$$

also den Durchschnitt der Sphäre  $S^2 \subseteq \mathbb{R}^3$  mit einem gewissen Kreiszyylinder mit Erzeugenden parallel zur  $x_3$ -Achse. Eine „lokale Auflösung“ lautet

$$x_1 = 1 - x_3^2, \quad x_2 = x_3 \sqrt{1 - x_3^2}, \quad |x_3| < 1,$$

eine andere

$$x_1 = 1 - x_3^2, \quad x_2 = -x_3 \sqrt{1 - x_3^2}, \quad |x_3| < 1.$$

Diese Darstellungen können zur lokalen Untersuchung der Schnittkurve  $K$  benutzt werden.

Im allgemeinen wird es nicht möglich sein, eine „lokale Auflösung“ explizit (formelmäßig) zu bewerkstelligen. Hier sind dann Aussagen von Interesse, die die grundsätzliche Möglichkeit einer solchen Auflösung (und die Differenzierbarkeit der darstellenden Funktionen) gewährleisten. Zunächst soll aber die Fragestellung allgemeiner und genauer gefaßt werden.

Eine Punktmenge  $N$  des  $\mathbb{R}^n$  sei definiert und beschrieben durch ein Gleichungssystem

$$\left. \begin{aligned} f_1(x_1, \dots, x_n) &= 0, \\ f_2(x_1, \dots, x_n) &= 0, \\ &\vdots \\ f_k(x_1, \dots, x_n) &= 0. \end{aligned} \right\} \quad (4.7)$$

Dabei sei  $M \subseteq \mathbb{R}^n$  eine offene Teilmenge, und

$$f_i : M \rightarrow \mathbb{R}, \quad i = 1, \dots, k < n,$$

seien reellwertige Funktionen der Klasse  $C^1$ . Sodann sei

$$N := \{X \in M \mid X = (x_1, \dots, x_n) \text{ erfüllt (4.7)}\}.$$

(Der Buchstabe  $N$  soll an „Nullstellenmenge“ erinnern.) Das Problem besteht nun darin, die Punktmenge  $N$  in einer Umgebung eines gegebenen Punktes  $X_0 \in N$  zu beschreiben, indem  $k$  Koordinaten als Funktionen der übrigen  $n - k$  Koordinaten dargestellt werden, etwa

$$\begin{aligned} x_{n-k+1} &= g_1(x_1, \dots, x_{n-k}), \\ &\vdots \\ x_n &= g_k(x_1, \dots, x_{n-k}) \end{aligned}$$

für  $(x_1, \dots, x_{n-k})$  aus einer passenden Menge  $V \subseteq \mathbb{R}^{n-k}$ . Das ist so zu verstehen: Setzt man die so berechneten Koordinaten in (4.7) ein, so soll dieses System erfüllt sein, also

$$f_i(x_1, \dots, x_{n-k}, g_1(x_1, \dots, x_{n-k}), \dots, g_k(x_1, \dots, x_{n-k})) = 0$$

für  $(x_1, \dots, x_{n-k}) \in V$  und  $i = 1, \dots, k$ . Genauer gesagt: Jeder Punkt

$$(x_1, \dots, x_{n-k}, g_1(x_1, \dots, x_{n-k}), \dots, g_k(x_1, \dots, x_{n-k}))$$

mit  $(x_1, \dots, x_{n-k}) \in V$  soll zu  $N$  gehören; umgekehrt soll aber auch der Durchschnitt von  $N$  mit einer ganzen Umgebung von  $X_0$  auf diese Weise erhalten werden. Man sagt dann, die Funktionen  $g_1, \dots, g_k$  seien „implizit definiert“ durch die Gleichungen

$$f_1(X) = 0, \dots, f_k(X) = 0.$$

Nützlich ist die hierdurch gegebene lokale Auflösung im Allgemeinen nur, wenn die Funktionen  $g_1, \dots, g_k$  ihrerseits so oft differenzierbar sind wie die Funktionen  $f_1, \dots, f_k$ . Für die Möglichkeit einer solchen Auflösung gibt Satz 4.8 eine hinreichende Bedingung.

Es ist im folgenden zweckmäßig, die Funktionen  $f_1, \dots, f_k$  als Koordinatenfunktionen einer Abbildung  $F : M \rightarrow \mathbb{R}^k$  und ebenso die Funktionen  $g_1, \dots, g_k$  als Koordinatenfunktionen einer Abbildung  $G : V \rightarrow \mathbb{R}^k$  aufzufassen. Unter dem *Graphen* von  $G$  versteht man die Menge

$$\text{Graph } G := \{(X, G(X)) \in \mathbb{R}^{n-k} \times \mathbb{R}^k \mid X \in V\}.$$

Der zu beweisende Satz läßt sich dann so formulieren, dass man eine Nullstellenmenge unter passenden Voraussetzungen lokal als Graph darstellen kann.

**4.8 Satz** (über implizit definierte Funktionen). *Sei  $k < n$ ,  $M \subseteq \mathbb{R}^n$  offen,  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung der Klasse  $C^r$  (für ein  $r \in \mathbb{N}$ ), sei*

$$N := \{X \in M \mid F(X) = 0\}.$$

*Sei  $X_0 \in N$  und  $DF_{X_0}$  vom Rang  $k$ . Dann gibt es (nach passender Identifizierung von  $\mathbb{R}^n$  mit  $\mathbb{R}^{n-k} \times \mathbb{R}^k$ ) eine offene Umgebung  $U$  von  $X_0$  in  $M$ , eine offene Menge  $V$  in  $\mathbb{R}^{n-k}$  und eine Abbildung  $G : V \rightarrow \mathbb{R}^k$  der Klasse  $C^r$  mit*

$$N \cap U = \text{Graph } G.$$

*Beweis.* Das Differential  $DF_{X_0}$  ist nach Voraussetzung vom Rang  $k$ . Die Funktionalmatrix von  $F$  in  $X_0$  hat also  $k$  linear unabhängige Spalten. Wir dürfen o.B.d.A. (d.h. nach passender Ummumerierung der Basisvektoren) annehmen, dass dies die letzten  $k$  Spalten sind. Wir identifizieren dann  $\mathbb{R}^{n-k}$  mit dem von den ersten  $n-k$  Basisvektoren und  $\mathbb{R}^k$  mit dem von den letzten  $k$  Basisvektoren des  $\mathbb{R}^n$  aufgespannten Unterraum, und wir identifizieren  $\mathbb{R}^n$  mit  $\mathbb{R}^{n-k} \times \mathbb{R}^k$ . Jeder Punkt  $X$  aus  $\mathbb{R}^n$  läßt sich also eindeutig in der Form  $(X', X'')$  mit  $X' \in \mathbb{R}^{n-k}$  und  $X'' \in \mathbb{R}^k$  schreiben. Wir definieren die Abbildung

$$\Phi : M \rightarrow \mathbb{R}^{n-k} \times \mathbb{R}^k : (X', X'') \mapsto (X', F(X)).$$

Dann ist  $\Phi$  eine Abbildung der Klasse  $C^r$ . Wegen

$$J\Phi = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \partial_1 f_1 & \cdots & \partial_{n-k} f_1 & \partial_{n-k+1} f_1 & \cdots & \partial_n f_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \partial_1 f_k & \cdots & \partial_{n-k} f_k & \partial_{n-k+1} f_k & \cdots & \partial_n f_k \end{pmatrix}$$

ist  $J\Phi$  an der Stelle  $X_0$  vom Rang  $n$ . Nach Satz 4.4 ist daher  $\Phi$  in einer Umgebung  $U_0$  von  $X_0$  ein Diffeomorphismus der Klasse  $C^r$ . Sei  $\Psi$  seine Umkehrabbildung.  $\Psi$  ist definiert auf einer offenen Umgebung  $W$  von  $\Phi(X_0) = (X'_0, F(X_0)) = (X'_0, 0)$ . Wir setzen

$$V := \{X' \in \mathbb{R}^{n-k} \mid (X', 0) \in W\},$$

dann ist  $V \subseteq \mathbb{R}^{n-k}$  eine offene Umgebung von  $X'_0$ . Wegen  $\Phi(X', X'') = (X', F(X', X''))$  ist

$$\Psi(X', Z) = (X', \Psi_2(X', Z)),$$

wodurch eine Abbildung  $\Psi_2 : W \rightarrow \mathbb{R}^k$  definiert wird. Wir setzen nun

$$G(X') := \Psi_2(X', 0) \quad \text{für } X' \in V.$$

Dann ist  $G : V \subseteq \mathbb{R}^{n-k} \rightarrow \mathbb{R}^k$  eine Abbildung der Klasse  $C^r$ . Anwendung von  $\Psi$  auf die Gleichung

$$\Phi(X'_0, X''_0) = (X'_0, F(X_0)) = (X'_0, 0)$$

ergibt

$$(X'_0, X''_0) = \Psi(X'_0, 0) = (X'_0, \Psi_2(X'_0, 0)) = (X'_0, G(X'_0)),$$

also  $G(X'_0) = X''_0$ . Für  $X' \in V$  gilt  $\Psi(X', 0) = (X', G(X'))$  (nach Definition von  $\Psi_2$  und  $G$ ). Anwendung von  $\Phi$  ergibt

$$(X', 0) = \Phi(X', G(X')) = (X', F(X', G(X'))),$$

also  $F(X', G(X')) = 0$  und damit

$$(X', G(X')) \in N.$$

Wir können o.B.d.A.  $V \times G(V) \subseteq U_0$  annehmen, denn dies läßt sich durch Verkleinerung von  $V$  erreichen, da  $G$  in  $X'_0$  stetig und  $G(X'_0) = X''_0$  ist. Setze  $U := (V \times \mathbb{R}^k) \cap U_0$ . Dann gilt

$$N \cap U = \text{Graph } G.$$

Zum Beweis sei  $(X', Y) \in \text{Graph } G$ , also  $X' \in V$  und  $Y = G(X')$ . Dann ist  $(X', Y) = (X', G(X')) \in V \times G(V) \subseteq U$  und, wie eben gezeigt,  $(X', G(X')) \in N$ .

Ist umgekehrt  $(X', Y) \in N \cap U$ , so ist

$$(X', F(X', Y)) = (X', 0) = (X', F(X', G(X'))),$$

also nach Definition von  $\Phi$

$$\Phi(X', Y) = \Phi(X', G(X')).$$

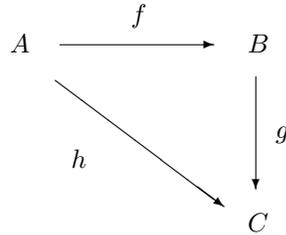
Da  $\Phi$  injektiv ist, folgt  $Y = G(X')$ , also  $(X', Y) \in \text{Graph } G$ . ■

Satz 4.4 gibt die Möglichkeit, auch für  $k \neq n$  differenzierbare Abbildungen  $F : M \rightarrow \mathbb{R}^k$  mit nichtausgeartetem Differential in einer Weise zu beschreiben, die zeigt, dass solche Abbildungen sich lokal qualitativ so wie ihr Differential verhalten. Für  $k < n$  ist diese Beschreibung im wesentlichen schon in Satz 4.8 enthalten; für  $k > n$  wird sie auf analoge Weise erhalten.

**4.9 Definition.** Sei  $F : M \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  differenzierbar in  $X_0 \in M$ . Dann wird der Rang des Differentials  $DF_{X_0}$  (= Rang der Funktionalmatrix  $JF(X_0)$ ) als Rang der Abbildung  $F$  in  $X_0$  bezeichnet.

Sei nun  $M \subseteq \mathbb{R}^n$  offen und  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung der Klasse  $C^1$ . Der Rang von  $F$  in  $X \in M$  kann dann höchstens gleich  $\min\{n, k\}$  sein. Ist er gleich dieser Zahl, so sagt man,  $F$  sei in  $X$  von maximalem Rang. Eine Teilaussage von Satz 4.4 kann man auch so formulieren: Im Fall  $k = n$  ist jede  $C^1$ -Abbildung, die in einem Punkt von maximalem Rang ist, in einer Umgebung dieses Punktes ein Diffeomorphismus. Wir wollen analoge Aussagen für  $k \neq n$  gewinnen. Sie besagen, dass man differenzierbare Abbildungen maximalen Ranges lokal in sehr übersichtlicher Weise durch Diffeomorphismen und lineare Abbildungen beschreiben kann.

**Bemerkung.** Sind  $f : A \rightarrow B, g : B \rightarrow C, h : A \rightarrow C$  ( $A, B, C$  Mengen) Abbildungen mit  $g \circ f = h$ , so drückt man dies auch aus durch die Sprechweise: „Das Diagramm



ist kommutativ.“ [Anschaulich: Man kommt, ausgehend von einem  $a \in A$ , zu demselben Element von  $C$ , gleichgültig auf welchem Wege man den Pfeilen folgt.]

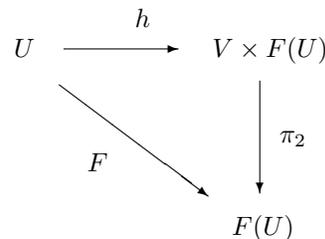
**Bemerkung.** Sind  $A, B$  Mengen, so sind auf dem kartesischen Produkt  $A \times B$  die Projektionen  $\pi_1, \pi_2$  definiert durch

$$\begin{aligned}
 \pi_1 : A \times B &\rightarrow A : (a, b) \mapsto a \\
 \pi_2 : A \times B &\rightarrow B : (a, b) \mapsto b.
 \end{aligned}$$

Sind  $A, B$  Vektorräume, so wird die kanonische Injektion  $i$  von  $A$  in  $A \times B$  definiert durch

$$\begin{aligned}
 i : A &\rightarrow A \times B \\
 a &\mapsto (a, 0).
 \end{aligned}$$

**4.10 Satz** (über lokal surjektive Abbildungen). Sei  $k < n$ . Sei  $M \subseteq \mathbb{R}^n$  offen, sei  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung der Klasse  $C^r$  ( $r \geq 1$ ). Sei  $X_0 \in M$  und  $F$  in  $X_0$  vom Rang  $k$  (also  $DF_{X_0}$  surjektiv). Dann gibt es eine offene Umgebung  $U$  von  $X_0$  in  $M$ , eine offene Menge  $V$  in  $\mathbb{R}^{n-k}$  und einen  $C^r$ -Diffeomorphismus  $h : U \rightarrow V \times F(U)$ , so dass das Diagramm



kommutativ ist.

**Bemerkung.** Die Abbildung  $F$  ist also in einer Umgebung des Punktes  $X_0$ , in dem das Differential surjektiv ist, dargestellt als Komposition eines  $C^r$ -Diffeomorphismus und einer surjektiven linearen Abbildung.

von Satz 4.10. Nach Voraussetzung hat die Funktionaldeterminante  $JF(X_0)$   $k$  linear unabhängige Spalten. Wir können o.B.d.A. annehmen, dass dies die letzten  $k$  Spalten sind. Wir fassen jetzt wieder  $\mathbb{R}^n$  als das kartesische Produkt  $\mathbb{R}^{n-k} \times \mathbb{R}^k$  auf und bezeichnen mit  $\pi_1, \pi_2$  die zugehörigen Projektionen. Aus Satz dem Beweis von Satz 4.8 folgt, dass die Abbildung

$$\varphi : M \rightarrow \mathbb{R}^{n-k} \times \mathbb{R}^k : X \mapsto (\pi_1(X), F(X))$$

auf einer offenen Umgebung  $U'$  von  $X_0$  ein  $C^r$ -Diffeomorphismus ist. Das Bild  $\varphi(U')$  enthält eine offene Umgebung von  $\varphi(X_0) = (\pi_1(X_0), F(X_0))$  der Form  $V \times W$ , dabei ist also  $V$  offen in  $\mathbb{R}^{n-k}$ . Setze  $U := \varphi^{-1}(V \times W)$  und  $h := \varphi|_U$ . Dann ist  $F(U) = W$ , und für  $X \in U$  gilt  $h(X) = (\pi_1(X), F(X))$ , also  $\pi_2 \circ h = F$ . ■

Der folgende Satz ist das Gegenstück zu Satz 4.10 für den Fall  $k > n$ .

**4.11 Satz** (über lokal injektive Abbildungen). *Sei  $k > n$ . Sei  $M \subseteq \mathbb{R}^n$  offen, sei  $F : M \rightarrow \mathbb{R}^k$  eine Abbildung der Klasse  $C^r$  ( $r \geq 1$ ). Sei  $X_0 \in M$  und  $F$  in  $X_0$  vom Rang  $n$  (also  $DF_{X_0}$  injektiv). Dann gibt es eine offene Umgebung  $U$  von  $X_0$  in  $M$ , eine offene Umgebung  $V$  von  $0$  in  $\mathbb{R}^{k-n}$ , eine offene Umgebung  $W$  von  $F(X_0)$  in  $\mathbb{R}^k$  und einen  $C^r$ -Diffeomorphismus  $h : U \times V \rightarrow W$ , so dass das Diagramm*

$$\begin{array}{ccc} U & \xrightarrow{i} & U \times V \\ & \searrow F & \downarrow h \\ & & W \end{array}$$

(wo  $i$  die kanonische Injektion bezeichnet) kommutativ ist.

**Bemerkung.** Die Abbildung  $F$  ist also in einer Umgebung des Punktes  $X_0$ , in dem das Differential injektiv ist, dargestellt als Komposition einer injektiven linearen Abbildung und eines  $C^r$ -Diffeomorphismus.

*Beweis (Satz 4.11).* Nach Voraussetzung hat die Funktionalmatrix  $JF(X_0)$   $n$  linear unabhängige Zeilen. Wir können o.B.d.A. annehmen, dass dies die ersten  $n$  Zeilen sind. Wir fassen den Raum  $\mathbb{R}^k$  als das kartesische Produkt  $\mathbb{R}^n \times \mathbb{R}^{k-n}$  auf und definieren

$$\psi : M \times \mathbb{R}^{k-n} \rightarrow \mathbb{R}^k : (X, Y) \mapsto F(X) + (0, Y).$$

Dann ist  $\psi$  von der Klasse  $C^r$ , und die Funktionalmatrix von  $\psi$  ist von der Form

$$\left( \begin{array}{c|c} A & 0 \\ \hline B & E \end{array} \right),$$

wo  $A$  aus den ersten  $n$  Zeilen der Funktionalmatrix von  $F$  gebildet ist und  $E$  die  $(k-n)$ -reihige Einheitsmatrix ist. Im Punkt  $(X_0, 0)$  ist also die Funktionaldeterminante von  $\psi$  ungleich Null. Nach Satz 4.4 gibt es daher offene Umgebungen  $U \times V$  (o.B.d.A. in dieser Produktform) von  $(X_0, 0)$  und  $W$  von  $\psi(X_0, 0) = F(X_0)$ , so dass  $h := \psi|_{U \times V}$  ein  $C^r$ -Diffeomorphismus von  $U \times V$  auf  $W$  ist. Für  $X \in U$  gilt  $h \circ i(X) = h(X, 0) = F(X)$ . ■

Es ist nun Gelegenheit, die Untersuchung von Extremwerten reeller Funktionen fortzuführen. Wir wollen eine notwendige Bedingung für das Vorliegen eines lokalen Extremums unter Nebenbedingungen aufstellen, die sogenannte *Multiplikatorenregel von Lagrange*. Zunächst betrachten wir als Beispiel eine typische Aufgabe dieser Art. Es seien die Extrema der Funktion  $f(x, y) = x + y$  zu finden unter der Nebenbedingung  $5x^2 + 6xy + 2y^2 = 1$ . Mit anderen Worten, es sei

$$N := \{(x, y) \in \mathbb{R}^2 \mid 5x^2 + 6xy + 2y^2 = 1\},$$

und es sollen die Extrema der Einschränkung  $f|_N$  gefunden werden. In diesem einfachen Fall kann man natürlich so vorgehen, dass man die Gleichung  $5x^2 + 6xy + 2y^2 = 1$  nach  $x$  oder  $y$  auflöst, zum Beispiel

$$y = -\frac{3}{2}x \pm \sqrt{\frac{1}{2} - \frac{1}{4}x^2}.$$

Wir setzen daher

$$N^\pm := \left\{ (x, y) \in \mathbb{R}^2 \mid |x| < \sqrt{2}, y = -\frac{3}{2}x \pm \sqrt{\frac{1}{2} - \frac{1}{4}x^2} \right\},$$

$$g^\pm(x) := x + \left( -\frac{3}{2}x \pm \sqrt{\frac{1}{2} - \frac{1}{4}x^2} \right) \quad \text{für } |x| < \sqrt{2}.$$

Für  $(x, y) \in N^+$  gilt dann  $f(x, y) = g^+(x)$ . Für  $(x, y) \in N^+$  hat also  $f$  in  $(x, y)$  genau dann ein lokales Extremum, wenn  $g^+$  in  $x$  ein lokales Extremum

hat. Nun ist  $(g^+)'(x) = 0$  nur für  $x = -1$ . Hat also  $f$  in  $N^+$  ein lokales Extremum, so im Punkt  $(-1, 2)$ . Analog gilt: Hat  $f$  in  $N^-$  ein lokales Extremum, so im Punkt  $(1, -2)$ . Man muß allerdings noch nachweisen, z.B. durch Auflösung nach  $x$ , dass  $f$  in den nicht erfaßten Punkten mit  $|x| = \sqrt{2}$  keine lokalen Extrema hat. Da  $f$  ein Maximum und ein Minimum annimmt, hat  $f$  also genau in den beiden Punkten  $(-1, 2)$  und  $(1, -2)$  lokale Extrema.

Im allgemeinen wird eine solche explizite Auflösung nicht möglich sein. Sie ist aber auch nicht erforderlich, da man folgendermaßen argumentieren kann. Wir können (wie später allgemeiner gezeigt und verwendet wird), die Menge  $N$  lokal parametrisieren, das heißt die Punkte  $(x, y) \in N$  darstellen in der Form  $x = \varphi(t)$ ,  $y = \psi(t)$  mit stetig differenzierbaren Funktionen  $\varphi, \psi$ . Dann setzen wir  $h(t) = f(\varphi(t), \psi(t))$  und bestimmen die kritischen Stellen von  $h$ . Eine notwendige Bedingung ist  $h'(t) = 0$ , also

$$\partial_1 f(\varphi(t), \psi(t))\varphi'(t) + \partial_2 f(\varphi(t), \psi(t))\psi'(t) = 0.$$

Ist  $g(x, y) = 0$  die Nebenbedingung, so gilt  $g(\varphi(t), \psi(t)) = 0$ , also

$$\partial_1 g(\varphi(t), \psi(t))\varphi'(t) + \partial_2 g(\varphi(t), \psi(t))\psi'(t) = 0.$$

An einer kritischen Stelle von  $h$  sind also, wenn dort nicht gerade  $\varphi' = \psi' = 0$  ist, die Spaltenvektoren  $(\partial_1 f, \partial_1 g)$  und  $(\partial_2 f, \partial_2 g)$  linear abhängig. Somit sind also auch die Zeilenvektoren  $(\partial_1 f, \partial_2 f)$  und  $(\partial_1 g, \partial_2 g)$  linear abhängig. Es gibt also, wenn an der kritischen Stelle  $(\partial_1 g, \partial_2 g) \neq (0, 0)$  ist, eine Zahl  $\lambda$  mit

$$\begin{aligned}\partial_1 f - \lambda \partial_1 g &= 0, \\ \partial_2 f - \lambda \partial_2 g &= 0.\end{aligned}$$

Zusammen mit  $g = 0$  sind das drei Gleichungen, aus denen man im Prinzip die Koordinaten des kritischen Punktes und den Wert von  $\lambda$  bestimmen kann. Dabei ist die Zahl  $\lambda$ , der Multiplikator von Lagrange, in der Regel nicht von Interesse, sondern hat nur eine Hilfsfunktion.

Im obigen Beispiel ist  $f(x, y) = x + y$ ,  $g(x, y) = 5x^2 + 6xy + 2y^2 - 1$ . Man erhält also die Gleichungen

$$\begin{aligned}1 - \lambda(10x + 6y) &= 0, \\ 1 - \lambda(6x + 4y) &= 0, \\ 5x^2 + 6xy + 2y^2 - 1 &= 0\end{aligned}$$

mit den Lösungen  $(x, y, \lambda) = (-1, 2, \frac{1}{2})$  und  $(1, -2, -\frac{1}{2})$ .

Wir wollen dieses Verfahren nun auf mehrere Nebenbedingungen ausdehnen und exakt begründen.

**4.12 Satz** (Multiplikatorenregel von Lagrange). *Sei  $M \subseteq \mathbb{R}^n$  offen, seien  $f : M \rightarrow \mathbb{R}$  und  $F : M \rightarrow \mathbb{R}^k$  mit  $1 \leq k < n$  Abbildungen der Klasse  $C^1$ , sei*

$$N := \{X \in M \mid F(X) = 0\}.$$

*Die Einschränkung  $f|_N$  habe an der Stelle  $X_0 \in N$  ein lokales Extremum, und die Abbildung  $F$  mit Koordinatenfunktionen  $f_1, \dots, f_k$  sei an der Stelle  $X_0$  vom Rang  $k$ . Dann gibt es reelle Zahlen  $\lambda_1, \dots, \lambda_k$  mit*

$$\partial_i \left( f - \sum_{j=1}^k \lambda_j f_j \right) (X_0) = 0 \quad \text{für } i = 1, \dots, n.$$

*Beweis.* Nach Satz 4.8 gibt es eine Umgebung  $U \subseteq \mathbb{R}^n$  von  $X_0$ , eine offene Menge  $V \subseteq \mathbb{R}^{n-k}$  und eine injektive Abbildung  $\varphi : V \subseteq \mathbb{R}^{n-k} \rightarrow \mathbb{R}^n$  der Klasse  $C^1$  vom Rang  $n - k$  mit  $U \cap N = \varphi(V)$  (nämlich  $\varphi(Z) = (Z, G(Z))$  mit  $G$  wie in Satz 4.8). Sei  $Z_0$  das Urbild von  $X_0$  unter der Funktion  $\varphi$ . Die Funktion  $h = f \circ \varphi$  hat an der Stelle  $Z_0$  ein lokales Extremum. Daher gilt  $Dh_{Z_0} = 0$ , nach der Kettenregel also

$$Df_{X_0}(D\varphi_{Z_0}(Y)) = 0 \quad \text{für alle } Y \in \mathbb{R}^{n-k}.$$

Wegen  $F \circ \varphi = 0$  gilt auch  $DF_{X_0}(D\varphi_{Z_0}(Y)) = 0$ , also

$$(Df_i)_{X_0}(D\varphi_{Z_0}(Y)) = 0 \quad \text{für alle } Y \in \mathbb{R}^{n-k}.$$

Die im folgenden auftretenden Differentiale von  $f$  und  $f_i$  sind sämtlich an der Stelle  $X_0$  genommen. Wir haben gezeigt, dass

$$Df(H) = 0 \quad \text{und} \quad Df_i(H) = 0 \quad \text{für } i = 1, \dots, k \text{ und } H \in D\varphi_{Z_0}(\mathbb{R}^{n-k})$$

gilt. Sei  $L := D\varphi_{Z_0}(\mathbb{R}^{n-k})$  und  $L^\perp$  der Orthogonalraum in  $\mathbb{R}^n$ . Dann ist  $\dim L^\perp = k$ . Sei  $(W_1, \dots, W_k)$ ,  $W_i \in \mathbb{R}^n$ , eine Basis von  $L^\perp$ . Die Zeilen der Matrix

$$\begin{pmatrix} Df_1(W_1) & \cdots & Df_k(W_1) \\ \vdots & & \vdots \\ Df_1(W_k) & \cdots & Df_k(W_k) \end{pmatrix}$$

sind die Koordinatenvektoren der Bildvektoren  $DF_{X_0}(W_1), \dots, DF_{X_0}(W_k)$ , sind also linear unabhängig. Daher sind auch die Spalten der Matrix linear unabhängig, und es gibt reelle Zahlen  $\lambda_1, \dots, \lambda_k$  mit

$$\begin{pmatrix} Df(W_1) \\ \vdots \\ Df(W_k) \end{pmatrix} = \lambda_1 \begin{pmatrix} Df_1(W_1) \\ \vdots \\ Df_1(W_k) \end{pmatrix} + \cdots + \lambda_k \begin{pmatrix} Df_k(W_1) \\ \vdots \\ Df_k(W_k) \end{pmatrix}.$$

Da jeder Vektor  $E \in \mathbb{R}^n$  in der Form  $E = H + c_1 W_1 + \cdots + c_k W_k$  mit  $H \in L$  und  $c_1, \dots, c_k \in \mathbb{R}$  dargestellt werden kann, folgt

$$Df(E) = \lambda_1 Df_1(E) + \cdots + \lambda_k Df_k(E).$$

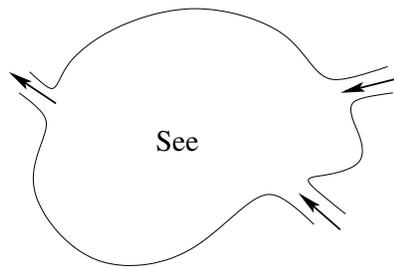
Mit  $E = E_i$ ,  $i = 1, \dots, n$ , folgt die Behauptung. ■

# 12 Gewöhnliche Differentialgleichungen

## 12.1 Motivation

Gewöhnliche Differentialgleichungen (ODE's) beschreiben Prozesse, die zeitabhängig sind. Sie liefern Einsichten in diese Prozesse, die oft anders nicht zu erhalten sind. Allerdings sind ODE's nur Modelle für die Realität. Selbst wenn man ODE's exakt lösen kann, muss dies nicht exakt die Realität sein.

**Beispiel (Fischpopulation in einem See).**



Wir betrachten einen See voller Fische. Dabei bezeichnen wir die “Anzahl” der Fische (in Tonnen) zum Zeitpunkt  $t$  mit  $y(t)$ . Wie viele Fische kann die Fischindustrie fangen, ohne dass die Fischpopulation ausstirbt?

**Annahme:** Der See ist isoliert, d.h. entweder ist die Anzahl der Fische, die durch Flüsse in den See kommen gleich der Anzahl der Fische, die den See durch Flüsse verlassen oder es existieren keine zu- bzw. abfließenden Flüsse.

Wir modellieren die Fischpopulation über ein *Erhaltungsgesetz*

$$\text{Änderung der Population} = \text{“Zuwachs”} - \text{“Abgang”}.$$

Dabei identifizieren wir

$$\begin{aligned}\text{Änderung der Population} &\cong y'(t), \\ \text{Zuwachs} &\cong \text{Geburtenrate } G(t), \\ \text{Abgang} &\cong \text{Todesrate } T(t) + \text{Fischfangrate } H(t),\end{aligned}$$

und nehmen an, dass die Geburtenrate und die Todesrate proportional zur Population sind, d.h.  $G(t) = b y(t)$  und  $T(t) = m y(t)$  für  $b, m > 0$ . Damit erhalten wir die Gleichung

$$y'(t) = (b - m) y(t) - H(t).$$

Weiter nehmen wir an, dass  $b - m =: a > 0$  gilt. Dabei kann die Größe  $a$  über Beobachtungen gemessen werden und  $H(t)$  kann kontrolliert werden. Zu einem gegebenen Anfangszustand  $y(t_0) = y_0$  zum Zeitpunkt  $t_0$  können wir nun berechnen, ob sich das Fischen lohnt. Betrachten wir die einfache Situation  $H(t) = H$ , d.h. die Fischfangrate ist konstant in der Zeit, und wählen  $t_0 = 0$ , so erhalten wir das erste Modell

$$\begin{aligned}y' &= a y - H, \\ y(0) &= y_0.\end{aligned}\tag{1.1}$$

Dies ist eine ähnliche Gleichung wie bei der Zinsrechnung, wir erwarten also eine Exponentialfunktion als Lösung. Diese leiten wir im folgenden heuristisch her, eine rigorose Herleitung folgt später. Multiplizieren wir die erste Zeile von (1.1) mit dem positiven Wert  $e^{-at}$  so folgt

$$\begin{aligned}-e^{-at}H &= e^{-at}(y'(t) - a y(t)) \\ &= e^{-at}(-a) y(t) + e^{at}y'(t) = (e^{-at}y(t))'.\end{aligned}$$

Integrieren wir diese Gleichung bzgl.  $t$  (bzw. bilden wir die Stammfunktion), so erhalten wir

$$e^{-at}y(t) = \frac{H}{a} e^{-at} + C.$$

Multiplizieren wir noch mit  $e^{at}$ , so folgt

$$y(t) = \frac{H}{a} + C e^{at},$$

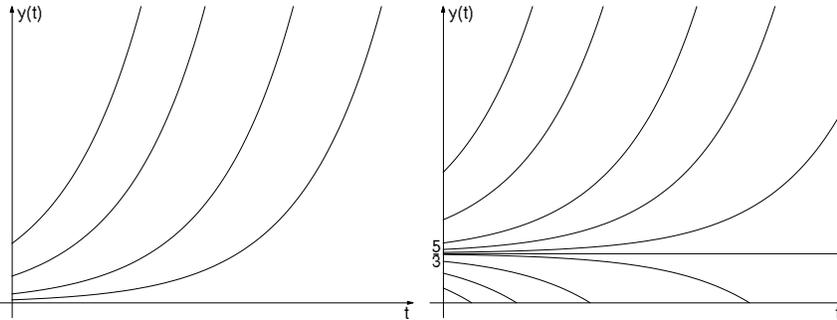
d.h. falls  $y$  eine Lösung von (1.1) ist, hat sie diese Form. Da in der Rechnung alle Operationen umkehrbar sind, löst das so definierte  $y(t)$  die erste Gleichung in (1.1). Setzen wir  $t = 0$  in die Gleichung für  $y(t)$  ein und verwenden die Bedingung  $y(0) = y_0$  für den Anfangszustand, so folgt

$$\frac{H}{a} + C = y_0 \quad \text{bzw.} \quad C = y_0 - \frac{H}{a}.$$

Also ist

$$y(t) = \frac{H}{a} + \left(y_0 - \frac{H}{a}\right)e^{at}$$

für  $t \geq 0$  Lösung von (1.1). Betrachten wir diese Lösung, so wächst die Population bei einer Fischfangrate echt größer der Anfangspopulation exponentiell, bei einer Fischfangrate gleich der Anfangspopulation bleibt die Population stabil und sonst sterben die Fische aus (vergleiche Abbildung 1.1). Dies ist



**Abb. 1.1.** Lösungen von (1.1) zu verschiedenen Anfangswerten. Die Parameter wurden im linken Bild mit  $H = 0$ ,  $a = 1$ , im rechten Bild mit  $H = \frac{5}{3}$ ,  $a = 1$  gewählt.

offensichtlich unrealistisch und somit unser Modell zu einfach gewählt. Insbesondere die Annahme der Proportionalität zwischen der Todesrate und der Population ist problematisch. In abgeschlossenen Seen ist die Todesrate höher als linear abhängig zur Population, z.B.  $T(t) = cy^2(t)$ . Dies führt zu einem neuen Modell

$$\begin{aligned} y'(t) &= by(t) - cy^2(t) - H(t), \\ y(0) &= y_0. \end{aligned} \tag{1.2}$$

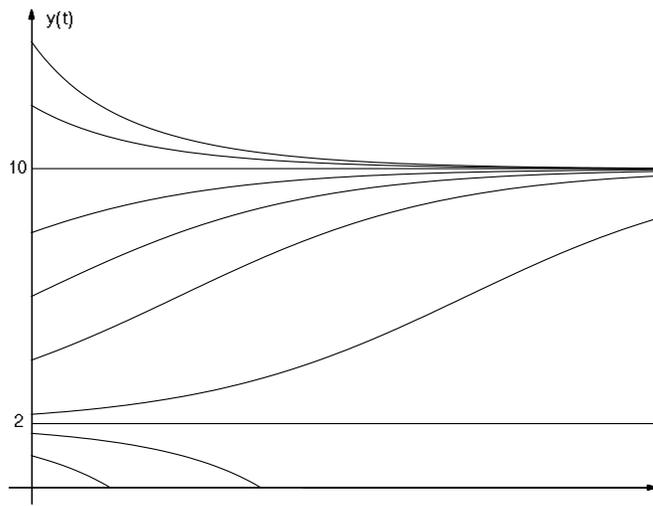
Wählen wir speziell  $b = 1$ ,  $c = \frac{1}{12}$  und  $H = \frac{5}{3}$ , so gilt

$$\begin{aligned} y'(t) &= y(t) - \frac{1}{12}y^2(t) - \frac{5}{3} \\ &= -\frac{1}{12}(y(t) - 10)(y(t) - 2) =: f(y(t)). \end{aligned}$$

Wir betrachten zuerst die Gleichung

$$y' = f(y)$$

heuristisch. Für  $y \in (2, 10)$  ist  $f(y) > 0$ , d.h.  $y' > 0$  und somit  $y$  wachsend, für  $y \in (0, 2)$  oder  $y > 10$  ist  $f(y) < 0$ , d.h.  $y' < 0$  und somit  $y$  fallend. Für



**Abb. 1.2.** Lösungen von (1.2) zu verschiedenen Anfangswerten. Die Parameter wurden als  $b = 1$ ,  $c = \frac{1}{12}$  und  $H = \frac{5}{3}$  gewählt.

$y = 2$  und  $y = 10$  ist  $f(y) = 0$  und somit  $y$  konstant. Dies ist in Abbildung 1.2 graphisch dargestellt.

Wir verwenden nun die Methode der *Separation der Variablen*, um diese Heuristik zu untermauern. So folgt aus

$$\frac{dy}{dt} = -\frac{1}{12}(y-10)(y-2)$$

für  $y \neq 2, 10$  die Gleichung

$$\frac{dy}{(y-10)(y-2)} = -\frac{dt}{12}$$

bzw.

$$\frac{1}{8} \left( \frac{1}{y-10} - \frac{1}{y-2} \right) dy = -\frac{dt}{12}.$$

Da auf jeder Seite der Gleichung nur eine Variable vorkommt, können wir auf beiden Seiten die Stammfunktion (bezüglich der jeweiligen Variablen) bilden und erhalten

$$\frac{1}{8} (\ln |y-10| - \ln |y-2|) = -\frac{t}{12} + c.$$

Verwenden wir nun die Rechenregeln für den Logarithmus folgt

$$\ln \left| \frac{y-10}{y-2} \right| = 8c - \frac{2}{3}t.$$

Nun wenden wir die Exponentialfunktion auf diese Gleichung an und erhalten

$$\left| \frac{y-10}{y-2} \right| = e^{8c} e^{-\frac{2}{3}t}.$$

Setzen wir  $c_0 := \exp(8c)$ , impliziert diese Gleichung  $c_0 > 0$ . Lösen wir nun den Betrag per Fallunterscheidung auf, so müssen wir  $c_0 \in \mathbb{R}$  zulassen. Somit können wir die Gleichung nach  $y$  auflösen und erhalten

$$y(t) = \frac{10 - 2c_0 e^{-\frac{2}{3}t}}{1 - c_0 e^{-\frac{2}{3}t}}.$$

Aus der Forderung  $y(0) = y_0$  folgt

$$c_0 = \frac{y_0 - 10}{y_0 - 2}.$$

Setzen wir also

$$y(t) := \frac{10 - 2 \frac{y_0 - 10}{y_0 - 2} e^{-\frac{2}{3}t}}{1 - \frac{y_0 - 10}{y_0 - 2} e^{-\frac{2}{3}t}}$$

für  $y_0 \neq 2, 10$  (und somit  $y(t) \neq 2, 10$ ), so löst  $y$  die Gleichung (1.2). Weiter sind die konstanten Funktionen  $y(t) := 2$  bzw.  $y(t) := 10$  Lösungen der Gleichung (1.2) zu den Anfangswerten  $y_0 = 2$  bzw.  $y_0 = 10$  und stellen somit ein Gleichgewicht zwischen Fischfang und Populationswachstum dar. Dieses Modell ist realistischer als das erste und erlaubt Vorhersagen: So stabilisiert sich der Fischbestand trotz fischen bei einer Anfangspopulation größer gleich 2 Tonnen, bei einer geringeren Anfangspopulation stirbt diese aus (vergleiche Abbildung 1.2).

Für allgemeine, aber konstante  $b, c$  und  $H$  kann man ebenfalls Lösungsformeln für (1.2) herleiten, bei zeitabhängigen Fischfangraten sind diese nicht bekannt.

## 12.2 Existenztheorie

Für  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  heißt die Gleichung

$$y^{(n)}(t) = f(t, y(t), y'(t), \dots, y^{(n-1)}(t)) \quad (2.1)$$

explizite *Differentialgleichung n-ter Ordnung*. Fordert man zusätzlich

$$y(t_0) = y_0, y'(t_0) = y_1, \dots, y^{(n-1)}(t_0) = y_{n-1}$$

für  $(t_0, y_1, \dots, y_{n-1}) \in \mathbb{R}^{n+1}$ , so spricht man von einem *Anfangswertproblem*.

**2.2 Definition.** Sei  $I$  ein Intervall. Eine Funktion  $y : I \rightarrow \mathbb{R}$  heißt Lösung von (2.1) im Intervall  $I$ , falls  $y$  in  $I$   $n$ -mal differenzierbar ist und

$$y^n(t) = f(t, y(t), y'(t), \dots, y^{n-1}(t))$$

für alle  $t \in I$  identisch erfüllt ist.

Man beachte, dass die Mengen  $I = (a, b)$ ,  $I = [a, b)$  oder  $I = [a, b]$ , mit  $a, b \in \mathbb{R}$  und  $a < b$ , alle zulässige Intervalle sind.

Sei  $F : \Omega \subseteq \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ . Das Gleichungssystem

$$Y'(t) = F(t, Y(t)) \quad (2.3)$$

heißt *System von Differentialgleichungen 1. Ordnung*. Falls  $F = (f_1, \dots, f_n)$  und  $Y = (y_1, \dots, y_n)$ , so können wir (2.3) komponentenweise schreiben

$$\begin{aligned} y_1'(t) &= f_1(t, y_1(t), \dots, y_n(t)) \\ &\vdots \\ y_n'(t) &= f_n(t, y_1(t), \dots, y_n(t)). \end{aligned} \quad (2.4)$$

Ergänzt man (2.3) durch die Bedingung

$$Y(t_0) = Y_0$$

für ein  $(t_0, Y_0) \in \Omega$  bzw. (2.4) durch

$$y_1(t_0) = y_1^0, \dots, y_n(t_0) = y_n^0$$

für ein  $(t_0, y_1^0, \dots, y_n^0) \in \Omega$ , so spricht man wieder von einem *Anfangswertproblem*.

**2.5 Definition.** Sei  $I$  ein Intervall. Eine vektorwertige Funktion  $Y : I \rightarrow \mathbb{R}^n$  heißt Lösung des Systems (2.3) in dem Intervall  $I$ , falls  $Y$  in  $I$  differenzierbar ist und

$$Y'(t) = F(t, Y(t))$$

für alle  $t \in I$  identisch erfüllt ist.

**Bemerkung.** (i) Für eine Lösung von (2.3) gilt insbesondere  $(t, Y(t)) \in \Omega$  für alle  $t \in I$ .

(ii) Die gesamte Existenztheorie wird direkt für Systeme von Differentialgleichungen bewiesen, da kein Mehraufwand im Vergleich zur Behandlung einer Differentialgleichung entsteht. Man sollte sich die Aussagen trotzdem für den anschaulich einfacheren Fall von einer Differentialgleichung klar machen.

(iii) Die explizite Differentialgleichung  $n$ -ter Ordnung (2.1) lässt sich in ein System von  $n$  Differentialgleichung 1. Ordnung umschreiben:

$$\begin{aligned} y_1'(t) &= y_2(t), \\ y_2'(t) &= y_3(t), \\ &\vdots \\ y_{n-1}'(t) &= y_n(t), \\ y_n'(t) &= f(t, y_1(t), \dots, y_n(t)). \end{aligned} \tag{2.6}$$

Denn sei  $y$  Lösung von (2.1). Dann ist  $Y := (y, y', \dots, y^{n-1})$  eine Lösung von (2.6). Sei umgekehrt  $Y = (y_1, \dots, y_n)$  Lösung von (2.6), so ist  $y := y_1$  Lösung von (2.1). Analog transformieren sich die entsprechenden Anfangswertprobleme.

Wir formulieren nun eine Voraussetzung an  $F$ , um die Lösbarkeit eines Systems von Differentialgleichungen 1. Ordnung zu sichern.

**Grundvoraussetzung (S).** *Es existiert eine offene Menge  $\Omega \subseteq \mathbb{R}^{n+1}$  so, dass  $F : \Omega \subseteq \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$  stetig ist.*

**2.7 Lemma.** *Sei die Voraussetzung (S) erfüllt und sei  $(t_0, Y_0) \in \Omega$ ,  $I$  ein Intervall und  $t_0 \in I$ . Dann ist  $Y$  genau dann eine Lösung von (2.3) mit  $Y(t_0) = Y_0$  auf dem Intervall  $I$ , wenn  $Y$  eine stetige Lösung der Integralgleichung*

$$Y(t) = Y_0 + \int_{t_0}^t F(s, Y(s)) ds, \quad t \in I \tag{2.8}$$

ist.

Das Integral für vektorwertige Funktionen ist komponentenweise definiert, d.h. für  $Y(t) = (y_1(t), \dots, y_n(t))$  gilt

$$\int_a^b Y(s) ds := \left( \int_a^b y_1(s) ds, \dots, \int_a^b y_n(s) ds \right),$$

falls die Integrale  $\int_a^b y_i(s) ds$ ,  $i = 1, \dots, n$  existieren.

*Beweis (Lemma 2.7). „ $\Rightarrow$ “* Sei  $Y$  Lösung von (2.3) mit  $Y(t_0) = Y_0$ . Somit ist  $Y$  differenzierbar in  $I$  und insbesondere stetig. Damit ist auch die Abbildung  $t \mapsto F(t, Y(t))$  stetig. Aufgrund der Gleichung  $Y'(t) = F(t, Y(t))$  für  $t \in I$  und dem Hauptsatz der Differential- und Integralrechnung (Analysis I, Kapitel 7, Satz 3.1) folgt für alle  $t \in I$

$$Y(t) - Y_0 = Y(t) - Y(t_0) = \int_{t_0}^t Y'(s) ds = \int_{t_0}^t F(s, Y(s)) ds.$$

Also ist  $Y$  eine stetige Lösung der Integralgleichung (2.8).

„ $\Leftarrow$ “ Sei  $Y$  stetige Lösung von (2.8). Nach dem Hauptsatz der Differential- und Integralrechnung ist  $Y$  stetig differenzierbar in  $I$  und es gilt

$$Y'(t) = \frac{d}{dt} \int_{t_0}^t F(s, Y(s)) ds = F(t, Y(t)),$$

d.h.  $Y$  ist stetig differenzierbar. Weiter gilt für den Anfangswert

$$Y(t_0) = Y_0 + \int_{t_0}^{t_0} F(s, Y(s)) ds = Y_0.$$

Somit ist  $Y$  eine Lösung von (2.3) auf dem Intervall  $I$  mit  $Y(t_0) = Y_0$ . ■

Wir wollen im Folgenden zeigen, dass die Bedingung (S) für die Existenz einer Lösung von (2.3) ausreicht. Diese Voraussetzung ist schwächer als die Lipschitz-Stetigkeit, die wir im Existenzsatz von Picard-Lindelöf (Satz 3.5 aus Kapitel 9) benötigen haben. Doch zunächst benötigen wir einige Hilfsmittel.

**2.9 Definition.** Eine Menge  $M = \{f_\alpha\}_{\alpha \in A}$  von stetigen Funktionen auf dem Intervall  $[a, b]$ , d.h.  $M \subseteq C([a, b])$ , heißt gleichgradig stetig, wenn zu jedem  $\varepsilon > 0$  ein  $\delta > 0$  existiert, so dass für alle  $s, t \in [a, b]$  mit  $|s - t| < \delta$  und alle  $\alpha \in A$  gilt:  $|f_\alpha(s) - f_\alpha(t)| < \varepsilon$ .

Eine stetige Funktion auf einem kompakten Intervall  $[a, b]$  ist gleichmäßig stetig (Satz 3.5 im Kapitel 4), d.h.

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall s, t \in [a, b] \text{ mit } |s - t| < \delta : |f(s) - f(t)| < \varepsilon.$$

Bei der gleichgradigen Stetigkeit ist neu, dass  $\delta$  unabhängig von  $f_\alpha$  ist.

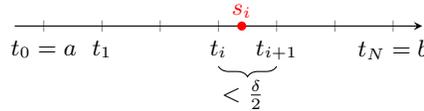
**2.10 Lemma.** Sei  $(f_n)_{n \in \mathbb{N}} \subseteq C([a, b])$  eine Folge gleichgradig stetiger Funktionen, die auf einer dichten Menge  $M \subseteq [a, b]$  gegen eine Funktion  $f : M \rightarrow \mathbb{R}$  konvergiert. Dann existiert eine Fortsetzung  $\tilde{f} : [a, b] \rightarrow \mathbb{R}$  der Funktion  $f$  (d.h. es gilt  $\tilde{f}(x) = f(x)$  für alle  $x \in M$ ), so dass  $(f_n)$  gleichmäßig auf  $[a, b]$  gegen  $\tilde{f}$  konvergiert. Insbesondere ist  $\tilde{f}$  stetig auf  $[a, b]$ .

Zur Erinnerung: (i)  $M$  dicht in  $[a, b]$  bedeutet, dass für alle  $s \in [a, b]$  eine Folge  $(s_n)_{n \in \mathbb{N}} \subseteq M$  mit  $s_n \rightarrow s$  existiert (z.B.  $\mathbb{Q}$  ist dicht in  $\mathbb{R}$ ). (ii) Die Funktionenfolge  $(f_n)_{n \in \mathbb{N}}$  konvergiert gleichmäßig gegen eine Funktion  $f$  auf der Menge  $D$ , falls für alle  $\varepsilon > 0$  ein  $N \in \mathbb{N}$  existiert, so dass für alle  $n \geq N$  und für alle  $x \in D$  gilt:  $|f_n(x) - f(x)| < \varepsilon$ .

*Beweis (Lemma 2.10).* Sei  $\varepsilon > 0$  und  $(f_n)_{n \in \mathbb{N}}$  gleichgradig stetig. Somit existiert ein  $\delta > 0$ , so dass

$$\forall n \in \mathbb{N} \forall s, t \in [a, b] \text{ mit } |s - t| < \delta : |f_n(s) - f_n(t)| < \varepsilon.$$

Wir wählen nun ein  $N \in \mathbb{N}$  mit  $\frac{|b-a|}{N} < \frac{\delta}{2}$  und zerlegen das Intervall  $[a, b]$  in die Intervalle  $[t_i, t_{i+1}]$ ,  $i = 0, \dots, N-1$ , wobei  $t_0 := a$  und  $t_i := a + i \frac{|b-a|}{N}$  gewählt werden (vergleiche Abbildung 2.1). Somit gilt  $t_i < t_{i+1}$ ,  $t_N = b$  sowie



**Abb. 2.1.** Zerlegung des Intervalls  $[a, b]$  in  $N$  gleichgroße Teilintervalle.

$|t_i - t_{i+1}| < \frac{\delta}{2}$ . Da  $M$  dicht in  $[a, b]$  liegt, existieren zu den Mittelpunkten der Intervalle  $[t_i, t_{i+1}]$  eine Folge aus Elementen von  $M$ , die gegen diese konvergieren. Da die Mittelpunkte einen positiven Abstand zum Rand der Intervalle haben, existieren insbesondere  $s_i \in [t_i, t_{i+1}] \cap M$  für  $i = 0, \dots, N-1$ . Weiter konvergiert nach Voraussetzung  $(f_n)_{n \in \mathbb{N}}$  auf  $M$  gegen  $f$  und somit ist  $(f_n)_{n \in \mathbb{N}}$  eine Cauchyfolge auf  $M$ . Da wir nur  $N$  viele  $s_i$  haben, folgern wir

$$\exists n_0 \in \mathbb{N} \forall m, n \geq n_0 \forall i = 0, \dots, N-1 : |f_n(s_i) - f_m(s_i)| < \varepsilon.$$

Wählen wir nun ein beliebiges  $s \in [a, b]$ . Nach Konstruktion existiert ein  $i \in \{0, \dots, N-1\}$  mit  $|s_i - s| < \delta$ . Für  $n, m \geq n_0$  gilt dann

$$\begin{aligned} |f_n(s) - f_m(s)| &\leq |f_n(s) - f_n(s_i)| + |f_n(s_i) - f_m(s_i)| + |f_m(s_i) - f_m(s)| \\ &< \varepsilon + \varepsilon + \varepsilon = 3\varepsilon. \end{aligned}$$

Bilden wir das Supremum über alle  $s \in [a, b]$  folgt, dass  $(f_n)_{n \in \mathbb{N}}$  eine Cauchyfolge bezüglich der Maximumsnorm ist. Somit konvergiert  $(f_n)_{n \in \mathbb{N}}$  auf  $[a, b]$  gleichmäßig gegen ein  $\tilde{f} \in C([a, b])$  (siehe Analysis I, Kapitel 8, Satz 1.1 und Satz 1.2). Da gleichmäßige Konvergenz die punktweise Konvergenz impliziert und der Grenzwert eindeutig ist, folgt  $\tilde{f}|_M = f$ . ■

**2.11 Satz (Arzela-Ascoli).** Sei  $(f_n)_{n \in \mathbb{N}} \subseteq C([a, b])$  eine Folge gleichmäßig stetiger, gleichmäßig beschränkter Funktionen, d.h. es existiert eine Konstante  $M$ , so dass für alle  $n \in \mathbb{N}$  und  $s \in [a, b]$  gilt  $|f_n(s)| \leq M$ . Dann existiert eine Teilfolge, die gleichmäßig gegen eine Funktion  $f \in C([a, b])$  konvergiert, d.h.

$$\exists (f_{n_k})_{k \in \mathbb{N}} \subseteq (f_n)_{n \in \mathbb{N}}, \exists f \in C([a, b]) \text{ mit } f_{n_k} \rightrightarrows f \text{ für } k \rightarrow \infty.$$

*Beweis.* Sei  $M = \{x_i\}_{i \in \mathbb{N}}$  eine Aufzählung der rationalen Zahlen aus  $[a, b]$ . Da nach Voraussetzung die Folge  $(f_n(x_1))_{n \in \mathbb{N}}$  beschränkt ist, existiert nach dem Satz von Bolzano–Weierstraß (Analysis I, Kapitel 4, Satz 1.3) eine Teilfolge  $(f_{n_k^1})_{k \in \mathbb{N}} \subseteq (f_n)_{n \in \mathbb{N}}$  und ein  $f(x_1) \in \mathbb{R}$  mit

$$f_{n_k^1}(x_1) \rightarrow f(x_1) \text{ für } k \rightarrow \infty.$$

Nun ist die Folge  $(f_{n_k^1}(x_2))_{k \in \mathbb{N}}$  beschränkt und es existiert wiederum nach dem Satz von Bolzano–Weierstraß eine Teilfolge  $(f_{n_k^2})_{k \in \mathbb{N}} \subseteq (f_{n_k^1})_{k \in \mathbb{N}}$  und ein  $f(x_2) \in \mathbb{R}$  mit

$$f_{n_k^2}(x_2) \rightarrow f(x_2) \text{ für } k \rightarrow \infty.$$

Führen wir dies für alle  $i \in \mathbb{N}$  fort, so erhalten wir ein System von Teilfolgen, die aufgrund der Schachtelung  $(f_{n_k^i})_{k \in \mathbb{N}} \subseteq (f_{n_k^{i-1}})_{k \in \mathbb{N}} \subseteq \dots \subseteq (f_n)_{n \in \mathbb{N}}$  für immer weitere Punkte konvergieren:

$$\begin{array}{ll} f_{n_1^1}, f_{n_2^1}, f_{n_3^1}, \dots & \text{konvergiert für } x = x_1 \\ f_{n_1^2}, f_{n_2^2}, f_{n_3^2}, \dots & \text{konvergiert für } x = x_1, x_2 \\ \vdots & \\ f_{n_1^i}, f_{n_2^i}, f_{n_3^i}, \dots & \text{konvergiert für } x = x_1, x_2, \dots, x_i \\ \vdots & \end{array}$$

Wählen wir nun die Diagonalfolge  $(f_{n_k^k})_{k \in \mathbb{N}}$ , so konvergiert diese für alle  $x_i$  gegen  $f(x_i)$ , denn ab dem  $i$ -ten Folgenglied ist  $(f_{n_k^k})_{k \in \mathbb{N}}$  eine Teilfolge von  $(f_{n_k^i})_{k \in \mathbb{N}}$  und somit konvergent. Wir haben also eine Teilfolge gefunden, die auf einer dichten Menge gegen die Funktion  $f$  (gegeben durch  $f(x_i)$  am Punkt  $x_i$ ) konvergiert. Wenden wir nun Lemma 2.10 an, so ist der Satz bewiesen. ■

**2.12 Folgerung.** Sei  $F_k : [a, b] \rightarrow \mathbb{R}^n$ ,  $k \in \mathbb{N}$ , eine Folge gleichgradig stetiger vektorwertiger Funktionen, d.h.

$$\forall \varepsilon > 0 \exists \delta > 0 : \forall k \in \mathbb{N}, \forall s, t \in [a, b] \text{ mit } |s - t| < \delta : \|F_k(s) - F_k(t)\| < \varepsilon,$$

die gleichmäßig beschränkt sind. Dann existiert eine Teilfolge  $(F_{k_\ell})_{\ell \in \mathbb{N}}$ , die gleichmäßig konvergiert.

*Beweis.* Schreiben wir  $F_k = (f_1^k, \dots, f_n^k)$ ,  $F = (f_1, \dots, f_n)$ , so ist die gleichmäßige Konvergenz von  $F_k$  gegen  $F$  äquivalent zur gleichmäßigen Konvergenz der  $f_i^k$  gegen  $f_i$ ,  $i = 1, \dots, n$ . Da nach Voraussetzung die  $f_i^k$ ,  $i = 1, \dots, n$ , ebenfalls gleichgradig stetig und gleichmäßig beschränkt sind, folgt die Behauptung mit Satz 2.11, sukzessive angewendet auf die Koordinatenfunktionen  $(f_i^k)$  und die sich ergebenden Teilfolgen  $(f_i^{k_\ell})_{\ell \in \mathbb{N}}$ ,  $i = 1, \dots, n$ . ■

**2.13 Satz (Peano).** Erfülle  $F$  die Voraussetzung (S), sei  $(t_0, Y_0) \in \Omega$  und seien  $a, b > 0$  so, dass

$$Q := \{(t, Y) \in \mathbb{R}^{n+1} \mid |t - t_0| \leq a, \|Y - Y_0\| \leq b\}$$

eine Teilmenge von  $\Omega$  ist, d.h.  $Q \subseteq \Omega$ . Sei  $K > 0$  derart, dass  $\|F(t, Y)\| \leq K$  für alle  $(t, Y) \in Q$  gilt. Dann besitzt das System gewöhnlicher Differentialgleichungen (2.3) mit  $Y(t_0) = Y_0$  eine stetig differenzierbare Lösung  $Y(\cdot)$  auf dem Intervall  $[t_0 - c, t_0 + c] =: I$  wobei die Konstante  $c = \min(a, \frac{b}{K})$  gewählt werden kann.

*Beweis.* Nach Lemma 2.7 genügt es, eine stetige Funktion  $Y$  mit

$$Y(t) = Y_0 + \int_{t_0}^t F(s, Y(s)) ds, \quad t \in I$$

zu finden. Wir wollen eine Lösung der Integralgleichung per Approximation finden: Dazu definieren wir für  $t \in I$  die Funktion  $Y_0(t) := Y_0$  und für  $n \in \mathbb{N}$  die Funktionen

$$Y_{n+1}(t) := Y_0 + \int_{t_0}^t F(s, Y_n(s)) ds \quad t \in I.$$

Für die Wohldefiniertheit von  $Y_{n+1} : I \rightarrow \mathbb{R}^n$  reicht es die Stetigkeit von  $Y_n$  und  $(t, Y_n(t)) \in Q \subseteq \Omega$  für  $t \in I$  zu zeigen. Dies machen wir per vollständiger Induktion:

**Induktionsanfang:** Offenbar ist  $Y_0$  stetig und es gilt  $(t, Y_0) \in Q$  für  $t \in I$  nach der Wahl von  $c$  und der Definition von  $Q$ .

**Induktionshypothese:** Gelte  $Y_n$  ist stetig und  $(t, Y_n(t)) \in Q$  für  $t \in I$ .

**Induktionsschritt:** Für  $t \in I$  folgt einerseits  $|t - t_0| \leq a$  wegen  $c \leq a$  und andererseits

$$\begin{aligned} \|Y_{n+1}(t) - Y_0\| &= \left\| \int_{t_0}^t F(s, Y_n(s)) ds \right\| \\ &\leq \int_{t_0}^t \|F(s, Y_n(s))\| ds \\ &\leq K |t - t_0| \leq K c \leq K \frac{b}{K} = b. \end{aligned}$$

Somit gilt  $(t, Y_{n+1}(t)) \in Q$ . Da nach Induktionshypothese  $Y_n$  stetig ist und  $F$  nach Bedingung (S) auf  $\Omega$  stetig ist, folgt mit dem Hauptsatz der Differential- und Integralrechnung (Analysis I, Kapitel 7, Satz 3.1) die Stetigkeit von  $Y_{n+1}$  und der Induktionsschritt ist bewiesen. Mit der gleichen Rechnung wie im Beweis der Wohldefiniertheit folgt für alle  $n \in \mathbb{N}$

$$\sup_{t \in I} \|Y_n(t)\| \leq \sup_{t \in I} \|Y_n(t) - Y_0\| + \|Y_0\| \leq b + \|Y_0\|$$

bzw. für  $n = 0$

$$\sup_{t \in I} \|Y_0(t)\| \leq \sup_{t \in I} \|Y_0\| \leq b + \|Y_0\|.$$

Also ist die Folge  $(Y_n)_{n \in \mathbb{N}_0}$  gleichmäßig beschränkt. Wir zeigen nun, dass diese Folge auch gleichgradig stetig ist: Sei dazu  $\varepsilon > 0$  und  $\delta := \frac{\varepsilon}{K}$ . Dann gilt für alle  $q, t \in I$  mit  $|q - t| < \delta$  und alle  $n \in \mathbb{N}$

$$\begin{aligned} \|Y_n(q) - Y_n(t)\| &= \left\| Y_0 + \int_{t_0}^q F(s, Y_{n-1}(s)) \, ds - Y_0 + \int_{t_0}^t F(s, Y_{n-1}(s)) \, ds \right\| \\ &\leq \left| \int_q^t \|F(s, Y_{n-1}(s))\| \, ds \right| \\ &\leq K |t - q| < K \delta = K \frac{\varepsilon}{K} = \varepsilon. \end{aligned}$$

Für  $n = 0$  folgt trivialerweise

$$\|Y_0(q) - Y_0(t)\| = \|Y_0 - Y_0\| = 0 < \varepsilon.$$

Wir können also den **Satz von Arzela-Ascoli** (bzw. Folgerung 2.12) anwenden und erhalten eine Teilfolge  $(Y_{n_k})_{k \in \mathbb{N}} \subseteq (Y_n)_{n \in \mathbb{N}}$  und eine stetige Funktion  $Y : I \rightarrow \mathbb{R}^n$  mit

$$Y_{n_k} \rightrightarrows Y \quad \text{für } k \rightarrow \infty.$$

Wir bemerken zunächst, dass nach Definition  $Q$  abgeschlossen ist und wegen  $(s, Y_n(s)) \in Q$  für alle  $s \in I$ ,  $n \in \mathbb{N}$  sowie der Konvergenz der  $Y_n$  gegen  $Y$  damit  $(s, Y(s)) \in Q$  für alle  $s \in I$  folgt. Da weiter  $F$  stetig auf der kompakten Menge  $Q$  ist, folgt die gleichmäßige Stetigkeit von  $F$  auf  $Q$ , d.h. für alle  $\varepsilon > 0$  existiert ein  $\delta > 0$ , so dass für alle  $(t, Y), (s, Z) \in Q$  mit  $\|(t, Y) - (s, Z)\| < \delta$  folgt:  $\|F(s, Z) - F(t, Y)\| < \varepsilon$ . Wählen wir nun zu  $\varepsilon > 0$  das  $\delta > 0$  aus der gleichmäßigen Stetigkeit und das  $k_0$  aus der gleichmäßigen Konvergenz von  $Y_{n_k}$  gegen  $Y$  mit

$$\|Y_{n_k}(s) - Y(s)\| < \delta \quad \forall s \in I, k \geq k_0,$$

so haben wir die gleichmäßige Konvergenz von  $F(\cdot, Y_{n_k}(\cdot))$  gegen  $F(\cdot, Y(\cdot))$  gezeigt. Nach Analysis I, Kapitel 7, Satz 2.4 folgt die Konvergenz des Integrals, d.h. mit der Definition der Folge  $Y_n$  gilt

$$\begin{aligned} Y(t) &= \lim_{k \rightarrow \infty} Y_{n_{k+1}}(t) \\ &= \lim_{k \rightarrow \infty} Y_0 + \int_{t_0}^t F(s, Y_{n_k}(s)) \, ds \\ &= Y_0 + \int_{t_0}^t F(s, Y(s)) \, ds. \end{aligned}$$

Somit löst  $Y$  die Integralgleichung und nach Lemma 2.7 haben wir die Existenz einer Lösung von (2.3) bewiesen. ■

Der Satz von Peano ist ein lokaler Existenzsatz, d.h. er liefert nur die Existenz einer Lösung auf einem kleinen Intervall um den Anfangswert. Wir stellen uns nun die Frage, ob wir diese Lösung fortsetzen können, in dem wir den Satz von Peano am Ende des Intervalls erneut anwenden, also für den Anfangswert die Lösung am Ende des Intervalls auswerten.

**2.14 Lemma.** *Erfülle  $F$  die Voraussetzung (S) und sei  $Y : (a, b) \rightarrow \mathbb{R}^n$  eine Lösung von (2.3). Weiter existiere der linksseitige Grenzwert*

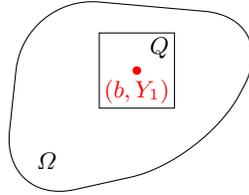
$$\lim_{t \nearrow b} Y(t) =: Y_1$$

und es gelte  $(b, Y_1) \in \Omega$ . Dann existiert ein  $\delta > 0$ , so dass man die Lösung  $Y$  zu einer Lösung auf dem Intervall  $(a, b + \delta]$  fortsetzen kann.

*Beweis.* Da  $(b, Y_1) \in \Omega$  gilt und  $\Omega$  offen ist, existiert eine offene Umgebung von  $(b, Y_1)$ , die vollständig in  $\Omega$  liegt. Insbesondere existieren also  $\alpha, \beta > 0$ , so dass

$$Q := \{(t, Y) \in \mathbb{R}^{n+1} \mid |t - b| \leq \alpha, \|Y - Y_1\| \leq \beta\} \subseteq \Omega$$

gilt (vergleiche Abbildung 2.2). Setzen wir weiter  $K := \sup_{(t, Y) \in Q} \|F(t, Y)\|$ ,



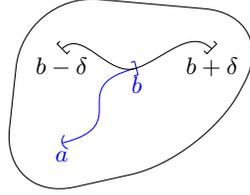
**Abb. 2.2.** Schachtelung der Mengen um den Punkt  $(b, Y_1)$ .

so folgt mit dem Satz von Peano die Existenz einer Lösung  $\tilde{Y}$  von

$$\begin{aligned} \tilde{Y}'(t) &= F(t, \tilde{Y}(t)), \\ \tilde{Y}(b) &= Y_1 \end{aligned}$$

auf dem Intervall  $I = [b - \delta, b + \delta]$ , wobei  $\delta$  durch  $\delta := \min(\alpha, \frac{\beta}{K})$  gegeben ist (siehe Abbildung 2.3). Wir definieren nun für  $t \in (a, b + \delta]$

$$Z(t) := \begin{cases} Y(t) & t \in (a, b) \\ \tilde{Y}(t) & t \in [b, b + \delta]. \end{cases}$$



**Abb. 2.3.** Lösungen  $Y, \tilde{Y}$  auf den Intervallen  $(a, b]$  bzw.  $[b - \delta, b + \delta]$

Offensichtlich ist  $Z$  stetig in  $(a, b + \delta]$  und stetig differenzierbar in  $(a, b)$  sowie  $(b, b + \delta)$ . Weiter erfüllt  $Z$  die Gleichung (2.3) in den Intervallen  $(a, b)$  sowie  $(b, b + \delta]$ . Wir müssen also nur die Existenz von  $Z'(b)$  und die Identität  $Z'(b) = F(b, Z(b))$  zeigen. Dazu bestimmen wir die Rechts- und Linksseitige Ableitung von  $Z$ : Nach Lemma 2.7 ist die Gleichung (2.3) äquivalent zur Integralgleichung

$$Y(t) = Y(t_0) + \int_{t_0}^t F(s, Y(s)) ds \quad t_0, t \in (a, b).$$

Da man nach Voraussetzung  $Y$  durch  $Y(b) := Y_1$  stetig auf das Intervall  $(a, b]$  fortsetzen kann, folgt die Stetigkeit von  $F(\cdot, Y(\cdot))$  auf dem Intervall  $(a, b]$ . Somit gilt die obige Integralgleichung für alle  $t \in (a, b]$ . Nach dem Hauptsatz der Differential- und Integralrechnung (Analysis I, Kapitel 7, Satz 3.1) gilt

$$Y'(t) = F(t, Y(t)) \quad \forall t \in (a, b].$$

Wegen  $Y(b) = Y_1 = \tilde{Y}(b)$  folgt  $Z = Y$  auf  $(a, b]$  und somit die Existenz der linksseitigen Ableitung im Punkt  $b$  mit

$$Z'_-(b) = F(b, Y_1) = F(b, Z(b)).$$

Wenden wir analog Lemma 2.7 auf  $\tilde{Y}$  und das Intervall  $[b, b + \delta]$  an, so folgt

$$\tilde{Y}(t) = Y_1 + \int_b^t F(s, \tilde{Y}(s)) ds \quad t \in [b, b + \delta]$$

Wie oben erhalten wir damit

$$\tilde{Y}'(t) = F(t, \tilde{Y}(t)) \quad t \in [b, b + \delta]$$

und somit die Existenz der rechtsseitigen Ableitung im Punkt  $b$  mit

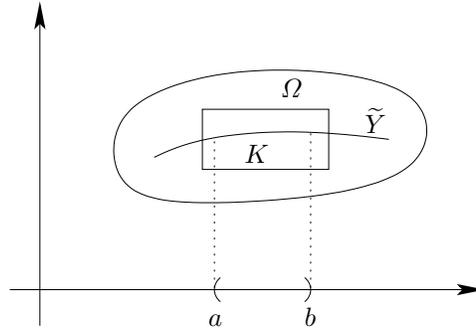
$$Z'_+(b) = F(b, \tilde{Y}(b)) = F(b, Z(b)).$$

Somit existiert  $Z'(b)$  und es gilt  $Z'(b) = F(b, Z(b))$ . Wir haben also eine passende Fortsetzung der Lösung gefunden. ■

**Bemerkung.** Eine analoge Aussage gilt für das linksseitige Fortsetzen: Erfülle  $F$  die Voraussetzung (S) und sei  $Y : (a, b) \rightarrow \mathbb{R}^n$  eine Lösung von (2.3). Weiter existiere der linksseitige Grenzwert  $\lim_{t \searrow a} Y(t) =: Y_1$  und es gelte  $(a, Y_1) \in \Omega$ . Dann existiert ein  $\delta > 0$ , so dass man die Lösung  $Y$  zu einer Lösung auf dem Intervall  $[a - \delta, b)$  fortsetzen kann.

Nun stellt sich die Frage, wie weit man diese Lösungen fortsetzen kann. Es könnte passieren, dass bei der Anwendung von Peano die Intervalle immer kleiner werden und der Prozess stoppt. Dies werden wir im Folgenden untersuchen.

**2.15 Lemma.** Erfülle  $F$  die Voraussetzung (S), sei  $Y : (a, b) \rightarrow \mathbb{R}^n$  eine Lösung von (2.3) und sei  $K$  eine kompakte Teilmenge von  $\Omega$ . Dann lässt sich  $Y$  über  $K$  hinaus fortsetzen, d.h. es existiert eine Funktion  $\tilde{Y} : (c, d) \rightarrow \mathbb{R}^n$  mit  $c \leq a < b \leq d$ , so dass  $\tilde{Y}$  eine Lösung von (2.3) auf dem Intervall  $(c, d)$  ist,  $\tilde{Y} = Y$  auf  $(a, b)$  gilt und  $c < \tau_1 < \tau_2 < d$  existieren, so dass für alle  $t \in (c, \tau_1) \cup (\tau_2, d)$  gilt:  $(t, \tilde{Y}(t)) \notin K$ . Insbesondere ist  $\text{graph}(\tilde{Y}) \not\subseteq K$  (vergleiche Abbildung 2.4).



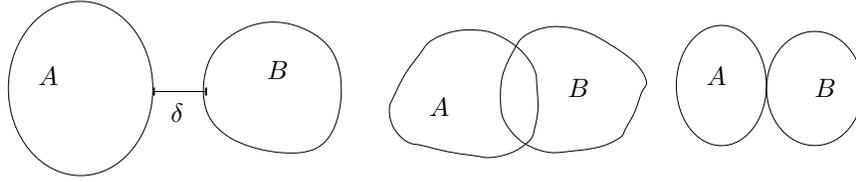
**Abb. 2.4.** Fortsetzung der Lösung über eine kompakte Menge  $K$  hinaus.

*Beweis.* Für zwei Mengen  $A, B \subseteq \mathbb{R}^{n+1}$  ist der Abstand voneinander durch

$$\text{dist}(A, B) := \inf_{X \in A, Y \in B} \|X - Y\|$$

definiert. Im Fall  $\text{dist}(A, B) = \delta > 0$  gilt folglich für alle  $X \in A$  und alle  $Y \in B$  die Ungleichung  $\|X - Y\| \geq \delta > 0$  (vergleiche Abbildung 2.5).

Da  $K$  kompakt in  $\Omega$  liegt, existiert ein  $\delta > 0$  mit  $\text{dist}(K, \partial\Omega) = 3\delta$ , denn sonst wäre  $\text{dist}(K, \partial\Omega) = 0$ . Somit würden Folgen  $(x_n)_{n \in \mathbb{N}} \subseteq K$  und  $(y_n)_{n \in \mathbb{N}} \subseteq \partial\Omega$  mit  $\|x_n - y_n\| \rightarrow 0$  für  $n \rightarrow \infty$  existieren. Da  $K$  kompakt ist, gäbe es ein



**Abb. 2.5.** Abstand zwischen zwei Mengen  $A, B$ . Links:  $\text{dist}(A, B) = \delta$ . Mitte und Rechts:  $\text{dist}(A, B) = 0$ .

$x_0 \in K$  und eine Teilfolge mit  $x_{n_k} \rightarrow x_0$  für  $k \rightarrow \infty$ . Damit würde auch folgen, dass  $y_{n_k} \rightarrow x_0$  für  $k \rightarrow \infty$  und somit wäre  $x_0 \in \partial\Omega$  aufgrund der Abgeschlossenheit des Randes. Da  $\Omega$  offen ist, ist die Menge  $\Omega^c$  abgeschlossen und insbesondere gilt  $\partial\Omega \subseteq \Omega^c$ . Nun haben wir einen Widerspruch, denn  $K \subseteq \Omega$  impliziert  $K \cap \Omega^c = \emptyset$ .

Definieren wir nun

$$K_{2\delta} := \{X \in \Omega \mid \text{dist}(X, K) \leq 2\delta\},$$

so folgt  $K \subseteq K_{2\delta} \subseteq \Omega$  nach der Wahl von  $\delta$  und man sieht leicht, dass  $K_{2\delta}$  beschränkt und abgeschlossen ist. Damit ist  $K_{2\delta}$  kompakt und

$$\sup_{(t,Y) \in K_{2\delta}} \|F(t, Y)\| =: C$$

ist endlich. Wir definieren

$$Q(t_0, Y_0) := \{(t, Y) \in \mathbb{R}^{n+1} \mid |t_0 - t| \leq \delta, \|Y - Y_0\| \leq \delta\}.$$

Für  $(t_0, Y_0) \in K$  folgt für alle  $(t, Y) \in Q(t_0, Y_0)$

$$\|(t_0, Y_0) - (t, Y)\|^2 = |t_0 - t|^2 + \|Y - Y_0\|^2 \leq 2\delta^2$$

und somit  $\text{dist}(Q(t_0, Y_0), K) \leq \sqrt{2}\delta \leq 2\delta$ , d.h.  $Q(t_0, Y_0) \subseteq K_{2\delta}$ .

Wir wählen nun ein  $t_0 \in (a, b)$  und betrachten die Lösung  $Y : (a, b) \rightarrow \mathbb{R}^n$  von (2.3) eingeschränkt auf die Intervalle  $[t_0, b)$  und  $(a, t_0]$  separat. Wir beginnen mit  $Y : [t_0, b) \rightarrow \mathbb{R}^n$  und zeigen, dass eine "rechtsseitige" Fortsetzung  $\tilde{Y}$  der Lösung mit  $\text{graph}(\tilde{Y}) \not\subseteq K$  existiert. Sei  $\text{graph}(Y) \subseteq K$ , da sonst  $\tilde{Y} = Y$  gewählt werden kann. Damit gilt:  $(t, Y(t)) \in \text{graph}(Y) \subseteq K \subseteq K_{2\delta}$  für  $t \in [t_0, b)$  und es folgt mit der Lösungseigenschaft

$$\|Y'(t)\| = \|F(t, Y(t))\| \leq C \quad t \in [t_0, b).$$

Wählen wir eine Folge  $t_n \nearrow b$ , so folgt mit dem Mittelwertsatz

$$\|Y(t_n) - Y(t_m)\| \leq \|Y'(x_{n,m})(t_n - t_m)\| \leq C |t_n - t_m|.$$

Somit ist  $(Y(t_n))_{n \in \mathbb{N}}$  eine Cauchyfolge in  $\mathbb{R}^n$  und es existiert ein  $Y_1 \in \mathbb{R}^n$  mit  $Y(t_n) \rightarrow Y_1$  für  $n \rightarrow \infty$ . Für eine weitere Folge  $\tilde{t}_n \nearrow b$  mit  $Y(\tilde{t}_n) \rightarrow \tilde{Y}_1$  für  $n \rightarrow \infty$  folgt wieder mit dem Mittelwertsatz

$$\begin{aligned} \|Y_1 - \tilde{Y}_1\| &\leq \|Y_1 - Y(t_n)\| + \|Y(t_n) - Y(\tilde{t}_n)\| + \|Y(\tilde{t}_n) - \tilde{Y}_1\| \\ &\leq \|Y_1 - Y(t_n)\| + C(|t_n - b| + |b - \tilde{t}_n|) + \|Y(\tilde{t}_n) - \tilde{Y}_1\| \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Somit ist der Grenzwert  $Y_1$  unabhängig von der gewählten Folge und

$$\lim_{t \nearrow b} Y(t) = Y_1$$

existiert. Da  $K$  abgeschlossen ist, folgt mit der Konvergenz  $(t_n, Y(t_n)) \rightarrow (b, Y_1)$ , dass  $(b, Y_1) \in K$  und somit  $(b, Y_1) \in \Omega$  gilt. Verwenden wir nun Lemma 2.14, so existiert eine Fortsetzung  $\tilde{Y}$  der Lösung auf einem Intervall  $[t_0, b + \rho]$ . Betrachten wir den Beweis von Lemma 2.14 genauer, so kann in unserer Situation  $\rho = \min(\delta, \frac{\delta}{C})$  gewählt werden. Gilt nun  $\text{graph}(\tilde{Y}) \not\subseteq K$  sind wir fertig, ansonsten wiederholen wir den Prozess. Da jede weitere Fortsetzung wieder um das feste  $\rho$  verlängert wird und  $K$  beschränkt ist, muss der Prozess nach endlich vielen Schritten abbrechen, d.h.  $\text{graph}(\tilde{Y}) \not\subseteq K$ . Für das Fortsetzen über den linken Rand betrachten wir  $Y : (a, t_0] \rightarrow \mathbb{R}^n$  und argumentieren analog. ■

**2.16 Lemma.** *Erfülle  $F$  die Voraussetzung (S) und sei  $(t_0, Y_0) \in \Omega$ . Sei  $(Y_n)_{n \in \mathbb{N}}$  eine Folge von Lösungen von (2.3), wobei  $Y_n$  auf dem Intervall  $I_n$  definiert ist. Gelte weiter  $t_0 \in I_n$  und  $Y_n(t_0) = Y_0$  sowie*

$$Y_n(t) = Y_m(t) \quad \forall t \in I_n \cap I_m$$

und alle  $n, m \in \mathbb{N}$ . Dann existiert im Intervall  $I := \bigcup_{n \in \mathbb{N}} I_n$  eine Lösung  $Y$  von (2.3) mit  $Y(t_0) = Y_0$ , für die  $Y(t) = Y_n(t)$  für alle  $t \in I_n$  und alle  $n \in \mathbb{N}$  gilt.

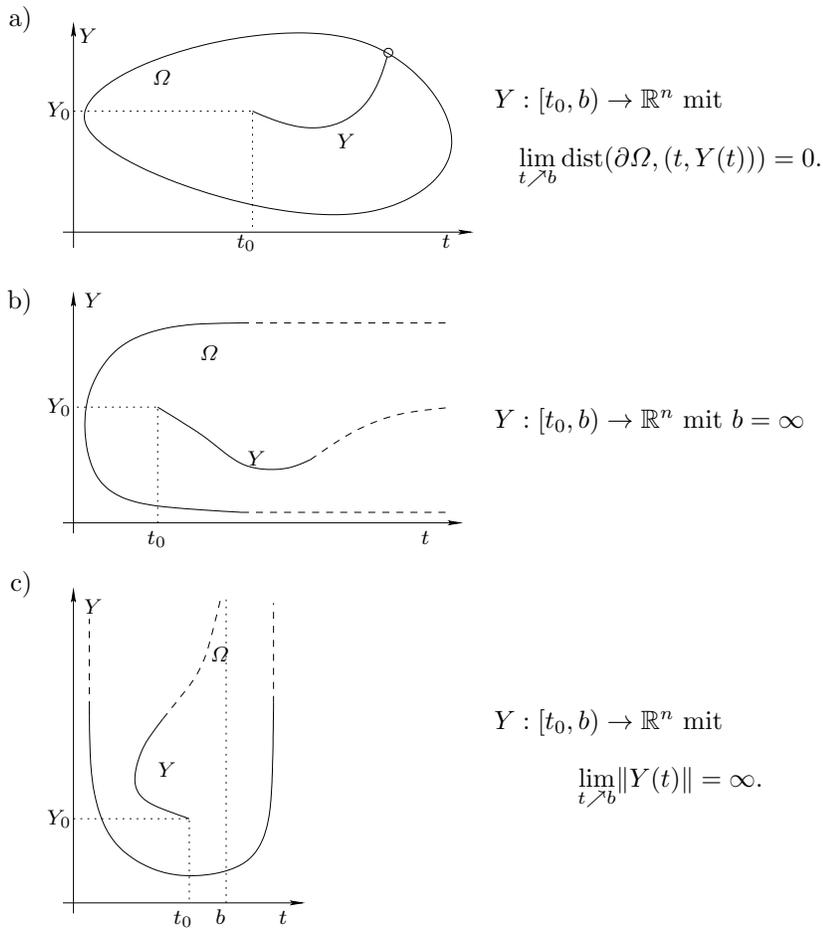
*Beweis.* Sei  $t \in I$ . Dann existiert ein  $n \in \mathbb{N}$  mit  $t \in I_n$ . Wir definieren  $Y(t) := Y_n(t)$ . Diese Definition ist wohldefiniert, d.h. unabhängig von der Wahl von  $n$ , denn für ein  $m \in \mathbb{N}$  mit  $t \in I_m$  folgt nach Voraussetzung  $Y_n(t) = Y_m(t)$ . Weiter ist  $Y$  auch eine Lösung von (2.3), denn für  $s \in I$  existiert ein  $n \in \mathbb{N}$  mit  $s \in I_n$  und insbesondere  $[t_0, s] \subseteq I_n$  (bzw.  $[s, t_0] \subseteq I_n$ ). Da  $Y_n$  eine Lösung von (2.3) auf dem Intervall  $I_n$  ist und  $Y(t) = Y_n(t)$  für alle  $t \in [t_0, s]$  (bzw.  $t \in [s, t_0]$ ) gilt, ist also auch  $Y$  eine Lösung auf diesem Intervall. Da  $s$  beliebig war, ist  $Y$  also eine gewünschte Lösung. ■

**2.17 Definition.** *Sei  $Y$  eine Lösung von (2.3) mit  $Y(t_0) = Y_0$  und sei  $G$  der abgeschlossene Graph von  $Y$ , d.h.  $G := \text{graph}(Y)$ . Wir sagen, dass  $Y$  nach rechts dem Rand beliebig nahe kommt, wenn*

$$G^+ := \{(t, Z) \in G \mid t \geq t_0\}$$

keine kompakte Teilmenge von  $\Omega$  ist. Analog ist nach links dem Rand beliebig nahe kommen definiert.

**Bemerkung.** Man kann zeigen, dass für eine Lösung  $Y : [t_0, b) \rightarrow \mathbb{R}^n$  der "rechtsseitige" Graph  $G^+$  keine kompakte Teilmenge von  $\Omega$  ist, falls einer folgenden Fälle eintritt: a)  $\lim_{t \nearrow b} \text{dist}(\partial\Omega, (t, Y(t))) = 0$ , oder b)  $b = \infty$ , oder c)  $\lim_{t \nearrow b} \|Y(t)\| = \infty$ . Diese drei Fälle kann man sich wie folgt vorstellen, wobei wir nur die Zusammenhangskomponente von  $\Omega$ , die  $(t_0, Y_0)$  enthält, betrachten:



**2.18 Satz.** Erfülle  $F$  die Voraussetzung (S) und sei  $(t_0, Y_0) \in \Omega$ . Dann existiert eine Lösung  $Y$  von (2.3) mit  $Y(t_0) = Y_0$ , die sich so fortsetzen lässt,

dass  $Y$  links und rechts dem Rand  $\partial\Omega$  beliebig nahe kommt. Eine derartig fortgesetzte Lösung nennen wir maximale Lösung.

*Beweis.* Da  $\Omega$  offen ist und  $(t_0, Y_0) \in \Omega$  gilt, existieren  $a, b > 0$  mit

$$Q := \{(t, Y) \in \mathbb{R}^{n+1} \mid |t - t_0| \leq a, \|Y - Y_0\| \leq b\} \subseteq \Omega.$$

Offenbar ist  $Q$  kompakt und somit  $\sup_{(t,Y) \in Q} \|F(t, Y)\|$  endlich. Nach dem **Satz von Peano** existiert also eine lokale Lösung  $Y$  von (2.3) mit  $Y(t_0) = Y_0$ . Definieren wir für  $n \in \mathbb{N}$  die Mengen

$$M_n := \{(t, Y) \in \Omega \mid \text{dist}((t, Y), \partial\Omega) \geq \frac{1}{n}, \|(t, Y)\| \leq n\},$$

so sind diese kompakt und es gilt  $M_n \subseteq M_{n+1}$  sowie  $\bigcup_{n \in \mathbb{N}} M_n = \Omega$ . Falls  $Y$  rechts dem Rand  $\partial\Omega$  nicht beliebig nahe kommt, also falls

$$G^+ = \{(t, Z) \in \overline{\text{graph}(Y)} \mid t \geq t_0\}$$

eine kompakte Teilmenge von  $\Omega$  ist, folgt wie im Beweis von Lemma 2.15  $\text{dist}(G^+, \partial\Omega) > 0$ . Somit existiert ein  $n \in \mathbb{N}$  mit  $G^+ \subseteq M_n$ . Da  $M_n$  kompakt ist, existiert nach Lemma 2.15 eine Fortsetzung  $Y_n : [t_0, b_n] \rightarrow \mathbb{R}^n$  von  $Y$  über  $M_n$  hinaus. Nun gibt es zwei Möglichkeiten: Entweder  $\text{graph}(Y_n) \not\subseteq M_{n+1}$ , dann setzen wir  $Y_{n+1} := Y_n$ . Oder  $\text{graph}(Y_n) \subseteq M_{n+1}$ . Da  $M_{n+1}$  kompakt ist, erhalten wir wieder nach Lemma 2.15 eine Fortsetzung  $Y_{n+1} : [t_0, b_{n+1}] \rightarrow \mathbb{R}^n$  über  $M_{n+1}$  hinaus. Insgesamt haben wir eine Folge  $(Y_k)_{k \in \mathbb{N}}$  von Lösungen auf den Intervallen  $I_k := [t_0, b_k]$  konstruiert, die  $b_{k+1} \geq b_k$  und  $Y_k = Y_\ell$  auf  $I_k$  für  $\ell \geq k$  erfüllen. Nach Lemma 2.16 existiert eine Lösung  $Y$  auf  $[t_0, b)$  mit  $b = \lim_{k \rightarrow \infty} b_k$ <sup>1</sup> und  $Y|_{I_k} = Y_k$ . Nach Konstruktion ist  $\text{graph}(Y)$  in keiner der Mengen  $M_n$  und somit in keiner kompakten Teilmenge von  $\Omega$  enthalten. Also kommt  $Y$  recht dem Rand  $\partial\Omega$  beliebig nahe. Analog zeigt man die Existenz einer Fortsetzung, die dem Rand  $\partial\Omega$  nach links beliebig nahe kommt. ■

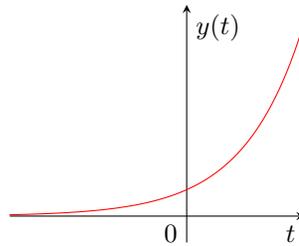
**Beispiele.** Die folgenden Lösungen von (2.3) kommen dem Rand  $\partial\Omega$  beliebig nahe.

1. Betrachte  $\Omega = \mathbb{R}^2$ ,  $f(t, y) = y$  und  $y(0) = 1$ , d.h. (2.3) entspricht

$$\begin{aligned} y'(t) &= y(t) \\ y(0) &= 1. \end{aligned}$$

Dann ist  $y(t) := e^t$  eine Lösung auf dem Intervall  $I = (-\infty, \infty)$ , d.h.

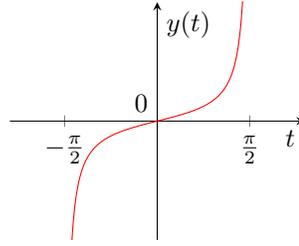
$$\lim_{t \rightarrow \infty} y(t) = \infty.$$



<sup>1</sup> Hier ist  $b = \infty$  zugelassen.

2. Betrachte  $\Omega = \mathbb{R}^2$ ,  $f(t, y) = 1 + y^2$  und  $y(0) = 0$ , d.h. (2.3) entspricht

$$\begin{aligned} y'(t) &= 1 + y^2(t) \\ y(0) &= 0. \end{aligned}$$



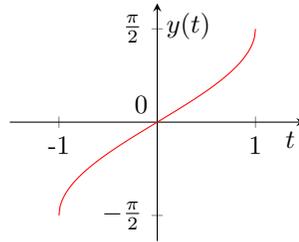
Dann ist  $y : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$ ,  $y(t) = \tan(t)$  eine Lösung, denn es gilt

$$y'(t) = \frac{1}{\cos^2(t)} = \frac{\sin^2(t) + \cos^2(t)}{\cos^2(t)} = 1 + \tan^2(t) = 1 + y(t)^2$$

und offenbar ist  $y(0) = 0$ . Hier haben wir ein endliches Zeitintervall, aber es gilt  $\lim_{t \rightarrow \pm \frac{\pi}{2}} y(t) = \pm \infty$ .

3. Betrachte  $\Omega = \mathbb{R} \times (-\frac{\pi}{2}, \frac{\pi}{2})$ ,  $f(t, y) = \frac{1}{\cos(y)}$  und  $y(0) = 0$ , d.h. (2.3) entspricht

$$\begin{aligned} y'(t) &= \frac{1}{\cos(y(t))} \\ y(0) &= 0. \end{aligned}$$



Dann ist  $y : (-1, 1) \rightarrow \mathbb{R}$ ,  $y(t) = \arcsin(t)$  eine Lösung, denn es gilt unter Verwendung von  $t = \sin(y(t))$

$$y'(t) = \frac{1}{\sqrt{1-t^2}} = \frac{1}{\sqrt{1-\sin^2(y(t))}} = \frac{1}{\cos(y(t))}$$

und offenbar ist  $y(0) = 0$ . Hier haben wir ein endliches Zeitintervall und es gilt  $\lim_{t \rightarrow \pm 1} y(t) = \pm \frac{\pi}{2}$ , aber  $(\pm 1, \pm \frac{\pi}{2}) \in \partial\Omega$ .

Als nächstes rechtfertigen wir den Namen “maximale Lösung”.

**2.19 Folgerung.** Erfülle  $F$  die Voraussetzung (S) und sei  $(t_0, Y_0) \in \Omega$ . Eine maximale Lösung von (2.3) mit  $Y(t_0) = Y_0$  kann nicht fortgesetzt werden.

*Beweis.* Sei  $Y : (a, b) \rightarrow \mathbb{R}^n$  eine maximale Lösung. Wir nehmen an, dass für  $\delta > 0$  die Funktion  $\tilde{Y} : (a, b + \delta) \rightarrow \mathbb{R}^n$  eine Fortsetzung der Lösung  $Y$  ist. Da  $\tilde{Y}$  stetig ist und  $\tilde{Y}|_{(a,b)} = Y$  gilt, folgt

$$\lim_{t \nearrow b} Y(t) = \lim_{t \nearrow b} \tilde{Y}(t) = \tilde{Y}(b).$$

Da  $\tilde{Y}$  eine Lösung ist, gilt  $(b, \tilde{Y}(b)) \in \Omega$ . Also folgt

$$G^+ = \{(t, Z) \in \overline{\text{graph}(Y)} \mid t \geq t_0\} \subseteq [t_0, b] \times \overline{\text{Bild}(\tilde{Y}|_{[t_0, b]})}.$$

Aufgrund der Stetigkeit von  $\tilde{Y}$  ist die rechte Seite kompakt und somit  $G^+$  beschränkt. Da weiter  $G^+$  abgeschlossen ist, folgt die Kompaktheit von  $G^+$ . Da  $Y$  eine Lösung ist, folgt  $(t, Y(t)) \in \Omega$ ,  $t \in [t_0, b)$ , und mit dem schon gezeigten  $(b, \tilde{Y}(b)) \in \Omega$  von oben folgt weiter  $G^+ \subseteq \Omega$ . Dies ist aber ein Widerspruch, da  $Y$  maximale Lösung ist. Analog zeigt man, dass die Lösung nicht nach links fortgesetzt werden kann. ■

Als nächstes stellen wir die Frage nach der Eindeutigkeit der Lösungen. Zunächst betrachten wir dies an einem Beispiel.

**Beispiel.** Sei  $\Omega = \mathbb{R}^2$  und  $f(t, y) = |y|^{\frac{1}{2}}$ , d.h. (2.3) entspricht

$$y'(t) = |y(t)|^{\frac{1}{2}}.$$

Sei  $y : I \rightarrow \mathbb{R}$  eine Lösung auf dem Intervall  $I$ , so ist  $z(t) := -y(-t)$  eine Lösung auf dem gespiegelten Intervall, denn

$$z'(t) = (-y(-t))' = -y'(-t)(-1) = y'(-t) = |y(-t)|^{\frac{1}{2}} = |z(t)|^{\frac{1}{2}}.$$

Es genügt also, positive Lösungen zu betrachten - negative Lösungen erhält man dann durch Substitution. Wir verwenden die Methode der Separation der Variablen:

1. Für positive Lösungen betrachten wir

$$\frac{dy}{dt} = y^{\frac{1}{2}}.$$

Separieren wir die Variablen, erhalten wir

$$\frac{dy}{y^{\frac{1}{2}}} = dt.$$

Bilden wir nun die Stammfunktion, so folgt

$$2y^{\frac{1}{2}} = t + c.$$

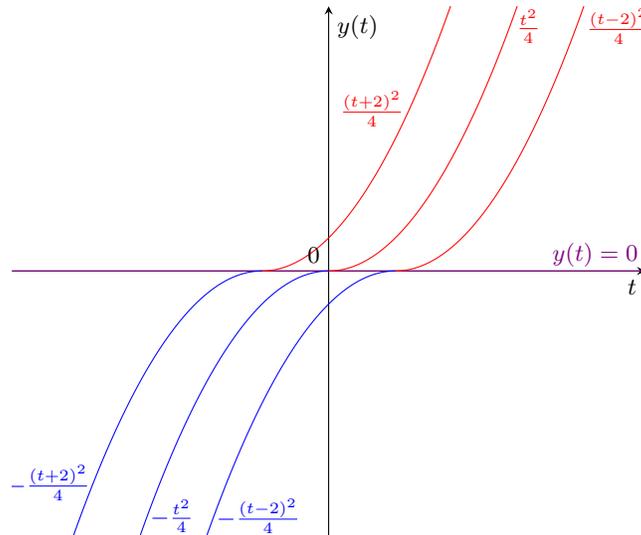
Da wir eine positive Lösung angenommen hatten, muss weiter  $t + c \geq 0$  und somit  $t \geq -c$  gelten. Somit ist

$$y(t) := \frac{(t+c)^2}{4}, \quad t \geq -c$$

eine Lösung.

2. Offenbar ist  $y(t) = 0$  eine Lösung
3. Durch Substitution erhalten wir die negative Lösung

$$y(t) := -\frac{(c-t)^2}{4} \quad t \leq c$$



**Abb. 2.6.** Lösungen von  $y'(t) = \sqrt{|y(t)|}$  zu verschiedenen Konstanten  $c$ .

Betrachten wir also das Anfangswertproblem  $y'(t) = |y(t)|^{\frac{1}{2}}$  mit  $y(2) = 1$ , so sind sowohl

$$y(t) := \begin{cases} \frac{t^2}{4} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

als auch

$$y(t) := \begin{cases} \frac{t^2}{4} & t \geq 0 \\ 0 & -2 < t < 0 \\ -\frac{(t+2)^2}{4} & t \leq -2 \end{cases}$$

Lösungen. Für “nur” stetige  $f$  ist also das Anfangswertproblem (2.3) nicht eindeutig.

**2.20 Satz.** Erfülle  $F$  die Voraussetzung (S) und sei  $F : \Omega \rightarrow \mathbb{R}^n$  lokal Lipschitz stetig bezüglich  $Y$ , d.h. für alle  $(t_0, Y_0) \in \Omega$  existiert eine Umgebung  $U$  des Punktes  $(t_0, Y_0)$  und eine Konstante  $L > 0$  so, dass für alle  $(t, Z), (t, Y) \in U$  gilt

$$\|F(t, Y) - F(t, Z)\| \leq L \|Y - Z\|.$$

Dann ist die maximale Lösung aus Satz 2.18 eindeutig.

*Beweis.* Sei  $(t_0, Y_0) \in \Omega$  und seien  $Y_1 : I_1 \rightarrow \mathbb{R}^n$ ,  $Y_2 : I_2 \rightarrow \mathbb{R}^n$  maximale Lösungen von (2.3) mit

$$Y_1(t_0) = Y_2(t_0) = Y_0.$$

Gelte  $Y_1 \neq Y_2$ , d.h. oBdA existiert ein  $\tau > t_0$  mit  $Y_1(\tau) \neq Y_2(\tau)$ . Wir definieren die Menge

$$A := \{t \in I_1 \cap I_2 \mid t \geq t_0, Y_1(t) \neq Y_2(t)\}.$$

Wegen  $\tau \in A$  folgt  $A \neq \emptyset$  und offensichtlich ist  $A$  durch  $t_0$  nach unten beschränkt. Somit gilt  $t_0 \leq \inf A =: \tilde{t} \leq \tau$  und für alle  $t_0 \leq t < \tilde{t}$  folgt  $Y_1(t) = Y_2(t)$ . Aufgrund der Stetigkeit von  $Y_1, Y_2$  gilt also auch  $Y_1(\tilde{t}) = Y_2(\tilde{t})$ . Insbesondere folgt  $(\tilde{t}, Y_1(\tilde{t})) \in \Omega$ . Aufgrund der lokalen Lipschitz Stetigkeit von  $F$  existiert eine Umgebung  $U$  von  $(\tilde{t}, Y_1(\tilde{t}))$  und ein  $L > 0$  mit

$$\|F(t, Y) - F(t, Z)\| \leq L \|Y - Z\| \quad \forall (t, Y), (t, Z) \in U.$$

Da  $U$  offen ist, existieren  $r, \delta > 0$  mit  $(\tilde{t} - 2\delta, \tilde{t} + 2\delta) \times B_{2r}(Y_1(\tilde{t})) \subseteq U$ . Aufgrund der Stetigkeit der  $Y_1, Y_2$  im Punkt  $\tilde{t}$  existieren weiter  $\delta_i > 0$ ,  $i = 1, 2$ , so dass für alle  $t \in \mathbb{R}$  mit  $|t - \tilde{t}| < \delta_i$  die Abschätzung  $|Y_i(\tilde{t}) - Y_i(t)| < r$  folgt. Setzen wir also  $\delta := \min\{\delta_1, \delta_2, \tilde{\delta}, \frac{1}{2L}\}$ , so gilt  $(t, Y_i(t)) \in U$  für  $t \in [\tilde{t}, \tilde{t} + \delta]$ . Nach Lemma 2.7 lösen die  $Y_i$ ,  $i = 1, 2$ , die Integralgleichungen

$$Y_i(t) = Y_0 + \int_{t_0}^t F(s, Y_i(s)) ds.$$

Verwenden wir  $Y_1(s) = Y_2(s)$  für  $t_0 \leq s < \tilde{t}$ , so folgt für  $t \in [\tilde{t}, \tilde{t} + \delta]$

$$\begin{aligned} \|Y_1(t) - Y_2(t)\| &= \left\| \int_{t_0}^t F(s, Y_1(s)) - F(s, Y_2(s)) ds \right\| \\ &= \left\| \int_{\tilde{t}}^t F(s, Y_1(s)) - F(s, Y_2(s)) ds \right\| \\ &\leq \int_{\tilde{t}}^t \|F(s, Y_1(s)) - F(s, Y_2(s))\| ds \\ &\leq L \int_{\tilde{t}}^t \|Y_1(s) - Y_2(s)\| ds \\ &\leq L \sup_{s \in [\tilde{t}, \tilde{t} + \delta]} \|Y_1(s) - Y_2(s)\| |t - \tilde{t}| \\ &\leq L \delta \sup_{s \in [\tilde{t}, \tilde{t} + \delta]} \|Y_1(s) - Y_2(s)\| \\ &\leq \frac{1}{2} \sup_{s \in [\tilde{t}, \tilde{t} + \delta]} \|Y_1(s) - Y_2(s)\|. \end{aligned}$$

Bilden wir nun das Supremum über alle  $t \in [\tilde{t}, \tilde{t} + \delta]$ , so folgt

$$\sup_{s \in [\tilde{t}, \tilde{t} + \delta]} \|Y_1(s) - Y_2(s)\| \leq \frac{1}{2} \sup_{s \in [\tilde{t}, \tilde{t} + \delta]} \|Y_1(s) - Y_2(s)\|.$$

Da  $Y_1, Y_2$  auf dem Intervall  $[\tilde{t}, \tilde{t} + \delta]$  stetig sind, sind die Suprema endlich und nach Konstruktion ungleich 0. Somit haben wir einen Widerspruch und die Aussage ist bewiesen. ■

### 12.3 Spezialfälle für Gleichungen 1. Ordnung

In diesem Abschnitt betrachten wir spezielle Funktionen  $f : \Omega \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ , welche die Voraussetzung (S) erfüllen und suchen Lösungen des Anfangswertproblems

$$\begin{aligned} y'(t) &= f(t, y(t)) \\ y(t_0) &= y_0. \end{aligned} \tag{3.1}$$

#### 12.3.1 Ortsunabhängige rechte Seiten $f = f(t)$

Falls die rechte Seite  $f$  unabhängig von  $y$  ist, so können wir die Lösung leicht berechnen. Sei  $I$  ein offenes Intervall mit  $t_0 \in I$ . Erfülle  $f : I \times \mathbb{R} \rightarrow \mathbb{R}$  die Voraussetzung (S) und gelte  $f(t, y) = f(t)$ . Wir suchen also nur eine Stammfunktion von  $f$ . Da für  $t \in I$  die Funktion  $f$  stetig auf dem Intervall  $[t_0, t]$  ist, ist sie dort integrierbar und

$$\varphi(t) := \int_{t_0}^t f(s) ds, \quad t \in I$$

ist wohldefiniert. Mit dem Hauptsatz der Differential und Integralrechnung folgt

$$\varphi'(t) = f(t).$$

Somit sind die Stammfunktionen von  $f$  durch

$$y(t) := \varphi(t) + c, \quad t \in I$$

mit  $c \in \mathbb{R}$  gegeben. Für den Anfangswert ergibt sich weiter  $y(t_0) = c$ . Somit hat (3.1) für alle  $y_0 \in \mathbb{R}$  eine Lösung. Wir fassen dies im folgendem Satz zusammen.

**3.2 Satz.** Sei  $I$  ein offenes Intervall, erfülle  $f : I \times \mathbb{R} \rightarrow \mathbb{R}$  die Voraussetzung (S) und gelte  $f(t, y) = f(t)$  für alle  $(t, y) \in I \times \mathbb{R}$ . Dann existiert für alle  $(t_0, y_0) \in I \times \mathbb{R}$  eine eindeutige, maximale Lösung  $y : I \rightarrow \mathbb{R}$  von (3.1) mit  $y(t_0) = y_0$ .

*Beweis.* Wir haben oben gezeigt, dass eine Lösung  $y$  von (3.1) mit  $y(t_0) = y_0$  auf dem Intervall  $I$  existiert. Wegen  $\Omega = I \times \mathbb{R}$  ist dies bereits eine maximale Lösung. Da  $f(t, y) = f(t)$  Lipschitzstetig bezüglich  $y$  ist, folgt mit Satz 2.20 die Eindeutigkeit der Lösung und somit die Behauptung. ■

### 12.3.2 Zeitunabhängige rechte Seiten $f = f(y)$

Nun vertauschen wir die Rolle von  $t$  und  $y$ , wir betrachten also ein Intervall  $I = (a, b)$  und eine Funktion  $f : \mathbb{R} \times I \rightarrow \mathbb{R}$ , welche die Voraussetzung (S) sowie  $f(t, y) = f(y)$  erfüllt. Nehmen wir weiter an, dass  $f(y) \neq 0$  für alle  $y \in (a, b)$  gilt. Falls  $\varphi : (\alpha, \beta) \rightarrow I$  eine Lösung von (3.1) ist, dann folgt

$$\varphi'(t) = f(\varphi(t)) \neq 0, \quad t \in (\alpha, \beta)$$

bzw.

$$\frac{\varphi'(t)}{f(\varphi(t))} = 1, \quad t \in (\alpha, \beta).$$

Aufgrund unserer Voraussetzungen an  $f$  ist  $\frac{1}{f}$  stetig und es existiert somit eine Stammfunktion  $F$  von  $\frac{1}{f}$  auf dem Intervall  $(\alpha, \beta)$ . Es folgt für  $t \in (\alpha, \beta)$

$$\frac{d}{dt} F(\varphi(t)) = F'(\varphi(t)) \varphi'(t) = \frac{\varphi'(t)}{f(\varphi(t))} = 1.$$

Nach dem Hauptsatz der Differential- und Integralrechnung existiert also ein  $c \in \mathbb{R}$  mit

$$F(\varphi(t)) = t + c, \quad t \in (\alpha, \beta).$$

Aufgrund der Stetigkeit von  $f$  und

$$F'(y) = \frac{1}{f(y)} \neq 0, \quad y \in (\alpha, \beta)$$

ist  $F$  strikt monoton. Insbesondere existiert die Umkehrfunktion  $F^{-1}$ . Somit erhalten wir als Gleichung für  $\varphi$

$$\varphi(t) = F^{-1}(t + c), \quad t \in (\alpha, \beta).$$

Wir fassen dieses Ergebnis im folgenden Lemma zusammen.

**3.3 Lemma.** Sei  $I$  ein offenes Intervall und sei  $f : \mathbb{R} \times I \rightarrow \mathbb{R}$  durch  $f(t, y) = f(y)$  gegeben, wobei die Funktion  $f : I \rightarrow \mathbb{R}$  stetig ist und  $f \neq 0$  auf  $I$  erfüllt. Falls  $\varphi : (\alpha, \beta) \rightarrow I$  eine Lösung von (3.1) ist, dann existiert ein  $c \in \mathbb{R}$  so, dass für alle  $t \in (\alpha, \beta)$  gilt:

$$\varphi(t) = F^{-1}(t + c),$$

wobei  $F$  eine Stammfunktion von  $\frac{1}{f}$  ist.

*Beweis.* Siehe Rechnungen vorher. ■

Falls für ein  $\hat{y} \in (a, b)$  gilt  $f(\hat{y}) = 0$ , ist  $y(t) := \hat{y}$ ,  $t \in \mathbb{R}$ , eine maximale Lösung von (3.1) im Falle von zeitunabhängigen rechten Seiten. Es stellt sich die Frage, ob man in der Situation von Lemma 3.3 immer Lösungen in obiger Form finden kann.

**3.4 Satz.** Erfülle  $f$  die Voraussetzungen von Lemma 3.3. Dann existiert für alle  $(t_0, y_0) \in \mathbb{R} \times I$  eine eindeutige maximale Lösung  $y : J_0 \rightarrow \mathbb{R}$  von (3.1) mit  $y(t_0) = y_0$ . Diese Lösung ist von der Form

$$y(t) = F^{-1}(t - t_0), \quad t \in J_0,$$

wobei  $F$  durch

$$F(y) := \int_{y_0}^y \frac{1}{f(x)} dx, \quad y \in I$$

gegeben ist.

**Bemerkung.** Das Intervall  $J_0$  aus Satz 3.4 ist durch  $J_0 := \text{Bild}(F) + t_0$  gegeben.

*Beweis von Satz 3.4.* Sei  $I = (a, b)$ . Wir definieren  $F : I \rightarrow \mathbb{R}$  durch

$$F(y) := \int_{y_0}^y \frac{1}{f(x)} dx, \quad y \in I.$$

Aufgrund der Eigenschaften von  $f$  ist  $F$  wohldefiniert. Nach dem Hauptsatz der Differential- und Integralrechnung gilt

$$F'(y) = \frac{1}{f(y)} \neq 0, \quad y \in I.$$

Wieder mit der Stetigkeit von  $\frac{1}{f}$  folgt die strikte Monotonie von  $F$ . Setzen wir  $J := \text{Bild}(F)$ , so ist  $J$  aufgrund der strikten Monotonie ein offenes Intervall und die Umkehrfunktion  $F^{-1} : J \rightarrow I$  existiert. Wegen  $F(y_0) = 0$  folgt  $0 \in J$ .

Verschieben wir das Intervall  $J$  um  $t_0$ , so existieren  $\alpha, \beta \in \mathbb{R}$  mit  $\alpha < \beta$  und  $(\alpha, \beta) = J + t_0$ , d.h. es gilt  $t \in (\alpha, \beta)$  genau dann, wenn  $t - t_0 \in J$  gilt. Wir definieren nun

$$y(t) := F^{-1}(t - t_0), \quad t \in (\alpha, \beta).$$

Dann gilt nach dem Satz über Ableitungen inverser Funktionen (Analysis I, Kapitel 6, Satz 1.4) für  $t \in (\alpha, \beta)$

$$y'(t) = (F^{-1})'(t - t_0) = \frac{1}{F'(F^{-1}(t - t_0))} = \frac{1}{F'(y(t))} = f(y(t)).$$

Wegen  $0 \in J$  und somit  $t_0 \in (\alpha, \beta)$ , sowie  $F(0) = y_0$  folgt weiter

$$y(t_0) = F^{-1}(0) = y_0.$$

Somit ist  $y$  eine Lösung der gesuchten Form von (3.1) mit  $y(t_0) = y_0$  auf dem Intervall  $(\alpha, \beta)$ . Nach Satz 2.18 existiert eine maximale Lösung  $\varphi : (\tilde{\alpha}, \tilde{\beta}) \rightarrow I$  von (3.1) mit  $t_0 \in (\tilde{\alpha}, \tilde{\beta})$  und  $\varphi(t_0) = y_0$ . Nach Lemma 3.3 folgt (mit der spezieller Wahl der Stammfunktion als unser konkretes  $F$ ) die Existenz eines  $c \in \mathbb{R}$  mit

$$\varphi(t) = F^{-1}(t + c), \quad t \in (\tilde{\alpha}, \tilde{\beta}).$$

Wegen  $t_0 \in (\alpha, \beta) \cap (\tilde{\alpha}, \tilde{\beta})$  folgt insbesondere

$$F^{-1}(t_0 + c) = \varphi(t_0) = y_0 = y(t_0) = F^{-1}(0).$$

Aufgrund der Injektivität von  $F^{-1}$  folgt  $0 = t_0 + c$  bzw.  $c = -t_0$ . Es gilt also  $y = \varphi$  auf  $(\alpha, \beta) \cap (\tilde{\alpha}, \tilde{\beta})$  und nach Konstruktion der  $\alpha, \beta$ , d.h.  $(\alpha, \beta) = \text{Bild}(F) + t_0$ , bzw. den Eigenschaften einer maximalen Lösung folgt  $\alpha = \tilde{\alpha}$  sowie  $\beta = \tilde{\beta}$ . Damit ist  $y$  eine maximale Lösung. Analog zeigt man die Eindeutigkeit dieser Lösung. ■

**Beispiele.** Wir verwenden nun Satz 3.4, um Lösungen von (3.1) zu finden.

- a) Wählen wir  $f(t, y) = 1 + y^2$  und  $(t_0, y_0) \in \mathbb{R} \times \mathbb{R}$ . Wir suchen also eine Lösung von

$$\begin{aligned} y'(t) &= 1 + y^2, \\ y(t_0) &= y_0. \end{aligned}$$

Da  $\arctan$  die Stammfunktion von  $(1 + y^2)^{-1}$  ist und  $\text{Bild}(\tan) = \mathbb{R}$  gilt, existiert für alle  $y_0 \in \mathbb{R}$  ein  $c \in (-\frac{\pi}{2}, \frac{\pi}{2})$  mit  $y_0 = \tan(c)$ . Wir berechnen für  $y \in \mathbb{R}$

$$F(y) = \int_{y_0}^y \frac{1}{1+x^2} dx = \arctan(y) - \arctan(y_0) = \arctan(y) - c.$$

Wegen  $\text{Bild}(\arctan) = (-\frac{\pi}{2}, \frac{\pi}{2})$  folgt

$$\text{Bild}(F) = \left(-\frac{\pi}{2} - c, \frac{\pi}{2} - c\right)$$

und  $F^{-1}$  ist auf  $\text{Bild}(F)$  durch

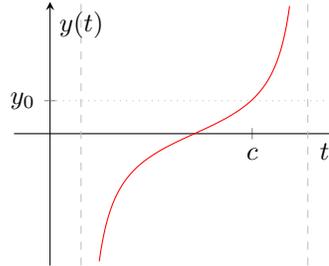
$$F^{-1}(t) = \tan(t + c)$$

gegeben. Somit ist die maximale Lösung nach Satz 3.4 auf dem Intervall

$$J_0 := \left(-\frac{\pi}{2} + t_0 - c, \frac{\pi}{2} + t_0 - c\right)$$

durch

$$y(t) = \tan(t - t_0 + c)$$



gegeben. Die maximale Lösung ist folglich nur auf einem beschränkten Zeitintervall definiert. Obwohl  $f$  also auf ganz  $\mathbb{R}^2$  definiert ist, kann hier keine Lösung für alle Zeiten  $t \in \mathbb{R}$  gefunden werden.

- b) Wählen wir  $f(t, y) = 3y^{\frac{2}{3}}$ , so gilt  $f(y) = 0$  genau dann, wenn  $y = 0$  gilt. Somit erfüllt  $f$  nicht die Voraussetzungen von Satz 3.4 auf dem Intervall  $I = \mathbb{R}$ . Wenden wir für  $(t_0, y_0) \in \mathbb{R} \times (0, \infty)$  stattdessen Satz 3.4 auf  $f|_{(0, \infty)}$  an, so ist  $F$  auf  $(0, \infty)$  durch

$$F(y) = \int_{y_0}^y \frac{1}{3x^{\frac{2}{3}}} dx = y^{\frac{1}{3}} - y_0^{\frac{1}{3}}.$$

gegeben. Damit ist für  $t \in (-y_0^{\frac{1}{3}}, \infty)$

$$F^{-1}(t) = (t + y_0^{\frac{1}{3}})^3$$

und die (zu  $f|_{(0, \infty)} : \mathbb{R} \times (0, \infty) \rightarrow \mathbb{R}$  maximale) Lösung gemäß Satz 3.4 durch

$$y(t) = (t - t_0 + y_0^{\frac{1}{3}})^3, \quad t \in (-y_0^{\frac{1}{3}} + t_0, \infty)$$

gegeben. Allerdings ist dies keine maximale Lösung des ursprünglichen Problems, denn

$$y_0(t) := \begin{cases} 0 & t \leq -y_0^{\frac{1}{3}} + t_0, \\ (t - t_0 + y_0^{\frac{1}{3}})^3 & t > -y_0^{\frac{1}{3}} + t_0 \end{cases}$$

ist eine (maximale) Lösung auf  $\mathbb{R}$ . Betrachten wir analog  $f|_{(-\infty,0)}$ , so ist

$$y_1(t) = (t - t_1 + y_1^{\frac{1}{3}})^3, \quad t \in (-\infty, -y_1^{\frac{1}{3}} + t_1)$$

eine Lösung für  $(t_1, y_1) \in \mathbb{R} \times (-\infty, 0)$ . Weiter ist für alle  $(t_0, y_0) \in \mathbb{R} \times (0, \infty)$  und  $(t_1, y_1) \in \mathbb{R} \times (-\infty, 0)$  mit  $y_0^{\frac{1}{3}} + t_0 > -y_1^{\frac{1}{3}} + t_1$

$$y_2(t) := \begin{cases} (t - t_1 + y_1^{\frac{1}{3}})^3 & t < -y_1^{\frac{1}{3}} + t_1 \\ 0 & -y_1^{\frac{1}{3}} + t_1 \leq t \leq -y_0^{\frac{1}{3}} + t_0, \\ (t - t_0 + y_0^{\frac{1}{3}})^3 & t > -y_0^{\frac{1}{3}} + t_0 \end{cases}$$

eine Lösung auf dem Intervall  $\mathbb{R}$ . Es gibt folglich zu dem Anfangswert  $(t_0, y_0)$  unendlich viele maximale Lösungen.

- c) Das Fischpopulationsmodell aus der Einführung hatte die Form  $f(y) = y(b - cy) - H$  für  $b, c, H > 0$  und kann folglich mit dem Satz 3.4 gelöst werden. Die Lösungen haben wir schon dort mit derselben Methode berechnet.

### 12.3.3 Separierbare rechte Seiten $f = h(t)g(y)$

Eine weitere Klasse von Gleichungen sind separierbare Differentialgleichungen erster Ordnung. Hierbei besitzt  $f : J \times I \rightarrow \mathbb{R}$  die Form  $f(t, y) = h(t)g(y)$ , wobei  $I = (a, b)$  und  $J = (c, d)$  offene Intervalle sind und die Funktionen  $g, h$  dort stetig sind. Weiter gelte  $g(y) \neq 0$  für alle  $y \in I$ . Wir können dabei ähnlich wie in Abschnitt 12.3.2 vorgehen. Falls  $\varphi : (\alpha, \beta) \subseteq (c, d) \rightarrow (a, b)$  eine Lösung von (3.1) ist, d.h.

$$\varphi'(t) = h(t)g(\varphi(t)), \quad t \in (\alpha, \beta)$$

gilt, so folgt

$$\frac{\varphi'(t)}{g(\varphi(t))} = h(t), \quad t \in (\alpha, \beta).$$

Aufgrund der Stetigkeit der Funktionen  $g, h$  auf  $I$  bzw.  $J$ , sowie  $g \neq 0$  auf  $I$  existieren Stammfunktionen  $G$  von  $\frac{1}{g}$  und  $H$  von  $h$ . Damit folgt für  $t \in (\alpha, \beta)$

$$\frac{d}{dt} G(\varphi(t)) = G'(\varphi(t))\varphi'(t) = \frac{\varphi'(t)}{g(\varphi(t))} = h(t) = \frac{d}{dt} H(t).$$

Nach dem Hauptsatz der Differential- und Integralrechnung existiert eine Konstante  $c \in \mathbb{R}$  mit

$$G(\varphi(t)) = H(t) + c, \quad t \in (\alpha, \beta).$$

Schließlich ist  $G$  aufgrund der Stetigkeit von  $g$  und  $G' = \frac{1}{g} \neq 0$  strikt monoton und die Umkehrfunktion  $G^{-1}$  existiert. Somit folgt

$$\varphi(t) = G^{-1}(H(t) + c), \quad t \in (\alpha, \beta).$$

**3.5 Lemma.** Seien  $I, J$  offene Intervalle und sei  $f : J \times I \rightarrow \mathbb{R}$  durch  $f(t, y) = h(t)g(y)$  gegeben, wobei die Funktionen  $g : I \rightarrow \mathbb{R}$ ,  $g \neq 0$  und  $h : J \rightarrow \mathbb{R}$  stetig sind. Falls  $\varphi : (\alpha, \beta) \subseteq J \rightarrow I$  eine Lösung von (3.1) ist, dann existiert ein  $c \in \mathbb{R}$  so, dass für alle  $t \in (\alpha, \beta)$

$$\varphi(t) = G^{-1}(H(t) + c)$$

gilt, wobei  $G$  eine Stammfunktion von  $\frac{1}{g}$  und  $H$  eine Stammfunktion von  $h$  ist.

*Beweis.* Siehe obige Rechnung. ■

**3.6 Satz.** Erfülle  $f$  die Voraussetzungen von Lemma 3.5. Dann existiert für alle  $(t_0, y_0) \in J \times I$  eine eindeutige maximale Lösung  $y : J_0 \rightarrow I$  von (3.1) mit  $y(t_0) = y_0$ . Diese Lösung ist von der Form

$$y(t) = G^{-1}(H(t)),$$

wobei  $G(y) = \int_{y_0}^y \frac{1}{g(x)} dx$  für  $y \in I$  und  $H(t) = \int_{t_0}^t h(s) ds$  für  $t \in J$  ist.

**Bemerkung.** Das Intervall  $J_0$  aus Satz 3.6 ist dabei das größte offene Intervall, welches in  $H^{-1}(\text{Bild}(G))$  liegt und  $t_0$  beinhaltet.

*Beweis von Satz 3.6.* Sei  $I = (a, b)$ ,  $J = (c, d)$  und  $(t_0, y_0) \in J \times I$  sowie

$$G(y) := \int_{y_0}^y \frac{1}{g(x)} dx, \quad y \in I$$

$$H(t) := \int_{t_0}^t h(s) ds, \quad t \in J.$$

Wir definieren  $K := \text{Bild}(G)$ . Es gilt  $G' = \frac{1}{g} \neq 0$  auf  $I$  und folglich ist  $G : I \rightarrow K$  strikt monoton sowie  $K$  ein offenes Intervall. Insbesondere existiert die Umkehrfunktion  $G^{-1} : K \rightarrow I$  und wegen  $G(y_0) = 0$  folgt  $0 \in K$ . Definieren wir

$$M := \{t \in J \mid H(t) \in K\} = H^{-1}(K),$$

so ist aufgrund der Stetigkeit von  $H$  die Menge  $M$  offen. Weiter folgt wegen  $H(t_0) = 0$ , dass  $t_0 \in M$  gilt. Sei nun  $(\alpha, \beta)$  das größte offene Intervall, welches in  $M$  liegt und  $t_0$  beinhaltet. Damit ist

$$y(t) := G^{-1}(H(t)), \quad t \in (\alpha, \beta)$$

wohldefiniert und es folgt mit dem Satz über die Ableitung einer inversen Funktion für  $t \in (\alpha, \beta)$

$$y'(t) = (G^{-1})'(H(t))H'(t) = \frac{h(t)}{G'(G^{-1}(H(t)))} = \frac{h(t)}{G'(y(t))} = g(y(t))h(t).$$

Weiter gilt

$$y(t_0) = G^{-1}(H(t_0)) = G^{-1}(0) = y_0.$$

Somit ist  $y$  eine Lösung von (3.1) mit  $y(t_0) = y_0$  auf dem Intervall  $(\alpha, \beta)$ . Nach Satz 2.18 existiert eine maximale Lösung  $\varphi : (\tilde{\alpha}, \tilde{\beta}) \rightarrow I$  von (3.1) mit  $y(t_0) = y_0$ . Nach Lemma 3.5 existiert ein  $c \in \mathbb{R}$  mit

$$\varphi(t) = G^{-1}(H(t) + c), \quad t \in (\tilde{\alpha}, \tilde{\beta}),$$

wobei  $G$  und  $H$  wie oben definiert sind. Insbesondere gilt  $t_0 \in (\tilde{\alpha}, \tilde{\beta}) \cap (\alpha, \beta)$  und somit

$$G^{-1}(H(t_0) + c) = \varphi(t_0) = y_0 = y(t_0) = G^{-1}(H(t_0)) = G^{-1}(0).$$

Aufgrund der Injektivität von  $G^{-1}$  folgt

$$0 = H(t_0) + c = c$$

und somit  $c = 0$ . Damit gilt  $\varphi = y$  auf  $(\tilde{\alpha}, \tilde{\beta}) \cap (\alpha, \beta)$ . Nach der Definition des Intervalls  $(\alpha, \beta)$  und  $(\tilde{\alpha}, \tilde{\beta}) \subseteq M$  folgt  $(\tilde{\alpha}, \tilde{\beta}) \subseteq (\alpha, \beta)$  und somit  $\alpha = \tilde{\alpha}$  bzw.  $\beta = \tilde{\beta}$ . Wir haben also gezeigt, dass  $y$  eine maximale Lösung ist. Analog zeigt man die Eindeutigkeit dieser Lösung. ■

Falls für ein  $\hat{y} \in I$  gilt  $g(\hat{y}) = 0$ , ist  $y(t) := \hat{y}$ ,  $t \in J$ , eine maximale Lösung von (3.1) im Falle von separierbaren rechten Seiten.

**Beispiel.** Wir betrachten das Anfangswertproblem

$$\begin{aligned} y'(t) &= 2t(1 + y^2), \\ y(t_0) &= y_0 \end{aligned}$$

mit  $(t_0, y_0) \in \mathbb{R}^2$ . Mit der Notation von Satz 3.6 gilt  $g : \mathbb{R} \rightarrow \mathbb{R} : y \mapsto 1 + y^2$  sowie  $h : \mathbb{R} \rightarrow \mathbb{R} : t \mapsto 2t$ . Damit folgt

$$G(y) = \int_{y_0}^y \frac{1}{1+x^2} dx = \arctan(y) - \arctan(y_0) =: \arctan(y) - c_0,$$

$$H(t) = \int_{t_0}^t 2s ds = t^2 - t_0^2.$$

Weiter gilt  $\text{Bild}(G) = (-\frac{\pi}{2} - c_0, \frac{\pi}{2} - c_0)$  mit  $c_0 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ . Zu der Bestimmung von  $H^{-1}(\text{Bild}(G))$  benötigen wir nun eine Fallunterscheidung:

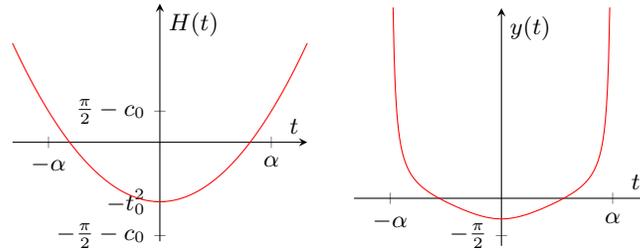
1. **Fall  $-t_0^2 > -\frac{\pi}{2} - c_0$ :** Die Urbildmenge ist in diesem Fall durch ein Intervall gegeben (vergleiche Abbildung 3.1). Setzen wir  $\alpha = \sqrt{\frac{\pi}{2} - c_0 + t_0^2}$ , so gilt  $H^{-1}(\text{Bild}(G)) = (-\alpha, \alpha)$ . Die maximale Lösung gemäß Satz 3.6 ist also durch

$$y(t) = \tan(t^2 - t_0^2 + \arctan(y_0)) \quad t \in (-\alpha, \alpha)$$

gegeben. Offenbar gilt

$$\lim_{t \nearrow \alpha} y(t) = \infty \quad \text{und} \quad \lim_{t \searrow -\alpha} y(t) = \infty$$

(vergleiche Abbildung 3.1).



**Abb. 3.1.** Die Graphen von  $H$  und  $y$  mit  $c_0 = \frac{\pi}{4}$  und  $t_0^2 = 1.5$ .

2. **Fall  $-t_0^2 \leq -\frac{\pi}{2} - c_0$ :** Die Urbildmenge besteht aus zwei Intervallen (vergleiche Abbildung 3.2). Setzen wir

$$\alpha := \sqrt{-\frac{\pi}{2} - c_0 + t_0^2} \quad \text{und} \quad \beta := \sqrt{\frac{\pi}{2} - c_0 + t_0^2},$$

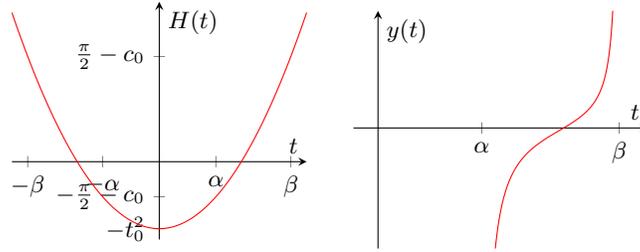
so ist  $H^{-1}(\text{Bild}(G)) = (-\beta, -\alpha) \cup (\alpha, \beta)$ . Sei oBdA  $t_0 \in (\alpha, \beta)$ , so ist die maximale Lösung gemäß Satz 3.6 durch

$$y(t) = \tan(t^2 - t_0^2 + \arctan(y_0)) \quad t \in (\alpha, \beta)$$

gegeben. Offenbar gilt

$$\lim_{t \nearrow \beta} y(t) = \infty \quad \text{und} \quad \lim_{t \searrow \alpha} y(t) = -\infty$$

(vergleiche Abbildung 3.2).



**Abb. 3.2.** Die Graphen von  $H$  und  $y$  mit  $c_0 = -\frac{\pi}{4}$  und  $t_0 = \sqrt{1.5}$ .

### 12.3.4 Lineare Gleichungen

Wir betrachten rechte Seiten der Form  $f(t, y) = h(t)y + p(t)$ , wobei die Funktionen  $p, h : I = (a, b) \rightarrow \mathbb{R}$  stetig sind und somit  $f : I \times \mathbb{R} \rightarrow \mathbb{R}$  die Voraussetzung (S) erfüllt. Dabei unterscheiden wir zwei Typen von Gleichungen: Gilt

- (i)  $p(t) \equiv 0$ , so nennen wir (3.1) eine *homogene Gleichung*,
- (ii)  $p(t) \not\equiv 0$ , so nennen wir (3.1) eine *inhomogene Gleichung*.

Der Name *lineare Gleichung* entspringt dabei folgende Anschauung: Schreiben wir (3.1) zu

$$y'(t) - h(t)y(t) = p(t)$$

um, so erhalten wir Lösungen dieser Gleichung, wenn wir das Urbild der rechten Seite  $p \in C^0(I)$  bezüglich des Operators

$$L : C^1(I) \rightarrow C^0(I) : y \mapsto y' - hy,$$

d.h.  $L(y)(t) := y'(t) - h(t)y(t)$ ,  $t \in I$ ,  $y \in C^1(I)$ , suchen. Der Operator  $L$  ist linear, es gilt also

$$L(\alpha y + \beta z) = \alpha L(y) + \beta L(z) \quad \alpha, \beta \in \mathbb{R}, y, z \in C^1(I).$$

Suchen wir nun eine Lösung der homogenen Gleichung, so ist dies ein Spezialfall der separierbaren Differentialgleichungen. Mit der Notation aus Abschnitt 12.3.3 gilt “ $h(t) = h(t)$ ” sowie  $g(y) = y$ , allerdings können wir aufgrund von  $g(0) = 0$  den Satz 3.6 nicht anwenden. Wir argumentieren dennoch ähnlich, um einen Kandidaten für die Lösung zu erhalten. Eine Stammfunktion von  $g$  ist auf  $\mathbb{R} \setminus \{0\}$  durch  $G(y) := \ln |y|$  gegeben und es folgt  $G^{-1}(t) = e^t$ . Weiter sei  $H$  eine Stammfunktion von  $h$ . Für  $y \neq 0$  können wir (3.1) wieder zu

$$\frac{dy}{y} = h dt$$

umschreiben. Bilden wir die Stammfunktionen auf beiden Seiten, so existiert ein  $c \in \mathbb{R}$  mit

$$\ln |y(t)| = H(t) + c$$

bzw.

$$|y(t)| = e^c e^{H(t)}.$$

Wir können den Betrag vernachlässigen, falls wir  $e^c$  durch eine Konstante  $\tilde{c} \in \mathbb{R} \setminus \{0\}$  ersetzen. Wir suchen also eine Lösung der Form

$$y(t) = \tilde{c} e^{H(t)}.$$

Daher definieren wir zu einem Anfangswert  $(t_0, y_0) \in I \times (\mathbb{R} \setminus \{0\})$  die Funktion

$$y(t) := y_0 e^{\tilde{H}(t)}, \quad t \in I$$

mit der speziellen Stammfunktion

$$\tilde{H}(t) := \int_{t_0}^t h(s) ds, \quad t \in I$$

von  $h$ . Da  $y$  auf ganz  $I$  definiert ist, sieht man sofort, dass die Funktion  $y : I \rightarrow \mathbb{R}$  eine maximale Lösung von (3.1) mit  $y(t_0) = y_0$  und  $y \neq 0$  ist. Für den Anfangswert  $(t_0, 0) \in I \times \{0\}$  ist offensichtlich die Funktion

$$y : I \rightarrow \mathbb{R}, \quad y(t) := 0$$

maximale Lösung von (3.1) mit  $y(t_0) = 0 = y_0$ . Tatsächlich wird dieser Fall ebenfalls von obiger Formel abgedeckt. Schließlich bemerken wir, dass die Funktion  $g$  Lipschitz-Stetig in  $y$  ist und nach Satz 2.20 die gefundenen Lösungen eindeutig sind. Bevor wir uns den inhomogenen Gleichungen zuwenden betrachten wir folgendes Lemma.

**3.7 Lemma.** *Sei  $I = (a, b)$  und für  $f : I \times \mathbb{R} \rightarrow \mathbb{R}$  gelte  $f(t, y) = h(t)y + p(t)$ , wobei  $p$  und  $h$  auf  $I$  stetig sind. Seien  $y_1, y_2$  Lösungen der inhomogenen Gleichung (3.1) und sei  $y_0$  eine Lösung der homogenen Gleichung (3.1). Dann gilt:*

(i)  $y_1 - y_2$  ist eine Lösung der homogenen Gleichung.

(ii)  $y_1 + y_0$  ist eine Lösung der inhomogenen Gleichung.

*Beweis.* Dies folgt durch triviales Einsetzen. ■

Anders formuliert bedeutet dies

$$\begin{aligned} \text{Lösung inhomogene Gleichung} &= \text{Lösung homogene Gleichung} \\ &+ \text{Lösung inhomogene Gleichung,} \end{aligned}$$

man kann also eine *spezielle* Lösung der inhomogenen Gleichung suchen und dazu alle Lösungen der homogenen Gleichung addieren.

Um die inhomogene Gleichung zu lösen, verwenden wir folgende, auf Lagrange zurückzuführende Idee der "Variation der Konstanten": Wir benutzen die Formel für die homogene Gleichung, wobei wir die Konstante  $\tilde{c}$  durch eine Funktion  $c(t)$  ersetzen. Wir betrachten also folgenden Ansatz

$$y(t) = c(t) e^{H(t)}.$$

Damit gilt

$$y'(t) = c'(t) e^{H(t)} + c(t) e^{H(t)} h(t).$$

Um die Gleichung (3.1) zu erfüllen, muss also

$$p(t) + h(t) y = y'(t) = c'(t) e^{H(t)} + c(t) e^{H(t)} h(t)$$

gelten. Wir müssen also  $c'(t) = p(t) e^{-H(t)}$  wählen, d.h. wir wählen  $c$  als Stammfunktion von  $p(t) e^{-H(t)}$ . Dies fassen wir in folgendem Satz zusammen.

**3.8 Satz.** Sei  $I = (a, b)$  und für  $f : I \times \mathbb{R} \rightarrow \mathbb{R}$  gelte  $f(t, y) = h(t) y + p(t)$ , wobei  $p$  und  $h$  auf  $I$  stetig sind. Dann existiert für alle  $(t_0, y_0) \in I \times \mathbb{R}$  eine eindeutige maximale Lösung  $y : I \rightarrow \mathbb{R}$  von (3.1) mit  $y(t_0) = y_0$ . Diese Lösung ist von der Form

$$y(t) = e^{H(t)} \left( y_0 + \int_{t_0}^t p(s) e^{-H(s)} ds \right),$$

wobei  $H : I \rightarrow \mathbb{R}$  durch

$$H(t) = \int_{t_0}^t h(s) ds$$

gegeben ist.

*Beweis.* Es gilt

$$y(t_0) = e^{H(t_0)}(y_0 + 0) = e^0 y_0 = y_0.$$

Zusammen mit den obigen Rechnungen ist  $y : I \rightarrow \mathbb{R}$  folglich eine maximale Lösung von (3.1). Seien weiter  $y_1, y_2$  Lösungen von (3.1) mit  $y_i(t_0) = y_0$ ,  $i = 1, 2$ , so ist  $y_0 := y_1 - y_2$  eine Lösung der homogenen Gleichung mit  $y_0(t_0) = 0$ . Wir hatten bereits gesehen, dass  $y_0(t) = 0$  die eindeutige Lösung der homogenen Gleichung mit  $y_0(t_0) = 0$  ist. Damit folgt  $y_1 = y_2$  und die Lösung  $y$  ist eindeutig. ■

# 13 Systeme linearer Differentialgleichungen

## 13.1 Grundlagen

Wir betrachten Systeme von Differentialgleichungen 1. Ordnung der Form

$$Y'(t) = \mathbf{A}(t) Y(t) + B(t) \quad (1.1)$$

mit  $\mathbf{A}(t) \in \mathbb{R}^{n \times n}$ ,  $B(t) \in \mathbb{R}^n$  sowie  $Y(t) \in \mathbb{R}^n$  auf einem Intervall  $I = (a, b)$ <sup>1</sup>. Definieren wir  $F : I \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  durch  $F(t, Y) := \mathbf{A}(t) Y + B(t)$ , so ist dies ein Spezialfall der Gleichung (2.3) aus Kapitel 12. Fordern wir also die Stetigkeit der Funktionen  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$  und  $B : I \rightarrow \mathbb{R}^n$ , so erfüllt  $F$  die Voraussetzung (S) auf  $I \times \mathbb{R}^n$  und ist insbesondere lokal Lipschitz stetig bezüglich  $Y$ . Nach Satz 2.20 aus Kapitel 12 existiert also für alle  $t_0 \in I$  und  $Y_0 \in \mathbb{R}^n$  genau eine maximale Lösung von (1.1) mit  $Y(t_0) = Y_0$ . Wir wollen nun zeigen, dass diese Lösung auf ganz  $I$  definiert ist.

**1.2 Lemma** (Gronwall). *Sei  $J$  ein Intervall,  $t_0 \in J$  und  $\alpha, \beta \in [0, \infty)$ . Ferner sei  $x : J \rightarrow [0, \infty)$  stetig und erfülle*

$$x(t) \leq \alpha + \beta \left| \int_{t_0}^t x(s) ds \right| \quad t \in J.$$

Dann gilt

$$x(t) \leq \alpha e^{\beta |t-t_0|} \quad t \in J.$$

*Beweis.* Wir betrachten zunächst den Fall  $t \geq t_0$  für  $t \in J$ . Dazu definieren wir

$$h(s) := \beta e^{\beta(t_0-s)} \int_{t_0}^s x(\tau) d\tau \quad s \in [t_0, t].$$

Dann gilt mit dem Hauptsatz der Differential- und Integralrechnung

---

<sup>1</sup> Wir lassen hier  $a = -\infty$  sowie  $b = \infty$  zu.

$$\begin{aligned}
h'(s) &= \beta e^{\beta(t_0-s)} (-1) \beta \int_{t_0}^s x(\tau) d\tau + \beta e^{\beta(t_0-s)} x(s) \\
&= -\beta h(s) + \beta e^{\beta(t_0-s)} x(s) \\
&\leq -\beta h(s) + \alpha \beta e^{\beta(t_0-s)} + \beta \beta e^{\beta(t_0-s)} \left| \int_{t_0}^s x(\tau) d\tau \right| \\
&= -\beta h(s) + \alpha \beta e^{\beta(t_0-s)} + \beta h(s) \\
&= \frac{d}{ds} (-\alpha e^{\beta(t_0-s)}).
\end{aligned}$$

Integrieren wir nun über das Intervall  $[t_0, t]$  so folgt

$$\int_{t_0}^t h'(s) ds = h(t) - h(t_0) = h(t) = \beta e^{\beta(t_0-t)} \int_{t_0}^t x(\tau) d\tau$$

bzw.

$$\int_{t_0}^t \frac{d}{ds} (-\alpha e^{\beta(t_0-s)}) ds = -\alpha e^{\beta(t_0-t)} + \alpha e^{\beta(t_0-t_0)} = \alpha - \alpha e^{\beta(t_0-t)}.$$

Damit folgt mit der Monotonie des Integrals

$$\beta e^{\beta(t_0-t)} \int_{t_0}^t x(\tau) d\tau \leq \alpha - \alpha e^{\beta(t_0-t)}$$

und somit

$$\beta \int_{t_0}^t x(\tau) d\tau \leq \alpha e^{\beta(t-t_0)} - \alpha.$$

Verwenden wir noch die Voraussetzung, so haben wir

$$x(t) \leq \alpha + \beta \int_{t_0}^t x(\tau) d\tau \leq \alpha e^{\beta(t-t_0)}$$

gezeigt. Analog folgt der Fall  $t < t_0$  für  $t \in J$ . ■

**1.3 Satz.** Seien  $\mathbf{A}, B$  stetige Funktionen auf dem Intervall  $I = (a, b)$ . Dann ist jede maximale Lösung von (1.1) auf ganz  $(a, b)$  definiert.

*Beweis.* Sei  $Y : (\alpha, \beta) \rightarrow \mathbb{R}^n$  eine maximale Lösung von (1.1) mit  $Y(t_0) = Y_0$  für ein  $t_0 \in (\alpha, \beta)$  und ein  $Y_0 \in \mathbb{R}^n$ . Da  $Y$  eine maximale Lösung ist kann  $\lim_{t \nearrow \beta} Y(t)$  nicht existieren. Anderenfalls gäbe es nach Lemma 2.14 aus Kapitel 12 eine Fortsetzung von  $Y$  über  $\beta$  hinaus. Dies ist aber nach Folgerung 2.19 aus Kapitel 12 ein Widerspruch zur Maximalität der Lösung. Gelte nun  $(\alpha, \beta) \subsetneq (a, b)$ , wobei wir oBdA  $\beta < b$  und somit insbesondere  $\beta < \infty$  annehmen können. Zu  $n \in \mathbb{N}$  definieren wir

$$K_n := \{(t, Y) \in \mathbb{R}^{n+1} \mid t \in [t_0, \beta], \|Y\| \leq n\}.$$

Offenbar ist  $K_n$  eine kompakte Teilmenge von  $(a, b) \times \mathbb{R}^n$ . Nach der Definition der maximalen Lösung (siehe Satz 2.18 aus Kapitel 12) ist

$$G^+ = \{(t, Z) \in \overline{\text{graph}(Y)} \mid t \geq t_0\}$$

keine kompakte Teilmenge von  $\Omega = (a, b) \times \mathbb{R}^n$ . Es gilt also  $G^+ \not\subseteq K_n$ . Da  $\lim_{t \nearrow \beta} Y(t)$  nicht existiert, muss  $\tau_n \in [t_0, \beta)$  existieren mit  $\|Y(\tau_n)\| = n$ . Somit existiert der Grenzwert

$$\lim_{n \rightarrow \infty} \|Y(\tau_n)\| = \infty.$$

Dies wollen wir zum Widerspruch führen. Dafür definieren wir

$$\begin{aligned} n(t) &:= \|Y(t)\|, & t \in (\alpha, \beta), \\ \delta &:= \max_{t \in [t_0, \beta]} \|\mathbf{A}(t)\|_{\mathbb{R}^n \times \mathbb{R}^n}, \\ \gamma &:= \max_{t \in [t_0, \beta]} \|B(t)\|_{\mathbb{R}^n}. \end{aligned}$$

Aufgrund der Stetigkeit von  $Y$ ,  $\mathbf{A}$  und  $B$  ist  $n$  stetig sowie  $\delta, \gamma$  wohldefiniert und endlich. Da  $Y$  Lösung von (1.1) ist, löst  $Y$  nach Lemma 2.7 aus Kapitel 12 die Integralgleichung

$$Y(t) = Y(t_0) + \int_{t_0}^t \mathbf{A}(s) Y(s) + B(s) ds, \quad t \in [t_0, \beta).$$

Mit der Cauchy-Schwarzschen Ungleichung und der Monotonie des Integrals folgt

$$\|Y(t)\| \leq \|Y(t_0)\| + \int_{t_0}^t \|\mathbf{A}(s)\| \|Y(s)\| + \|B(s)\| ds, \quad t \in [t_0, \beta)$$

und somit

$$n(t) \leq \|Y(t_0)\| + (\beta - t_0)\gamma + \delta \int_{t_0}^t n(s) ds, \quad t \in [t_0, \beta).$$

Wenden wir nun das **Gronwallsche Lemma** mit  $x(t) = n(t)$ ,  $\beta = \delta$  und  $\alpha = \|Y(t_0)\| + (\beta - t_0)\gamma$  an, so folgt

$$\|Y(t)\| = n(t) \leq \left( \|Y(t_0)\| + (\beta - t_0)\gamma \right) e^{\delta|t-t_0|}, \quad t \in [t_0, \beta).$$

Dabei ist die rechte Seite stetig auf dem Intervall  $[t_0, \beta]$  und somit gleichmäßig beschränkt, ein Widerspruch zu  $\lim_{n \rightarrow \infty} \|Y(\tau_n)\| = \infty$ . Die Annahme  $(\alpha, \beta) \subsetneq (a, b)$  war somit falsch und die maximale Lösung ist folglich auf dem ganzen Intervall  $(a, b)$  definiert. ■

**1.4 Definition.** Das System (1.1) heißt homogen, falls  $B(t) \equiv 0$  gilt. Das homogene System (1.1) reduziert sich also zu

$$Y'(t) = \mathbf{A}(t) Y(t). \quad (1.5)$$

Ist  $B(t) \not\equiv 0$ , so heißt das System inhomogen.

Wir wenden uns zunächst den homogenen Systemen zu.

## 13.2 Homogene Systeme

Die (maximalen) Lösungen des homogenen Systems (1.5) bilden einen Vektorraum. In der Tat, seien  $Y_1, Y_2$  Lösungen von (1.5), so folgt für  $\alpha \in \mathbb{R}$

$$\begin{aligned} \frac{d}{dt}(Y_1(t) + Y_2(t)) &= Y_1'(t) + Y_2'(t) \\ &= \mathbf{A}(t) Y_1(t) + \mathbf{A}(t) Y_2(t) \\ &= \mathbf{A}(t) (Y_1(t) + Y_2(t)) \end{aligned}$$

sowie

$$\frac{d}{dt}(\alpha Y_1(t)) = \alpha Y_1'(t) = \alpha \mathbf{A}(t) Y_1(t) = \mathbf{A}(t) (\alpha Y_1(t)).$$

Dieser Vektorraum ist ein Teilraum von  $C^1(I)$ .

**2.1 Lemma.** Sei  $I = (a, b)$  und  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$  eine stetige Funktion. Dann hat der Vektorraum der maximalen Lösungen des homogenen Systems (1.5) die Dimension  $n$ .

*Beweis.* Sei  $I = (a, b)$  und  $t_0 \in I$  fest und sei  $E_i, i = 1, \dots, n$  die Standardbasis von  $\mathbb{R}^n$ . Dann gilt:

1. Es existieren  $n$  linear unabhängige Lösungen:  
Seien  $Z_i : I \rightarrow \mathbb{R}^{n \times n}$  die maximalen Lösungen von

$$\begin{aligned} Z_i'(t) &= \mathbf{A}(t) Z_i(t) \\ Z_i(t_0) &= E_i \end{aligned}$$

für  $i = 1, \dots, n$ . Dann sind die  $Z_i$  als Funktionen auf  $I$  linear unabhängig, da diese in  $t_0$  linear unabhängige Vektoren sind.

2. Die  $Z_i$ ,  $i = 1, \dots, n$  bilden eine Basis:  
Sei  $Y$  eine Lösung des homogenen Systems mit

$$Y(t_0) = (y_1(t_0), \dots, y_n(t_0)).$$

Dann ist

$$Z(t) := \sum_{i=1}^n y_i(t_0) Z_i(t) \quad t \in I$$

Lösung des homogenen Systems und es gilt

$$Z(t_0) = \sum_{i=1}^n y_i(t_0) E_i = Y(t_0).$$

Aufgrund der Eindeutigkeit der Lösung mit Anfangswert  $(t_0, Y(t_0))$  folgt  $Y = Z$ . Die  $Z_i$ ,  $i = 1, \dots, n$  bilden also eine Basis. ■

**2.2 Definition.** Sei  $I = (a, b)$  und  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$  eine stetige Funktion. Jede Basis des Lösungsraumes des homogenen Systemes (1.5) nennen wir Fundamentalsystem von (1.5). Sei  $Y^{(i)}$ ,  $i = 1, \dots, n$  ein Fundamentalsystem von (1.5), dann heißt die Matrix

$$\mathbf{Y} := (Y^{(1)}, \dots, Y^{(n)}) = \begin{pmatrix} Y_1^{(1)} & \dots & Y_1^{(n)} \\ \vdots & & \vdots \\ Y_n^{(1)} & \dots & Y_n^{(n)} \end{pmatrix}$$

Fundamentalmatrix von (1.5).

**Bemerkungen.** 1. Sei  $\mathbf{Y} = (Y_{i,j})_{i,j=1,\dots,n}$  eine Fundamentalmatrix von (1.5). Setzen wir  $\mathbf{Y}' := (Y'_{i,j})_{i,j=1,\dots,n}$ , so folgt  $\mathbf{Y}' = \mathbf{A} \mathbf{Y}$ .

2. Sei  $t_0 \in I$  und  $\mathbf{Y}$  eine Fundamentalmatrix von (1.5). Gilt  $\mathbf{Y}(t_0) = \mathbf{E}$ , wobei  $\mathbf{E}$  die Einheitsmatrix bezeichnet, so ist die Lösung von

$$Y' = \mathbf{A} Y, \quad Y(t_0) = Y_0$$

durch  $Y(t) := \mathbf{Y}(t) Y_0$  gegeben (vergleiche Beweis von Lemma 2.1).

3. Sei  $t_0 \in I$  und  $\mathbf{Y}$  eine Fundamentalmatrix von (1.5). Für alle  $t_0$  ist die Matrix  $\mathbf{Y}(t_0)$  invertierbar. Anderenfalls gäbe es eine nichttriviale Linearkombination der Spaltenvektoren  $Y^{(i)}(t_0)$ , d.h.  $\sum_{i=1}^n \lambda_i Y^{(i)}(t_0) = 0$ . Dann würde  $Y(t) := \sum_{i=1}^n \lambda_i Y^{(i)}(t)$  das System  $Y' = \mathbf{A} Y$  mit  $Y(t_0) = 0$  lösen. Da die Lösungen von (1.5) eindeutig sind, folgt  $Y(t) = 0$ , was ein Widerspruch zur Definition des Fundamentalsystems wäre. Definieren wir

$$\mathbf{Z}(t) := \mathbf{Y}(t) \mathbf{Y}^{-1}(t_0),$$

so gilt

$$\mathbf{Z}' = \mathbf{Y}' \mathbf{Y}^{-1}(t_0) = \mathbf{A} \mathbf{Y} \mathbf{Y}^{-1}(t_0) = \mathbf{A} \mathbf{Z}.$$

Insbesondere gilt  $\mathbf{Z}(t_0) = \mathbf{E}$ . Damit ist  $\mathbf{Z}$  wieder eine Fundamentalmatrix von (1.5). Die Lösung von

$$Y' = \mathbf{A} Y, \quad Y(t_0) = Y_0$$

ist also durch

$$Y(t) = \mathbf{Y}(t) \mathbf{Y}^{-1}(t_0) Y_0$$

gegeben.

4. Im Allgemeinen existiert keine Methode die Fundamentalmatrix zu bestimmen, falls  $n > 1$  gilt und  $\mathbf{A}(t)$  nicht konstant ist.

**2.3 Definition.** Sei  $I = (a, b)$  und  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$  eine stetige Funktion. Sei  $\mathbf{Y}$  eine Fundamentalmatrix des homogenen Systems (1.5). Wir nennen

$$W(t) := \det \mathbf{Y}(t), \quad t \in I$$

die Wronski-Determinante.

**2.4 Satz.** Sei  $I = (a, b)$ ,  $t_0 \in I$  und  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$  eine stetige Funktion. Sei  $\mathbf{Y}$  eine Fundamentalmatrix des homogenen Systems (1.5). Dann gilt für die zugehörige Wronski-Determinante

$$W(t) = W(t_0) e^{\int_{t_0}^t \operatorname{tr} \mathbf{A}(s) ds}, \quad t \in I,$$

wobei für  $\mathbf{A} = (a_{ij})_{i,j=1,\dots,n}$  die Spur der Matrix durch  $\operatorname{tr} \mathbf{A} = \sum_{i=1}^n a_{ii}$  definiert ist.

*Beweis.* Es gilt  $\operatorname{tr} \mathbf{A} \in \mathbb{R}$  sowie  $W = \det \mathbf{Y} \in \mathbb{R}$ , wir betrachten also eine skalare Gleichung. Nach der Theorie der linearen Gleichungen (siehe Satz 3.8 aus Kapitel 12) gilt die gewünschte Formel (als Lösungsformel von (3.1) aus Kapitel 12 mit Anfangswert  $W(t_0)$ ) genau dann, wenn

$$W'(t) = \operatorname{tr} \mathbf{A}(t) W(t), \quad t \in I$$

gilt. Wir benötigen also eine Formel für die Ableitung der Determinante. Es gilt für  $\mathbf{B} = (B_1, \dots, B_n) = (b_{ij})_{i,j=1,\dots,n}$

$$\det \mathbf{B} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)},$$

wobei  $S_n$  die symmetrische Gruppe vom Grad  $n$  ist und das Signum der Permutation  $\sigma$  durch  $\operatorname{sgn}(\sigma) := \det(E_{\sigma(1)}, \dots, E_{\sigma(n)}) \in \{-1, 1\}$  gegeben ist. Somit gilt

$$\begin{aligned} (\det \mathbf{B})' &= \sum_{i=1}^n \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} \cdots b'_{i\sigma(i)} \cdots b_{n\sigma(n)} \\ &= \sum_{i=1}^n \det(B_1, \dots, B'_i, \dots, B_n). \end{aligned}$$

Wählen wir nun ein beliebiges  $t_1 \in I$ , so können wir wie in den obigen Bemerkungen eine Fundamentalmatrix  $\mathbf{Z}(t) = (Z_1(t), \dots, Z_n(t))$  des homogenen Systems (1.5) mit  $\mathbf{Z}(t_1) = \mathbf{E}$ , also  $Z_i(t_1) = E_i$  konstruieren. Wie wir gesehen hatten gilt  $Z'_i(t_1) = \mathbf{A}(t_1) Z_i(t_1) = \mathbf{A}(t_1) E_i$  sowie

$$\mathbf{Y}(t) = \mathbf{Z}(t) \mathbf{Y}(t_1), \quad t \in I.$$

Für die Wronski-Determinante folgt

$$W(t) = \det \mathbf{Y}(t) = \det \mathbf{Z}(t) \det \mathbf{Y}(t_1) = \det \mathbf{Z}(t) W(t_1), \quad t \in I$$

und somit

$$W'(t) = (\det \mathbf{Z}(t))' W(t_1), \quad t \in I.$$

Mit der Formel für die Ableitung der Determinante erhalten wir

$$\begin{aligned} (\det \mathbf{Z}(t_1))' &= \sum_{i=1}^n \det(Z_1(t_1), \dots, Z'_i(t_1), \dots, Z_n(t_1)) \\ &= \sum_{i=1}^n \det(E_1, \dots, \mathbf{A}(t_1) E_i, \dots, E_n) \\ &= \sum_{i=1}^n a_{ii}(t_1) \\ &= \operatorname{tr} \mathbf{A}(t_1). \end{aligned}$$

Es gilt folglich

$$W'(t_1) = \operatorname{tr} \mathbf{A}(t_1) W(t_1).$$

Da  $t_1 \in I$  beliebig war, folgt also mit Satz 3.8 die Behauptung. ■

**2.5 Folgerung.** Die Wronski-Determinante einer Fundamentalmatrix ist überall ungleich Null.

### 13.3 Inhomogene Systeme

Man sieht leicht, dass eine analoge Version von Lemma 3.7 aus Kapitel 12 auch für Systeme von linearen Differentialgleichungen 1. Ordnung gilt: Falls  $Y_1, Y_2$  Lösungen des inhomogenen Systemes sind, so ist  $Y_1 - Y_2$  eine Lösung des homogenen Systemes. Falls weiter  $Y_0$  eine Lösung des homogenen Systemes ist, so ist  $Y_0 + Y_1$  eine Lösung des inhomogenen Systemes. Anders ausgedrückt:

$$\begin{aligned} & \text{Lösungen des inhomogenen Systems} \\ &= \text{spezielle Lösung des inhomogenen Systems} \\ &+ \text{Lösungen des homogenen Systems.} \end{aligned}$$

Wir können also wie im Fall  $n = 1$  die Methode der Variation der Konstanten verwenden. Ersetzen wir die Konstante  $Y_0$  mit  $C(t)$  in der Lösungsformel, betrachten wir also

$$Z(t) = \mathbf{Y}(t) \mathbf{Y}^{-1}(t_0) C(t),$$

so kommen wir auf die Lösungsformel im folgenden Satz.

**3.1 Satz.** Sei  $I = (a, b)$  und  $\mathbf{A} : I \rightarrow \mathbb{R}^{n \times n}$ ,  $B : I \rightarrow \mathbb{R}^n$  stetige Funktionen. Sei  $\mathbf{Y}$  eine Fundamentalmatrix des homogenen Systems (1.5). Dann gibt es für alle  $(t_0, Y_0) \in I \times \mathbb{R}^n$  eine eindeutige maximale Lösung von (1.1) mit  $Y(t_0) = Y_0$ . Diese Lösung hat die Form

$$Y(t) = \mathbf{Y}(t) \mathbf{Y}^{-1}(t_0) Y_0 + \mathbf{Y}(t) \int_{t_0}^t \mathbf{Y}^{-1}(s) B(s) ds, \quad t \in I.$$

*Beweis.* Die Existenz einer eindeutigen maximalen Lösung wurde schon bewiesen (vergleiche die Einleitung des Kapitels bzw. Satz 2.20 aus Kapitel 12). Ist  $Y$  wie im Satz definiert, so gilt

$$Y(t_0) = \mathbf{Y}(t_0) \mathbf{Y}^{-1}(t_0) Y_0 + \mathbf{Y}(t_0) \int_{t_0}^{t_0} \mathbf{Y}^{-1}(s) B(s) ds = Y_0.$$

Weiter ist aufgrund der Eigenschaften der Fundamentalmatrix die Funktion  $Y(t) = \mathbf{Y}(t) \mathbf{Y}^{-1}(t_0) Y_0$  eine Lösung des homogenen Systems (1.5) mit  $Y(t_0) = Y_0$ . Wir müssen also nur zeigen, dass

$$Z(t) := \mathbf{Y}(t) \int_{t_0}^t \mathbf{Y}^{-1}(s) B(s) ds, \quad t \in I$$

eine Lösung des inhomogenen Systems (1.1) ist. Dazu berechnen wir

$$\begin{aligned}
Z'(t) &= \mathbf{Y}'(t) \int_{t_0}^t \mathbf{Y}^{-1}(s) B(s) ds + \mathbf{Y}(t) \mathbf{Y}^{-1}(t) B(t) \\
&= \mathbf{A}(t) \mathbf{Y}(t) \int_{t_0}^t \mathbf{Y}^{-1}(s) B(s) ds + B(t) \\
&= \mathbf{A}(t) Z(t) + B(t).
\end{aligned}$$

Damit ist  $Z$  eine spezielle Lösung von (1.1) mit  $Z(t_0) = 0$  und folglich  $Y$  die maximale Lösung von (1.1) mit  $Y(t_0) = Y_0$ . ■

### 13.4 Systeme mit konstantem $\mathbf{A}$

Wir betrachten Systeme von Differentialgleichungen 1. Ordnung der Form

$$Y'(t) = \mathbf{A} Y(t) \quad (4.1)$$

mit konstanter Matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . Betrachten wir die lineare Abbildung

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^n : Y \mapsto \mathbf{A} Y,$$

so gilt

$$\mathbf{A} = \mathbf{M}_E^E(A),$$

wobei  $\mathbf{M}_E^E(A)$  die Darstellungsmatrix von  $A$  bezüglich der Standardbasis ist. Dabei ist für zwei Basen  $B = \{B_1, \dots, B_n\}$ ,  $C = \{C_1, \dots, C_n\}$  mit  $B_i, C_i \in \mathbb{R}^n$ ,  $i = 1, \dots, n$ , die Matrix  $\mathbf{M}_C^B(A)$  durch die Koordinatenspaltenvektoren bezüglich  $C$ , die sich aus den Bildern  $A(B_1), \dots, A(B_n)$  der Basis  $B$  ergeben, gegeben. Bezüglich der Basis  $B$  hat die Abbildung  $A$  also eine andere Darstellungsmatrix  $\mathbf{M}_B^B(A)$ . Wir suchen nun eine solche Basis, so dass  $\mathbf{M}_B^B(A)$  möglichst einfach ist. Aus der Linearen Algebra ist bekannt, dass

$$\mathbf{M}_B^B(A) = \mathbf{M}_B^E(id) \mathbf{M}_E^E(A) \mathbf{M}_E^B(id)$$

gilt. Weiter ist die Matrix  $\mathbf{M}_E^E(id) =: \mathbf{B}$  invertierbar und es gilt  $\mathbf{M}_B^E(id) = (\mathbf{M}_E^B(id))^{-1}$ . Es folgt also

$$\mathbf{M}_B^B(A) = \mathbf{B}^{-1} \mathbf{A} \mathbf{B} =: \mathbf{D}.$$

Setzen wir

$$Z(t) := \mathbf{B}^{-1} Y(t) \quad \text{und somit} \quad Y(t) = \mathbf{B} Z(t),$$

so folgt aus (4.1) die Gleichung

$$Z'(t) = \mathbf{B}^{-1} Y'(t) = \mathbf{B}^{-1} \mathbf{A} Y(t) = \mathbf{B}^{-1} \mathbf{A} \mathbf{B} Z(t) = \mathbf{D} Z(t). \quad (4.2)$$

### 13.4.1 Symmetrische Matrizen $\mathbf{A}$

Betrachten wir nun den speziellen Fall, dass  $\mathbf{A}$  symmetrisch ist. Dann existiert, wie in der Linearen Algebra gezeigt wurde, eine Matrix  $\mathbf{B}$ , so dass  $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$  gilt. Die Gleichung

$$Z'(t) = \text{diag}(\lambda_1, \dots, \lambda_n) Z(t)$$

lässt sich also als System

$$\begin{aligned} z_1'(t) &= \lambda_1 z_1(t), \\ &\vdots \\ z_n'(t) &= \lambda_n z_n(t) \end{aligned}$$

von  $n$  linearen Differentialgleichungen 1. Ordnung schreiben. Diese haben nach Satz 3.8 aus Kapitel 12 die Lösungen  $z_i(t) = c_i e^{\lambda_i t}$ ,  $i = 1, \dots, n$ . Mit der Wahl von  $Z(t) := (z_1(t), \dots, z_n(t))$  haben wir somit eine Lösung der Gleichung (4.2) gefunden. Mit der Notation  $Z_i(t) := e^{\lambda_i t} E_i$  ist

$$\mathbf{Z}(t) = (Z_1(t), \dots, Z_n(t)) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & e^{\lambda_n t} \end{pmatrix}, \quad t \in \mathbb{R} \quad (4.3)$$

eine Fundamentalmatrix von  $Z' = \mathbf{D}Z$ . Um zurück zur Gleichung (4.1) zu kommen, setzen wir mit der Notation  $\mathbf{B} = (B_1, \dots, B_n)$

$$\mathbf{Y} = \mathbf{B}\mathbf{Z} = (\mathbf{B}Z_1, \dots, \mathbf{B}Z_n) = (B_1 e^{\lambda_1 t}, \dots, B_n e^{\lambda_n t}).$$

Offenbar ist  $\mathbf{Y}$  eine Fundamentalmatrix von (4.1) und wir können die Lösungsformel aus Satz 3.1 für inhomogene Systeme anwenden.

### 13.4.2 Matrizen $\mathbf{A}$ mit nur reellen Eigenwerten

Betrachten wir nun den allgemeineren Fall einer Matrix  $\mathbf{A}$ , welche nur reelle Eigenwerte besitzt. In diesem Fall liefert die Lineare Algebra die Existenz einer Matrix  $\mathbf{B}$ , so dass

$$\mathbf{D} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{J}_m \end{pmatrix}, \quad \text{wobei} \quad \mathbf{J}_i = \begin{pmatrix} \lambda_i & 0 & \dots & \dots & 0 \\ 1 & \lambda_i & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & \lambda_i \end{pmatrix}$$

für  $i = 1, \dots, m$  gilt. Dabei wird  $\mathbf{J}_i \in \mathbb{R}^{k \times k}$  *Jordan Block der Ordnung  $k$*  genannt. Insbesondere sind mehrfache Eigenwerte, also  $\lambda_i = \lambda_j$  für  $i \neq j$  möglich. Analog zum Fall der symmetrischen Matrizen reicht es, sich die einzelnen Jordan Blöcke anzusehen: Für einen Jordan Block müssen wir also die Gleichungen

$$\begin{aligned} z_1'(t) &= \lambda z_1(t), \\ z_2'(t) &= z_1(t) + \lambda z_2(t), \\ &\vdots \\ z_k'(t) &= z_{k-1}(t) + \lambda z_k(t) \end{aligned}$$

lösen. Wir hatten schon gesehen, dass  $z_1(t) = c_1 e^{\lambda t}$  eine Lösung der ersten Gleichung ist. Für die zweite Gleichung verwenden wir wieder die Methode der Variation der Konstanten: Betrachten wir

$$z_2(t) = c(t) e^{\lambda t},$$

so folgt

$$z_2'(t) = c'(t) e^{\lambda t} + \lambda c(t) e^{\lambda t} = c'(t) e^{\lambda t} + \lambda z_2(t).$$

Wählen wir also  $c'(t) = c_1$  bzw.  $c(t) = c_1 t + c_2$ , so ist  $z_2(t) = (c_1 t + c_2) e^{\lambda t}$  eine Lösung der zweiten Gleichung. Analog erhalten wir, dass

$$\begin{aligned} z_1(t) &= c_1 e^{\lambda t}, \\ z_2(t) &= (c_1 t + c_2) e^{\lambda t}, \\ &\vdots \\ z_k(t) &= \left( c_1 \frac{t^{k-1}}{(k-1)!} + \dots + c_k \right) e^{\lambda t} \end{aligned}$$

eine Lösung des obigen Gleichungssystems ist. Eine (Teil-) Fundamentalmatrix von  $Z' = \mathbf{D} Z$  bezüglich dieses Jordan Blocks  $\mathbf{J}_i$  ist also durch

$$\mathbf{Z}_i = (Z_1^i, \dots, Z_k^i) = \begin{pmatrix} e^{\lambda_i t} & 0 & \dots & 0 \\ t e^{\lambda_i t} & e^{\lambda_i t} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \frac{t^{k-1}}{(k-1)!} e^{\lambda_i t} & \frac{t^{k-2}}{(k-2)!} e^{\lambda_i t} & \dots & e^{\lambda_i t} \end{pmatrix} \quad (4.4)$$

gegeben. Setzt man nun die Jordan Blöcke zusammen, so ist

$$\mathbf{Z} = (Z_1, \dots, Z_n) = \begin{pmatrix} \mathbf{Z}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{Z}_m \end{pmatrix}$$

eine Fundamentalmatrix von  $Z' = \mathbf{D}Z$ . Beispielsweise erhalten wir aus der unten gegebenen Matrix  $\mathbf{D}$  die Fundamentalmatrix  $\mathbf{Z}$

$$\mathbf{D} = \left( \begin{array}{c|cc} \mu & 0 & 0 \\ \hline 0 & \lambda & 0 \\ 0 & 1 & \lambda \end{array} \right) \quad \mathbf{Z} = \left( \begin{array}{c|cc} e^{\mu t} & 0 & 0 \\ \hline 0 & e^{\lambda t} & 0 \\ 0 & t e^{\lambda t} & e^{\lambda t} \end{array} \right).$$

Die Rücktransformation

$$\mathbf{Y} = \mathbf{B}\mathbf{Z}$$

ist im Allgemeinen schwer aufzuschreiben, kann aber in jedem konkreten Fall ausgeführt werden.

**Beispiele.** (a) Sei die Matrix  $\mathbf{A}$  durch

$$\mathbf{A} = \begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix}$$

gegeben. Wir müssen zunächst die Matrix  $\mathbf{A}$  in Jordan Normalform bringen. Dazu berechnen wir das charakteristische Polynom

$$\det(\mathbf{A} - \lambda \mathbf{E}) = \begin{vmatrix} 1 - \lambda & 4 \\ -1 & -3 - \lambda \end{vmatrix} = -(1 - \lambda)(3 + \lambda) + 4 = (\lambda + 1)^2$$

und erhalten  $\lambda = -1$  als doppelten Eigenwert von  $\mathbf{A}$ . Um eine Basis des Eigenraumes  $E(-1) = \text{Ker}(\mathbf{A} + \mathbf{E})$  zu bestimmen, betrachten wir

$$\begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dieses System wird z.B. von  $u = -2$  sowie  $v = 1$  erfüllt und wir erhalten den Eigenvektor  $V_1 = (-2, 1)^\top$ , der eine Basis von  $E(-1)$  ist. Insbesondere ist  $\dim E(-1) = 1$ . Zur Bestimmung der Nilpotenzordnung des Eigenwertes  $\lambda = -1$  betrachten wir

$$(\mathbf{A} + \mathbf{E})^2 = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Die Nilpotenzordnung ist also gleich 2 und der verallgemeinerte Eigenraum  $F(-1) = \text{Ker}(\mathbf{A} + \mathbf{E})^2$ . Dieser enthält den Eigenraum  $E(-1)$ . Eine Basis von  $F(-1)$  ist z.B. durch  $(V_2, V_1)$  mit  $V_2 = E_1$  gegeben. Bezüglich dieser Basis muss die Darstellungsmatrix von  $F(-1)$  allerdings nicht die Gestalt eines Jordan Blockes haben. Um dies sicherzustellen modifizieren wir den Basisvektor  $V_1$  wie folgt:

$$\tilde{V}_1 = (\mathbf{A} + \mathbf{E})V_2 = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Somit ist  $(V_2, \tilde{V}_1)$  eine Basis von  $F(-1)$  bezüglich derer die Darstellungsmatrix von  $F(-1)$  die Gestalt eines Jordan Blockes hat. Wir erhalten somit die Basiswechselmatrix  $\mathbf{B}$  und die Darstellungsmatrix  $\mathbf{D}$  von  $\mathbf{A}$  bezüglich dieser Basis als

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}.$$

Die Fundamentalmatrix von  $Z' = \mathbf{D}Z$  ist also durch

$$\mathbf{Z}(t) = \begin{pmatrix} e^{-t} & 0 \\ t e^{-t} & e^{-t} \end{pmatrix}$$

gegeben. Die Rücktransformation ergibt

$$\mathbf{Y}(t) = \mathbf{B}\mathbf{Z}(t) = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} e^{-t} & 0 \\ t e^{-t} & e^{-t} \end{pmatrix} = \begin{pmatrix} e^{-t} + 2t e^{-t} & 2e^{-t} \\ -t e^{-t} & -e^{-t} \end{pmatrix}$$

als Fundamentalmatrix von  $Y' = \mathbf{A}Y$ . Um die Lösungsformel mit  $t_0 = 0$  anwenden zu können, müssen wir noch  $\mathbf{Y}(0)^{-1}$  berechnen. Offenbar gilt

$$\mathbf{Y}(0) = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} = \mathbf{Y}(0)^{-1}.$$

Damit ist für alle  $Y_0 \in \mathbb{R}^2$

$$Y(t) := \mathbf{Y}(t)\mathbf{Y}(0)^{-1}Y_0$$

Lösung von (4.1) mit  $Y(0) = Y_0$ .

(b) Sei die Matrix  $\mathbf{A}$  durch

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & -3 \\ 1 & 4 & 4 \\ 0 & 0 & -1 \end{pmatrix}$$

gegeben. Wir berechnen wieder das charakteristische Polynom

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{E}) &= \begin{vmatrix} 2 - \lambda & -1 & -3 \\ 1 & 4 - \lambda & 4 \\ 0 & 0 & -1 - \lambda \end{vmatrix} \\ &= (-1 - \lambda) \begin{vmatrix} 2 - \lambda & -1 \\ 1 & 4 - \lambda \end{vmatrix} \\ &= -(1 + \lambda) \left( (2 - \lambda)(4 - \lambda) + 1 \right) \\ &= -(1 + \lambda)(\lambda - 3)^2 \end{aligned}$$

und erhalten  $\lambda_1 = -1$  als einfachen Eigenwert sowie  $\lambda_2 = 3$  als doppelten Eigenwert. Zur Bestimmung des Eigenraumes  $E(3)$  und des verallgemeinerten Eigenraumes  $F(3)$  betrachten wir

$$(\mathbf{A} - 3\mathbf{E}) = \begin{pmatrix} -1 & -1 & -3 \\ 1 & 1 & 4 \\ 0 & 0 & -4 \end{pmatrix}, \quad (\mathbf{A} - 3\mathbf{E})^2 = \begin{pmatrix} 0 & 0 & 11 \\ 0 & 0 & -15 \\ 0 & 0 & 16 \end{pmatrix},$$

sowie

$$(\mathbf{A} - 3\mathbf{E})^3 = \begin{pmatrix} 0 & 0 & -44 \\ 0 & 0 & 60 \\ 0 & 0 & -64 \end{pmatrix}.$$

Die Nilpotenzordnung ist also 2. Weiterhin gilt:  $E(3) = \text{Ker}(\mathbf{A} - 3\mathbf{E})$ ,  $F(3) = \text{Ker}(\mathbf{A} - 3\mathbf{E})^2$  sowie  $\dim E(3) = 1$ ,  $\dim F(3) = 2$ . Eine Basis von  $E(3)$  ist z.B. durch

$$V_1 := \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

gegeben. Diese ergänzen wir durch

$$V_2 := \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

zu einer Basis von  $F(3)$ . Um sicherzustellen, dass die Darstellungsmatrix ein Jordan Block ist, modifizieren wir  $V_1$  wie folgt:

$$\tilde{V}_1 = (\mathbf{A} - 3\mathbf{E})V_2 = \begin{pmatrix} -1 & -1 & -3 \\ 1 & 1 & 4 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix}.$$

Weiter benötigen wir eine Basis von  $E(-1)$ . Mit

$$\mathbf{A} - (-1)\mathbf{E} = \begin{pmatrix} 3 & -1 & -3 \\ 1 & 5 & 4 \\ 0 & 0 & 0 \end{pmatrix}$$

sieht man leicht, dass

$$V_3 = \begin{pmatrix} 11 \\ -15 \\ 16 \end{pmatrix}$$

eine Basis von  $\text{Ker}(\mathbf{A} + \mathbf{E}) = E(-1)$  ist. Bezüglich der Basis  $(V_3, V_2, \tilde{V}_1)$  ist nun die Darstellungsmatrix  $\mathbf{D}$  von  $\mathbf{A}$  sowie die Basiswechselform  $\mathbf{B}$  durch

$$\mathbf{B} = \begin{pmatrix} 11 & 1 & -2 \\ -15 & 1 & 2 \\ 16 & 0 & 0 \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 1 & 3 \end{pmatrix}$$

gegeben. Damit erhalten wir die Fundamentalmatrix von  $Z' = \mathbf{D}Z$  als

$$\mathbf{Z}(t) = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{3t} & 0 \\ 0 & t e^{3t} & e^{3t} \end{pmatrix}.$$

Die Rücktransformation ergibt

$$\mathbf{Y}(t) = \begin{pmatrix} 11 & 1 & -2 \\ -15 & 1 & 2 \\ 16 & 0 & 0 \end{pmatrix} \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{3t} & 0 \\ 0 & t e^{3t} & e^{3t} \end{pmatrix} = \begin{pmatrix} 11 e^t & (e^{3t} - 2t e^{3t}) & -2 e^{3t} \\ -15 e^t & (e^{3t} + 2t e^{3t}) & 2 e^{3t} \\ 16 e^t & 0 & 0 \end{pmatrix}.$$

Für eine Lösung von (4.1) können wir also wieder die Lösungsformel anwenden.

### 13.4.3 Matrizen $\mathbf{A}$ mit komplexen Eigenwerten

Falls  $\lambda = a + ib$  ein (komplexer) Eigenwert der (reellen) Matrix  $\mathbf{A}$  ist, so ist auch  $\bar{\lambda} = a - ib$  ein Eigenwert der Matrix  $\mathbf{A}$ . Falls die Dimension des Systems  $n \leq 3$  ist, kann es folglich höchstens die einfachen komplexen Eigenwerte  $\lambda = a + ib$  und  $\bar{\lambda} = a - ib$  geben. Wir werden im weiteren nur den speziellen Fall  $n = 2$  betrachten. Gehen wir nun wie im Fall **symmetrischer Matrizen** vor (und erlauben dabei komplexe Koeffizienten), so transformieren wir die Matrix  $\mathbf{A}$  mithilfe der komplexen Matrix  $\tilde{\mathbf{B}}^2$  in die *komplexe Jordanmatrix*

$$\tilde{\mathbf{A}} := \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix}$$

und erhalten die *komplexe Fundamentalmatrix*

$$\tilde{\mathbf{Z}}(t) = \begin{pmatrix} e^{\lambda t} & 0 \\ 0 & e^{\bar{\lambda} t} \end{pmatrix}$$

von  $Z' = \tilde{\mathbf{A}}Z$ . Wir suchen nun eine reelle Darstellung dieser Fundamentalmatrix bzw. der Fundamentalmatrix des ursprünglichen Systems  $Y' = \mathbf{A}Y$ . Sei dafür  $V \in \mathbb{C}^2$  ein (komplexer) Eigenvektor von  $\mathbf{A}$  zum Eigenwert  $\lambda$ , d.h.  $\mathbf{A}V = \lambda V$ . Für den (komponentenweise) komplex konjugierten Vektor  $\bar{V}$  gilt die Gleichung  $\mathbf{A}\bar{V} = \bar{\lambda}\bar{V}$ . Somit ist  $\bar{v}$  ein Eigenvektor von  $\mathbf{A}$  zum Eigenwert  $\bar{\lambda}$ . Wir betrachten nun die Darstellung des komplexen Eigenvektors

<sup>2</sup> Die Matrix  $\tilde{\mathbf{B}}$  besteht aus den Spaltenvektoren  $V$  und  $\bar{V}$ , wobei  $V$  ein komplexer Eigenvektor von  $\mathbf{A}$  ist.

und Eigenwertes  $V = U + iW$  mit  $U, W \in \mathbb{R}^2$  und  $\lambda = a + ib$  für  $b > 0$ . Damit zerlegen wir die Gleichung  $\mathbf{A}V = \lambda V$  in den Imaginär- und Realteil:

$$\begin{aligned}\mathbf{A}U &= aU - bW, \\ \mathbf{A}W &= aW + bU.\end{aligned}$$

Dies können wir auch als Matrixmultiplikation schreiben:

$$\mathbf{A} \begin{pmatrix} U & W \end{pmatrix} = \begin{pmatrix} U & W \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Da  $V$  und  $\bar{V}$  linear unabhängig sind, folgt auch die lineare Unabhängigkeit von  $U$  und  $W$ . Somit ist  $\mathbf{B} := \begin{pmatrix} U & W \end{pmatrix}$  regulär und es gilt

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} =: \hat{\mathbf{A}}. \quad (4.5)$$

Die Matrix  $\mathbf{B}$  transformiert also  $\mathbf{A}$  in die *reelle Normalform*  $\hat{\mathbf{A}}$ . Wir können jetzt direkt die Fundamentalmatrix des Systems  $Z' = \hat{\mathbf{A}}Z$  ausrechnen. Aufgrund folgender Überlegung kann man diese aber direkt aus der komplexen Fundamentalmatrix des Systems  $Z' = \tilde{\mathbf{A}}Z$  ablesen. Wenn man das Transformationsverhalten der Matrix  $\mathbf{A}$  beachtet, d.h.

$$\begin{pmatrix} a + ib & 0 \\ 0 & a - ib \end{pmatrix} = \tilde{\mathbf{A}} \xleftarrow[\text{mit } \tilde{\mathbf{B}}]{\text{Transformation}} \mathbf{A} \xrightarrow[\text{mit } \mathbf{B}]{\text{Transformation}} \hat{\mathbf{A}} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

und vermutet, dass sich dieses Transformationsverhalten auf die Fundamentalmatrix  $\mathbf{Y}(t)$  des Systems  $Y' = \mathbf{A}Y$  überträgt, so erwarten wir

$$\begin{aligned}e^{at} \begin{pmatrix} \cos(bt) + i \sin(bt) & 0 \\ 0 & \cos(bt) - i \sin(bt) \end{pmatrix} &= \tilde{\mathbf{Z}}(t) \\ \xleftarrow[\text{mit } \tilde{\mathbf{B}}]{\text{Transformation}} \text{Fundamentalmatrix } \mathbf{Y}(t) &\xrightarrow[\text{mit } \mathbf{B}]{\text{Transformation}} \\ \mathbf{Z}(t) &:= e^{at} \begin{pmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{pmatrix}.\end{aligned}$$

Durch Nachrechnen kann man leicht überprüfen, dass

$$\mathbf{Z}(t) = e^{at} \begin{pmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{pmatrix}$$

in der Tat die reelle Fundamentalmatrix von  $Z' = \hat{\mathbf{A}}Z$  ist. Um zum ursprünglichen Problem zurückzukommen, beachten wir die Identität  $\mathbf{B}^{-1}\mathbf{A}\mathbf{B} = \hat{\mathbf{A}}$ . Setzen wir also

$$\mathbf{Y}(t) := \mathbf{B}\mathbf{Z}(t), \quad t \in I,$$

so ist  $\mathbf{Y}$  eine *reelle Fundamentalmatrix* des Systems  $Y' = \mathbf{A} Y$ . Da die Matrix  $\mathbf{B} = (U, W)$  aus dem Realteil sowie Imaginärteil des Eigenvektors  $V$  aufgebaut ist, haben wir also eine reelle Darstellung einer Fundamentalmatrix von  $Y' = \mathbf{A} Y$  gefunden.

**Beispiel.** Sei die Matrix  $\mathbf{A}$  durch

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$$

gegeben. Für das charakteristische Polynom berechnen wir

$$\begin{aligned} \det(\mathbf{A} - \lambda \mathbf{E}) &= \det \begin{pmatrix} 1 - \lambda & 2 \\ -1 & 3 - \lambda \end{pmatrix} \\ &= (1 - \lambda)(3 - \lambda) + 2 = \lambda^2 - 4\lambda + 5. \end{aligned}$$

Die Eigenwerte sind also durch  $\lambda_{1,2} = (2 \pm \sqrt{4-5}) = 2 \pm i$  gegeben. Wir setzen  $\lambda := 2 + i$  und  $\bar{\lambda} = 2 - i$ . Wir suchen nun Eigenvektoren zu diesen Eigenwerten. Dazu betrachten wir  $(\mathbf{A} - \lambda \mathbf{E})V = 0$ :

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 - 2 - i & 2 \\ -1 & 3 - 2 - i \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -1 - i & 2 \\ -1 & 1 - i \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

Diese Gleichung wird offenbar von  $V = \begin{pmatrix} 2 \\ 1 + i \end{pmatrix}$  erfüllt. Es folgt

$$U = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad W = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{und somit} \quad \mathbf{B} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}.$$

Die Fundamentalmatrix von  $Z' = \hat{\mathbf{A}} Z$

$$\mathbf{Z}(t) = e^{2t} \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}$$

transformiert sich also zur Fundamentalmatrix

$$\mathbf{Y}(t) = \mathbf{B} \mathbf{Z}(t) = e^{2t} \begin{pmatrix} 2 \cos(t) & 2 \sin(t) \\ \cos(t) - \sin(t) & \cos(t) + \sin(t) \end{pmatrix}$$

von  $Y' = \mathbf{A} Y$ . Für den Anfangswert zum Zeitpunkt  $t_0 = 0$  benötigen wir eine Fundamentalmatrix  $\tilde{\mathbf{Y}}(t)$  mit  $\tilde{\mathbf{Y}}(0) = \mathbf{E}$ . Dazu bestimmen wir  $\mathbf{Y}(0)^{-1}$ :

$$\mathbf{Y}(0) = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{und somit} \quad \mathbf{Y}(0)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix}.$$

Damit erhalten wir als gewünschte Fundamentalmatrix

$$\tilde{\mathbf{Y}}(t) = \mathbf{Y}(t) \mathbf{Y}(0)^{-1} = e^{2t} \begin{pmatrix} \cos(t) - \sin(t) & 2 \sin(t) \\ -\sin(t) & \cos(t) + \sin(t) \end{pmatrix}.$$

### 13.4.4 Reelle Systeme für $n = 2$

Sei  $\mathbf{A} \in \mathbb{R}^{n \times n}$  mit  $\det \mathbf{A} \neq 0$ . Für die Jordan'sche Normalform gibt es folgende Möglichkeiten:

1.  $\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$  mit  $\lambda, \mu \in \mathbb{R}$ ,  $\lambda, \mu \neq 0$  ( $\lambda = \mu$  möglich)
2.  $\begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$  mit  $\lambda \in \mathbb{R} \setminus \{0\}$
3.  $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$  mit  $\lambda \in \mathbb{C}$ ,  $\lambda = a + ib$

Wir hatten bereits gesehen, dass für alle  $t \in \mathbb{R}$  und alle Anfangswerte  $Y_0 \in \mathbb{R}^2$  maximale Lösungen existieren. Es stellt sich nun die Frage nach dem asymptotischen Verhalten der Lösungen für  $t \rightarrow \infty$ . Wir werden dabei die Fälle 1. - 3. einzeln betrachten.

**Zu Fall 1:** Die Lösung der auf eine Basis von Eigenvektoren transformierten Gleichung (4.1) ist durch

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 e^{\lambda t} \\ c_2 e^{\mu t} \end{pmatrix} \quad (4.6)$$

gegeben. Um uns einen besseren Überblick über das qualitative Lösungsverhalten zu verschaffen, betrachten wir das *Phasenportrait* der Lösung. Dazu interpretieren wir die Lösungsformel (4.6) als Parameterdarstellung zu verschiedenen  $c_1, c_2$  einer Kurve im  $\mathbb{R}^2$ , den sogenannten *Phasenkurven* oder *Trajektorien* der Gleichung. Fassen wir nun mehrere Trajektorien in einem Diagramm zusammen und orientieren diese mit der üblichen Orientierung von  $\mathbb{R}$ , so sprechen wir von einem *Phasenportrait*. Dazu unterscheiden wir die folgenden vier Fälle.

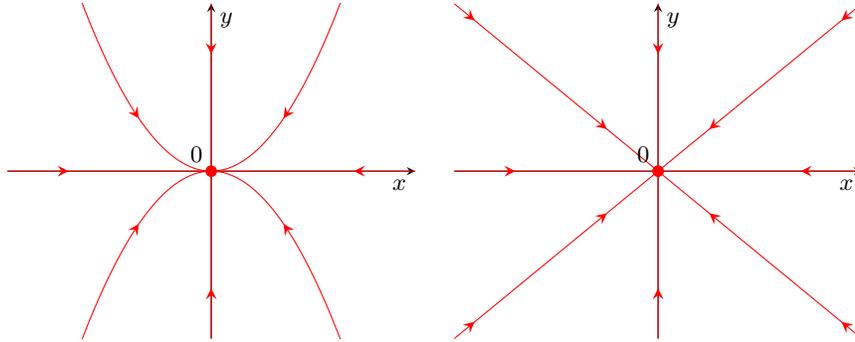
(a) Gelte  $\lambda, \mu < 0$ . Direkt aus der Lösungsformel sehen wir

$$\lim_{t \rightarrow \infty} x(t) = 0, \quad \lim_{t \rightarrow \infty} y(t) = 0. \quad (4.7)$$

Für den Fall  $c_1, c_2 \neq 0$  erhalten wir aus der Lösungsformel die Trajektorien

$$y(t) = c_2 e^{\mu t} = c_2 (e^{\lambda t})^{\frac{\mu}{\lambda}} = c_2 \left( \frac{x(t)}{c_1} \right)^{\frac{\mu}{\lambda}}.$$

Man beachte, dass  $\frac{\mu}{\lambda} > 0$ . Für  $c_1 = 0$  ist die Trajektorie  $x(t) = 0$ , für  $c_2 = 0$  erhalten wir  $y(t) = 0$  und für  $c_1 = c_2 = 0$  ist die Trajektorie durch  $x(t) = y(t) = 0$  gegeben. Damit erhalten wir das Phasenportrait in Abbildung 4.1. Die Richtung der Pfeile ist durch (4.7) gegeben. Insbesondere laufen alle Trajektorien in den Ursprung  $(0, 0)$ . Deshalb nennen wir  $(0, 0)$  einen *stabilen Knoten*. Das Phasenportrait (Abbildung 4.1) ist bezüglich der Basis aus Eigenwerten gezeichnet.



**Abb. 4.1.** Phasenportrait im Fall 1 und  $\lambda \leq \mu < 0$ . Links  $\frac{\mu}{\lambda} = 2$ , Rechts  $\frac{\mu}{\lambda} = 1$ .

Betrachten wir die Standardbasis (und somit das ursprüngliche Problem (4.1)), so ist das Phasenportrait durch eine affine Deformation des obigen gegeben (siehe Abbildung 4.2).

- (b) Gelte  $\lambda, \mu > 0$ . Dann gilt für  $c_1, c_2 \neq 0$

$$\lim_{t \rightarrow \infty} |x(t)| = \infty, \quad \lim_{t \rightarrow \infty} |y(t)| = \infty. \tag{4.8}$$

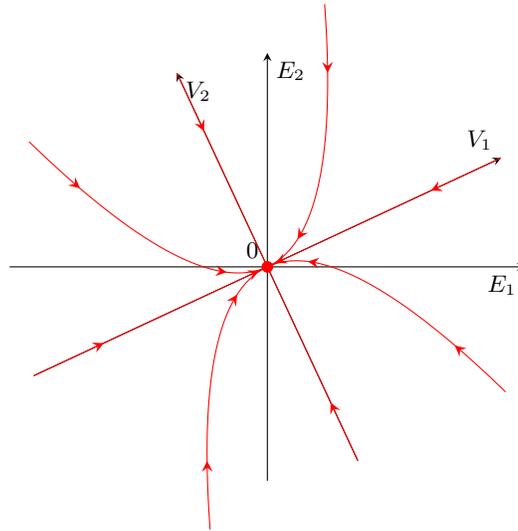
Ansonsten sind die Rechnungen aus dem vorherigen Fall auch hier gültig. Insbesondere ist  $\frac{\mu}{\lambda} > 0$ . Durch (4.8) drehen sich allerdings die Pfeile um. Das Phasenportrait ist in Abbildung 4.3 veranschaulicht. Da jetzt alle Trajektorien aus  $(0, 0)$  heraus laufen, nennen wir  $(0, 0)$  einen *instabilen Knoten*.

- (c) Gelte  $\lambda < 0 < \mu$ . Dann gilt für  $c_1, c_2 \neq 0$

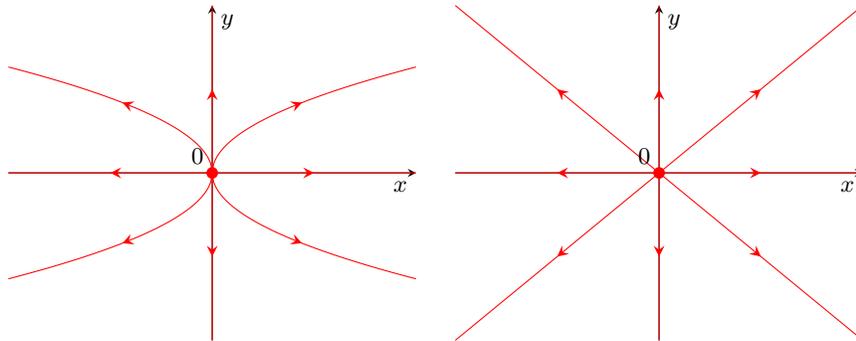
$$\lim_{t \rightarrow \infty} x(t) = 0, \quad \lim_{t \rightarrow \infty} |y(t)| = \infty.$$

Man beachte, dass in der Darstellung

$$y(t) = c_2 \left( \frac{x(t)}{c_1} \right)^{\frac{\mu}{\lambda}} \tag{4.9}$$



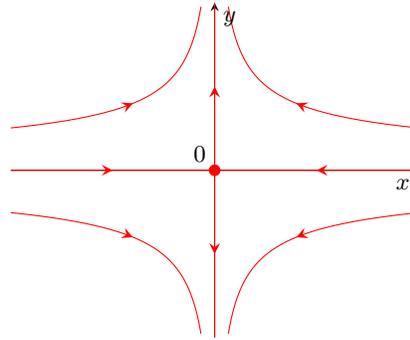
**Abb. 4.2.** Phasenportrait im Fall 1 und  $\lambda \leq \mu < 0$  bezüglich der Standardbasis ( $\frac{\mu}{\lambda} = 2$ ,  $V_1, V_2$  Eigenvektoren).



**Abb. 4.3.** Phasenportrait im Fall 1 und  $\lambda \geq \mu > 0$  mit  $\frac{\mu}{\lambda} = 0.5$  (links) und  $\frac{\mu}{\lambda} = 1$  (rechts).

der Exponent  $\frac{\mu}{\lambda}$  jetzt negativ ist. Dadurch ergibt sich ein völlig anderes Phasenportrait. Die Richtung der Pfeile ist durch (4.9) gegeben. Das Phasenportrait ist dann in Abbildung 4.4 veranschaulicht. Der Ursprung wird als *instabiler Sattelpunkt* bezeichnet, da die Trajektorie auf der "x"-Achse in ihn hineinläuft und die Trajektorie auf der "y"-Achse aus ihm heraus läuft.

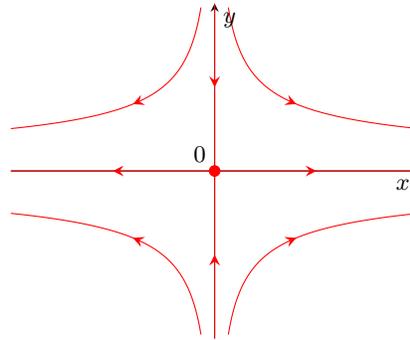
(d) Gelte  $\mu < 0 < \lambda$ . Dann gilt für  $c_1, c_2 \neq 0$



**Abb. 4.4.** Phasenportrait im Fall **1** und  $\lambda < 0 < \mu$  mit  $\frac{\mu}{\lambda} = -0.5$ .

$$\lim_{t \rightarrow \infty} |x(t)| = \infty, \quad \lim_{t \rightarrow \infty} y(t) = 0.$$

In diesem Fall vertauschen sich gerade die Rollen von  $x$  und  $y$ . Das Phasenportrait ist dann in Abbildung 4.5 dargestellt. Auch hier wird der Ursprung als ein *instabiler Sattelpunkt* bezeichnet.



**Abb. 4.5.** Phasenportrait im Fall **1** und  $\mu < 0 < \lambda$  mit  $\frac{\mu}{\lambda} = -0.5$ .

**Zu Fall 2:** Wie wir im Abschnitt 13.4.2 gesehen haben, ist die Lösung der auf eine Basis von Eigenvektoren transformierten Gleichung (4.1) durch

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 \\ c_1 t + c_2 \end{pmatrix} e^{\lambda t} \quad t \in \mathbb{R} \quad (4.10)$$

gegeben. Auch hier betrachten wir unterschiedliche Fälle.

(a) Gelte  $\lambda < 0$ . Dann folgt aus (4.10)

$$\lim_{t \rightarrow \infty} x(t) = 0, \quad \lim_{t \rightarrow \infty} y(t) = 0. \quad (4.11)$$

Weiter erhalten wir für  $c_1 \neq 0$

$$\frac{y(t)}{x(t)} = \frac{c_1 t e^{\lambda t} + c_2 e^{\lambda t}}{c_1 e^{\lambda t}} = t + \frac{c_2}{c_1}$$

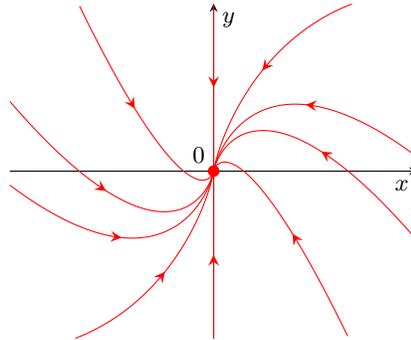
und somit

$$y(t) = t x(t) + \frac{c_2}{c_1} x(t).$$

Weiter folgt aus  $\frac{x(t)}{c_1} = e^{\lambda t} > 0$  die Gleichung  $\ln\left(\frac{x(t)}{c_1}\right) = \lambda t$ . Somit können wir die explizite  $t$ -Abhängigkeit eliminieren und erhalten für die Trajektorien

$$y(t) = \frac{1}{\lambda} x(t) \ln\left(\frac{x(t)}{c_1}\right) + \frac{c_2}{c_1} x(t).$$

Das Phasenportrait ist in Abbildung 4.6 veranschaulicht. Die Richtung der Pfeile ist durch (4.11) gegeben. Insbesondere ist der Ursprung ein stabiler Knoten.



**Abb. 4.6.** Phasenportrait im Fall 2 und  $\lambda = -1$ .

(b) Gelte  $\lambda > 0$ . Dann folgt für  $c_1 \neq 0$  aus (4.10)

$$\lim_{t \rightarrow \infty} |x(t)| = \infty, \quad \lim_{t \rightarrow \infty} |y(t)| = \infty. \quad (4.12)$$

Das Phasenportrait ist in Abbildung 4.7 dargestellt. Die Richtung der Pfeile dreht sich aufgrund von (4.12) im Vergleich zum vorherigen Fall um. Insbesondere ist der Ursprung jetzt ein instabiler Knoten.

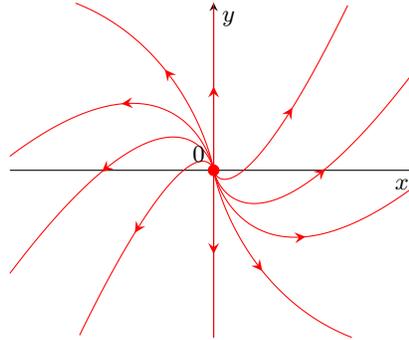


Abb. 4.7. Phasenportrait im Fall 2 und  $\lambda = 1$ .

**Zu Fall 3:** Wie wir in Abschnitt 13.4.3 gesehen haben, ist die reelle Fundamentalmatrix für das transformierte System durch

$$\mathbf{Z}(t) = e^{at} \begin{pmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{pmatrix}$$

gegeben. Zur Veranschaulichung der Lösungen ist es sinnvoll  $\mathbb{R}^2$  und  $\mathbb{C}$  zu identifizieren, d.h.  $(x, y) \in \mathbb{R}^2$  wird mit  $x + iy \in \mathbb{C}$  identifiziert. Unter Berücksichtigung von

$$\cos(bt) - i \sin(bt) = e^{-ibt}$$

erhalten wir die komplexen Basislösungen

$$\begin{aligned} z_1(t) &= e^{at} e^{-ibt}, \\ z_2(t) &= i z_1(t). \end{aligned}$$

Es reicht also die Basislösung  $z_1$  zu veranschaulichen, da  $z_2$  durch eine Drehung aus  $z_1$  hervorgeht. Es können nun die folgenden Fälle auftreten.

(a) Gilt  $a = 0$ , so folgt

$$|z_1(t)| = \text{const.}$$

Das Phasenportrait ist dann durch Abbildung 4.8 gegeben und wir nennen den Ursprung ein *stabiles Zentrum*.

(b) Gilt  $a > 0$ , so folgt für Anfangswerte ungleich Null

$$\lim_{t \rightarrow \infty} |z_1(t)| = \infty.$$

Das Phasenportrait ist dann in Abbildung 4.9 veranschaulicht. Der Ursprung wird *instabiler Strudel* genannt.

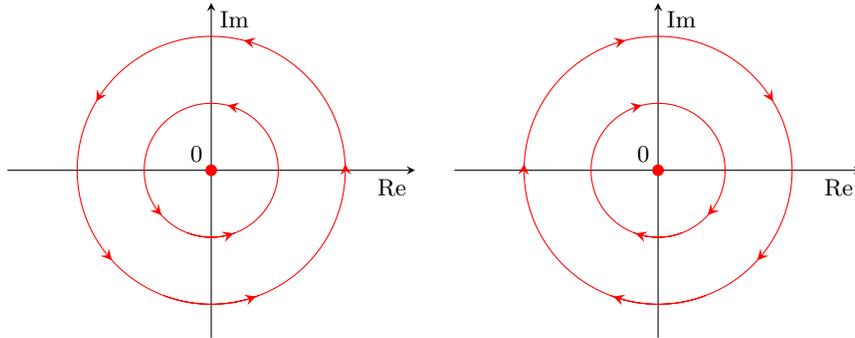


Abb. 4.8. Phasenportrait im Fall 3 für  $a = 0$  und links  $b < 0$ , rechts  $b > 0$ .

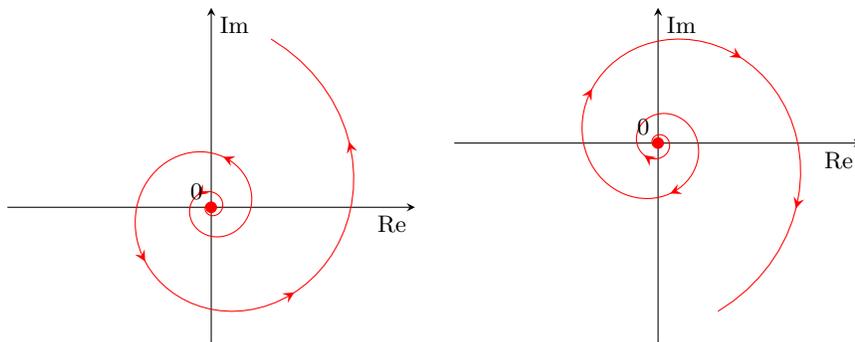


Abb. 4.9. Phasenportrait im Fall 3 für  $a > 0$  und links  $b < 0$ , rechts  $b > 0$ .

(c) Gilt  $a < 0$ , so folgt

$$\lim_{t \rightarrow \infty} z_1(t) = 0.$$

Das Phasenportrait ist dann in Abbildung 4.10 dargestellt. Der Ursprung wird *stabiler Strudel* genannt.

### 13.5 Exponentialfunktion für Matrizen

Für eine lineare homogene Differentialgleichung mit konstanten Koeffizienten  $y' = a y$  haben wir eine explizite Lösungsformel mithilfe der Exponentialfunktion hergeleitet, nämlich  $y(t) = c e^{at}$ . Um diese Formel auf lineare Systeme mit konstanten Koeffizienten  $Y' = \mathbf{A} Y$  zu verallgemeinern benötigen wir die Exponentialfunktion für Matrizen.

**5.1 Definition.** Sei  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . Wir definieren die Exponentialfunktion für  $\mathbf{A}$  durch

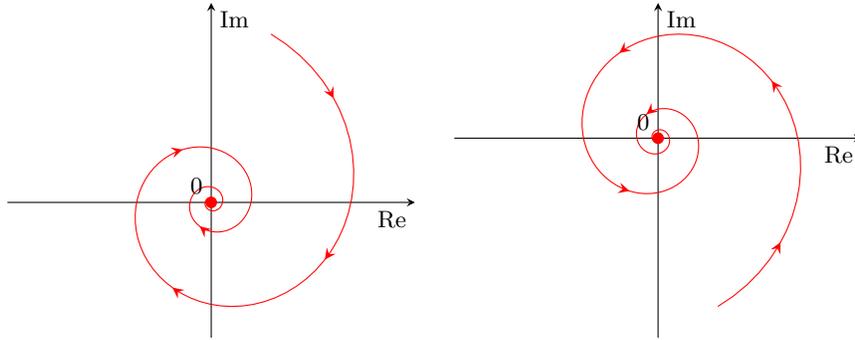


Abb. 4.10. Phasenportrait im Fall 3 für  $a < 0$  und links  $b < 0$ , rechts  $b > 0$ .

$$e^{\mathbf{A}} := \sum_{n=0}^{\infty} \frac{\mathbf{A}^n}{n!}. \tag{5.2}$$

**Bemerkungen.** 1. Für  $\mathbf{A} \in \mathbb{R}^{n \times n}$  gilt

$$\|\mathbf{A}^n\| = \|\mathbf{A}\mathbf{A} \dots \mathbf{A}\| \leq \|\mathbf{A}\| \cdot \dots \cdot \|\mathbf{A}\| = \|\mathbf{A}\|^n.$$

Somit folgt

$$\left\| \sum_{n=0}^N \frac{\mathbf{A}^n}{n!} \right\| \leq \sum_{n=0}^N \frac{\|\mathbf{A}\|^n}{n!}.$$

Also konvergiert die Reihe in (5.2) absolut, da die reelle Exponentialreihe  $\sum_{n=0}^{\infty} \frac{\|\mathbf{A}\|^n}{n!}$  eine Majorante ist.

2. Analog kann man Funktionen von Matrizen für beliebige Potenzreihen definieren, z.B. für  $\sin(\mathbf{A})$ ,  $\tan(\mathbf{A})$  oder  $\ln(\mathbf{A})$ . Diese konvergieren absolut innerhalb des Konvergenzradius der Reihe.

**5.3 Lemma.** *Es gilt:*

- (i)  $e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}} \cdot e^{\mathbf{B}} = e^{\mathbf{B}} \cdot e^{\mathbf{A}}$ , falls  $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$ .
- (ii)  $e^{\mathbf{B}^{-1}\mathbf{A}\mathbf{B}} = \mathbf{B}^{-1} e^{\mathbf{A}} \mathbf{B}$ , falls  $\det \mathbf{B} \neq 0$ .
- (iii)  $e^{\text{diag}(\lambda_1, \dots, \lambda_n)} = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})$ .

*Beweis.* (i): Aufgrund unserer Voraussetzung folgt

$$(\mathbf{A} + \mathbf{B})(\mathbf{A} + \mathbf{B}) = \mathbf{A}^2 + \mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{A} + \mathbf{B}^2 = \mathbf{A}^2 + 2\mathbf{A}\mathbf{B} + \mathbf{B}^2.$$

Da die Reihen absolut konvergieren ist weiter insbesondere das Cauchyprodukt dieser Reihen wohldefiniert. Somit folgt analog zum Beweis für die reelle Exponentialreihe

$$\begin{aligned} e^{\mathbf{A}+\mathbf{B}} &= \sum_{n=0}^{\infty} \frac{(\mathbf{A}+\mathbf{B})^n}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{\mathbf{B}^k \mathbf{A}^{n-k}}{k!(n-k)!} \\ &= \left( \sum_{n=0}^{\infty} \frac{\mathbf{B}^n}{n!} \right) \left( \sum_{\ell=0}^{\infty} \frac{\mathbf{A}^{\ell}}{\ell!} \right) = e^{\mathbf{B}} \cdot e^{\mathbf{A}}. \end{aligned}$$

Vertauschen wir die Rolle von  $\mathbf{A}$  und  $\mathbf{B}$ , so folgt  $e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{B}} \cdot e^{\mathbf{A}}$ .

(ii): Es gilt

$$(\mathbf{B}^{-1} \mathbf{A} \mathbf{B})^2 = (\mathbf{B}^{-1} \mathbf{A} \mathbf{B})(\mathbf{B}^{-1} \mathbf{A} \mathbf{B}) = \mathbf{B}^{-1} \mathbf{A}^2 \mathbf{B}$$

uns somit induktiv

$$(\mathbf{B}^{-1} \mathbf{A} \mathbf{B})^k = \mathbf{B}^{-1} \mathbf{A}^k \mathbf{B}.$$

Also folgt

$$\begin{aligned} e^{\mathbf{B}^{-1} \mathbf{A} \mathbf{B}} &= \sum_{k=0}^{\infty} \frac{(\mathbf{B}^{-1} \mathbf{A} \mathbf{B})^k}{k!} \\ &= \sum_{k=0}^{\infty} \mathbf{B}^{-1} \frac{\mathbf{A}^k}{k!} \mathbf{B} \\ &= \mathbf{B}^{-1} \left( \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \right) \mathbf{B} = \mathbf{B}^{-1} e^{\mathbf{A}} \mathbf{B}. \end{aligned}$$

(iii): Es gilt offenbar  $\text{diag}(\lambda_1, \dots, \lambda_n)^2 = \text{diag}(\lambda_1^2, \dots, \lambda_n^2)$ . Somit folgt analog zum Beweis von (ii) die Behauptung. ■

**5.4 Satz.** Für  $\mathbf{A} \in \mathbb{R}^{n \times n}$  setzen wir

$$\mathbf{S}(t) := e^{\mathbf{A}t}, \quad t \in \mathbb{R}. \quad (5.5)$$

Dann gilt:

1.  $\mathbf{S}(0) = \mathbf{E}$ ,
2.  $\mathbf{S}(s+t) = \mathbf{S}(s)\mathbf{S}(t) = \mathbf{S}(t)\mathbf{S}(s)$ ,
3.  $\frac{d}{dt}\mathbf{S}(t) = \mathbf{A}\mathbf{S}(t) = \mathbf{S}(t)\mathbf{A}$ .

Mit anderen Worten:  $\mathbf{S}$  ist eine Fundamentalmatrix von  $Y' = \mathbf{A}Y$ .

*Beweis.* Nach Definition gilt

$$\mathbf{S}(t) = \sum_{n=0}^{\infty} \frac{t^n \mathbf{A}^n}{n!}.$$

Somit folgt

$$\mathbf{S}(0) = \sum_{n=0}^{\infty} \frac{0^n \mathbf{A}^n}{n!} = \frac{0^0 \mathbf{A}^0}{0!} + \mathbf{0} + \dots = \frac{1 \mathbf{E}}{1} = \mathbf{E}.$$

Es gilt also 1. Weiter folgt mit der Binomialformel und dem Cauchyprodukt

$$\begin{aligned} \mathbf{S}(s+t) &= \sum_{n=0}^{\infty} \frac{(s+t)^n \mathbf{A}^n}{n!} \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{t^k s^{n-k} \mathbf{A}^{n-k} \mathbf{A}^k}{k!(n-k)!} \\ &= \sum_{k=0}^{\infty} \frac{t^k \mathbf{A}^k}{k!} \sum_{\ell=0}^{\infty} \frac{s^\ell \mathbf{A}^\ell}{\ell!} \\ &= \mathbf{S}(t) \cdot \mathbf{S}(s). \end{aligned}$$

Vertauschen wir die Rolle von  $s$  und  $t$ , so haben wir 2. gezeigt. Verwenden wir diese Aussage, so erhalten wir

$$\begin{aligned} \left. \frac{d}{dt} \mathbf{S}(t) \right|_{t=t_0} &= \lim_{h \rightarrow 0} \frac{\mathbf{S}(t_0+h) - \mathbf{S}(t_0)}{h} \\ &= \mathbf{S}(t_0) \lim_{h \rightarrow 0} \frac{\mathbf{S}(h) - \mathbf{E}}{h} \\ &= \mathbf{S}(t_0) \lim_{h \rightarrow 0} \frac{1}{h} \left( \sum_{k=0}^{\infty} \frac{h^k \mathbf{A}^k}{k!} - \mathbf{E} \right) \\ &= \mathbf{S}(t_0) \lim_{h \rightarrow 0} \left( \mathbf{A} + \sum_{k=2}^{\infty} \frac{h^{k-1} \mathbf{A}^k}{k!} \right) \\ &= \mathbf{S}(t_0) \mathbf{A}, \end{aligned}$$

wobei wir die absolute Konvergenz der Reihe verwendet haben. Im letzten Schritt haben wir auch benutzt, dass

$$\begin{aligned}
\left\| \sum_{k=2}^{\infty} \frac{h^{k-1} \mathbf{A}^k}{k!} \right\| &= \left\| \lim_{N \rightarrow \infty} \sum_{k=2}^N \frac{h^{k-1} \mathbf{A}^k}{k!} \right\| \\
&= \lim_{N \rightarrow \infty} \left\| \sum_{k=2}^N \frac{h^{k-1} \mathbf{A}^k}{k!} \right\| \\
&\leq \lim_{N \rightarrow \infty} \sum_{k=2}^N |h|^{k-1} \frac{\|\mathbf{A}\|^k}{k!} \\
&\leq \lim_{N \rightarrow \infty} |h| \|\mathbf{A}\|^2 \sum_{k=0}^{N-2} \frac{h^k \|\mathbf{A}\|^k}{(k+2)!} \\
&\leq \lim_{N \rightarrow \infty} |h| \|\mathbf{A}\|^2 \sum_{k=0}^{N-2} \frac{h^k \|\mathbf{A}\|^k}{k!} \\
&\leq |h| \|\mathbf{A}\|^2 e^{h \|\mathbf{A}\|} \xrightarrow{h \rightarrow 0} 0.
\end{aligned}$$

Wir haben also 3. gezeigt und somit den Satz bewiesen.  $\blacksquare$

Wir haben somit die Analogie zum Fall  $n = 1$  gerechtfertigt. In diesem Fall wussten wir schon, dass die Gleichung

$$y'(t) = a y(t)$$

die Lösung

$$y(t) = c e^{a t}$$

besitzt. Jetzt haben wir gezeigt, dass

$$\mathbf{Y}(t) := e^{\mathbf{A} t},$$

eine Fundamentalmatrix des Systems  $Y' = \mathbf{A} Y$  ist und  $\mathbf{Y}(0) = \mathbf{E}$  erfüllt. Somit hat das Anfangswertproblem

$$Y'(t) = \mathbf{A} Y(t), \quad Y(0) = Y_0$$

die Lösung

$$Y(t) = e^{\mathbf{A} t} Y_0.$$

Wir haben auch eine neue Berechtigung für (4.4) gefunden. Wir können einen Jordan Block der Ordnung  $k$  wie folgt zerlegen

$$\mathbf{J} = \begin{pmatrix} \lambda & 0 & \cdots & \cdots & 0 \\ 1 & \lambda & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \lambda \end{pmatrix} = \lambda \mathbf{E} + \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} =: \lambda \mathbf{E} + \mathbf{F}.$$

Somit folgt

$$\mathbf{F}^2 = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ 1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

und insbesondere  $\mathbf{F}^k = \mathbf{0}$ . Für die Exponentialfunktion erhalten wir also

$$\begin{aligned} e^{\mathbf{F}t} &= \sum_{n=0}^{\infty} \frac{\mathbf{F}^n t^n}{n!} = \mathbf{E} + \frac{\mathbf{F}^1 t^1}{1} + \cdots + \frac{\mathbf{F}^{k-1} t^{k-1}}{(k-1)!} + \mathbf{0} + \cdots \\ &= \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ t & 1 & \ddots & \ddots & \vdots \\ \frac{1}{2} t^2 & t & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \frac{t^{k-1}}{(k-1)!} & \cdots & \frac{1}{2} t^2 & t & 1 \end{pmatrix}. \end{aligned}$$

Da  $\mathbf{E}\mathbf{F} = \mathbf{F}\mathbf{E}$ , gilt mithilfe von Lemma 5.3

$$e^{\mathbf{J}t} = e^{\lambda \mathbf{E}t} e^{\mathbf{F}t} = e^{\lambda t} \mathbf{E} e^{\mathbf{F}t} = \begin{pmatrix} e^{\lambda t} & 0 & \cdots & 0 \\ t e^{\lambda t} & e^{\lambda t} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \frac{t^{k-1}}{(k-1)!} e^{\lambda t} & \frac{t^{k-2}}{(k-2)!} e^{\lambda t} & \cdots & e^{\lambda t} \end{pmatrix},$$

also die Formel (4.4). Gilt  $\mathbf{J} = \mathbf{B}^{-1} \mathbf{A} \mathbf{B}$ , so folgt  $\mathbf{A} = \mathbf{B} \mathbf{J} \mathbf{B}^{-1}$  und wir erhalten mit Lemma 5.3

$$e^{\mathbf{A}t} = \mathbf{B} e^{\mathbf{J}t} \mathbf{B}^{-1}.$$



# A Anhang

## A.1 Die Jordan'sche Normalform

In Kapitel 13 haben wir mehrfach die Jordan'sche Normalform verwendet und berechnet. Diese wurde bereits in der Linearen Algebra behandelt, der Vollständigkeit halber werden wir sie an dieser Stelle ebenfalls einführen. Die grundlegende Zielsetzung ist dabei, einen nilpotenten Endomorphismus, d.h. eine lineare Abbildung  $f : K^n \rightarrow K^n$  eines Vektorraumes in sich selbst mit

$$f^k \equiv 0$$

für ein  $k \in \mathbb{N}$  bzw. deren Darstellungsmatrix in eine möglichst einfache Gestalt zu transformieren.

**1.1 Definition.** Sei  $K$  eine Körper und  $n \in \mathbb{N}$ . Die Matrix

$$J_n := \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 1 & 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \in K^{n \times n}$$

heißt Jordanmatrix der Größe  $n$  zum Eigenwert 0.

**1.2 Folgerung.** (a) Für das charakteristische Polynom gilt  $\chi_{J_n} = T^n$ . Insbesondere ist 0 der einzige Eigenwert von  $J_n$ .

(b) Sei  $\mathbf{A} = J_n$ . Der zugehörige Endomorphismus  $f_{\mathbf{A}}$  ist durch

$$f_{\mathbf{A}} : K^n \rightarrow K^n, \quad e_i \mapsto \begin{cases} e_{i+1} & i < n, \\ 0 & i = n \end{cases}$$

gegeben. Es folgt für  $\ell \in \mathbb{N}$



- i) **Idee:** Finde einen Basisvektor  $w_1 \in \text{Ker}(g^d) \setminus \text{Ker}(g^{d-1})$ . Setzen wir  $w_i := g^{i-1}(w_1)$ ,  $1 \leq i \leq d$  so erhalten wir  $J_d$ .  
 Finde nun  $w_{d+1} \in \text{Ker}(g^d) \setminus \text{Ker}(g^{d-1})$ , welches linear unabhängig von  $(w_1, \dots, w_d)$  ist und setze  $w_{d+i} := g^{i-1}(w_{d+1})$ ,  $1 \leq i \leq d$ . Damit erhalten wir die nächste Jordanmatrix  $J_d$ . Führe dies solange fort, bis alle Jordanmatrizen gefunden sind.
- ii) Betrachte  $d := \min\{i \in \mathbb{N} \mid g^i \equiv 0\}$ . Aufgrund der Nilpotenz von  $g$  ist  $d$  wohldefiniert und endlich. Weiter gilt

$$\{0\} \subset \text{Ker}(g) \subset \text{Ker}(g^2) \subset \dots \subset \text{Ker}(g^{d-1}) \subset \text{Ker}(g^d) = V.$$

iii) **Algorithmus:**

- a) Wähle eine Hilfsbasis von  $\text{Ker}(g^{d-1})$  und ergänze diese zu einer Basis von  $V$ , also wähle eine Basis  $(\bar{v}_d^{(1)}, \dots, \bar{v}_d^{(s_d)})$  von  $\text{Ker}(g^d)/\text{Ker}(g^{d-1})$  und wähle  $v_d^{(1)}, \dots, v_d^{(s_d)} \in \text{Ker}(g^d)$  mit  $v_d^{(j)} + \text{Ker}(g^{d-1}) = \bar{v}_d^{(j)}$ . Definiere  $W_d := K v_d^{(1)} \oplus \dots \oplus K v_d^{(s_d)}$ .
- b) Setze  $v_i^{(j)} := g^{d-i}(v_d^{(j)})$  für alle  $1 \leq i < d$  und  $1 \leq j \leq s_d$ . Definiere weiter  $\mathcal{B}_d := (v_d^{(1)}, \dots, v_1^{(1)}, \dots, v_d^{(s_d)}, \dots, v_1^{(s_d)})$ .  
 Damit gilt  $v_i^j \in \text{Ker}(g^i) \setminus \text{Ker}(g^{i-1})$  für alle  $1 \leq i \leq d$ ,  $1 \leq j \leq s_d$ , denn

$$g^{i-1}(v_i^{(j)}) = g^{i-1}(g^{d-i}(v_d^{(j)})) = g^{d-1}(v_d^{(j)}) \neq 0$$

$$g^i(v_i^{(j)}) = g^d(v_d^{(j)}) = 0.$$

Es gilt also

$$\{0\} \subset \text{Ker}(g) \subset \dots \subset \text{Ker}(g^\ell) \subset \dots \subset \text{Ker}(g^{d-1}) \subset \text{Ker}(g^d) = V.$$

$\Psi$	$\Psi$	$\Psi$	$\Psi$
$v_1^{(1)}$	$v_\ell^{(1)}$	$v_{d-1}^{(1)}$	$v_d^{(1)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v_1^{(s_d)}$	$v_\ell^{(s_d)}$	$v_{d-1}^{(s_d)}$	$v_d^{(s_d)}$

- c) Führe für  $\ell = d - 1, \dots, 1$  die folgenden Schritte aus:

- 3i) Wähle eine Hilfsbasis von  $\text{Ker}(g^{\ell-1})$  und ergänze diese durch die schon bekannten  $v_\ell^{(1)}, \dots, v_\ell^{(s_{\ell+1})}$  sowie neuen  $v_\ell^{(1+s_{\ell+1})}, \dots, v_\ell^{s_{\ell+1}}$

zu einer Basis von  $\text{Ker}(g^\ell)$ . Wähle also eine Basis  $(\bar{v}_\ell^{(1+s_{\ell+1})}, \dots, \bar{v}_\ell^{(s_\ell)})$  von  $\text{Ker}(g^\ell)/(\text{Ker}(g^{\ell-1}) + K v_\ell^{(1)} + \dots + K v_\ell^{(s_{\ell+1})})$  und wähle  $v_\ell^{(j)} \in \text{Ker}(g^\ell)$  mit

$$v_\ell^{(j)} + \text{Ker}(g^{\ell-1}) + K v_\ell^{(1)} + \dots + K v_\ell^{(s_{\ell+1})} = \bar{v}_\ell^{(j)}$$

für  $j = 1 + s_{\ell+1}, \dots, s_\ell$  (hier gilt  $s_\ell > s_{\ell+1}$ ). Beachte dabei, dass die benötigte lineare Unabhängigkeit von  $v_\ell^{(1)}, \dots, v_\ell^{(s_{\ell+1})}$  später gezeigt wird. Definiere nun  $W_\ell := K v_\ell^{(1+s_{\ell+1})} \oplus \dots \oplus K v_\ell^{(s_\ell)}$ .

3ii) Setze  $v_i^{(j)} := g^{\ell-i}(v_\ell^{(j)})$  für alle  $1 \leq i \leq \ell$ ,  $1 + s_{\ell+1} \leq j \leq s_\ell$  und  $\mathcal{B}_\ell := \mathcal{B}_{\ell+1} \cup \{v_\ell^{(1+s_{\ell+1})}, \dots, v_\ell^{(1+s_{\ell+1})}, \dots, v_\ell^{(s_\ell)}, \dots, v_\ell^{(s_\ell)}\}$ .

Wir zeigen, dass  $\mathcal{B} := \mathcal{B}_1$  die gesuchte Basis ist.

iv) Wir behaupten, dass für alle  $1 \leq \ell \leq d$

$$\text{Ker}(g^\ell) = \text{Ker}(g^{\ell-1}) \oplus g^{d-\ell}(W_d) \oplus \dots \oplus g(W_{\ell+1}) \oplus W_\ell$$

gilt. Dies zeigen wir mit Hilfe der absteigenden Induktion nach  $\ell$ .

$\ell = d$ ,  $\text{Ker}(g^\ell) = V = \text{Ker}(g^{\ell-1}) \oplus W_d$  nach **iii)a**).

**Induktionshypothese:**  $\text{Ker}(g^{\ell+1}) = \text{Ker}(g^\ell) \oplus g^{d-\ell-1}(W_d) \oplus \dots \oplus W_{\ell+1}$ .

**Induktionsschritt:**

- Aus  $W_i \subset \text{Ker}(g^i)$  folgt  $g^{i-\ell}(W_i) \subset \text{Ker}(g^\ell)$ .
- Die Abbildung  $g : (g^{d-\ell-1}(W_d) \oplus \dots \oplus W_{\ell+1}) \rightarrow V$ ,  $v \mapsto g(v)$  ist injektiv, denn sei  $v \in \text{Ker}(g) \cap (g^{d-\ell-1}(W_d) \oplus \dots \oplus W_{\ell+1})$ , so folgt mit  $\text{Ker}(g) \subset \text{Ker}(g^\ell)$  und der Induktionshypothese  $v \in \text{Ker}(g^\ell) \cap (g^{d-\ell-1}(W_d) \oplus \dots \oplus W_{\ell+1}) = \{0\}$ . Somit zerfällt das Bild  $g^{d-\ell}(W_d) \oplus \dots \oplus g(W_{\ell+1})$  in die entsprechende direkte Summe.
- Es gilt  $\text{Ker}(g^{\ell-1}) \cap (g^{d-\ell}(W_d) \oplus \dots \oplus g(W_{\ell+1})) = \{0\}$ , denn sei  $w \in (g^{d-\ell}(W_d) \oplus \dots \oplus g(W_{\ell+1}))$  mit  $v = g(w) \in \text{Ker}(g^{\ell-1})$ , so folgt  $w \in \text{Ker}(g^\ell)$  und mit der Induktionshypothese  $w = 0$  sowie  $v = 0$ . Somit ist die Summe  $\text{Ker}(g^{\ell-1}) \oplus g^{d-\ell}(W_d) \oplus \dots \oplus g(W_\ell)$  direkt.
- Nach Konstruktion ist  $\text{Ker}(g^\ell) = (\text{Ker}(g^{\ell-1}) \oplus g^{d-\ell}(W_d) \oplus \dots \oplus g(W_{\ell+1})) \oplus W_\ell$ .

v)  $W_i$  hat die Basis  $v_i^{k(i+1)+\ell}, \dots, v_i^{k(i)}$ , denn für alle  $\ell \geq 1$  ist

$$g^{i-\ell} : W_i \xrightarrow{g} g(W_i) \xrightarrow{g} \dots \xrightarrow{g} g^{i-\ell}(W_i)$$

ein Isomorphismus. Somit hat  $g^{i-\ell}(W_i)$  die Basis  $v_i^{k(i+1)+\ell}, \dots, v_i^{k(i)}$ .

vi) Der Vektorraum

$$\begin{aligned} V &= \text{Ker}(g^d) \\ &= \text{Ker}(g^{d-1}) \oplus W_d \\ &= \text{Ker}(g^{d-2}) \oplus g(W_d) \oplus W_d \oplus W_{d-1} \\ &= \dots \\ &= (W_d \oplus g(W_d) \oplus \dots \oplus g^{d-1}(W_d)) \oplus \dots \oplus (W_2 \oplus g(W_2)) \oplus W_1 \end{aligned}$$

hat Basis  $\mathcal{B}$ .

Nun zeigen wir die Eindeutigkeit der  $s_d, \dots, s_1$ : Es gilt

$$\begin{aligned} s_d &= r_d = \dim(W_d) = \dim(V) - \dim(\text{Ker}(g^{d-1})), \\ s_{d-1} &= r_{d-1} + r_d = \dim(W_{d-1} \oplus g(W_d)) = \dim(\text{Ker}(g^{d-1})) - \dim(\text{Ker}(g^{d-2})), \\ &\vdots \\ s_1 &= r_1 + \dots + r_d = \dim(W_1 \oplus \dots \oplus g^{d-1}(W_d)) = \dim(\text{Ker}(g^1)). \end{aligned}$$

Somit sind rekursiv  $r_d, \dots, r_1$  eindeutig durch  $g$  bestimmt:

$$r_\ell = s_\ell - s_{\ell+1} = 2 \dim(\text{Ker}(g^\ell)) - \dim(\text{Ker}(g^{\ell-1})) - \dim(\text{Ker}(g^{\ell+1}))$$

für alle  $1 \leq \ell \leq d$ . ■

Wir haben also folgenden Algorithmus zur Bestimmung der Jordan-Normalform: Sei  $\mathbf{A} \in K^{n \times n}$  nilpotent.

0. Berechne  $\mathbf{A}^2, \mathbf{A}^3, \dots, \mathbf{A}^d$  mit  $d := \min\{i \in \mathbb{N} \mid \mathbf{A}^i \equiv 0\}$ . Bestimme die Dimension und Basis von  $\text{Ker}(\mathbf{A}^\ell)$ ,  $1 \leq \ell \leq d-1$ . Setze

$$s_\ell := \dim(\text{Ker}(\mathbf{A}^\ell)) - \dim(\text{Ker}(\mathbf{A}^{\ell-1})) \quad 1 \leq \ell \leq d$$

(beachte dabei  $\mathbf{A}^0 := Id_n$ ). Setze weiter  $r_\ell := s_\ell - s_{\ell+1}$  für  $1 \leq \ell \leq d$ .

1. Ergänze die Basis von  $\text{Ker}(\mathbf{A}^{d-1})$  durch  $v_d^{(1)}, \dots, v_d^{(s_d)}$  zu einer Basis von  $\text{Ker}(\mathbf{A}^d) = K^n$ .
2. Setze  $v_i^{(j)} := \mathbf{A}^{d-i} v_d^{(j)}$  für alle  $1 \leq i < d$  und alle  $1 \leq j \leq s_d$ . Setze  $\mathcal{B}_d := \{v_d^{(1)}, \dots, v_1^{(1)}, \dots, v_d^{(s_1)}, \dots, v_1^{(s_1)}\}$ .
3. Für  $\ell = d-1, \dots, 1$  tue folgendes:
  - Falls  $r_\ell = 0$  gilt, mache nichts und gehe zum nächsten  $\ell$

– Falls  $r_\ell > 0$  gilt:

3 a) Ergänze die Basis von  $\text{Ker}(g^{\ell-1})$  durch die schon bekannten  $v_\ell^{(1)}, \dots, v_\ell^{(s_{\ell+1})}$  sowie die neuen  $v_\ell^{(1+s_{\ell+1})}, \dots, v_\ell^{(s_\ell)}$  zu einer Basis von  $\text{Ker}(g^\ell)$ .

3 b) Setze  $\mathcal{B}_\ell := \mathcal{B}_{\ell+1} \cup \{v_\ell^{(1+s_{\ell+1})}, \dots, v_\ell^{(s_\ell)}, \dots, v_1^{(s_\ell)}\}$

– Erhalte  $\mathcal{B} = \mathcal{B}_1$  Basis von  $K^n$  mit  $M_{\mathcal{B}}^{\mathcal{B}}(f_{\mathbf{A}}) = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$  wie in Satz 1.3.