Grundlagen der Mathematik

Andreas Gathmann

Vorlesungsskript RPTU Kaiserslautern 2022/23

Inhaltsverzeichnis (Grundlagen der Mathematik 1)

0. Einleitung und Motivation 4

1. Etwas Logik und Mengenlehre 7
1.A Logik 7 1.B Mengenlehre 12

2. Relationen und Funktionen 15
2.A Funktionen 15 2.B Äquivalenzrelationen 21

3. Erste Eigenschaften der reellen Zahlen 24
3.A Gruppen und Körper 24 3.B Vollständige Induktion 30 3.C Polynomfunktionen 31

Grundlagen der Mathematik 1: Analysis

4. Weitere Eigenschaften der reellen Zahlen 35
4.A Potenzen in Körpern 35 4.B Geordnete Körper 38 4.C Supremum und Infimum 41

5. Folgen und Grenzwerte 47
5.A Grenzwerte von Folgen 47 5.B Konvergenzkriterien für Folgen 54 5.C Limes superior und inferior 59 5.D Mächtigkeiten von Mengen 61

6. Komplexe Zahlen 65
6.A Die Konstruktion der komplexen Zahlen 65 6.B Eigenschaften der komplexen Zahlen 68 6.C Reelle und komplexe Folgen 71

7. Reihen 75
7.A Grenzwerte von Reihen 75 7.B Konvergenzkriterien für Reihen 77 7.C Potenzreihen 84

8. Stetigkeit 90
8.A Grenzwerte von Funktionen 90 8.B Eigenschaften stetiger Funktionen 96 8.C Gleichmäßige Konvergenz und Stetigkeit 99

9. Spezielle Funktionen 105
9.A Logarithmen und allgemeine Potenzen 105 9.B Winkelfunktionen 108

10. Differentialrechnung 116
10.A Ableitungen von Funktionen 116 10.B Extremwerte und der Mittelwertsatz 122

11. Anwendungen der Differentialrechnung 128
11.A Die Regel von de l'Hôpital 128 11.B Taylor-Entwicklung 130

12. Integralrechnung 137
12.A Das Riemann-Integral 137 12.B Stammfunktionen 145 12.C Integrationsregeln 149

Grundlagen der Mathematik 1: Lineare Algebra

13. Vektorräume 155
13.A Der Vektorraumbegriff 155 13.B Lineare Abbildungen 161

14. Basen und Dimension 167
14.A Linearkombinationen und Basen 167 14.B Existenz von Basen 171

15. Quotientenräume und Dimensionsformeln 177
15.A Komplemente und Quotientenräume 177 15.B Isomorphiesätze und Dimensionsformeln 183

16. Lineare Abbildungen als Matrizen 187
16.A Matrizen 187 16.B Lineare Abbildungen zwischen endlich erzeugten Vektorräumen 194 16.C Äquivalente Matrizen und Normalformen 199

17. Das Gauß-Verfahren 203
17.A Lineare Gleichungssysteme 203 17.B Weitere Algorithmen mit dem Gauß-Verfahren 211

18. Determinanten 216
18.A Konstruktion der Determinante 216 18.B Eigenschaften der Determinante 223

Inhaltsverzeichnis (Grundlagen der Mathematik 2)

Grundlagen der Mathematik 2: Lineare Algebra

19. Endomorphismen 227
19.A Ähnliche Matrizen 227 19.B Eigenwerte 230 19.C Diagonalisierbarkeit 235

20. Die Jordansche Normalform 243
20.A Haupträume 243 20.B Jordandiagramme 247 20.C Minimalpolynome 255

21. Euklidische und unitäre Räume 260
21.A Bilinearformen 261 21.B Skalarprodukte 264 21.C Orthogonalität 271

22. Endomorphismen euklidischer und unitärer Räume 279
22.A Orthogonale und unitäre Abbildungen 279 22.B Selbstadjungierte Abbildungen 283 22.C Der Spektralsatz 286 22.D Die Singulärwertzerlegung 294

Grundlagen der Mathematik 2: Analysis

23. Topologische Grundbegriffe 299
23.A Normierte und metrische Räume 299 23.B Konvergenz in metrischen Räumen 304 23.C Offene und abgeschlossene Mengen 309 23.D Kompaktheit 315

24. Stetigkeit in metrischen Räumen 320
24.A Stetige Abbildungen 320 24.B Eigenschaften stetiger Abbildungen 324 24.C Stetige Bilder kompakter Mengen 329 24.D Peano-Kurven 334

25. Differenzierbarkeit im Mehrdimensionalen 338
25.A Differenzierbare Abbildungen 338 25.B Eigenschaften differenzierbarer Abbildungen 346

26. Höhere Ableitungen 352
26.A Die mehrdimensionale Taylor-Entwicklung 353 26.B Extremwerte 360

27. Implizite Funktionen 363
27.A Umkehrfunktionen 364 27.B Der Satz über implizite Funktionen 367 27.C Extrema unter Nebenbedingungen 372

28. Integralrechnung im Mehrdimensionalen 376
28.A Das mehrdimensionale Riemann-Integral 377 28.B Der Satz von Fubini 382

29. Messbare Mengen 387
29.A Integrale über beschränkte Mengen 387 29.B Nullmengen 392 29.C Normalbereiche 398 29.D Uneigentliche Integrale 403

30. Der Transformationssatz für mehrdimensionale Integrale 407
30.A Transformation von Volumina 408 30.B Transformation von Integralen 414

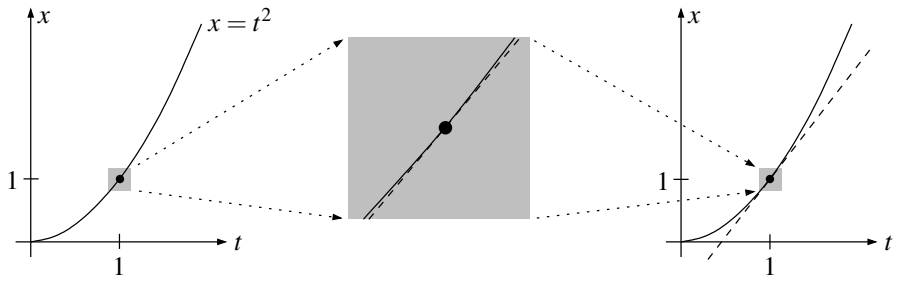
Literatur 420
Index 421

0. Einleitung und Motivation

In diesem Skript – so verspricht es der Titel – wollen wir uns die Grundlagen der Mathematik erarbeiten. Aber was ist das überhaupt, die „Grundlagen der Mathematik“? Bei uns an der TU Kaiserslautern verstehen wir unter dieser Vorlesung die Kombination zweier Themengebiete, die in der Tat das grundlegende Handwerkszeug für nahezu die gesamte Mathematik darstellen, nämlich

- der *Analysis*, d. h. der Untersuchung von Folgen und Grenzwerten, Stetigkeit, sowie der Differential- und Integralrechnung (zunächst in einer und im zweiten Semester dann auch in mehreren Variablen), und
- der *linearen Algebra*, d. h. der Theorie der Vektorräume, linearen Abbildungen und Gleichungssysteme.

Von beiden Gebieten habt ihr ja aus der Schule wahrscheinlich schon eine ungefähre Vorstellung. In der *Analysis* geht es grob gesagt darum, reelle Funktionen *lokal*, also in der Umgebung eines gewählten Punktes, zu untersuchen, und aus diesen Untersuchungen dann wieder Aussagen über die gesamte Funktion zurückzugewinnen. Betrachten wir z. B. ein Auto, das sich entlang einer geraden Strecke bewegt. Der Einfachheit halber nehmen wir an, dass die Position x des Autos nach der Zeit t (in geeigneten Einheiten) durch die Gleichung $x = t^2$ beschrieben werden kann, so dass die Bewegung durch die Kurve im folgenden Bild links dargestellt wird:



Wir wollen diese Bewegung nun nur in einer kleinen Umgebung eines fest gewählten Zeitpunkts, z. B. des Zeitpunkts $t = 1$ (und damit auch $x = 1$) betrachten. Im Bild oben haben wir diese Umgebung grau markiert und in der Mitte stark vergrößert dargestellt. Wir sehen, dass die Kurve in dieser Umgebung fast wie eine *Gerade* aussieht; wir haben diese Gerade gestrichelt eingezeichnet und im Bild rechts auch außerhalb der gewählten Umgebung fortgesetzt. Geometrisch ist diese Gerade natürlich einfach die *Tangente* an die Kurve an der Stelle $t = 1$. Physikalisch repräsentiert die Steigung dieser Geraden die *Geschwindigkeit* des Autos zum betrachteten Zeitpunkt, denn sie gibt ja gerade an, in welchem Verhältnis sich dort die Strecke x mit der Zeit t verändert. Aus der Schule wisst ihr auch schon, wie man diese Steigung ausrechnet: Man muss dazu die gegebene Funktion t^2 *differenzieren* – so dass man die Ableitung $2t$ erhält – und dort die betrachtete Stelle $t = 1$ einsetzen. Die Steigung ist in unserem Fall also gerade $2 \cdot 1 = 2$, und man rechnet sofort nach, dass $x = 2t - 1$ die Gleichung der oben eingezeichneten Tangente ist. Man sagt, dass die Gerade $x = 2t - 1$ eine *lineare Approximation* der ursprünglich gegebenen Funktion $x = t^2$ im Punkt $t = 1$ ist.

Wir können aus der Kenntnis der zurückgelegten Strecke zu jedem Zeitpunkt also die Geschwindigkeit des Autos durch Differenzieren bestimmen. Man kann sich natürlich auch die umgekehrte Frage stellen: Angenommen, der Kilometerzähler eures Autos ist kaputt, aber ihr beobachtet auf eurer Fahrt ständig eure Geschwindigkeit. Könnt ihr dann am Ende der Fahrt trotzdem ausrechnen, wie weit ihr gefahren seid? Dies ist offensichtlich die „Umkehrung“ des Differenzierens – und auch hier wisst ihr aus der Schule natürlich schon, dass dies auf die *Integralrechnung* führen wird.

In der Praxis bewegt man sich mit dem Auto aber nicht immer nur auf einer geraden Strecke, und demzufolge braucht man für die Beschreibung solch einer Bewegung (und natürlich auch vieler anderer natürlich auftretender Prozesse) mehrere Variablen. Wir werden daher auch Abbildungen betrachten, die mehrere Variablen auf mehrere andere abbilden, wie z. B. die folgende Vorschrift, die zwei reelle Zahlen y_1 und y_2 in Abhängigkeit von zwei anderen x_1 und x_2 ausdrückt:

$$\begin{aligned} y_1 &= 2x_1 + x_2 \cos x_1 \\ y_2 &= x_1 \mathrm{e}^{x_2} - x_2 \end{aligned}$$

Genau wie oben werden wir uns auch hier wieder die Frage stellen, ob wir diese (in diesem Fall recht komplizierte) Funktion in der Nähe eines gegebenen Punktes nicht vielleicht durch eine einfache *lineare* Abhängigkeit annähern können. In der Tat ist dies möglich: Wir werden sehen, dass z. B. in einer kleinen Umgebung des Nullpunkts $(x_1, x_2) = (0,0)$ die obige Vorschrift näherungsweise die gleichen Ergebnisse liefert wie das lineare Gleichungssystem

$$\begin{aligned} y_1 &= 2x_1 + x_2 \\ y_2 &= x_1 - x_2. \end{aligned}$$

Auch wenn diese Gleichungen natürlich viel einfacher als die ursprünglichen sind, sollte offensichtlich sein, dass auch solche linearen Gleichungssysteme bei wachsender Zahl von Variablen (und in der Praxis sind Hunderte oder Tausende von Variablen keine Seltenheit) recht kompliziert werden können. Wir werden daher einen wesentlichen Teil dieser Vorlesung mit der *linearen Algebra*, also dem Studium derartiger linearer Gleichungssysteme, verbringen. Um dabei überhaupt erst einmal den Notationsaufwand in Grenzen zu halten, tut man dabei gut daran, die Start- und Zielvariablen nicht alle einzeln hinzuschreiben, sondern sie zu sogenannten *Vektoren* zusammenzufassen. In der Tat kann man in dieser Sichtweise die lineare Algebra als das Studium von linearen Abbildungen zwischen Vektorräumen beschreiben.

Wenn wir dann die linearen Abbildungen zwischen mehreren Variablen gut genug verstanden haben, können wir uns im zweiten Teil der Analysis schließlich daran machen, die Differential- und Integralrechnung auf den Fall von mehreren Variablen auszuweiten.

Wir haben damit jetzt relativ kurz umrissen, welcher mathematische Stoff uns in dieser Vorlesung erwartet. Es wird in diesem Skript aber nicht nur darum gehen, mathematische Resultate kennenzulernen. Mindestens ebenso wichtig ist es, das „mathematische Denken“ zu lernen, d. h. die Fähigkeit zu entwickeln, mit abstrakten Konzepten umzugehen, exakte logische Schlüsse zu ziehen und Beweise zu führen. Anders als in der Schule oder im Studium z. B. ingenieurswissenschaftlicher Fächer werden wir genau darauf achten, eine „wasserdichte“ Theorie aufzubauen: Jeder neue Begriff bzw. jeder neue Satz wird nur unter Verwendung des bisher Bekannten exakt definiert bzw. bewiesen. So sind z. B. Formulierungen in dem Stil „eine Funktion wird durch eine Gerade angenähert“ oder „eine Funktion heißt stetig, wenn man sie zeichnen kann, ohne den Stift abzusetzen“ als Veranschaulichung unserer Ideen zwar sehr sinnvoll, als exakte mathematische Formulierung jedoch schlichtweg unbrauchbar. Wir werden daher in dieser Vorlesung viel exakter arbeiten als ihr es wahrscheinlich aus der Schule gewohnt seid, und es ist wichtig, dass ihr diese exakte Denkweise verinnerlicht und anzuwenden lernt. Die Mathematik ist ein riesiges Gebäude – viel größer und komplexer als ihr es euch wahrscheinlich im Moment vorstellen könnt – das ständig immer höher gebaut wird, indem schon bewiesene Sätze auf neue Fälle angewendet oder wieder für neue Beweise verwendet werden. Wir starten gerade beim Fundament dieses Gebäudes und können es uns da wirklich nicht leisten herumzupfuschen.

Diese konsequent logische und exakte Herangehensweise ist zwar am Anfang wahrscheinlich ungewohnt, hat jedoch für euch auch einen Vorteil: Etwas überspitzt formuliert erzählen wir euch während des gesamten Studiums eigentlich nur Dinge, die logisch aus dem folgen, was ihr ohnehin schon wusstet. Dadurch ist die Mathematik wahrscheinlich das Studienfach, in dem man am wenigsten auswendig lernen muss – in dem es aber im Gegenzug auch am meisten auf das *Verständnis* des Stoffes ankommt. Je besser euer Verständnis für die Mathematik wird, um so mehr Dinge werden euch letztlich einfach „klar“ werden, so dass es euch dann auch viel leichter fällt, sie zu lernen.

Dieses Verständnis für die Mathematik bekommt man aber natürlich nur durch intensiven und *aktiven* Umgang mit dem Stoff, weswegen neben dem Studium der Vorlesung auch die Bearbeitung der Übungsaufgaben besonders wichtig ist.

Heißt das alles nun, dass wir nur mit logischen Argumenten die Mathematik sozusagen „aus dem Nichts“ aufbauen können? Nein, das geht natürlich nicht ... von nichts kommt nichts. Man muss am Anfang immer gewisse Dinge als gegeben annehmen, also Aussagen als wahr voraussetzen, die man nicht mehr beweist bzw. beweisen kann, und auf denen dann die gesamte Theorie beruht. Derartige Annahmen bezeichnet man als **Axiome**. Natürlich versucht man in der Mathematik, mit möglichst wenigen und sehr elementaren Axiomen auszukommen, die hoffentlich niemand anzweifeln würde. In der modernen Mathematik ist es üblich, hierfür die grundlegenden Prinzipien der Logik und Mengenlehre zu verwenden (zu denen wir auch gleich in Kapitel 1 Genaueres sagen werden).

In dieser Vorlesung wollen wir uns das Leben allerdings etwas leichter machen und zusätzlich auch die *Existenz und elementaren Eigenschaften der reellen Zahlen* axiomatisch voraussetzen (um welche Eigenschaften es sich hierbei handelt, werden wir natürlich genau angeben). Man kann zwar nur aus den Axiomen der Logik und Mengenlehre beweisen, dass die reellen Zahlen existieren und dass sie die erwarteten Eigenschaften haben, der Beweis wäre zu diesem frühen Zeitpunkt im Studium aber sehr verwirrend und würde euch auch keine großartigen neuen Erkenntnisse bringen. In diesem Sinne starten wir also sozusagen doch nicht ganz beim Fundament unseres „Gebäudes Mathematik“, sondern bereits im ersten Stock.

Im weiteren Verlauf ist dieses Skript dann wie im Inhaltsverzeichnis angegeben in mehrere Teile gegliedert. Diese bauen der Reihe nach aufeinander auf, mit einer Ausnahme: Nach den in jedem Fall benötigten grundlegenden Anfangskapiteln 1 bis 3 sind die Teile „Grundlagen der Mathematik 1: Analysis“ (Kapitel 4 bis 12) und „Grundlagen der Mathematik 1: Lineare Algebra“ (Kapitel 13 bis 18) *unabhängig voneinander* und können somit in beliebiger Reihenfolge oder auch (wie in den Studienplänen in der Regel vorgesehen) parallel studiert werden.

Aber jetzt genug der Vorrede ... beginnen wir nun also unser Studium der Mathematik mit den „Grundlagen der Grundlagen“, den für uns wesentlichen Prinzipien der Logik und Mengenlehre.

1. Etwas Logik und Mengenlehre

Bevor wir mit dem eigentlichen Inhalt der Vorlesung beginnen, müssen wir in diesem Kapitel kurz die exakte mathematische Sprache beschreiben, in der wir unsere Ergebnisse formulieren werden: die der Logik und Mengenlehre. Zentral hierbei sind die Begriffe der *Aussage* (in der Logik) und der *Menge* (in der Mengenlehre).

Da wir es hier mit den ersten beiden Begriffen überhaupt zu tun haben, die in der Mathematik vorkommen, können wir sie natürlich nicht durch bereits bekannte Dinge definieren oder mit bereits bekannten Resultaten ihre Eigenschaften herleiten. Wir müssen sie daher (wie schon in der Einleitung erwähnt) axiomatisch voraussetzen. Wir müssen *voraussetzen*, dass es sinnvoll ist, über logische Aussagen und deren Wahrheit zu reden, dass Mengen überhaupt existieren, dass man Mengen vereinigen und schneiden kann, aus ihnen Elemente auswählen kann, und noch einiges mehr. Wenn ihr euch zum Beispiel auf den Standpunkt stellt, dass ihr nicht an die Existenz von Mengen glaubt, wird euch niemand widerlegen können. Allerdings zweifelt ihr damit dann auch die Existenz der gesamten Mathematik an, wie sie heutzutage betrieben wird – und aus der Tatsache, dass ihr in dieser Vorlesung sitzt, schließe ich einmal, dass das nicht der Fall ist.

Glücklicherweise sind die Dinge, die wir benötigen, jedoch allesamt anschaulich sofort einleuchtend und euch natürlich aus der Schule auch schon hinlänglich bekannt. Ich möchte es euch (und mir) daher ersparen, an dieser Stelle eine vollständige und präzise axiomatische Formulierung der Logik und Mengenlehre hinzuschreiben, zumal das momentan sicher mehr verwirren als helfen würde und außerdem gerade im Bereich der Logik auch zu sehr in die Philosophie abdriften würde. Stattdessen wollen wir uns hier damit begnügen, die für uns wichtigsten Prinzipien und Notationen sowie beliebte Fehlerquellen in verständlicher Sprache zu erklären, auch wenn ein paar Dinge (insbesondere die Begriffsfestlegung – „Definition“ möchte ich es eigentlich gar nicht nennen – einer Aussage und einer Menge) dadurch recht schwammig klingen werden. Außerdem werden wir in Beispielen zur besseren Verdeutlichung bereits hier die reellen Zahlen und ihre einfachsten Eigenschaften (die euch sicherlich bekannt sein werden) benutzen, auch wenn wir diese erst später formalisieren werden. Da es sicher niemanden von euch verwirren wird, werden wir auch die Schreibweise „$x \in \mathbb{R}$“ für „x ist eine reelle Zahl“ schon verwenden, bevor sie in den Notationen 1.13 und 1.15 offiziell eingeführt wird.

1.A Logik

Beginnen wir also mit der Logik. Unter einer **Aussage** verstehen wir (grob gesagt) ein sprachliches Gebilde, das entweder wahr oder falsch ist (wobei wir in der Mathematik natürlich letztlich daran interessiert sind, *wahre* Aussagen herzuleiten). Wichtig sind auch solche sprachlichen Gebilde, in denen freie **Variablen**, also Platzhalter, vorkommen, und die erst beim Einsetzen von Werten für diese Variablen Aussagen liefern. Man bezeichnet sie als **Aussageformen**.

Beispiel 1.1.

(a) $1+1=2$ ist eine wahre, $1+1=3$ eine falsche, und $1+1$ überhaupt keine Aussage.

(b) $x+1=2$ ist eine Aussageform, die beim Einsetzen von $x=1$ in eine wahre, beim Einsetzen jeder anderen reellen Zahl in eine falsche Aussage übergeht.

Bemerkung 1.2. Als Variablen in Aussageformen kann man beliebige Symbole benutzen. Üblich sind neben den normalen lateinischen Klein- und Großbuchstaben auch die griechischen Buchstaben, die wir zur Erinnerung hier auflisten:

A α alpha	B β beta	Γ γ gamma	Δ δ delta	E ε epsilon	Z ζ zeta	H η eta	Θ ϑ theta
I ι iota	K κ kappa	Λ λ lambda	M μ my	N ν ny	Ξ ξ xi	O o omikron	Π π pi
P ρ rho	Σ σ sigma	T τ tau	Y υ ypsilon	Φ φ phi	X χ chi	Ψ ψ psi	Ω ω omega

Oft verziert man Buchstaben auch noch mit einem Symbol oder versieht sie mit einem Index, um neue Variablen zu erhalten: So sind z. B. $x, x', \tilde{x}, \bar{x}, x_1, x_2, \ldots$ alles Symbole für verschiedene Variablen, die zunächst einmal nichts miteinander zu tun haben (aber tunlichst für irgendwie miteinander zusammenhängende Objekte eingesetzt werden sollten, wenn man den Leser nicht vollends verwirren will).

Notation 1.3 (Zusammengesetzte Aussagen)**.** Sind A und B Aussagen, so lassen sich daraus wie folgt neue bilden:

Symbol	Wahrheitstafel				Bedeutung
A	w	f	w	f	
B	w	w	f	f	
$\neg A$	f	w			nicht A
$A \wedge B$	w	f	f	f	A und B
$A \vee B$	w	w	w	f	A oder B (oder beides): „nicht-ausschließendes Oder“
$A \Leftrightarrow B$	w	f	f	w	A und B sind gleichbedeutend / *äquivalent*, bzw. A genau dann, wenn B
$A \Rightarrow B$	w	w	f	w	aus A folgt B, bzw. wenn A dann B

Die sogenannte **Wahrheitstafel** in den mittleren vier Spalten ist dabei die eigentliche Definition der neuen zusammengesetzten Aussagen. Sie gibt in Abhängigkeit der Wahrheit von A und B (in den ersten beiden Zeilen) an, ob die zusammengesetzte Aussage wahr oder falsch ist.

Bemerkenswert ist hierbei wohl nur die Folgerungsaussage $A \Rightarrow B$, die keine Aussage über die Richtigkeit von A oder B separat macht, sondern nur sagt, dass B wahr ist, wenn auch A es ist. Ist hingegen A falsch, so ist die Folgerungsaussage $A \Rightarrow B$ stets wahr („aus einer falschen Voraussetzung kann man alles folgern“). So ist z. B. $0 = 1 \;\Rightarrow\; 2 = 3$ eine wahre Aussage. In der Regel wollen wir uns in der Mathematik aber natürlich mit wahren Aussagen beschäftigen, und neue wahre Aussagen aus alten herleiten. Gerade in Beweisen ist die übliche Verwendung der Notation $A \Rightarrow B$ daher, dass A eine bereits als wahr erkannte Aussage ist, und wir damit nun schließen wollen, dass auch B wahr ist.

Bemerkung 1.4 (Beweise mit Wahrheitstafeln)**.** Wollen wir kompliziertere zusammengesetzte Aussagen miteinander vergleichen, so können wir dies wieder mit Hilfe von Wahrheitstafeln tun. So ist für zwei Aussagen A und B z. B.

$$A \Rightarrow B \quad \text{äquivalent zu} \quad (\neg A) \vee B,$$

denn wenn wir in der Wahrheitstafel

A	w	f	w	f
B	w	w	f	f
$\neg A$	f	w	f	w
$(\neg A) \vee B$	w	w	f	w

mit Hilfe der Definitionen von $\neg$ und $\vee$ aus Notation 1.3 zunächst $\neg A$ und dann $(\neg A) \vee B$ berechnen, sehen wir, dass das Ergebnis mit $A \Rightarrow B$ übereinstimmt. Nach der Bemerkung aus Notation 1.3 ist dies auch anschaulich klar: Die Folgerungsaussage $A \Rightarrow B$ ist ja genau dann wahr, wenn A falsch (also $\neg A$ wahr) ist, oder wenn B wahr ist (oder beides).

Genauso zeigt man die ebenfalls einleuchtende Aussage, dass

$$A \Leftrightarrow B \quad \text{äquivalent zu} \quad (A \Rightarrow B) \;\wedge\; (B \Rightarrow A)$$

ist – was auch die übliche Art ist, wie man eine Äquivalenz zeigt: Man zeigt separat die beiden Folgerungen $A \Rightarrow B$ und $B \Rightarrow A$.

Notation 1.5. Folgerungen („$\Rightarrow$“) und Äquivalenzen („$\Leftrightarrow$“) sind natürlich zwei verschiedene Dinge, die man nicht durcheinanderwerfen darf (auch wenn das in der Schule wahrscheinlich manchmal nicht so genau genommen wird). Es hat sich jedoch in der Mathematik eingebürgert, bei *Definitionen* von Begriffen durch eine äquivalente, definierende Eigenschaft die Sprechweise „wenn“ anstatt des eigentlich korrekten „genau dann, wenn“ zu verwenden: So würde man z. B. als Definition des Begriffs einer geraden Zahl hinschreiben

„Eine ganze Zahl x heißt gerade, wenn $\frac{x}{2}$ eine ganze Zahl ist“,

obwohl man genau genommen natürlich meint

„Eine ganze Zahl x heißt *genau dann* gerade, wenn $\frac{x}{2}$ eine ganze Zahl ist“.

Eine gewöhnliche Folgerungsaussage wie z. B. die wahre Aussage

„Wenn eine ganze Zahl x positiv ist, dann ist auch $x+1$ positiv“

ist dagegen immer nur in einer Richtung zu verstehen; hier wird also nicht behauptet, dass mit $x+1$ auch x immer positiv sein muss (was ja auch falsch wäre).

Notation 1.6 (Quantoren)**.** Ist A eine Aussageform, in der eine freie Variable x vorkommt – wir schreiben dies dann auch als $A(x)$ – so setzen wir

Symbol	Bedeutung
$\forall x : A(x)$	für alle x gilt $A(x)$
$\exists x : A(x)$	es gibt ein x mit $A(x)$

Die beiden Symbole $\forall$ und $\exists$ bezeichnet man als **Quantoren**. Beachte, dass diese beiden Quantoren *nicht* miteinander vertauschbar sind: So besagt z. B. die Aussage

$$\forall x \in \mathbb{R}\ \exists y \in \mathbb{R} : y > x$$

„zu jeder reellen Zahl x gibt es eine Zahl y, die größer ist“ (was offensichtlich wahr ist), während die Umkehrung der beiden Quantoren die Aussage

$$\exists y \in \mathbb{R}\ \forall x \in \mathbb{R} : y > x$$

„es gibt eine reelle Zahl y, die größer als jede reelle Zahl x ist“ liefern würde (was ebenso offensichtlich falsch ist). Der Unterschied besteht einfach darin, dass im ersten Fall zuerst das x gewählt werden muss und dann ein y dazu existieren muss (das von x abhängen darf), während es im zweiten Fall *dasselbe* y für alle x tun müsste.

Bemerkung 1.7. Jede Aussage lässt sich natürlich auf viele Arten aufschreiben, sowohl als deutscher Satz als auch als mathematische Formel. Die gerade eben betrachtete Aussage könnte man z. B. auf die folgenden (absolut gleichwertigen) Arten aufschreiben:

(a) $\forall x \in \mathbb{R}\ \exists y \in \mathbb{R} : y > x$.

(b) Es sei $x \in \mathbb{R}$. Dann gibt es ein $y \in \mathbb{R}$ mit $y > x$.

(c) Zu jeder reellen Zahl gibt es noch eine größere.

Welche Variante man beim Aufschreiben wählt ist weitestgehend Geschmackssache. Die Formulierung einer Aussage als deutscher Satz hat den Vorteil, dass wir sie oft leichter verstehen können, weil wir die deutsche Sprache schon länger kennen als die mathematische. Wenn wir uns jedoch erst einmal an die mathematische Sprache gewöhnt haben, wird auch sie ihre Vorzüge bekommen: Sie ist deutlich kürzer und besser logisch strukturiert. Wir werden im Folgenden beide Schreibweisen mischen und jeweils diejenige wählen, mit der unsere Aussagen (hoffentlich) am einfachsten verständlich werden.

Wenn wir mathematische Symbole verwenden, müssen wir diese aber auch stets in ihrer korrekten Notation und nicht als „Abkürzungen“ für deutsche Wörter verwenden: Man würde die Aussage

„2 und 4 sind gerade Zahlen“ sicher niemals schreiben als „$2 \wedge 4$ sind gerade Zahlen“, und analog genauso wenig „Es gilt $n \geq 0$ für alle $n \in \mathbb{N}$“ als „Es gilt $n \geq 0 \; \forall n \in \mathbb{N}$“.

Bemerkung 1.8 (Negationen)**.** Es ist wichtig zu wissen, wie man von einer Aussage das Gegenteil, also die „Verneinung“ bildet. Da hierbei oft Fehler gemacht werden, wollen wir die allgemeinen Regeln hierfür kurz auflisten:

(a) $\neg(\neg A) \Leftrightarrow A$: Ist es falsch, dass A falsch ist, so bedeutet dies genau, dass A wahr ist.

(b) $\neg(A \wedge B) \Leftrightarrow (\neg A) \vee (\neg B)$: Das Gegenteil von „$A$ und B sind richtig“ ist „A oder B ist falsch“.

(c) $\neg(A \vee B) \Leftrightarrow (\neg A) \wedge (\neg B)$: Das Gegenteil von „$A$ oder B ist richtig“ ist „A und B sind falsch“.

(d) $\neg(\forall x : A(x)) \Leftrightarrow \exists x : \neg A(x)$: Das Gegenteil von „für alle x gilt $A(x)$“ ist „es gibt ein x, für das $A(x)$ falsch ist“.

(e) $\neg(\exists x : A(x)) \Leftrightarrow \forall x : \neg A(x)$: Das Gegenteil von „es gibt ein x, für das $A(x)$ gilt“ ist „für alle x ist $A(x)$ falsch“.

Man kann also sagen, dass eine Verneinung dazu führt, dass „und“ mit „oder“ sowie „für alle“ mit „es gibt“ vertauscht wird. So ist z. B. das Gegenteil der Aussage

„In Frankfurt haben *alle* Haushalte Strom *und* fließendes Wasser“

die Aussage

„In Frankfurt *gibt es* einen Haushalt, der keinen Strom *oder* kein fließendes Wasser hat“.

Beispiel 1.9 (Negation einer Folgerung)**.** Wollen wir eine Folgerung $A \Rightarrow B$ verneinen, so können wir sie zunächst mit Bemerkung 1.4 zu $(\neg A) \vee B$ umformen, und erhalten nach Bemerkung 1.8 als Umkehrung dann $A \wedge \neg B$. Dies ist auch anschaulich einleuchtend: Die Folgerungsaussage „wenn A dann B“ ist genau dann falsch, wenn die Voraussetzung A zwar gilt, die Behauptung B aber nicht. Wir sehen also:

$$\text{Die Verneinung einer Folgerung } A \Rightarrow B \text{ ist } A \wedge \neg B.$$

Bemerkung 1.10 (Widerspruchsbeweis)**.** Eine oft vorkommende Anwendung der Regeln für die Verneinung von Aussagen ist der sogenannte **Widerspruchsbeweis** bzw. Beweis durch **Kontraposition**. Nach Bemerkung 1.4 gesehen ist die Folgerung $A \Rightarrow B$ („aus A folgt B“) gleichbedeutend mit $(\neg A) \vee B$. Damit ist diese Aussage nach Bemerkung 1.8 (a) auch äquivalent zu $(\neg(\neg B)) \vee (\neg A)$, also zu $\neg B \Rightarrow \neg A$. Mit anderen Worten: Haben wir eine Schlussfolgerung $A \Rightarrow B$ zu beweisen, so können wir genauso gut $(\neg B) \Rightarrow (\neg A)$ zeigen, d. h. *wir können annehmen, dass die zu zeigende Aussage B falsch ist und dies dann zu einem Widerspruch führen bzw. zeigen, dass dann auch die Voraussetzung A falsch sein muss.*

Beispiel 1.11. Hier sind zwei Beispiele für die Anwendung der Prinzipien aus Bemerkung 1.8 und 1.10 – und auch unsere ersten Beispiele dafür, wie man Beweise von Aussagen exakt aufschreiben kann.

(a) Einen Beweis durch Widerspruch könnte man z. B. so aufschreiben:

Behauptung: Für alle $x \in \mathbb{R}$ gilt $2x+1 > 0$ oder $2x-1 < 0$.

Beweis: Angenommen, die Behauptung wäre falsch, d. h. (nach Bemerkung 1.8 (c) und (d)) es gäbe ein $x \in \mathbb{R}$ mit

$$2x+1 \leq 0 \quad (1) \qquad \text{und} \qquad 2x-1 \geq 0 \quad (2).$$

Für dieses x würde dann folgen, dass

$$0 \overset{(1)}{\geq} 2x+1 = 2x-1+2 \overset{(2)}{\geq} 0+2 = 2.$$

Dies ist aber ein Widerspruch. Also war unsere Annahme falsch und somit die zu beweisende Aussage richtig. □

Das dabei verwendete Symbol „$\square$“ ist die übliche Art, das Ende eines Beweises zu kennzeichnen. Zur Verdeutlichung haben wir die beiden Ungleichungen mit (1) und (2) markiert, um später angeben zu können, wo sie verwendet werden.

(b) Manchmal weiß man von einer Aussage aufgrund der Aufgabenstellung zunächst einmal noch nicht, ob sie wahr oder falsch ist. In diesem Fall muss man sich dies natürlich zuerst überlegen – und, falls die Aussage falsch ist, ihre Negation beweisen. Als Beispiel dafür betrachten wir die Aufgabe

> Man beweise oder widerlege: Für alle $x \in \mathbb{R}$ gilt $2x+1 < 0$ oder $2x-1 > 0$.

In diesem Fall merkt man schnell, dass die Aussage falsch sein muss, weil die Ungleichungen schon für den Fall $x = 0$ nicht stimmen. Man könnte als Lösung der Aufgabe unter Beachtung der Negationsregeln aus Bemerkung 1.8 also aufschreiben:

> *Behauptung:* Die Aussage ist falsch, d. h. es gibt ein $x \in \mathbb{R}$ mit $2x+1 \geq 0$ und $2x-1 \leq 0$.
>
> *Beweis:* Für $x = 0$ ist $2x+1 = 1 \geq 0$ und $2x-1 = -1 \leq 0$. $\square$

Beachte, dass dies ein vollständiger Beweis ist: *Um eine allgemeine Aussage zu widerlegen, genügt es, ein Gegenbeispiel dafür anzugeben.*

Bemerkung 1.12. Bevor wir unsere kurze Auflistung der für uns wichtigen Prinzipien der Logik beenden, wollen wir noch kurz auf ein paar generelle Dinge eingehen, die man beim Aufschreiben mathematischer Beweise oder Rechnungen beachten muss.

Dass wir bei unseren logischen Argumenten sauber und exakt arbeiten – also z. B. nicht Folgerungen, die keine Äquivalenzen sind, in der falschen Richtung verwenden, „für alle“ mit „es gibt“ verwechseln oder ähnliches – sollte sich von selbst verstehen. Die folgende kleine Geschichte hilft vielleicht zu verstehen, was damit gemeint ist.

> Ein Ingenieur, ein Physiker und ein Mathematiker fahren mit dem Zug nach Frankreich und sehen dort aus dem Fenster des Zuges ein schwarzes Schaf.
>
> Da sagt der Ingenieur: „Oh, in Frankreich sind die Schafe schwarz!“
>
> Darauf der Physiker: „Nein ... wir wissen jetzt nur, dass es in Frankreich mindestens ein schwarzes Schaf gibt.“
>
> Der Mathematiker: „Nein ... wir wissen nur, dass es in Frankreich mindestens ein Schaf gibt, das auf mindestens einer Seite schwarz ist.“

Es gibt aber noch einen weiteren sehr wichtigen Punkt, der selbst von fortgeschritteneren Studenten leider oft nicht beachtet wird: In der Regel werden wir beim Aufschreiben sowohl Aussagen notieren wollen, die wir erst noch zeigen wollen (um schon einmal zu sagen, worauf wir hinaus wollen), als auch solche, von denen wir bereits wissen, dass sie wahr sind (z. B. weil sie für die zu zeigende Behauptung als wahr vorausgesetzt werden oder weil sie sich logisch aus irgendetwas bereits Bekanntem ergeben haben). Es sollte offensichtlich sein, dass wir Aussagen mit derartig verschiedenen Bedeutungen für die Argumentationsstruktur nicht einfach kommentarlos nebeneinander schreiben dürfen, wenn noch jemand in der Lage sein soll, die Argumente nachzuvollziehen. Betrachten wir z. B. noch einmal unseren Beweis aus Beispiel 1.11 (a) oben, so wäre eine Art des Aufschreibens in folgendem Stil (wie man es leider oft sieht)

$$\begin{aligned} &2x+1 > 0 \ \text{ oder } \ 2x-1 < 0 \\ &2x+1 \leq 0 \qquad\quad 2x-1 \geq 0 \\ &0 \geq 2x+1 = 2x-1+2 \geq 0+2 = 2 \end{aligned}$$

völlig inakzeptabel, obwohl hier natürlich letztlich die gleichen Aussagen stehen wie oben. Kurz gesagt:

> Von *jeder* aufgeschriebenen Aussage muss für den Leser *sofort* und *ohne eigenes Nachdenken* ersichtlich sein, welche Rolle sie in der Argumentationsstruktur spielt: Ist es z. B. eine noch zu zeigende Behauptung, eine Annahme oder eine Folgerung (und wenn ja, aus was)?

Dies bedeutet allerdings nicht, dass wir ganze Aufsätze schreiben müssen. Eine (schon recht platzoptimierte) Art, den Beweis aus Beispiel 1.11 (a) aufzuschreiben, wäre z. B.

> Angenommen, es gäbe ein $x \in \mathbb{R}$ mit $2x+1 \leq 0$ und $2x-1 \geq 0$.
>
> Dann wäre $0 \geq 2x+1 = 2x-1+2 \geq 0+2 = 2$, Widerspruch. $\square$

01

1.B Mengenlehre

Nachdem wir die wichtigsten Regeln der Logik behandelt haben, wenden wir uns jetzt der Mengenlehre zu. Die gesamte moderne Mathematik basiert auf diesem Begriff der Menge, der ja auch schon aus der Schule hinlänglich bekannt ist. Zur Beschreibung, was eine Menge ist, zitiert man üblicherweise die folgende Charakterisierung von Georg Cantor (1845–1918):

> „Eine **Menge** ist eine Zusammenfassung von bestimmten, wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen."

Die in einer Menge zusammengefassten Objekte bezeichnet man als ihre **Elemente**.

Notation 1.13.

(a) Wir schreiben $x \in M$, falls x ein Element der Menge M ist, und $x \notin M$ andernfalls.

(b) Die einfachste Art, eine Menge konkret anzugeben, besteht darin, ihre Elemente in geschweiften Klammern aufzulisten, wobei es auf die Reihenfolge und Mehrfachnennungen nicht ankommt. So sind z. B. $\{1,2,3\}$ und $\{2,3,3,1\}$ zwei Schreibweisen für dieselbe Menge. Man beachte, dass die Elemente einer Menge nicht unbedingt Zahlen sein müssen – wir könnten z. B. auch die Menge aller Studenten in dieser Vorlesung betrachten. In der Tat kommen in der Mathematik sogar oft Mengen vor, deren Elemente selbst wieder Mengen sind; so ist z. B. $\{\{1,2,3\},\{1,2,3,4\}\}$ eine gültige Menge (mit 2 Elementen).

(c) Man kann die Elemente einer Menge auch durch eine beschreibende Eigenschaft angeben: $\{x : A(x)\}$ bezeichnet die Menge aller Objekte x, für die die Aussage $A(x)$ wahr ist, wie z. B. in $\{x \in \mathbb{R} : x^2 = 1\} = \{-1,1\}$.

(d) Die Menge $\{\}$ ohne Elemente, die sogenannte **leere Menge**, bezeichnen wir mit $\emptyset$.

(e) Man schreibt $M \subset N$ (oder $N \supset M$), wenn jedes Element von M auch Element von N ist, bzw. in der Quantorenschreibweise von Notation 1.6 wenn

$$\forall x : x \in M \Rightarrow x \in N.$$

In diesem Fall sagt man, dass M eine **Teilmenge** von N (oder N eine **Obermenge** von M) ist. Beachte, dass M und N dabei auch gleich sein können; in der Tat ist offensichtlich

$$M = N \quad \text{genau dann, wenn} \quad M \subset N \text{ und } N \subset M.$$

Oft wird man eine Gleichheit $M = N$ von Mengen auch so beweisen, dass man separat $M \subset N$ und $N \subset M$ zeigt.

Wenn wir ausdrücken wollen, dass M eine Teilmenge von N und nicht gleich N ist, so schreiben wir dies als $M \subsetneq N$ und sagen, dass M eine **echte Teilmenge** von N ist. Es ist wichtig, dies von der Aussage $M \not\subset N$ zu unterscheiden, die bedeutet, dass M keine Teilmenge von N ist.

Achtung: Manchmal wird in der Literatur das Symbol „$\subset$" für *echte* Teilmengen und „$\subseteq$" für nicht notwendig echte Teilmengen verwendet.

(f) Hat eine Menge M nur endlich viele Elemente, so nennt man M eine **endliche Menge** und schreibt die Anzahl ihrer Elemente als $|M|$. Andernfalls setzt man formal $|M| = \infty$.

Bemerkung 1.14 (Russellsches Paradoxon)**.** Die oben gegebene Charakterisierung von Mengen von Cantor ist aus mathematischer Sicht natürlich sehr schwammig. In der Tat hat Bertrand Russell kurz darauf bemerkt, dass sie sogar schnell zu Widersprüchen führt. Er betrachtet dazu

$$M = \{A : A \text{ ist eine Menge mit } A \notin A\}, \tag{$*$}$$

also „die Menge aller Mengen, die sich nicht selbst als Element enthalten". Sicherlich ist es eine merkwürdige Vorstellung, dass eine Menge sich selbst als Element enthalten könnte – im Sinne von Cantors Charakterisierung wäre die Definition $(*)$ aber zulässig. Fragen wir uns nun allerdings, ob sich die so konstruierte Menge M selbst als Element enthält, so erhalten wir sofort einen Widerspruch: Wenn $M \in M$ gilt, so würde das nach der Definition $(*)$ ja gerade bedeuten, dass $M \notin M$ ist – und das wiederum, dass doch $M \in M$ ist. Man bezeichnet dies als das *Russellsche Paradoxon*.

Die Ursache für diesen Widerspruch ist, dass die Definition $(*)$ rückbezüglich ist: Wir wollen eine neue Menge M konstruieren, verwenden dabei aber auf der rechten Seite der Definition *alle Mengen*, also u. a. auch die Menge M, die wir gerade erst definieren wollen. Das ist in etwa so, als würdet ihr im Beweis eines Satzes die Aussage des Satzes selbst verwenden – und das ist natürlich nicht zulässig.

Man muss bei der Festlegung, was Mengen sind und wie man sie bilden kann, also eigentlich viel genauer vorgehen, als es Cantor getan hat. Heutzutage verwendet man hierzu in der Regel das im Jahre 1930 aufgestellte Axiomensystem von Zermelo und Fraenkel, das genau angibt, wie man aus bekannten Mengen neue konstruieren darf: z. B. indem man sie schneidet oder vereinigt, oder aus bereits bekannten Mengen Elemente mit einer bestimmten Eigenschaft auswählt. Wir wollen dies hier in dieser Vorlesung aber nicht weiter thematisieren und uns mit der naiven Mengencharakterisierung von Cantor begnügen (sowie der Versicherung meinerseits, dass schon alles in Ordnung ist, wenn wir neue Mengen immer nur aus alten konstruieren und keine rückbezüglichen Definitionen hinschreiben). Genaueres zum Zermelo-Fraenkel-Axiomensystem könnt ihr z. B. in [E, Kapitel 13] nachlesen.

Notation 1.15 (Reelle Zahlen)**.** Unser wichtigstes Beispiel für eine Menge ist die Menge der **reellen Zahlen**, die wir mit $\mathbb{R}$ bezeichnen werden. Wir wollen die Existenz der reellen Zahlen in dieser Vorlesung axiomatisch voraussetzen und begnügen uns daher an dieser Stelle damit zu sagen, dass man sie sich als die Menge der Punkte auf einer Geraden (der „Zahlengeraden") vorstellen kann. Zusätzlich werden wir in den nächsten beiden Kapiteln die mathematischen Eigenschaften von $\mathbb{R}$ exakt angeben (und ebenfalls axiomatisch voraussetzen) – und zwar genügend viele Eigenschaften, um $\mathbb{R}$ dadurch eindeutig zu charakterisieren.

Ich möchte hier noch einmal betonen, dass man die Existenz und die Eigenschaften der reellen Zahlen eigentlich nicht voraussetzen müsste: Man kann das auch allein aus den Axiomen der Logik und Mengenlehre herleiten! Dies wäre jedoch relativ aufwändig und würde euch im Moment mehr verwirren als helfen, daher wollen wir hier darauf verzichten. Wer sich trotzdem dafür interessiert, kann die Einzelheiten hierzu in [E, Kapitel 1 und 2] nachlesen.

Außer den reellen Zahlen sind vor allem noch die folgenden Teilmengen von $\mathbb{R}$ wichtig:

(a) die Menge $\mathbb{N} = \{0, 1, 2, \dots\}$ der **natürlichen Zahlen** (Achtung: Es gibt Bücher, in denen die 0 nicht mit zu den natürlichen Zahlen gezählt wird!);

(b) die Menge $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ der **ganzen Zahlen**;

(c) die Menge $\mathbb{Q} = \{\frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0\}$ der **rationalen Zahlen**.

Offensichtlich sind diese Mengen ineinander enthalten: Es gilt $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$. Teilmengen von $\mathbb{R}$, die durch Ungleichungen gegeben sind, schreiben wir in der Regel, indem wir die Ungleichungsbedingung als Index an das Symbol $\mathbb{R}$ schreiben, z. B. $\mathbb{R}_{\geq 0}$ für die Menge $\{x \in \mathbb{R} : x \geq 0\}$ aller nicht-negativen Zahlen.

Definition 1.16. Sind M und N Mengen, so bezeichnen wir mit ...

(a) $M \cap N := \{x : x \in M \text{ und } x \in N\}$ die **Schnittmenge** von M und N. Gilt $M \cap N = \emptyset$, so sagen wir, dass M und N **disjunkt** sind.

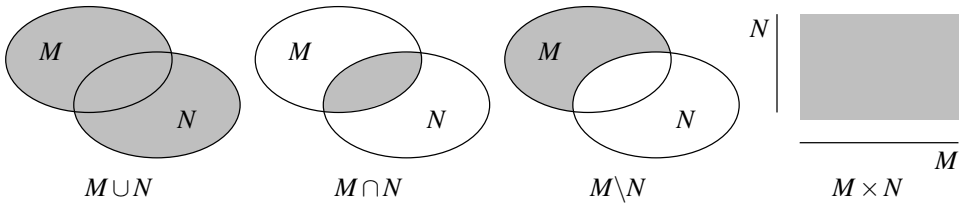
(b) $M \cup N := \{x : x \in M \text{ oder } x \in N\}$ die **Vereinigungsmenge** von M und N. im Fall einer **disjunkten Vereinigung** mit $M \cap N = \emptyset$ schreiben wir statt $M \cup N$ auch $M \dot\cup N$.

(c) $M \setminus N := \{x : x \in M \text{ und } x \notin N\}$ die **Differenzmenge** von M und N.

(d) $M \times N := \{(x,y) : x \in M, y \in N\}$ die **Produktmenge** bzw. das Produkt von M und N. Die Schreibweise (x,y) steht hierbei für ein **geordnetes Paar**, d. h. einfach für die Angabe eines Elements aus M und eines aus N (wobei es auch im Fall $M = N$ auf die Reihenfolge ankommt, d. h. (x,y) ist genau dann gleich (x',y') wenn $x = x'$ und $y = y'$). Im Fall $M = N$ schreibt man $M \times N = M \times M$ auch als M^2.

(e) $\mathscr{P}(M) := \{A : A \text{ ist Teilmenge von } M\}$ die **Potenzmenge** von M.

Das Symbol „:=“ bedeutet hierbei, dass der Ausdruck auf der linken Seite durch die rechte Seite definiert wird. Die Konstruktionen (a) bis (d) können durch die folgenden Bilder veranschaulicht werden. Natürlich sind sie auch für mehr als zwei Mengen möglich; aus der Schule kennt ihr zum Beispiel sicher den Fall $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.



Die Potenzmenge $\mathscr{P}(M)$ aller Teilmengen einer gegebenen Menge M lässt sich dagegen nicht so einfach durch ein Bild darstellen. Es ist z. B.

$$\mathscr{P}(\{0,1\}) = \big\{\emptyset, \{0\}, \{1\}, \{0,1\}\big\}.$$

Aufgabe 1.17. Wie lautet die Negation der folgenden Aussagen? Formuliere außerdem die Aussage (a) in Worten (also analog zu (b)) sowie die Aussage (b) mit Quantoren und anderen mathematischen Symbolen (also analog zu (a)).

(a) $\forall n \in \mathbb{N}\ \exists m \in \mathbb{N} : n = 2m$.

(b) Zwischen je zwei verschiedenen reellen Zahlen gibt es noch eine weitere reelle Zahl.

(c) Sind M, N, R Mengen mit $R \subset N \subset M$, so ist $M \setminus N \subset M \setminus R$.

Aufgabe 1.18. Es seien A, B, C Aussagen und M, N, R Mengen. Man zeige:

(a) $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$ und $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$.

(b) $M \cup (N \cap R) = (M \cup N) \cap (M \cup R)$ und $M \cap (N \cup R) = (M \cap N) \cup (M \cap R)$.

Aufgabe 1.19. Welche der folgenden Aussagen sind für beliebige gegebene x und M äquivalent zueinander? Zeige jeweils die Äquivalenz bzw. widerlege sie durch ein Gegenbeispiel.

(a) $x \in M$	(b) $\{x\} \subset M$	(c) $\{x\} \cap M \neq \emptyset$
(d) $\{x\} \in M$	(e) $\{x\} \setminus M = \emptyset$	(f) $M \setminus \{x\} = \emptyset$

Aufgabe 1.20. Man beweise oder widerlege: Für alle Mengen $A \subset M$ und $A' \subset M'$ gibt es Teilmengen B und C von M sowie B' und C' von M', so dass

$$(M \times M') \setminus (A \times A') = (B \times B') \dot\cup (C \times C').$$

Könnt ihr die Aussage durch eine Skizze veranschaulichen?

2. Relationen und Funktionen

Nachdem wir Mengen eingeführt haben, wollen wir nun auch mehrere von ihnen miteinander in Beziehung setzen können. Das Grundkonzept hierfür ist das einer Relation.

Definition 2.1 (Relationen)**.** Es seien M und N zwei Mengen. Eine **Relation** zwischen M und N ist eine Teilmenge R des Produkts $M \times N$. Für $x \in M$ und $y \in N$ mit $(x,y) \in R$ sagen wir dann „*x steht (bezüglich R) in Relation zu y*“. Ist $M = N$, so nennen wir R auch eine *Relation auf M*.

Bemerkung 2.2. Um eine Relation R anzugeben, also eine Teilmenge $R \subset M \times N$ zu definieren, müssen wir demzufolge einfach für alle Paare (x,y) mit $x \in M$ und $y \in N$ festlegen, ob $(x,y) \in R$ gelten, also ob x in Relation zu y stehen soll.

Wie wir in diesem Kapitel sehen werden, werden Relationen in der Mathematik für völlig unterschiedliche Konzepte verwendet – z. B. um Zahlen miteinander zu vergleichen wie in Beispiel 2.3 (b), um eine Menge auf eine andere abzubilden wie in Abschnitt 2.A, oder um die Elemente einer Menge nach bestimmten Kriterien zu Klassen zusammenzufassen wie in Abschnitt 2.B. Dementsprechend sind für die Aussage „x steht bezüglich R in Relation zu y“ auch je nach Anwendung ganz unterschiedliche Notationen üblich. Für allgemeine, nicht näher spezifizierte Relationen schreibt man hierfür oft xRy.

Beispiel 2.3.

(a) Es sei M die Menge aller derzeit an der TU Kaiserslautern eingeschriebenen Studenten und N die Menge aller in diesem Semester angebotenen Vorlesungen. Dann ist

$$R = \{(x,y) : x \text{ hört } y\}$$

eine Relation zwischen M und N.

(b) Für $M = N = \mathbb{R}$ betrachten wir die Relation

$$R = \{(x,y) : x,y \in \mathbb{R} \text{ mit } x < y\},$$

für die x also genau dann in Relation zu y steht, wenn $x < y$ gilt. Man nennt R deshalb auch die *Kleiner-Relation* auf $\mathbb{R}$. Die Notation „xRy“ aus Bemerkung 2.2 stimmt in diesem Fall also mit der Schreibweise „$x < y$“ überein, wenn man die Relation R direkt mit dem Symbol „$<$“ bezeichnet. In der Tat ist es aus diesem Grund bei manchen Relationen üblich, sie gleich mit Symbolen statt mit Buchstaben zu benennen.

2.A Funktionen

Die mit Abstand wichtigsten Relationen sind ohne Zweifel die Funktionen, die ihr natürlich bereits hinlänglich aus der Schule kennt. Wir wollen sie hier nun exakt einführen und ihre ersten Eigenschaften untersuchen.

Definition 2.4 (Funktionen)**.** Es seien M und N zwei Mengen.

(a) Eine **Funktion** oder **Abbildung** f von M nach N, geschrieben $f\colon M \to N$, ist eine Relation zwischen M und N, bezüglich der jedes Element x von M zu *genau einem* Element y von N in Relation steht. Wir schreiben dies dann als $x \mapsto y$ oder $y = f(x)$ und sagen, y ist das **Bild** von x unter f bzw. der **Wert** von f in x.

(b) Für eine Funktion $f\colon M \to N$ bezeichnet man die Menge M als **Definitionsmenge**, **Startmenge** oder **Startraum** von f. Die Menge N heißt **Zielmenge** oder **Zielraum** von f.

Bemerkung 2.5.

(a) Um eine Funktion komplett festzulegen, müssen wir zuerst einmal den Start- und Zielraum angeben, und dann schließlich noch von jedem Element des Startraums sagen, auf welches Element des Zielraums es abgebildet wird. In welcher Form wir diese Zuordnung angeben – ob durch eine Formel, durch explizites Auflisten der Funktionswerte aller Elemente des Startraums, oder irgendwie anders – spielt dabei keine Rolle. So sind z. B.

$$f\colon \{0,1\} \to \mathbb{R},\ x \mapsto 2x^2, \quad g\colon \{0,1\} \to \mathbb{R},\ x \mapsto 2x^3, \quad h\colon \{0,1\} \to \mathbb{R},\ x \mapsto \begin{cases} 0 & \text{für } x = 0, \\ 2 & \text{für } x = 1 \end{cases}$$

trotz ihrer ganz verschieden aussehenden Vorschriften dieselbe Funktion, da alle drei den gleichen Start- und Zielraum haben und aus den gleichen Zuordnungen $0 \mapsto 0$ und $1 \mapsto 2$ bestehen. Mit anderen Worten sind zwei Funktionen $f,g\colon M \to N$ also genau dann gleich, wenn sie an jedem Punkt die gleichen Werte besitzen, also wenn gilt

$$\forall x \in M : f(x) = g(x).$$

(b) Man sieht leider oft, dass eine Funktion $f\colon M \to N$ als $f(x)$ geschrieben wird. Es ist wichtig zu verstehen, dass diese Notation gemäß Definition 2.4 falsch ist: Mit $f(x)$ wird *der Wert der Funktion f in einem Punkt $x \in M$* bezeichnet. Somit ist $f(x)$ (für gegebenes x) ein Element von N, und damit ein ganz anderes mathematisches Objekt als die Funktion selbst, die wir nur mit f bezeichnen und die eine Relation zwischen M und N ist. Dies mag auf den ersten Blick spitzfindig erscheinen – wir werden aber später noch oft Mengen sehen, deren Elemente Funktionen sind, und dann ist es natürlich wichtig, dies von der Menge ihrer Funktions*werte* zu unterscheiden.

Beispiel 2.6.

(a) Die Zuordnungen

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \frac{1}{x} \qquad \text{und} \qquad g\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} x+1 & \text{für } x \le 0 \\ x & \text{für } x \ge 0 \end{cases}$$

sind in dieser Form keine zulässigen Funktionsdefinitionen, weil im Fall von f der Zahl 0 kein gültiger Funktionswert zugeordnet wird und im Fall g für die Zahl 0 zwei (sich widersprechende) Festlegungen des Funktionswertes gemacht werden. Dies lässt sich jedoch in beiden Fällen leicht reparieren, z. B. indem man die Festlegungen abändert in

$$f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R},\ x \mapsto \frac{1}{x} \qquad \text{und} \qquad g\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} x+1 & \text{für } x < 0 \\ x & \text{für } x \ge 0 \end{cases}.$$

(b) Zu jeder Menge M gibt es die **identische Abbildung**

$$\mathrm{id}_M\colon M \to M,\ x \mapsto x,$$

die jedes Element auf sich selbst abbildet.

(c) Ist $f\colon M \to N$ eine Abbildung und $A \subset M$ eine Teilmenge des Startraums, so erhält man durch die Einschränkung der Definitionsmenge von M auf A eine neue Abbildung, die wir mit

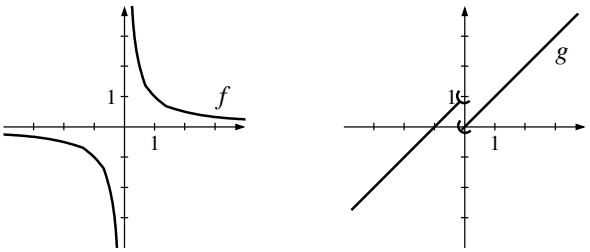
$$f|_A\colon A \to N,\ x \mapsto f(x)$$

bezeichnen und die die **Einschränkung** von f auf A genannt wird. Genauso kann man natürlich auch die Zielmenge N auf eine Teilmenge B einschränken, wenn f nur Werte in B annimmt. Es ist üblich, bei einer derartigen Einschränkung der Zielmenge immer noch den gleichen Namen für die Abbildung zu verwenden, also dann $f\colon M \to B$ zu schreiben (auch wenn es sich dabei um eine andere Funktion als das ursprüngliche $f\colon M \to N$ handelt).

Bemerkung 2.7 (Graph einer Abbildung)**.** Zu einer Abbildung $f\colon M \to N$ heißt die Menge

$$\{(x, f(x)) : x \in M\} \quad \subset M \times N$$

der **Graph** von f. Sind M und N Teilmengen von $\mathbb{R}$, so ist dieser Graph also eine Teilmenge von $\mathbb{R}^2$, und man kann ihn leicht zeichnen und dadurch die Abbildung veranschaulichen. Für die Abbildungen aus Beispiel 2.6 (a) sieht dies z. B. so aus:



Beachte, dass dieser Graph nach den Definitionen 2.1 und 2.4 eigentlich sogar genau das gleiche ist wie die Funktion selbst, nämlich die Teilmenge des Produkts $M \times N$, die aus den Paaren (x, y) besteht, für die x bezüglich f in Relation zu y steht, also $y = f(x)$ gilt. Der Begriff des Graphen soll hier also nur noch einmal deutlich machen, dass man sich die Funktion gerade wirklich als ein derart „grafisches“ Objekt vorstellt und nicht als eine „Zuordnung“ von M nach N.

In der Definition 2.4 einer Abbildung $f\colon M \to N$ verlangen wir, dass jedem Element von M genau ein Element von N zugeordnet wird. Wir fordern jedoch nicht auch umgekehrt, dass jedes Element des Zielraums N das Bild von genau einem Element von M ist, oder dass es überhaupt als Bild eines Elements von M auftritt. Abbildungen, die diese Eigenschaften dennoch besitzen, haben spezielle Namen, die wir jetzt einführen wollen.

Definition 2.8 (Eigenschaften von Abbildungen)**.** Es sei $f\colon M \to N$ eine Abbildung.

(a) Ist $y \in N$ und $x \in M$ mit $f(x) = y$, so heißt x ein **Urbild** von y unter f.

(b) Hat jedes $y \in N$ ...

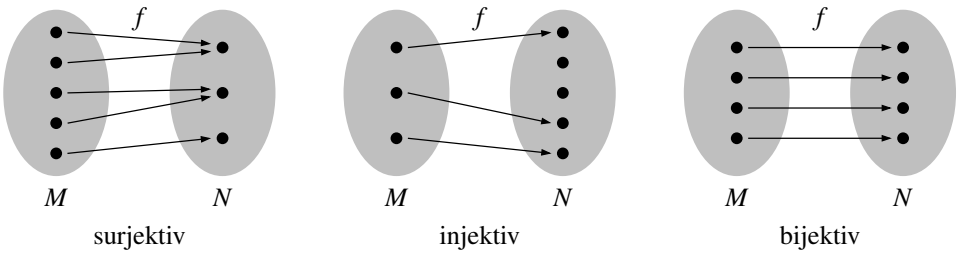
  (i) *mindestens* ein Urbild, so heißt f **surjektiv**.

  In Quantoren bedeutet dies: $\forall y \in N\ \exists x \in M\colon f(x) = y$.

  (ii) *höchstens* ein Urbild, so heißt f **injektiv**.

  In Quantoren bedeutet dies: $\forall x_1, x_2 \in M\colon f(x_1) = f(x_2) \Rightarrow x_1 = x_2$. (Also: Haben zwei Elemente des Startraums das gleiche Bild, so müssen sie bereits dasselbe Element sein.)

  (iii) *genau* ein Urbild, ist f also surjektiv und injektiv, so heißt f **bijektiv**.



Beispiel 2.9. Betrachten wir noch einmal die Funktionen aus Beispiel 2.6 (a) bzw. Bemerkung 2.7. Die Funktion f ist nicht surjektiv, da das Element 0 des Zielraums kein Urbild hat. Sie ist jedoch injektiv: Sind $x_1, x_2 \in \mathbb{R}\backslash\{0\}$ mit $f(x_1) = f(x_2)$, also $\frac{1}{x_1} = \frac{1}{x_2}$, so folgt durch Multiplikation mit $x_1 x_2$ sofort $x_1 = x_2$.

Die Funktion g dagegen ist surjektiv: Eine Zahl $y \in \mathbb{R}$ hat als Urbild $x = y$ für $y \geq 0$, und $x = y - 1$ für $y < 0$. Sie ist allerdings nicht injektiv, denn es ist $g(-1) = g(0) = 0$.

Beachte, dass diese Eigenschaften auch von der Wahl der Start- und Zielmenge abhängen: So wird z. B. f bijektiv, wenn man die Zielmenge $\mathbb{R}$ durch $\mathbb{R}\backslash\{0\}$ ersetzt.

Aufgabe 2.10. Wie viele Abbildungen gibt es zwischen den Mengen $\{1,2,3,4\}$ und $\{1,2,3,4,5,6\}$? Wie viele von ihnen sind injektiv?

Bilder und Urbilder unter Abbildungen betrachtet man oft auch von ganzen Mengen statt nur von Punkten:

Definition 2.11. Es sei $f\colon M \to N$ eine Abbildung.

(a) Für $A \subset M$ heißt die Menge
$$f(A) := \{f(x) : x \in A\} \quad \subset N$$
(also die Menge aller Bilder von Punkten in A) das **Bild** von A unter f.

(b) Ist $B \subset N$, so heißt die Menge
$$f^{-1}(B) := \{x \in M : f(x) \in B\} \quad \subset M$$
(also die Menge aller Urbilder von Punkten in B) das **Urbild** von B unter f.

Bemerkung 2.12. Die Grundidee der Notation in Definition 2.11 (a) ist: Schreiben wir als Argument einer Funktion $f\colon M \to N$ eine *Teilmenge* statt einem *Element* von M, so bedeutet dies, dass wir alle Werte $f(x)$ für $x \in M$ zusammen nehmen und diese wieder in einer Menge zusammenfassen. Diese Schreibweise verwendet man auch oft, wenn die Funktion aus einer Rechenverknüpfung besteht, wie z. B. in
$$\mathbb{N}+\tfrac{1}{2} := \left\{n+\tfrac{1}{2} : n \in \mathbb{N}\right\} = \left\{\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2},\dots\right\} \quad \text{oder} \quad 3\mathbb{Z} := \{3n : n \in \mathbb{Z}\} = \{\dots,-6,-3,0,3,6,\dots\}.$$

Beispiel 2.13. Für die Funktion $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $x \mapsto \frac{1}{x}$ aus Beispiel 2.6 (a) ist $f(\mathbb{R}_{>0}) = \mathbb{R}_{>0}$ und
02 $f^{-1}(\{0\}) = \emptyset$.

Beispiel 2.14. Zwischen den Konstruktionen von Bild und Urbild aus Definition 2.11 und den Mengenoperationen aus Abschnitt 1.B gibt es sehr viele Beziehungen. Um einmal exemplarisch zu sehen, wie derartige Beziehungen aussehen und bewiesen werden können, wollen wir nun zeigen, dass für jede Abbildung $f\colon M \to N$ und zwei beliebige Teilmengen $A,B \subset M$ stets
$$f(A)\backslash f(B) \subset f(A\backslash B) \tag{$*$}$$
gilt.

Zum Beweis müssen wir zeigen, dass jedes Element der linken Menge auch in der rechten Menge liegt. Es sei also $y \in f(A)\backslash f(B)$ beliebig. Insbesondere ist damit $y \in f(A)$, nach Definition 2.11 (a) also $y = f(x)$ für ein $x \in A$. Würde nun auch $x \in B$ gelten, so hätten wir wegen $y = f(x)$ auch $y \in f(B)$, im Widerspruch zu $y \in f(A)\backslash f(B)$. Also ist $x \notin B$, und damit $x \in A\backslash B$. Damit besagt $y = f(x)$ aber gerade $y \in f(A\backslash B)$. Insgesamt haben wir somit die behauptete Teilmengenbeziehung $(*)$ gezeigt.

Beachte allerdings, dass in $(*)$ im Allgemeinen keine Gleichheit gilt: Für $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ mit $A = \{-1,1\}$ und $B = \{-1\}$ ist
$$f(A)\backslash f(B) = \{1\}\backslash\{1\} = \emptyset, \quad \text{aber} \quad f(A\backslash B) = f(\{1\}) = \{1\}.$$

Aufgabe 2.15. Beweise die folgenden Teilmengenbeziehungen und untersuche jeweils, ob auch die Gleichheit gilt.

(a) Für alle Mengen M,A,B gilt $M\backslash(A\cup B) \subset (M\backslash A)\cap(M\backslash B)$.

(b) Ist $f\colon M \to N$ eine Abbildung und $A \subset N$, so ist $f\left(f^{-1}(A)\right) \subset A$.

Aufgabe 2.16. Es sei $f\colon M \to N$ eine Abbildung. Finde für das Symbol $\square$ jeweils eine der Mengenbeziehungen $\subset, =, \supset$, so dass die folgenden Aussagen wahr werden, und beweise die so entstandenen Aussagen!

(a) $f(A)\cap f(B) \;\square\; f(A\cap B)$ für alle $A,B \subset M$.

(b) $f^{-1}(A) \cap f^{-1}(B) \,\square\, f^{-1}(A \cap B)$ für alle $A, B \subset N$.

Als Nächstes wollen wir nun die euch sicher bereits bekannte Verkettung, also die Hintereinanderausführung von Funktionen einführen.

Definition 2.17 (Verkettung von Funktionen)**.** Es seien $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen (also so dass die Zielmenge von f gleich der Startmenge von g ist). Dann heißt die Abbildung

$$g \circ f\colon M \to R,\ x \mapsto g(f(x))$$

die **Verkettung** von f und g.

Bemerkung 2.18. Bei der Verkettung zweier Funktionen kommt es natürlich auf die Reihenfolge an, allein schon weil in der Situation von Definition 2.17 in der Regel der Zielraum von g ja nicht mit dem Startraum von f übereinstimmt und die „umgekehrte Verkettung“ $f \circ g$ damit gar nicht definierbar wäre. Beachte dabei, dass die Notation $g \circ f$ lautet, obwohl wir zuerst f (von M nach N) und dann g (von N nach R) anwenden. Diese vielleicht etwas merkwürdig erscheinende Notation kommt einfach daher, dass die Buchstaben in der gleichen Reihenfolge stehen sollen wie bei der Abbildungsvorschrift $x \mapsto g(f(x))$.

Wir wollen nun unser erstes *Lemma* beweisen – „Lemma“ ist griechisch und bedeutet eigentlich „Annahme“, aber in der Mathematik wird dieser Begriff für einen *Hilfssatz* verwendet, also für ein kleines Zwischenresultat, das vielleicht für sich genommen nicht übermäßig überraschend oder interessant ist, aber das in späteren Beweisen immer wieder nützlich sein wird. In unserem momentanen Fall geht es einfach darum, dass die Verkettung von Abbildungen *assoziativ* ist (siehe auch Definition 3.1):

Lemma 2.19 (Assoziativität der Verkettung)**.** *Sind $f\colon M \to N$, $g\colon N \to R$ und $h\colon R \to S$ drei Abbildungen, so gilt $h \circ (g \circ f) = (h \circ g) \circ f$. (Man schreibt für diese Abbildung daher oft auch einfach $h \circ g \circ f$.)*

Beweis. Nach Definition 2.4 können wir die Gleichheit zweier Funktionen zeigen, indem wir für jedes Element der Startmenge nachweisen, dass sein Bild unter beiden Funktionen übereinstimmt. Dies rechnen wir nun einfach durch wiederholtes Einsetzen von Definition 2.17 nach: Es gilt

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)))$$

und

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x))).$$

Da diese beiden Ausdrücke übereinstimmen, ist das Lemma bewiesen. □

Schließlich wollen wir nun noch sehen, wie sich die Begriffe der Surjektivität, Injektivität und Bijektivität aus Definition 2.8 (b) äquivalent mit Hilfe von Umkehrfunktionen ausdrücken lassen.

Lemma 2.20 (Umkehrfunktionen)**.** *Es sei $f\colon M \to N$ eine Abbildung zwischen nicht-leeren Mengen. Dann gilt:*

(a) *f ist surjektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$.*

(b) *f ist injektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $g \circ f = \mathrm{id}_M$.*

(c) *f ist bijektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$ und $g \circ f = \mathrm{id}_M$.*

*In diesem Fall ist diese Abbildung g eindeutig bestimmt. Wir nennen sie die **Umkehrfunktion** bzw. **Umkehrabbildung** von f und bezeichnen sie mit f^{-1}.*

Beweis. Wir beweisen die Richtungen „$\Rightarrow$“ und „$\Leftarrow$“ der drei Teile separat.

(a) „$\Rightarrow$“: Es sei f surjektiv. Dann können wir

$$g\colon N \to M,\ y \mapsto \text{ein Urbild von } y \text{ unter } f$$

setzen (wobei wir, falls mehrere solche Urbilder existieren, jeweils eines auswählen). Für alle $y \in N$ gilt dann $f(g(y)) = y$, da $g(y)$ ja ein Urbild von y unter f ist. Also ist $f \circ g = \mathrm{id}_N$.

„$\Leftarrow$“: Es gebe ein $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$. Ist dann $y \in N$ beliebig, so können wir $x = g(y)$ setzen, und es gilt $f(x) = f(g(y)) = \mathrm{id}_N(y) = y$. Also hat jedes solche y ein Urbild unter f, d. h. f ist surjektiv.

(b) „$\Rightarrow$“: Es sei f injektiv. Dann setzen wir

$$g\colon N \to M,\ y \mapsto \begin{cases} \text{das eindeutig bestimmte Urbild von } y \text{ unter } f\text{, falls eines existiert,} \\ \text{ein beliebiges Element von } M \text{ sonst} \end{cases}$$

(wobei wir in der zweiten Zeile die Voraussetzung $M \neq \emptyset$ brauchen). Für alle $x \in M$ ist dann $g(f(x))$ nach Konstruktion das eindeutige Urbild von $f(x)$ unter f, also x. Damit folgt $g \circ f = \mathrm{id}_M$.

„$\Leftarrow$“: Es gebe ein $g\colon N \to M$ mit $g \circ f = \mathrm{id}_M$. Wir müssen zeigen, dass f injektiv ist, also die Bedingung aus Definition 2.8 (b)(ii) überprüfen. Es seien dazu $x_1, x_2 \in M$ mit $f(x_1) = f(x_2)$. Anwenden von g liefert $g(f(x_1)) = g(f(x_2))$, wegen $g \circ f = \mathrm{id}_M$ also $x_1 = x_2$. Damit ist f injektiv.

(c) „$\Rightarrow$“: Ist f bijektiv, so können wir

$$g\colon N \to M,\ y \mapsto \text{das eindeutig bestimmte Urbild von } y \text{ unter } f$$

setzen. Da diese Definition zu denen aus (a) und (b) „$\Rightarrow$“ passt, zeigen die obigen Rechnungen also, dass f dann surjektiv und injektiv und damit bijektiv ist.

„$\Leftarrow$“: Dies folgt sofort aus Teil „$\Leftarrow$“ von (a) und (b).

Eindeutigkeit von g: Es seien $g_1, g_2\colon N \to M$ zwei Umkehrfunktionen von f, insbesondere ist also $f \circ g_2 = \mathrm{id}_N$ und $g_1 \circ f = \mathrm{id}_M$. Dann gilt

$$g_1 = g_1 \circ \mathrm{id}_N = g_1 \circ (f \circ g_2) \overset{2.19}{=} (g_1 \circ f) \circ g_2 = \mathrm{id}_M \circ g_2 = g_2. \qquad \square$$

Bemerkung 2.21 (Urbilder und Umkehrabbildungen)**.** Beachte, dass wir das Urbild einer Menge unter einer Abbildung $f\colon M \to N$ in Definition 2.11 (b) mit dem gleichen Symbol f^{-1} bezeichnet haben wie (im Fall einer bijektiven Abbildung) die Umkehrabbildung aus Lemma 2.20 (c). Das ist vielleicht etwas unglücklich gewählt, aber in der Literatur so fest verankert, dass wir hier nicht davon abweichen wollen. Bei genauem Hinschauen kann man aber auch immer feststellen, was gemeint ist:

Ist $f\colon M \to N$ eine Abbildung, so bezeichnet ...
... $f^{-1}(B)$ für eine *Menge* $B \subset N$ das *Urbild* von B wie in Definition 2.11 (b); es existiert für jede Abbildung f.
... $f^{-1}(y)$ für ein *Element* $y \in B$ den *Wert der Umkehrabbildung* bei y wie in Lemma 2.20 (c); er existiert nur für bijektives f.

Letztlich hängen diese beiden Notationen aber auch eng miteinander zusammen: Ist f bijektiv und ist $x \in M$ mit $f(x) = y$, so ist $f^{-1}(y) = x$ (mit f^{-1} im Sinne der Umkehrabbildung) und $f^{-1}(\{y\}) = \{x\}$ (mit f^{-1} im Sinne des Urbildes).

Aufgabe 2.22.

(a) Untersuche die Abbildung $f\colon \mathbb{Z} \to \mathbb{Z},\ x \mapsto 3x+2$ auf Injektivität und Surjektivität.

(b) Untersuche die Abbildung $f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ (x,y) \mapsto (xy, x+1)$ auf Injektivität und Surjektivität.

(c) Man zeige: Sind $f\colon M \to N$ und $g\colon N \to R$ surjektiv, so ist auch $g \circ f\colon M \to R$ surjektiv.

Aufgabe 2.23. Es seien $f\colon M \to N$ und $g\colon N \to R$ bijektiv. Zeige, dass dann auch $f^{-1}\colon N \to M$ und $g \circ f\colon M \to R$ bijektiv sind.

Aufgabe 2.24. Man beweise oder widerlege:

(a) Sind $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen und ist $g \circ f$ injektiv, so ist auch f injektiv.

(b) Sind $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen und ist $g \circ f$ injektiv, so ist auch g injektiv.

Aufgabe 2.25. Es sei $f : M \to N$ eine Abbildung zwischen zwei Mengen. Man beweise:

$$f \text{ ist surjektiv} \quad \Leftrightarrow \quad \text{für alle } A, B \subset N \text{ mit } f^{-1}(A) = f^{-1}(B) \text{ gilt } A = B.$$

2.B Äquivalenzrelationen

Am Anfang dieses Kapitels haben wir allgemeine Relationen eingeführt, als einzigen Spezialfall davon aber bisher nur die Funktionen ausführlicher betrachtet. Wir wollen daher nun noch einen ganz anderen wichtigen Typ von Relationen studieren, die sogenannten Äquivalenzrelationen.

Angenommen, wir möchten eine Menge M untersuchen, die uns zunächst einmal zu groß oder zu kompliziert erscheint. Es gibt dann zwei prinzipiell verschiedene Möglichkeiten, wie man daraus eine kleinere bzw. einfachere Menge machen kann:

- Wir können uns auf eine Teilmenge von M beschränken – dann schließen wir allerdings manche Elemente von M von unserer Untersuchung aus.
- Wir können zwar alle Elemente von M betrachten, aber manche von ihnen miteinander identifizieren bzw. zu Klassen zusammenfassen – d. h. sie als gleich bzw. „äquivalent“ ansehen, wenn sie für das zu untersuchende Problem ähnliche Eigenschaften haben.

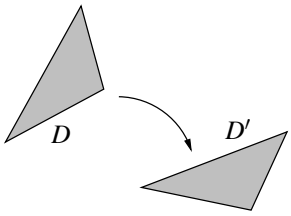
Diese zweite Idee der Identifizierung ähnlicher Elemente führt zum Begriff der Äquivalenzrelationen. Sie klingt vielleicht zunächst etwas abstrakt, ist euch aber sicher schon an vielen Stellen begegnet. Hier sind zwei einfache Beispiele dafür.

Beispiel 2.26.

(a) Eine analoge Uhr vereinfacht die recht große Menge aller (vergangenen und zukünftigen) Zeitpunkte, die wir uns als Zeitachse $M = \mathbb{R}$ vorstellen können, indem sie nach jeweils 12 Stunden wieder dasselbe anzeigt. Sie identifiziert also zwei Zeitpunkte $x, y \in \mathbb{R}$ (gemessen in Stunden) miteinander, wenn $x - y$ ein Vielfaches von 12 ist. Dadurch „verkleinert“ sie die ursprüngliche Zeitachse auf ein gut überschaubares Intervall von 12 Stunden – und wir alle wissen, dass uns ein Blick auf die Uhr in vielen Fällen ausreicht, wenn wir den aktuellen Zeitpunkt wissen wollen, auch wenn uns das nichts über das Datum oder die Tageszeit (vormittags oder nachmittags) sagt.

(b) Als „mathematischeres“ Beispiel können wir die Menge M aller Dreiecke in der Ebene $\mathbb{R}^2$ betrachten.

Bekanntlich heißen zwei solche Dreiecke $D, D' \in M$ zueinander *kongruent*, wenn sie wie im Bild rechts durch eine Drehung und / oder Verschiebung auseinander hervorgehen – wir schreiben dies im Folgenden als $D \sim D'$. Zueinander kongruente Dreiecke werden oft miteinander identifiziert, nämlich immer dann, wenn es uns nur auf die Form bzw. Größe der Dreiecke, aber nicht auf ihre Lage in der Ebene ankommt.



Wenn wir z. B. sagen, dass die drei Seitenlängen ein Dreieck eindeutig bestimmen, dann meinen wir damit in Wirklichkeit, dass sie das Dreieck *bis auf Kongruenz* eindeutig bestimmen, also nur die Form und Größe festlegen, aber nicht die Lage des Dreiecks in $\mathbb{R}^2$. Formal kann man dies so ausdrücken: zu einem Dreieck D nennt man

$$\overline{D} := \{D' \in M : D' \sim D\},$$

also die Menge aller zu D kongruenten Dreiecke, die *Kongruenzklasse* von D. Die Menge aller dieser Kongruenzklassen bezeichnen wir mit

$$M/\sim \; := \; \{\overline{D} : D \in M\}.$$

Man kann dann z. B. sagen, dass die Seitenlängen eines Dreiecks ein eindeutiges Element in $M/\sim$ bestimmen, also eine eindeutige Kongruenzklasse von Dreiecken festlegen – nicht aber ein eindeutiges Element von M.

Mit der Idee dieser Beispiele im Kopf wollen wir nun den Begriff der Äquivalenzrelation exakt definieren.

Definition 2.27 (Äquivalenzrelationen)**.** Es sei $\sim$ wie in Definition 2.1 eine Relation auf einer Menge M. Wie in Bemerkung 2.2 schreiben wir $x \sim y$, wenn x und y bezüglich $\sim$ in Relation stehen.

Man nennt $\sim$ eine **Äquivalenzrelation** auf M, wenn die folgenden Eigenschaften gelten:

(a) **Reflexivität**: Für alle $x \in M$ gilt $x \sim x$.

(b) **Symmetrie**: Sind $x, y \in M$ mit $x \sim y$, so gilt auch $y \sim x$.

(c) **Transitivität**: Sind $x, y, z \in M$ mit $x \sim y$ und $y \sim z$, so gilt auch $x \sim z$.

In diesem Fall sagt man statt $x \sim y$ auch, dass x (bezüglich dieser Relation) zu y **äquivalent** ist. Zu $x \in M$ heißt dann die Menge

$$\overline{x} := \{y \in M : y \sim x\}$$

aller Elemente, die zu x äquivalent sind, die **Äquivalenzklasse** bzw. einfach **Klasse** von x; jedes Element dieser Menge nennt man einen **Repräsentanten** dieser Klasse. Die Menge aller Äquivalenzklassen schreiben wir als

$$M/\sim \; := \; \{\overline{x} : x \in M\}.$$

Beispiel 2.28.

(a) Das Beispiel 2.26 (a) einer analogen Uhr lässt sich mathematisch exakt wie folgt definieren: Auf $M = \mathbb{R}$ betrachten wir die Relation

$$x \sim y \;:\Leftrightarrow\; \text{es gibt ein } k \in \mathbb{Z} \text{ mit } x - y = 12k. \qquad (*)$$

Dies ist in der Tat eine Äquivalenzrelation, denn für alle $x, y, z \in \mathbb{R}$ gilt:

- Reflexivität: Es gilt $x - x = 12 \cdot 0 = 12k$ mit $k = 0 \in \mathbb{Z}$, also $x \sim x$.
- Symmetrie: Es gelte $x \sim y$, also $x - y = 12k$ für ein $k \in \mathbb{Z}$. Durch Multiplikation mit -1 folgt dann auch $y - x = 12 \cdot (-k) = 12k'$ mit $k' := -k \in \mathbb{Z}$, und damit $y \sim x$.
- Transitivität: Es gelte $x \sim y$ und $y \sim z$, nach Definition der Relation also $x - y = 12k$ und $y - z = 12k'$ für gewisse $k, k' \in \mathbb{Z}$ (beachte, dass der Wert von k in $(*)$ von x und y abhängt und wir daher für die Differenz $y - z$ eine neue Variable k' brauchen). Durch Addition dieser beiden Gleichungen erhalten wir $x - z = 12(k + k') = 12k''$ mit $k'' := k + k' \in \mathbb{Z}$, und damit $x \sim z$.

Die Äquivalenzklasse z. B. von $2 \in \mathbb{R}$ ist

$$\overline{2} = \{x \in \mathbb{R} : x \sim 2\} = \{x \in \mathbb{R} : \text{es gibt ein } k \in \mathbb{Z} \text{ mit } x - 2 = 12k\} = \{2 + 12k : k \in \mathbb{Z}\},$$

also die Menge aller Zeitpunkte, zu denen die Uhr auf 2 steht. Jeder beliebige Zeitpunkt $x \in \mathbb{R}$, zu dem die Uhr auf 2 steht, ist ein Repräsentant dieser Klasse, und die Menge $M/\sim$ entspricht allen möglichen Ständen der Uhr.

(b) Die Kongruenz von Dreiecken aus Beispiel 2.26 (b) ist ebenfalls eine Äquivalenzrelation (es ist offensichtlich, dass sie die Eigenschaften aus Definition 2.27 erfüllt). Die Äquivalenzklassen sind in diesem Fall genau die Kongruenzklassen.

(c) Die „Kleiner-Relation“ auf $\mathbb{R}$ aus Beispiel 2.3 (b), also die Relation, für die für $x, y \in \mathbb{R}$ genau dann $x \sim y$ gilt, wenn $x < y$ ist, ist keine Äquivalenzrelation, da sie weder reflexiv noch symmetrisch ist.

Beachte, dass bei unseren Äquivalenzrelationen aus Beispiel 2.28 (a) und (b) jedes Element von M in genau einer Äquivalenzklasse liegt: Zu jedem Zeitpunkt hat eine analoge Uhr genau einen Stand, und jedes Dreieck in der Ebene liegt in genau einer Kongruenzklasse. Dies beschreibt genau unsere Idee, dass wir die Elemente von M auf eine bestimmte Art zu Klassen zusammenfassen wollen. Allgemein sind die Axiome einer Äquivalenzrelation aus Definition 2.27 anschaulich genau diejenigen, die man braucht, damit die Relation sinnvoll eine solche Identifizierung von Elementen zu Klassen beschreiben kann. Dies zeigt auch noch einmal der folgende zentrale Satz über Äquivalenzrelationen.

Satz 2.29 (Eigenschaften von Äquivalenzrelationen)**.** *Es sei $\sim$ eine Äquivalenzrelation auf einer Menge M.*

(a) *Für $x, y \in M$ gilt $x \sim y$ genau dann, wenn $\overline{x} = \overline{y}$. (Zwei Elemente sind also genau dann äquivalent zueinander, wenn sie die gleiche Äquivalenzklasse bestimmen.)*

(b) *Jedes Element $x \in M$ liegt in genau einer Äquivalenzklasse (nämlich in $\overline{x}$). Insbesondere ist M also die disjunkte Vereinigung aller Äquivalenzklassen. Man sagt auch, dass die Äquivalenzklassen eine* Partition *von M bilden.*

Beweis.

(a) Es seien $x, y \in M$.

„$\Rightarrow$“: Es gelte $x \sim y$. Ist dann $z \in M$ mit $z \in \overline{x}$, also $z \sim x$, so ist nach der Transitivität wegen $x \sim y$ auch $z \sim y$, also $z \in \overline{y}$. Damit gilt $\overline{x} \subset \overline{y}$. Da mit $x \sim y$ wegen der Symmetrie aber auch $y \sim x$ gilt, folgt analog auch umgekehrt $\overline{y} \subset \overline{x}$, und somit insgesamt $\overline{x} = \overline{y}$.

„$\Leftarrow$“: Es sei nun $\overline{x} = \overline{y}$. Wegen der Reflexivität ist $x \sim x$, also $x \in \overline{x} = \overline{y}$, und damit $x \sim y$.

(b) Wegen der Reflexivität liegt natürlich jedes $x \in M$ in seiner eigenen Äquivalenzklasse $\overline{x}$. Ist nun auch $x \in \overline{y}$ für ein $y \in M$, also $x \sim y$, so gilt nach (a) bereits $\overline{y} = \overline{x}$. Also liegt x in genau einer Äquivalenzklasse von $\sim$, nämlich in $\overline{x}$. □

Aufgabe 2.30. Es sei $M = \{n \in \mathbb{Z} : |n| \leq 100\} = \{-100, -99, \ldots, 0, \ldots, 99, 100\}$. Welche der folgenden Relationen sind Äquivalenzrelationen auf M? Gib im Fall einer Äquivalenzrelation außerdem die Äquivalenzklasse $\overline{-34}$ explizit an.

(a) $x \sim y \;:\Leftrightarrow\;$ es gibt ein $n \in \mathbb{Z}$ mit $x = 2^n y$.

(b) $x \sim y \;:\Leftrightarrow\; xy \geq 0$.

Aufgabe 2.31. Welche der folgenden Relationen sind Äquivalenzrelationen auf $\mathbb{R}^2$? Im Fall einer Äquivalenzrelation berechne und skizziere man außerdem die Äquivalenzklassen von $(0,1) \in \mathbb{R}^2$ und $(1,1) \in \mathbb{R}^2$.

(a) $(x,y) \sim (x',y') \;:\Leftrightarrow\;$ es gibt ein $a \in \mathbb{R}\backslash\{0\}$ mit $x = a^2 x'$ und $y = ay'$;

(b) $(x,y) \sim (x',y') \;:\Leftrightarrow\;$ es gibt ein $a \in \mathbb{R}\backslash\{0\}$ mit $x = ay'$ und $y = ax'$.

03

3. Erste Eigenschaften der reellen Zahlen

In Notation 1.15 haben wir bereits die reellen Zahlen $\mathbb{R}$ als „Menge der Punkte auf einer Geraden“ eingeführt. Man kann aber natürlich noch viel mehr Dinge mit den reellen Zahlen tun als sie als eine einfache Punktmenge zu betrachten: Man kann sie addieren, multiplizieren, die Größe von zwei Zahlen miteinander vergleichen, und noch einiges mehr. Wir wollen die Eigenschaften der reellen Zahlen in diesem und dem nächsten Kapitel exakt formalisieren, damit wir danach genau wissen, welche Eigenschaften von $\mathbb{R}$ wir in dieser Vorlesung axiomatisch voraussetzen. In der Tat werden diese Eigenschaften letztlich sogar ausreichen, um die reellen Zahlen eindeutig zu charakterisieren. Wir beginnen in diesem Kapitel aber zunächst einmal nur mit den „Grundrechenarten“, also mit der Addition und der Multiplikation sowie ihren Umkehrungen, der Subtraktion und Division.

3.A Gruppen und Körper

Die Eigenschaften von Verknüpfungen wie der Addition oder Multiplikation reeller Zahlen werden mathematisch durch die Begriffe einer Gruppe bzw. eines Körpers beschrieben, die wir jetzt einführen wollen.

Definition 3.1 (Gruppen)**.** Eine **Gruppe** ist eine Menge G zusammen mit einer „Verknüpfung“, d. h. einer Abbildung

$$*\colon G \times G \to G,\ (x,y) \mapsto x*y,$$

so dass die folgenden Eigenschaften (auch *Gruppenaxiome* genannt) gelten:

(a) (**Assoziativität**) Für alle $x,y,z \in G$ gilt $(x*y)*z = x*(y*z)$. Man schreibt diesen Ausdruck dann in der Regel auch einfach als $x*y*z$, weil die Reihenfolge der Klammerung ja egal ist.

(b) (Existenz eines neutralen Elements) Es gibt ein $e \in G$, für das $e*x = x*e = x$ für alle $x \in G$ gilt. Man nennt ein solches e ein **neutrales Element**, und verlangt davon zusätzlich:

(c) (Existenz von inversen Elementen) Für alle $x \in G$ gibt es ein $x' \in G$ mit $x'*x = x*x' = e$. Man nennt x' dann ein **inverses Element** zu x.

Wir bezeichnen eine solche Gruppe mit $(G,*)$. Wenn aus dem Zusammenhang klar ist, welche Verknüpfung gemeint ist, schreiben wir oft auch einfach nur G für die Gruppe.

Gilt zusätzlich zu den obigen Eigenschaften noch

(d) (**Kommutativität**) $x*y = y*x$ für alle $x,y \in G$,

so heißt $(G,*)$ eine **kommutative** oder **abelsche Gruppe**.

Bemerkung 3.2. Manchmal wird in der Definition einer Gruppe in Teil (b) lediglich $e*x = x$ und in Teil (c) lediglich $x'*x = e$ gefordert (man spricht dann auch von einem **linksneutralen** bzw. **linksinversen** Element). Man kann jedoch unter Verwendung der übrigen Gruppenaxiome zeigen, dass in diesem Fall automatisch auch $x*e = x$ und $x*x' = e$ gelten muss, also dass linksneutrale Elemente immer neutral und linksinverse Element bereits immer inverse Elemente sind [G, Satz 1.7]. Die beiden Varianten der Definition einer Gruppe stimmen also letztlich überein.

Beispiel 3.3.

(a) $(\mathbb{R},+)$ ist eine abelsche Gruppe, denn die Addition ist (wie wir axiomatisch voraussetzen werden) eine Verknüpfung auf $\mathbb{R}$ mit den Eigenschaften:

- $(x+y)+z = x+(y+z)$ für alle $x,y,z \in \mathbb{R}$;
- $0 \in \mathbb{R}$ ist ein neutrales Element, denn $0+x = x+0 = x$ für alle $x \in \mathbb{R}$;
- zu jedem $x \in \mathbb{R}$ ist $-x \in \mathbb{R}$ ein inverses Element, denn $(-x)+x = x+(-x) = 0$;

- $x+y=y+x$ für alle $x,y \in \mathbb{R}$.

Auf die gleiche Art sind auch $(\mathbb{Q},+)$ und $(\mathbb{Z},+)$ abelsche Gruppen, jedoch nicht $(\mathbb{N},+)$: Hier existiert zwar noch ein neutrales Element 0, aber die Zahl $1 \in \mathbb{N}$ hat kein Inverses mehr, denn es gibt kein $x \in \mathbb{N}$ mit $x+1=0$.

(b) $(\mathbb{R},\cdot)$ ist keine Gruppe: Die Multiplikation ist zwar assoziativ und kommutativ und hat das neutrale Element 1, aber die Zahl 0 hat kein Inverses – denn dies müsste ja eine Zahl $x \in \mathbb{R}$ sein mit $x \cdot 0 = 1$.

Nimmt man jedoch die 0 aus $\mathbb{R}$ heraus, so erhält man mit $(\mathbb{R}\backslash\{0\},\cdot)$ wieder eine abelsche Gruppe, bei der das neutrale Element 1 und das zu einem x inverse Element $\frac{1}{x}$ ist. Genauso funktioniert dies für $(\mathbb{Q}\backslash\{0\},\cdot)$, aber z. B. nicht für $(\mathbb{Z}\backslash\{0\},\cdot)$: Hier gibt es zwar noch ein neutrales Element 1, aber die Zahl $2 \in \mathbb{Z}\backslash\{0\}$ hat kein Inverses mehr, denn es gibt kein $x \in \mathbb{Z}\backslash\{0\}$ mit $2 \cdot x = 1$.

(c) Hier ist noch ein Beispiel von einem ganz anderen Typ: Es sei M eine beliebige Menge und $G = \{f\colon M \to M \text{ bijektiv}\}$ die Menge aller bijektiven Abbildungen von M nach M. Da die Verkettung bijektiver Abbildungen nach Aufgabe 2.23 wieder bijektiv ist, definiert sie eine Verknüpfung auf G. In der Tat wird G damit zu einer Gruppe, denn die Verkettung ist assoziativ nach Lemma 2.19, die Identität id_M ist ein neutrales Element, und zu einem $f \in G$ ist die Umkehrabbildung f^{-1} aus Lemma 2.20 (c) ein inverses Element: Sie ist nach Aufgabe 2.23 selbst wieder bijektiv (also in G) und erfüllt $f^{-1} \circ f = f \circ f^{-1} = \mathrm{id}_M$ nach Lemma 2.20 (c). Im Allgemeinen ist diese Gruppe jedoch nicht kommutativ.

Wir wollen nun ein paar einfache Eigenschaften von Gruppen beweisen, u. a. dass die in Definition 3.1 geforderten neutralen und inversen Elemente eindeutig sind und wir daher in Zukunft auch von *dem* neutralen und *dem* zu einem gegebenen Element inversen Element sprechen können.

Lemma 3.4 (Eigenschaften von Gruppen)**.** *Es seien $(G,*)$ eine Gruppe und $x,y \in G$.*

(a) *Es gibt genau ein neutrales Element (wie in Definition 3.1 (b)).*

(b) *Es gibt genau ein inverses Element zu x (wie in Definition 3.1 (c)).*

(c) *Sind x' und y' die inversen Elemente zu x bzw. y, so ist $y' * x'$ das inverse Element zu $x * y$.*

(d) *Ist x' das inverse Element zu x, so ist x das inverse Element zu x' („das Inverse des Inversen ist wieder das Ausgangselement“).*

Beweis.

(a) Sind e und $\tilde{e}$ neutrale Elemente, so folgt

$$\begin{aligned} e &= \tilde{e} * e && \text{(denn } \tilde{e} \text{ ist ein neutrales Element)} \\ &= \tilde{e} && \text{(denn } e \text{ ist ein neutrales Element).} \end{aligned}$$

(b) Sind x' und $\tilde{x}'$ inverse Elemente zu x, so gilt

$$\begin{aligned} x' &= e * x' && (e \text{ neutrales Element}) \\ &= (\tilde{x}' * x) * x' && (\tilde{x}' \text{ ist ein inverses Element zu } x) \\ &= \tilde{x}' * (x * x') && \text{(Assoziativität)} \\ &= \tilde{x}' * e && (x' \text{ ist ein inverses Element zu } x) \\ &= \tilde{x}' && (e \text{ neutrales Element}). \end{aligned}$$

(c) Es gilt

$$(y' * x') * (x * y) = y' * (x' * x) * y = y' * e * y = y' * y = e$$

und analog auch $(x * y) * (y' * x') = e$. Damit ist $y' * x'$ das inverse Element zu $x * y$.

(d) Die Gleichung $x' * x = x * x' = e$ besagt direkt, dass x das inverse Element zu x' ist. □

Wie wir in Beispiel 3.3 (a) und (b) gesehen haben, erlauben die reellen Zahlen *zwei* grundlegende Gruppenstrukturen: die Addition und (nach Herausnahme der 0) die Multiplikation. Diese beiden Strukturen sind jedoch nicht unabhängig voneinander, da sie durch das Distributivgesetz $(x+y)\cdot z = xz+yz$ für alle $x,y,z \in \mathbb{R}$ miteinander verbunden sind. Eine derartige Kombination zweier Gruppenstrukturen bezeichnet man als einen Körper.

Definition 3.5 (Körper)**.** Ein **Körper** ist eine Menge K zusammen mit zwei Verknüpfungen

$$+ : K \times K \to K \quad \text{(genannt } \textit{Addition}\text{)} \qquad \text{und} \qquad \cdot : K \times K \to K \quad \text{(genannt } \textit{Multiplikation}\text{)},$$

so dass die folgenden Eigenschaften (auch *Körperaxiome* genannt) gelten:

(a) $(K,+)$ ist eine abelsche Gruppe. Wir bezeichnen ihr neutrales Element mit 0 und das zu einem $x \in K$ inverse Element mit $-x$.

(b) $(K\backslash\{0\},\cdot)$ ist ebenfalls eine abelsche Gruppe. Wir bezeichnen ihr neutrales Element mit 1 und das zu einem $x \in K\backslash\{0\}$ inverse Element mit x^{-1}.

(c) (Distributivität) Für alle $x,y,z \in K$ gilt $(x+y)\cdot z = (x\cdot z)+(y\cdot z)$.

Mit dieser Definition wollen wir nun also axiomatisch voraussetzen:

$$\boxed{\mathbb{R} \text{ ist ein Körper.}}$$

Um Verwirrungen zu vermeiden, werden wir die beiden Verknüpfungen in einem Körper immer mit den Symbolen „+“ und „$\cdot$“ bezeichnen. Ebenso werden wir (wie ihr es natürlich gewohnt seid) vereinbaren, dass man den Punkt bei der Multiplikation auch weglassen darf und bei ungeklammerten Ausdrücken zuerst die Multiplikationen und dann die Additionen ausgeführt werden, so dass man also z. B. die Distributivität aus Definition 3.5 (c) auch als $(x+y)z = xz+yz$ schreiben kann.

Es ist jedoch wichtig zu verstehen, dass wir ab jetzt *nicht* mehr voraussetzen werden, dass Addition und Multiplikation in einem Körper wie z. B. $\mathbb{R}$ genau die Verknüpfungen sind, „an die man als Erstes denken würde“ – was auch immer das heißen mag. Stattdessen sind es einfach irgendwelche zwei Verknüpfungen, die die Eigenschaften aus Definition 3.5 haben. Unsere zukünftigen Beweise über Körper wie z. B. $\mathbb{R}$ müssen wir also ausschließlich auf diesen Eigenschaften aufbauen.

Dieser axiomatische Zugang hat zwei Vorteile:

- Zum einen wissen wir dadurch genau, welche Eigenschaften der Grundrechenarten auf den reellen Zahlen wir eigentlich voraussetzen. Es sollte schließlich klar sein, dass wir eine *exakte* Mathematik nicht auf einer *anschaulichen* Vorstellung von $\mathbb{R}$ aufbauen können. Solltet ihr euch also z. B. später einmal dafür interessieren, wie man die Existenz der reellen Zahlen beweisen kann, so wüsstet ihr dann genau, was eigentlich zu beweisen ist: nämlich die Existenz einer Menge mit genau den Eigenschaften, die wir jetzt axiomatisch voraussetzen.
- Zum anderen werdet ihr im Laufe eures Studiums noch viele weitere Körper kennenlernen, z. B. in Kapitel 6 den sehr wichtigen Körper der komplexen Zahlen. Alle Resultate, die nur auf den Körperaxiomen aufbauen, übertragen sich dann also sofort auf diese neuen Fälle, ohne dass man sich darüber noch einmal neu Gedanken machen muss.

Beispiel 3.6.

(a) Neben $\mathbb{R}$ ist auch $\mathbb{Q}$ (mit den gleichen Verknüpfungen wie auf $\mathbb{R}$) ein Körper. Die ganzen Zahlen $\mathbb{Z}$ bilden mit diesen Verknüpfungen jedoch keinen Körper, da $(\mathbb{Z}\backslash\{0\},\cdot)$ nach Beispiel 3.3 (b) keine Gruppe ist. Ebenso ist $\mathbb{N}$ mit diesen Verknüpfungen kein Körper, da hier nach Beispiel 3.3 (a) bereits die Addition keine Gruppenstruktur liefert.

(b) Hier ist ein Beispiel für einen Körper, der sich ganz anders verhält als $\mathbb{R}$ und $\mathbb{Q}$. Wir definieren auf der Menge $K = \{g,u\}$ zwei Verknüpfungen durch die folgenden Tabellen.

$+$	g	u
g	g	u
u	u	g

und

$\cdot$	g	u
g	g	g
u	g	u

Die Idee dieser Verknüpfungen ist, dass g für gerade und u für ungerade ganze Zahlen steht. So haben wir in der Tabelle z. B. $g+u$ als u definiert, weil die Addition einer geraden und einer ungeraden Zahl eine ungerade Zahl ergibt.

Man kann zeigen, dass K mit diesen beiden Verknüpfungen einen Körper bildet. Er wird in der Literatur mit $\mathbb{Z}_2$ bezeichnet, da seine Elemente die Reste ganzer Zahlen bei Division durch 2 beschreiben. Um zu beweisen, dass $\mathbb{Z}_2$ ein Körper ist, könnte man z. B. einfach die geforderten Eigenschaften für alle Elemente – es gibt ja nur zwei – explizit nachprüfen. In der Vorlesung „Algebraische Strukturen" zeigt man allerdings, dass man die Körperaxiome hier auch viel eleganter direkt aus den Eigenschaften von $\mathbb{Z}$ folgern kann [G, Satz 7.10]. Wir wollen uns hier damit begnügen, die neutralen und inversen Elemente anzugeben:

- Das additive neutrale Element ist g, wie man leicht aus der Tabelle abliest. Im Sinne der Notationen von Definition 3.5 ist also $0 = g$. Wegen $g+g = u+u = g = 0$ sind die additiven inversen Elemente $-g = g$ und $-u = u$. Dies stimmt natürlich auch mit der Interpretation als gerade und ungerade Zahlen überein, da das Negative von einer geraden bzw. ungeraden Zahl ebenfalls wieder gerade bzw. ungerade ist.
- Das multiplikative neutrale Element in $\mathbb{Z}_2\backslash\{0\}$ ist u – in der Tat ist es ja auch das einzige Element in $\mathbb{Z}_2\backslash\{0\}$. Gemäß der Notation von Definition 3.5 ist also $1 = u$.

Beachte, dass in diesem Körper $\mathbb{Z}_2$ die Gleichung $1+1 = u+u = g = 0$ gilt. Die Körperaxiome lassen es also zu, dass man bei fortgesetzter Addition der 1 irgendwann wieder zur 0 zurück kommt. Wir werden in dieser Vorlesung nicht viel mit dem Körper $\mathbb{Z}_2$ zu tun haben – wir haben ihn hier nur als Beispiel dafür angegeben, dass die Körperaxiome noch weit davon entfernt sind, die rationalen oder reellen Zahlen eindeutig zu charakterisieren.

Anschaulich kann man die Körperaxiome so interpretieren, dass ein Körper eine Menge ist, auf der „die vier Grundrechenarten existieren und die erwarteten Eigenschaften haben". Wir wollen nun noch ein paar weitere dieser erwarteten Eigenschaften zeigen, die bereits aus den Körperaxiomen folgen und die wir dann beim Rechnen z. B. in $\mathbb{R}$ natürlich ständig benutzen werden.

Bemerkung 3.7. Es seien K ein Körper und $x, y \in K$.

(a) Wenden wir Lemma 3.4 (c) und (d) auf die (kommutative) Addition und Multiplikation an, so sehen wir sofort, dass

$$-(x+y) = (-x)+(-y) \quad \text{und} \quad -(-x) = x$$

sowie für $x, y \neq 0$

$$(xy)^{-1} = x^{-1}\cdot y^{-1} \quad \text{und} \quad (x^{-1})^{-1} = x.$$

(b) Etwas versteckt in Definition 3.5 steht in Teil (b) u. a. die Aussage, dass die Multiplikation überhaupt eine Verknüpfung auf $K\backslash\{0\}$ ist, also dass für $x, y \in K\backslash\{0\}$ auch $xy \in K\backslash\{0\}$ gilt. Äquivalent dazu bedeutet das:

$$\text{Ist } xy = 0, \text{ so gilt } x = 0 \text{ oder } y = 0.$$

Lemma 3.8 (Eigenschaften von Körpern)**.** *In jedem Körper K gilt für alle $x, y \in K$:*

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

(b) $x\cdot(-y) = -(xy)$.

(c) *Für $x \neq 0$ ist $-(x^{-1}) = (-x)^{-1}$.*

Beweis.

(a) Es gilt

$$\begin{aligned} 0\cdot x &= (0+0)\cdot x && \text{(0 ist additives neutrales Element)} \\ &= 0\cdot x + 0\cdot x, && \text{(Distributivität)} \end{aligned}$$

woraus durch Addition des additiven Inversen von $0\cdot x$ auf beiden Seiten die gewünschte Gleichung $0 = 0\cdot x$ folgt.

(b) Es ist

$$\begin{aligned} x\cdot(-y)+xy &= x\cdot(-y+y) && \text{(Distributivität)}\\ &= x\cdot 0 && (-y \text{ ist additives Inverses zu } y)\\ &= 0 && \text{(nach (a))}, \end{aligned}$$

daher ist $x\cdot(-y)$ das additive Inverse zu xy, d. h. es gilt $x\cdot(-y) = -(xy)$.

(c) Doppeltes Anwenden von (b), einmal für den linken und einmal für den rechten Faktor, ergibt

$$(-(x^{-1}))\cdot(-x) = -(x^{-1}\cdot(-x)) = -(-(x^{-1}\cdot x)) = -(-1) \overset{3.7(a)}{=} 1.$$

Also ist $-(x^{-1})$ das multiplikative Inverse zu $-x$, d. h. es ist $-(x^{-1}) = (-x)^{-1}$. □

Notation 3.9. In einem Körper K verwendet man üblicherweise die folgenden Notationen, von denen euch die meisten sicher bekannt sein werden:

(a) Für $x,y \in K$ setzt man $x-y := x+(-y)$. Ist $y \neq 0$, so setzt man $\frac{x}{y} := x\cdot y^{-1}$.

(b) Für $x \in K$ und $n \in \mathbb{N}$ definiert man die n-te **Potenz** von x als

$$x^n := \underbrace{x\cdot\ \cdots\ \cdot x}_{n\text{-mal}},$$

wobei dieser Ausdruck für $n = 0$ als $x^0 := 1$ zu verstehen ist. Insbesondere legen wir also auch $0^0 := 1$ fest. Beachte, dass aus dieser Definition (und der Kommutativität der Multiplikation) unmittelbar die Potenzrechenregeln

$$x^m\cdot x^n = x^{m+n} \quad \text{und} \quad (xy)^n = x^n\cdot y^n$$

für alle $x,y \in K$ folgen. Ist $x \neq 0$, so definiert man zusätzlich Potenzen mit negativen ganzzahligen Exponenten durch $x^{-n} := (x^{-1})^n$.

Beachte, dass auch in einem beliebigen Körper K die Exponenten einer Potenz stets *ganze Zahlen* sind und keine Elemente aus K. Eine Potenz x^y für $x,y \in K$ lässt sich im Allgemeinen nicht definieren (auch wenn dies für $K = \mathbb{R}$ in vielen Fällen möglich ist, siehe Definition 9.7).

(c) Manchmal möchte man mehrere Elemente $x_m, x_{m+1}, x_{m+2}, \dots, x_n$ in einem Körper (oder allgemeiner in einer additiv geschriebenen abelschen Gruppe) aufsummieren, die durch eine ganzzahlige Laufvariable indiziert werden, die von einem $m \in \mathbb{Z}$ bis zu einem $n \in \mathbb{Z}$ (mit $n \geq m$) läuft. Man schreibt dies dann als

$$\sum_{i=m}^{n} x_i := x_m + x_{m+1} + x_{m+2} + \dots + x_n$$

(also mit einem großen griechischen Sigma, das an das Wort „Summe“ erinnern soll). So steht z. B.

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + 3^2 + \dots + n^2 \qquad (*)$$

für die Summe aller Quadratzahlen bis n^2. Natürlich ist der Name der Laufvariablen dabei egal, und der Ausdruck $(*)$ hängt nicht von einem i ab (wie man auf der rechten Seite ja auch sieht). Außerdem kann man die Laufvariable verschieben, ohne den eigentlichen Ausdruck zu ändern: Setzt man z. B. $i = j+1$, also $j = i-1$, in der obigen Summe $(*)$, so läuft j dort von 0 bis $n-1$, wenn i von 1 bis n läuft, und wir können dieselbe Summe auch schreiben als

$$\sum_{j=0}^{n-1} (j+1)^2 = 1^2 + 2^2 + 3^2 + \dots + n^2.$$

Natürlich kann man diesen Ausdruck nun auch wieder genauso gut mit dem Buchstaben i statt j als $\sum_{i=0}^{n-1}(i+1)^2$ schreiben, oder den Index um mehr als 1 in die eine oder andere Richtung verschieben. Also:

Der Wert einer Summe ändert sich nicht, wenn man zur Laufvariablen im zu summierenden Ausdruck eine Konstante addiert, und dafür von der Ober- und Untergrenze der Summe diese Konstante abzieht.

Wir sagen in diesem Fall, dass die neue Darstellung der Summe durch eine **Indexverschiebung** (im Beispiel oben $i \mapsto i+1$) aus der alten hervorgeht.

Analog schreibt man

$$\prod_{i=m}^{n} x_i := x_m \cdot x_{m+1} \cdot x_{m+2} \cdot \cdots \cdot x_n$$

(mit einem großen griechischen Pi für das Produkt), wenn man die Körperelemente multiplizieren statt addieren möchte. Ist schließlich die Obergrenze einer Summe oder eines Produkts kleiner als die Untergrenze (man spricht dann von der **leeren Summe** bzw. dem **leeren Produkt**), so definiert man dies als

$$\sum_{i=m}^{n} x_i := 0 \qquad \text{und} \qquad \prod_{i=m}^{n} x_i := 1 \qquad \text{für } n < m,$$

also als das additive bzw. multiplikative neutrale Element.

(d) Ist n eine natürliche Zahl, so fasst man diese oft auch als das Element

$$\sum_{i=1}^{n} 1 = \underbrace{1 + \cdots + 1}_{n\text{-mal}}$$

von K auf. Im Fall $K = \mathbb{R}$ ist dies dann einfach die natürliche Zahl $n \in \mathbb{N} \subset \mathbb{R}$ und liefert somit keine neue Notation, aber z. B. in $K = \mathbb{Z}_2$ aus Beispiel 3.6 (b) ist $2 = 1 + 1 = 0$.

Aufgabe 3.10. Zeige, dass in jedem Körper K die üblichen Rechenregeln

$$\frac{x}{y} + \frac{z}{w} = \frac{xw + yz}{yw} \qquad \text{und} \qquad \frac{x}{y} \cdot \frac{z}{w} = \frac{xz}{yw}$$

für Brüche gelten, wobei $x, y, z, w \in K$ mit $y, w \neq 0$.

Aufgabe 3.11. Es sei $a \in \mathbb{Q}$ fest gegeben. Wir definieren auf $\mathbb{Q}^2$ eine Addition und Multiplikation durch

$$(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2) \quad \text{und} \quad (x_1, x_2) \cdot (y_1, y_2) := (x_1 y_1 + a x_2 y_2, x_1 y_2 + x_2 y_1).$$

Man prüft leicht durch explizite Rechnung nach, dass $\mathbb{Q}^2$ mit dieser Addition eine kommutative Gruppe mit neutralem Element $(0,0)$ ist, dass auch die Multiplikation kommutativ ist, und dass diese beiden Operationen das Distributivgesetz erfüllen – dies soll in dieser Aufgabe nicht bewiesen werden. Man zeige stattdessen:

(a) Die Multiplikation ist assoziativ und besitzt ein neutrales Element.

(b) $(\mathbb{Q}^2, +, \cdot)$ ist genau dann ein Körper, wenn a kein Quadrat in $\mathbb{Q}$ ist, also wenn es kein $b \in \mathbb{Q}$ gibt mit $a = b^2$.

Aufgabe 3.12. Zu einem Körper K und einer Menge M mit $|M| \geq 2$ sei

$$V = \{f \colon f \text{ ist eine Abbildung von } M \text{ nach } K\}$$

die Menge aller reellwertigen Funktionen auf M. Für $f, g \in V$ definieren wir die Addition $f + g$ und Multiplikation $f \cdot g$ dieser Funktionen punktweise durch

$$f + g \colon M \to K,\ x \mapsto f(x) + g(x) \qquad \text{und} \qquad f \cdot g \colon M \to K,\ x \mapsto f(x) \cdot g(x).$$

(a) Zeige, dass V mit dieser Addition eine abelsche Gruppe ist.

(b) Ist V mit dieser Addition und Multiplikation ein Körper?

04

3.B Vollständige Induktion

Häufig möchte man in der Mathematik Aussagen beweisen, die von einer natürlichen Zahl abhängen – z. B. bei Formeln, die Summen oder Produkte wie in Notation 3.9 mit variablen Unter- oder Obergrenzen beinhalten. Die einfachste und bekannteste solcher Aussagen ist vermutlich die folgende Formel für die Summe aller natürlichen Zahlen bis zu einer gegebenen Obergrenze.

Satz 3.13 (Summenformel von Gauß). *Für alle $n \in \mathbb{N}$ gilt*

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

Beispiel 3.14. Für $n = 5$ ist z. B.

$$\sum_{k=1}^{5} k = 1+2+3+4+5 = 15 = \frac{5 \cdot 6}{2}.$$

Um derartige Aussagen zu beweisen, ist oft das Beweisverfahren der **(vollständigen) Induktion** nützlich, das wir jetzt einführen wollen.

Angenommen, wir wollen (wie z. B. in Satz 3.13) eine Aussage $A(n)$ für alle $n \in \mathbb{N}$ beweisen. Dann können wir dies tun, indem wir die folgenden beiden Dinge zeigen:

(a) (**Induktionsanfang**) Die Aussage $A(0)$ ist wahr.

(b) (**Induktionsschritt**) Für alle $n \in \mathbb{N}$ gilt $A(n) \Rightarrow A(n+1)$, d. h. *wenn* die Aussage $A(n)$ für ein gegebenes $n \in \mathbb{N}$ gilt (die „Induktionsannahme“ bzw. „Induktionsvoraussetzung“), *dann* gilt auch die Aussage $A(n+1)$ (der „Induktionsschluss“).

Haben wir diese beiden Dinge gezeigt, so folgt daraus nämlich die Gültigkeit von $A(n)$ für alle $n \in \mathbb{N}$: Die Aussage $A(0)$ haben wir mit dem Induktionsanfang gezeigt, und durch fortgesetztes Anwenden des Induktionsschritts $A(n) \Rightarrow A(n+1)$ für $n = 0, 1, 2, \dots$ erhalten wir dann auch

$$A(0) \Rightarrow A(1) \Rightarrow A(2) \Rightarrow A(3) \Rightarrow \cdots,$$

also die Gültigkeit von $A(n)$ für alle $n \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$.

Derartige Induktionsbeweise sind immer dann sinnvoll, wenn die Aussagen $A(n)$ und $A(n+1)$ „ähnlich genug“ sind, so dass es beim Beweis von $A(n+1)$ hilft, die Gültigkeit von $A(n)$ voraussetzen zu dürfen.

Mit diesem Verfahren können wir nun die Summenformel aus Satz 3.13 beweisen:

Beweis von Satz 3.13. Wir zeigen die Formel mit Induktion über n.

Induktionsanfang ($n = 0$): Für $n = 0$ stimmen die beiden Seiten der zu zeigenden Gleichung überein, denn es ist

$$\sum_{k=1}^{0} k = 0 = \frac{0 \cdot (0+1)}{2}.$$

Induktionsschritt ($n \to n+1$): Als Induktionsvoraussetzung nehmen wir an, dass die zu beweisende Formel für ein gegebenes $n \in \mathbb{N}$ richtig ist, d. h. dass

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

gilt. (Beachte, dass wir diese Gleichung nicht für alle $n \in \mathbb{N}$ voraussetzen – dies wäre ja schon die gesamte zu zeigende Aussage!) Wir müssen zeigen, dass die entsprechende Gleichung dann auch für $n+1$ gilt, also dass

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

Dies ergibt sich nun leicht aus der folgenden Rechnung:

$$\begin{aligned}
\sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) && \text{(Abspalten des letzten Summanden für } k = n+1\text{)} \\
&= \frac{n(n+1)}{2} + (n+1) && \text{(nach Induktionsvoraussetzung)} \\
&= \frac{n^2+n+2n+2}{2} \\
&= \frac{(n+1)(n+2)}{2}.
\end{aligned}$$

Damit ist der Satz mit vollständiger Induktion bewiesen. □

Bemerkung 3.15. Offensichtlich erlaubt das Beweisverfahren der vollständigen Induktion die folgenden Abwandlungen:

(a) Im Induktionsschritt kann man, wenn es hilfreich ist, beim Beweis der Aussage $A(n+1)$ nicht nur die direkt vorangegangene Aussage $A(n)$, sondern *alle bereits gezeigten Aussagen* $A(0), A(1), \ldots, A(n)$ voraussetzen.

(b) Möchte man die Aussage $A(n)$ nicht für alle $n \in \mathbb{N}$, sondern für alle $n \in \mathbb{Z}$ ab einem gewissen Startwert $n_0 \in \mathbb{Z}$ zeigen, so kann man als Induktionsanfang die Aussage $A(n_0)$ zeigen, und im Induktionsschritt dann die Folgerung $A(n) \Rightarrow A(n+1)$ für alle $n \geq n_0$.

Aufgabe 3.16. Zeige für alle $n \in \mathbb{N}$:

$$\text{(a)} \quad \sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}, \qquad \text{(b)} \quad \prod_{k=1}^{n} \left(1 + \frac{1}{n+k}\right) = 2 - \frac{1}{n+1}.$$

3.C Polynomfunktionen

Als erste Anwendung der Körpereigenschaften wollen wir zum Abschluss dieses Kapitels die euch sicher schon aus der Schule bekannten Polynomfunktionen behandeln – also die Funktionen, die sich aus den grundlegenden Körperoperationen Addition und Multiplikation bilden lassen.

Definition 3.17 (Polynomfunktionen und Nullstellen)**.** Es seien D eine Teilmenge eines Körpers K und $f\colon D \to K$ eine Funktion.

(a) Ist f von der Form

$$f(x) = \sum_{k=0}^{n} a_k x^k = a_n x^n + \cdots + a_1 x + a_0$$

für gewisse $a_0, \ldots, a_n \in K$, so sagt man, dass f eine **Polynomfunktion** ist. Ist n dabei so gewählt, dass der erste Koeffizient a_n ungleich Null ist, so heißt f eine Polynomfunktion vom **Grad** n und mit **Leitkoeffizient** a_n. Ist der Leitkoeffizient 1, so heißt f eine **normierte** Polynomfunktion.

Sind in der obigen Darstellung alle Koeffizienten $a_0, \ldots, a_n$ gleich 0 (und ist f damit die Nullfunktion), so nennen wir f formal eine Polynomfunktion vom Grad $-\infty$. In diesem Fall hat f keinen Leitkoeffizienten.

(b) Ist $x_0 \in D$ mit $f(x_0) = 0$, so nennt man x_0 eine **Nullstelle** von f.

Das Besondere an Nullstellen von Polynomfunktionen ist, dass man sie wie im folgenden Satz als Linearfaktoren abspalten kann.

Satz 3.18 (Abspalten von Nullstellen in Polynomfunktionen)**.** *Es seien K ein Körper, $D \subset K$ und $f\colon D \to K$ eine Polynomfunktion vom Grad $n \in \mathbb{N}$.*

(a) *Ist $x_0 \in D$ eine Nullstelle von f, so gibt es eine Polynomfunktion $g\colon D \to K$ vom Grad $n-1$ mit $f(x) = (x - x_0)\, g(x)$ für alle $x \in D$ (d. h. man kann „den Linearfaktor $x - x_0$ abspalten“).*

(b) *Die Funktion f hat höchstens n Nullstellen.*

Beweis. Wir zeigen die beiden Aussagen mit Induktion über n. Der Beweis von (a) ist dabei konstruktiv, d. h. er gibt auch ein Verfahren an, wie g berechnet werden kann (siehe Beispiel 3.19).

Der Induktionsanfang für $n = 0$ ist trivial, denn f ist dann eine Konstante ungleich 0 und hat somit keine Nullstellen. Für den Induktionsschluss nehmen wir an, dass die Aussagen des Satzes bis zu einem gegebenen n gelten, und betrachten $f\colon D \to K$, $x \mapsto a_{n+1}x^{n+1} + \dots + a_1 x + a_0$ vom Grad $n+1$, also mit $a_{n+1} \neq 0$.

(a) Wir definieren eine Polynomfunktion $\tilde{f}\colon D \to K$ durch

$$\begin{aligned}\tilde{f}(x) &:= f(x) - a_{n+1}x^n(x - x_0)\\ &= a_{n+1}x_0x^n + a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0\end{aligned}$$

für alle x. Ist $\tilde{f}$ die Nullfunktion, so sind wir fertig, da dann ja $f(x) = a_{n+1}x^n(x - x_0)$ für alle $x \in D$ gilt. Andernfalls ist $\tilde{f}$ nach Konstruktion eine Polynomfunktion von einem Grad kleiner als $n+1$ (der x^{n+1}-Term hebt sich ja gerade heraus), die immer noch die Nullstelle x_0 hat. Nach Induktionsvoraussetzung gibt es dann also eine Polynomfunktion $\tilde{g}\colon D \to K$ vom Grad kleiner als n mit $\tilde{f}(x) = (x - x_0)\,\tilde{g}(x)$ für alle $x \in D$, und somit ist

$$\begin{aligned}f(x) &= a_{n+1}x^n(x - x_0) + \tilde{f}(x)\\ &= (x - x_0) \cdot \underbrace{(a_{n+1}x^n + \tilde{g}(x))}_{=:g(x)}\end{aligned}$$

für alle $x \in D$, wobei g offensichtlich vom Grad n ist.

(b) Hat f keine Nullstelle, so sind wir fertig. Andernfalls wählen wir eine Nullstelle x_0 von f und schreiben $f(x) = (x - x_0)\,g(x)$ für alle $x \in D$ wie in (a) mit einer Polynomfunktion g vom Grad n. Nach Induktionsvoraussetzung hat g höchstens n Nullstellen, und nach Bemerkung 3.7 (b) sind die Nullstellen von f genau x_0 zusammen mit den Nullstellen von g. Also hat f höchstens $n+1$ Nullstellen. Damit ist die Behauptung mit Induktion bewiesen. □

Beispiel 3.19 (Polynomdivision)**.** Das Verfahren aus dem Beweis von Satz 3.18 (a) wird als *Polynomdivision* [G, Satz 10.19] bezeichnet: Man subtrahiert fortlaufend geeignete Vielfache von $x - x_0$ von f, so dass sich der jeweils höchste Term von f weghebt, und sammelt die dabei verwendeten Faktoren in g. Das folgende Schema, das genauso aussieht wie eine normale schriftliche Division, verdeutlicht dieses Verfahren am Beispiel der Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2 + 3x - 4$ mit Nullstelle $x_0 = 1$, die wir als $f(x) = (x-1)\,g(x)$ für alle $x \in \mathbb{R}$ schreiben wollen. Das Ergebnis ist in diesem Fall $g\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 4$.

$$\begin{array}{lrcl}
f \longrightarrow & (x^2 + 3x - 4) & : (x-1) & = x + 4 \longleftarrow g \\
 & -\ (x^2 - \ \ x) & & \cdot x \\
\hline
\tilde{f} \longrightarrow & 4x - 4 & & \\
 & -\ \ (4x - 4) & & \cdot 4 \\
\hline
 & 0 & &
\end{array}$$

Bemerkung 3.20. Satz 3.18 liefert uns zwar die neue Funktion nach dem Abspalten des Linearfaktors, er sagt uns hingegen nicht, wie wir überhaupt erst einmal eine Nullstelle von f finden können, oder ob es überhaupt Nullstellen gibt (die reelle Polynomfunktion $f(x) = x^2 + 1$ hat ja z. B. keine Nullstellen). In der Tat gibt es im Allgemeinen kein Verfahren, wie man Nullstellen von Polynomfunktionen exakt berechnen kann! Genauer gesagt gilt:

- Für Polynomfunktionen vom Grad höchstens 4 gibt es explizite Verfahren zur exakten Bestimmung der Nullstellen (für Grad 1 ist das klar, für Grad 2 gibt es die bekannte „p-q-Formel“ bzw. die quadratische Ergänzung, und für Grad 3 bzw. 4 sind die Formeln so lang, dass man im Allgemeinen nicht mehr mit ihnen arbeiten möchte).

- Für Polynomfunktionen vom Grad größer als 4 kann man beweisen(!), dass es keine derartigen Verfahren zur exakten Bestimmung der Nullstellen geben kann (das beweist man z. B. in der Vorlesung „Einführung in die Algebra“, die ihr im nächsten Studienjahr hören könnt). Aber:
- Für reelle Polynomfunktionen beliebigen Grades gibt es zumindest numerische Verfahren, die die Nullstellen (mit beliebiger Genauigkeit) näherungsweise bestimmen können.

Zum Schluss wollen wir nun noch zwei wichtige Konzepte für Polynomfunktionen untersuchen, die ihr beide im reellen Fall vielleicht schon aus der Schule kennt: den sogenannten Koeffizientenvergleich (also dass eine Polynomfunktion eindeutig ihre Koeffizienten bestimmt) und die Vielfachheit von Nullstellen. Es stellt sich jedoch heraus, dass man hierfür im allgemeinen Fall die Voraussetzung benötigt, dass die Definitionsmenge der betrachteten Funktionen unendlich viele Elemente besitzt.

Lemma 3.21 (**Koeffizientenvergleich**). *Es seien K ein Körper, $D \subset K$ mit $|D| = \infty$, und $f\colon D \to K$ eine Polynomfunktion mit zwei Darstellungen*

$$f(x) = a_n x^n + \cdots + a_1 x + a_0 = b_n x^n + \cdots + b_1 x + b_0 \quad \text{für alle } x \in D$$

für gewisse $a_0, \ldots, a_n, b_0, \ldots, b_n \in K$. (Beachte, dass wir dabei in beiden Darstellungen den gleichen höchsten Exponenten n wählen können, da wir nicht $a_n \neq 0$ und $b_n \neq 0$ vorausgesetzt haben.)

Dann gilt bereits $a_i = b_i$ für alle $i = 0, \ldots, n$. Es ist also nicht möglich, „eine Polynomfunktion auf zwei verschiedene Arten hinzuschreiben“.

Beweis. Nach Voraussetzung ist die Polynomfunktion

$$D \to K,\ x \mapsto (a_n - b_n)x^n + \cdots + (a_1 - b_1)x + (a_0 - b_0) = f(x) - f(x) = 0$$

die Nullfunktion auf D. Da sie damit wegen $|D| = \infty$ unendlich viele Nullstellen besitzt, muss sie nach Satz 3.18 (b) vom Grad $-\infty$ sein. Also sind alle Koeffizienten dieser Polynomfunktion gleich 0, d. h. es ist $a_i = b_i$ für alle $i = 0, \ldots, n$. □

Bemerkung und Notation 3.22 (Polynome). Die Voraussetzung $|D| = \infty$ in Lemma 3.21 ist wirklich notwendig: So sind für $D = K = \mathbb{Z}_2$ wie in Beispiel 3.6 (b) z. B. $x \mapsto x$ und $x \mapsto x^2$ dieselbe Funktion, da sie beide 0 auf 0 und 1 auf 1 abbilden und in $\mathbb{Z}_2$ keine weiteren Elemente existieren.

In der Literatur bezeichnet man einen formalen Ausdruck der Form $a_n x^n + \cdots + a_1 x + a_0$ mit $a_0, \ldots, a_n \in K$ als ein **Polynom** über K [G, Kapitel 9]. Jedes solche Polynom bestimmt natürlich eine Polynomfunktion von jeder Teilmenge D von K nach K, allerdings können verschiedene Polynome wie im eben angegebenen Beispiel durchaus dieselbe Polynomfunktion definieren: Über $\mathbb{Z}_2$ sind x und x^2 verschiedene Polynome, sie bestimmen aber dieselbe Polynomfunktion.

Mit dieser Notation ist die Aussage von Lemma 3.21 also, dass Polynome und Polynomfunktionen im Fall von unendlichen Definitionsmengen dasselbe sind. Da wir Polynomfunktionen im Folgenden in der Regel nur in diesem Fall unendlicher Definitionsmengen benötigen, werden wir die Begriffe Polynom und Polynomfunktion oft synonym verwenden. Wegen der Eindeutigkeit der Koeffizienten sind dann auch der Grad (und der Leitkoeffizient) einer Polynomfunktion f wie in Definition 3.17 (a) eindeutig bestimmt. Wir können daher eine Bezeichnung dafür einführen:

Definition 3.23 (Grad eines Polynoms). Wir bezeichnen den **Grad** einer Polynomfunktion f (mit unendlicher Definitionsmenge) mit $\deg f \in \mathbb{N} \cup \{-\infty\}$ (vom englischen Wort „degree“).

Satz und Definition 3.24 (Vielfachheit von Nullstellen). *Es seien K ein Körper, $D \subset K$ mit $|D| = \infty$ und $f\colon D \to K$ eine Polynomfunktion, die nicht die Nullfunktion ist. Dann lässt sich f (bis auf die Reihenfolge der Faktoren) eindeutig als*

$$f(x) = g(x) \cdot (x - x_1)^{a_1} \cdot \cdots \cdot (x - x_k)^{a_k} \qquad \text{für alle } x \in D$$

*schreiben, wobei $x_1, \ldots, x_k \in D$ die verschiedenen Nullstellen von f sind, $a_1, \ldots, a_k \in \mathbb{N}_{>0}$ gilt, und g eine Polynomfunktion ohne Nullstellen in D ist. In dieser Darstellung nennt man a_i für $i = 1, \ldots, k$ die **Vielfachheit** der Nullstelle x_i (in der Literatur sind auch die Bezeichnungen **Ordnung** und **Multiplizität** der Nullstelle üblich).*

Beweis. Die Existenz einer solchen Darstellung ergibt sich sofort durch fortgesetztes Abspalten von Nullstellen gemäß Satz 3.18 (a). Wir zeigen nun die Eindeutigkeit mit Induktion über den Grad der Polynomfunktion. Dabei ist der Induktionsanfang für Grad 0 trivial, denn dann hat f keine Nullstellen, und es ist zwangsläufig $k=0$ und $g=f$.

Für den Induktionsschritt bemerken wir zuerst, dass $x_1,\dots,x_k$ natürlich in jedem Fall als die Nullstellen von f eindeutig bestimmt sind. Wir nehmen also an, dass wir zwei Darstellungen

$$f(x) = g(x)\cdot(x-x_1)^{a_1}\cdot\ldots\cdot(x-x_k)^{a_k} = h(x)\cdot(x-x_1)^{b_1}\cdot\ldots\cdot(x-x_k)^{b_k}$$

wie in der Behauptung des Satzes haben. Im nullstellenfreien Fall $k=0$ sind wir natürlich bereits fertig. Andernfalls liefert Division durch $x-x_1$ für alle $x\in D\backslash\{x_1\}$ (wir müssen x_1 hier herausnehmen, da wir sonst durch 0 teilen würden!)

$$\begin{aligned} &g(x)\cdot(x-x_1)^{a_1-1}\cdot(x-x_2)^{a_2}\cdot\ldots\cdot(x-x_k)^{a_k} \\ =\;&h(x)\cdot(x-x_1)^{b_1-1}\cdot(x-x_2)^{b_2}\cdot\ldots\cdot(x-x_k)^{b_k}. \end{aligned} \tag{$*$}$$

Wir haben also wieder zwei Darstellungen einer Polynomfunktion auf der immer noch unendlichen Menge $D\backslash\{x_1\}$. Da der Grad dieser Polynomfunktion nun um 1 kleiner ist als der von f, müssen diese Darstellungen aber nach der Induktionsvoraussetzung bereits übereinstimmen. Also gilt $g=h$, $a_1-1=b_1-1$, $a_2=b_2$, …, $a_k=b_k$, und damit stimmen auch die beiden ursprünglichen Darstellungen von f überein. □

Aufgabe 3.25. Bestimme die Nullstellen des reellen Polynoms $x\mapsto x^4+3x^3-4x$ und ihre Vielfachheiten.

05

Grundlagen der Mathematik 1: Analysis

4. Weitere Eigenschaften der reellen Zahlen

Wir haben uns nun das elementare Handwerkszeug für diese Vorlesung erarbeitet und beginnen jetzt mit dem Studium der eindimensionalen Analysis. Wer sich auch (bzw. zurzeit nur) für die lineare Algebra interessiert, kann ab diesem Zeitpunkt auch zusätzlich (bzw. ausschließlich) den Teil „Grundlagen der Mathematik 1: Lineare Algebra“ in den Kapiteln 13 bis 18 durcharbeiten.

Hier in diesem Kapitel wollen wir zunächst nach den gerade behandelten elementaren Eigenschaften der reellen Zahlen noch ein paar weitere untersuchen, die vor allem in der Analysis nützlich sind. Unter anderem wird sich daraus am Ende dieses Kapitels auch eine vollständige Charakterisierung der reellen Zahlen ergeben.

4.A Potenzen in Körpern

Wir beginnen mit zwei weiteren oft vorkommenden Formeln zu Potenzen, die sich allein aus den Eigenschaften aus Abschnitt 3.A herleiten lassen und somit nicht nur in den reellen Zahlen, sondern sogar in beliebigen Körpern gelten. Die erste – die sogenannte (endliche) geometrische Reihe – ermöglicht es, den Wert einer Summe fortlaufender Potenzen eines Körperelements explizit zu berechnen.

Satz 4.1 (**Endliche geometrische Reihe**). *Es seien K ein Körper, $q \in K\backslash\{1\}$ und $n \in \mathbb{N}$. Dann gilt*

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

Beweis. Der Beweis ist sehr einfach und hilft auch dabei, sich die Formel zu merken: Multiplizieren wir die gesuchte Summe $\sum_{k=0}^{n} q^k = 1+q+q^2+\cdots+q^n$ mit $1-q$, so heben sich fast alle Terme weg und wir erhalten sofort das gewünschte Ergebnis: Es ist

$$\begin{aligned}(1+q+q^2+\cdots+q^n)\cdot(1-q) &= 1+q+q^2+\cdots+q^n \\ &\quad -q-q^2-\cdots-q^n-q^{n+1} \\ &= 1-q^{n+1},\end{aligned}$$

und damit für $q \neq 1$ wie behauptet

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

Beispiel 4.2. In $\mathbb{R}$ ist z. B.

$$1+2+4+8+16 = \sum_{k=0}^{4} 2^k \overset{4.1}{=} \frac{1-2^5}{1-2} = \frac{-31}{-1} = 31.$$

Die zweite Formel, die wir hier behandeln wollen, ist die sogenannte binomische Formel, die eine Verallgemeinerung der aus der Schule bekannten Formel $(x+y)^2 = x^2+2xy+y^2$ auf höhere Exponenten darstellt. Dazu benötigen wir zunächst die folgende Definition.

Definition 4.3 (Fakultät und Binomialkoeffizienten)**.**

(a) Für $n \in \mathbb{N}$ setzen wir

$$n! := \prod_{i=1}^{n} i = 1\cdot 2\cdot\cdots\cdot n \quad \in \mathbb{N} \qquad \text{(gesprochen „}n\text{-}\textbf{Fakultät}\text{“)},$$

wobei $0!$ gemäß Notation 3.9 (c) als 1 zu verstehen ist.

(b) Für $k,n \in \mathbb{N}$ mit $k \le n$ definiert man ferner die **Binomialkoeffizienten**

$$\binom{n}{k} := \frac{n!}{k!\,(n-k)!} \qquad \text{(gesprochen „}n\text{ über }k\text{“)},$$

die so genannt werden, weil sie in der binomischen Formel in Satz 4.7 auftreten. Sie sind aufgrund der Definition zunächst positive rationale Zahlen; wir werden aber in Bemerkung 4.6 sehen, dass sie sogar natürliche Zahlen sind.

Bemerkung 4.4.

(a) Offensichtlich ist $\binom{n}{0} = \binom{n}{n} = 1$ und $\binom{n}{k} = \binom{n}{n-k}$ für alle $k,n \in \mathbb{N}$ mit $k \le n$.

(b) Man kann im definierenden Ausdruck für $\binom{n}{k}$ die Faktoren von 1 bis $n-k$ kürzen und erhält damit die alternative Darstellung der Binomialkoeffizienten

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{1\cdot\dots\cdot(n-k)\cdot(n-k+1)\cdot\dots\cdot n}{(1\cdot\dots\cdot k)\cdot(1\cdot\dots\cdot(n-k))} = \frac{(n-k+1)\cdot\dots\cdot n}{1\cdot\dots\cdot k},$$

d. h. $\binom{n}{k}$ ist ein Bruch mit k Zahlen im Zähler und k Zahlen im Nenner, wobei man „im Zähler von n nach unten und im Nenner von 1 nach oben zählt“. So ist z. B. $\binom{n}{1} = \frac{n}{1} = n$ und $\binom{n}{2} = \frac{n(n-1)}{1\cdot 2} = \frac{n(n-1)}{2}$.

Die wichtigste Identität zwischen den Binomialkoeffizienten ist die folgende:

Lemma 4.5. *Für alle $n,k \in \mathbb{N}$ mit $1 \le k \le n$ gilt $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$.*

Beweis. Dies ergibt sich durch einfaches Nachrechnen mit der Darstellung aus Bemerkung 4.4 (b):

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

□

Bemerkung 4.6 (Pascalsches Dreieck)**.** Man kann die Binomialkoeffizienten $\binom{n}{k}$ wie folgt in einem dreieckigen Schema, dem sogenannten **Pascalschen Dreieck**, anordnen.

$$\begin{tikzcd}[column sep=small]
 & \binom{0}{0} & & & & & & 1 & & & & \\
 & \binom{1}{0} & \binom{1}{1} & & & & & 1 & & 1 & & \\
 & \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & & & 1 & & 2 & & 1 \\
 & \iddots & & \ddots & & & & 1 \arrow[dr, "+"] & 3 & & 3 & & 1 \\
 & \binom{n}{k-1} \arrow[dr, "+"] & \binom{n}{k} & & & & 1 & & 4 & & 6 & & 4 & & 1 \\
 & & \binom{n+1}{k} & & & & \iddots & & & & & & & & \ddots
\end{tikzcd}$$

Nach Bemerkung 4.4 (a) stehen auf den Schenkeln dieses Dreiecks nur Einsen, und nach Lemma 4.5 ergibt sich jede andere Zahl in diesem Diagramm als die Summe der beiden darüber stehenden. Insbesondere folgt daraus, dass alle Binomialkoeffizienten *natürliche* Zahlen sind – was aus der Definition aufgrund des Bruches ja nicht offensichtlich ist. Wir können sie damit für jeden Körper K gemäß Notation 3.9 (d) als Elemente von K auffassen (was wir gleich in Satz 4.7 auch tun werden).

Mit dieser Vorarbeit können wir nun die sehr wichtige binomische Formel beweisen. Ihr Name kommt übrigens von der lateinischen Vorsilbe „bi“ für „zwei“: Ein Binom ist eine Summe, die aus zwei Termen besteht, und die binomische Formel berechnet die Potenzen eines solchen Binoms.

Satz 4.7 (**Binomische Formel**). *Es seien K ein Körper, $x, y \in K$ und $n \in \mathbb{N}$. Dann gilt*

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

Beweis. Wir beweisen die Formel mit Induktion über n. Für $n = 0$ sind beide Seiten der Gleichung 1; die Aussage ist in diesem Fall also richtig. Für den Induktionsschritt nehmen wir nun an, dass die Gleichung für ein gegebenes $n \in \mathbb{N}$ richtig ist, und folgern daraus zunächst

$$\begin{aligned}
(x+y)^{n+1} &= (x+y) \cdot (x+y)^n \\
&= (x+y) \cdot \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} && \text{(nach Induktionsvoraussetzung)} \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k+1} && \text{(durch Ausmultiplizieren)} \\
&= \sum_{k=1}^{n+1} \binom{n}{k-1} x^k y^{n-k+1} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k+1} && \text{(Indexverschiebung } k \mapsto k-1 \text{ in der} \\
& && \text{ersten Summe, siehe Notation 3.9 (c)).}
\end{aligned}$$

Lösen wir hier nun aus der ersten Summe den Term für $k = n+1$ und aus der zweiten den für $k = 0$ heraus, so können wir diesen Ausdruck auch schreiben als

$$\begin{aligned}
(x+y)^{n+1} &= \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] \cdot x^k y^{n+1-k} + \binom{n}{n} x^{n+1} y^0 + \binom{n}{0} x^0 y^{n+1} \\
&= \sum_{k=1}^{n} \binom{n+1}{k} \cdot x^k y^{n+1-k} + x^{n+1} + y^{n+1} && \text{(nach Lemma 4.5).}
\end{aligned}$$

Die letzten beiden Summanden x^{n+1} und y^{n+1} sind hier aber genau diejenigen, die sich in der vorderen Summe ergeben, wenn man $k = n+1$ bzw. $k = 0$ setzt. Also können wir die Summe über k gleich über alle Werte von 0 bis $n+1$ laufen lassen und erhalten

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

also genau die zu zeigende Gleichung für die Potenz $n+1$. Die binomische Formel ist damit durch Induktion bewiesen. □

Beispiel 4.8. Für $n = 2$ ergibt Satz 4.7 wie erwartet die bekannte Formel

$$(x+y)^2 = \binom{2}{0} x^0 y^2 + \binom{2}{1} x^1 y^1 + \binom{2}{2} x^2 y^0 = y^2 + 2xy + x^2.$$

Bemerkung 4.9 (Kombinatorische Deutung der Binomialkoeffizienten)**.** Man kann sich die binomische Formel natürlich auch so entstanden denken, dass man den Ausdruck

$$(x+y)^n = \underbrace{(x+y) \cdot \cdots \cdot (x+y)}_{n\text{-mal}}$$

nach dem Distributivgesetz ausmultipliziert. Im Fall $n = 3$ erhalten wir z. B. zunächst ohne Verwendung der Kommutativität der Multiplikation

$$\begin{aligned}
(x+y)^3 &= (x+y) \cdot (x+y) \cdot (x+y) \\
&= (x+y) \cdot (xx + xy + yx + yy) \\
&= xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy. && (*)
\end{aligned}$$

Insgesamt bekommen wir also eine Summe aus Produkten mit jeweils n Faktoren x oder y. Jede Möglichkeit, alle diese Faktoren separat als x oder y zu wählen, kommt dabei genau einmal vor.

Fassen wir nun mit der Kommutativität der Multiplikation gleiche Terme zusammen, so erhalten wir den Term $x^k y^{n-k}$ also genau so oft, wie wir Möglichkeiten haben, aus den n Faktoren die k auszuwählen, die gleich x sein sollen. In $(*)$ oben bekommen wir z. B. den Term xy^2 dreimal, nämlich aus xyy, yxy und yyx. Nach der binomischen Formel ist der Vorfaktor von $x^k y^{n-k}$ in $(x+y)^n$ aber gerade $\binom{n}{k}$. Daher ist dieser Binomialkoeffizient genau die Anzahl der Möglichkeiten, aus n Objekten (hier: Faktoren) k auszuwählen (hier: diejenigen, bei denen wir x gewählt haben). Man kann sich die binomische Formel also als algebraische Formulierung dieser kombinatorischen Aussage vorstellen.

Aufgabe 4.10.

(a) Beweise für alle $k,n \in \mathbb{N}$ mit $k \le n$ die Gleichung
$$\sum_{m=k}^{n} \binom{m}{k} = \binom{n+1}{k+1}.$$

(b) Zeige mit Induktion über n, dass die Gleichung
$$x_1 + \cdots + x_n = d$$
für gegebenes $n \in \mathbb{N}_{>0}$ und $d \in \mathbb{N}$ genau $\binom{n+d-1}{n-1}$ Lösungen $(x_1,\dots,x_n)$ in *natürlichen* Zahlen $x_1,\dots,x_n$ besitzt.

Aufgabe 4.11. Für alle $n,p \in \mathbb{N}$ bezeichnen wir mit $s_p(n) := \sum_{k=1}^{n} k^p = 1^p + \cdots + n^p$ die Summe der p-ten Potenzen aller natürlichen Zahlen von 1 bis n.

(a) Beweise für alle $n,p \in \mathbb{N}$ die Formel
$$\binom{p+1}{0} s_0(n) + \binom{p+1}{1} s_1(n) + \cdots + \binom{p+1}{p} s_p(n) = (n+1)^{p+1} - 1.$$

(b) Zeige mit Hilfe von (a), dass s_2 ein Polynom in n ist, und berechne dieses Polynom explizit. Ist s_p für alle $p \in \mathbb{N}$ ein Polynom in n?

Aufgabe 4.12. Für ein gegebenes $n \in \mathbb{N}$ betrachten wir das Polynom
$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto x^n + x^{n-1} + \cdots + x + 1.$$
Man zeige:

(a) Ist n ungerade, so hat f genau eine Nullstelle. Was ist ihre Vielfachheit?

(b) Ist n gerade, so hat f keine Nullstelle.

4.B Geordnete Körper

Wir haben bisher von den reellen Zahlen nur die Körpereigenschaften, also die Eigenschaften der vier Grundrechenarten ausgenutzt – und dabei z. B. in Beispiel 3.6 (b) gesehen, dass es außer den reellen Zahlen auch noch ganz andere (und in der Tat sogar sehr viele) Körper gibt. Wir müssen also noch weitere Eigenschaften auflisten, um die reellen Zahlen eindeutig zu charakterisieren. Dies wollen wir im Rest dieses Kapitels tun.

Eine Eigenschaft der reellen Zahlen, die wir bisher völlig vernachlässigt haben, ist, dass man sie *ordnen* kann, also dass man zwei Zahlen der Größe nach vergleichen kann. Die Eigenschaften dieser Ordnungsrelation werden im Begriff des sogenannten *geordneten Körpers* formalisiert.

Definition 4.13 (Geordnete Körper)**.** Ein Körper K heißt **geordneter** oder **angeordneter Körper**, wenn in ihm eine Menge $P \subset K$ (die „Menge der positiven Zahlen“) ausgezeichnet ist, so dass die folgenden drei Eigenschaften gelten:

(a) Für alle $x \in K$ gilt *genau* eine der drei Eigenschaften $x = 0$, $x \in P$ oder $-x \in P$. (Im zweiten Fall nennt man x eine **positive** Zahl, im dritten eine **negative** Zahl.)

(b) Für alle $x,y \in P$ ist $x+y \in P$ („die Summe zweier positiver Zahlen ist positiv“).

(c) Für alle $x, y \in P$ ist $xy \in P$ („das Produkt zweier positiver Zahlen ist positiv“).

In diesem Fall schreibt man $x < y$ oder $y > x$ falls $y - x \in P$, und $x \leq y$ oder $y \geq x$ falls $y - x \in P$ oder $y = x$. (Insbesondere ist also $x > 0$ genau dann, wenn $x \in P$, und $x < 0$ genau dann, wenn $-x \in P$; außerdem gilt nach (a) für $x, y \in K$ stets genau eine der Aussagen $x = y$, $x < y$ oder $y < x$.)

Beispiel 4.14.

(a) $\mathbb{R}$ ist ein geordneter Körper (was wir hier wiederum axiomatisch voraussetzen wollen). Nennt man in der Teilmenge $\mathbb{Q}$ von $\mathbb{R}$ genau die Zahlen positiv, die es auch in $\mathbb{R}$ sind, so ist damit auch $\mathbb{Q}$ ein geordneter Körper.

(b) Der Körper $\mathbb{Z}_2$ aus Beispiel 3.6 (b) kann nicht zu einem geordneten Körper gemacht werden: Das Element $1 = u$ ist nicht gleich 0, also müsste nach Definition 4.13 (a) genau eine der beiden Eigenschaften $1 \in P$ und $-1 \in P$ gelten. Dies ist aber unmöglich, da wegen $1 + 1 = 0$ in $\mathbb{Z}_2$ die Gleichung $-1 = 1$ gilt.

Bemerkung 4.15 (Partielle und totale Ordnungen)**.** Ist K ein geordneter Körper, so ist $\leq$ wie in Definition 4.13 natürlich eine Relation auf K im Sinne von Definition 2.1. Sie besitzt die folgenden Eigenschaften für alle $x, y, z \in K$:

(a) **Reflexivität**: Es gilt $x \leq x$.

(b) **Antisymmetrie**: Ist $x \leq y$ und $y \leq x$, so folgt $x = y$.

(c) **Transitivität**: Gilt $x \leq y$ und $y \leq z$, so folgt auch $x \leq z$.

(d) **Totalität**: Es gilt (mindestens) eine der Aussagen $x \leq y$ und $y \leq x$. (Mit anderen Worten: „Zwei beliebige Elemente von K sind stets miteinander vergleichbar.“)

In der Tat folgt (a) unmittelbar aus Definition 4.13. Da $x \leq y$ nach Definition äquivalent zu $y - x \in P$ oder $y - x = 0$ ist, und $y \leq x$ zu $-(y - x) \in P$ oder $y - x = 0$, ergeben sich (b) und (d) außerdem aus Definition 4.13 (a). Die Aussage (c) schließlich ist trivial falls $x = y$ oder $y = z$; andernfalls gilt $y - x \in P$ oder $z - y \in P$, und damit $z - x = (y - x) + (z - y) \in P$, also $x < z$, nach Definition 4.13 (b).

Auf einer beliebigen Menge K (die also nicht notwendig ein Körper ist) nennt man eine Relation $\leq$ mit den Eigenschaften (a), (b) und (c) eine **partielle Ordnung**. Gilt zusätzlich noch (d), so heißt $\leq$ eine **(totale) Ordnung** auf K. Jeder geordnete Körper K liefert also eine totale Ordnung auf K.

Das Standardbeispiel für eine partielle Ordnung auf einer Menge ist die Teilmengenrelation auf der Potenzmenge $\mathscr{P}(M)$ einer beliebigen Menge M: Sind $A, B, C \in \mathscr{P}(M)$ Teilmengen von M, so gilt natürlich $A \subset A$; aus $A \subset B$ und $B \subset A$ folgt $A = B$; und aus $A \subset B$ und $B \subset C$ folgt $A \subset C$. Diese partielle Ordnung ist aber in der Regel nicht total: Für $M = \mathbb{N}$ sind die Teilmengen $A = \{0\}$ und $B = \{1\}$ nicht vergleichbar, denn es gilt weder $A \subset B$ noch $B \subset A$.

Wie schon bei den Körpern wollen wir nun auch hier für einen geordneten Körper kurz die wichtigsten Eigenschaften ableiten, die aus der Definition folgen (und die euch für die reellen Zahlen sicher bekannt sind). Wir werden sie im Folgenden verwenden, ohne jedesmal darauf hinzuweisen.

Lemma 4.16 (Eigenschaften geordneter Körper)**.** *Für alle x, y, z in einem geordneten Körper K gilt:*

(a) *Ist $x < y$, so folgt $x + z < y + z$.*

(b) *Ist $x < y$ und $z > 0$, so gilt auch $xz < yz$. Ist dagegen $x < y$ und $z < 0$, so folgt $xz > yz$.*
(Ungleichungen drehen sich also bei der Multiplikation mit einer negativen Zahl um.)

(c) *Gilt $x \neq 0$, so ist $x^2 > 0$. Insbesondere ist also $1 > 0$.*

(d) *Wenn $0 < x < y$, dann folgt $0 < y^{-1} < x^{-1}$.*

Entsprechende Aussagen gelten natürlich auch für die nicht-strikten Ungleichungen $\leq$ bzw. $\geq$.

Beweis.

(a) Ist $x < y$, also $y - x \in P$, so ist auch $(y + z) - (x + z) = y - x \in P$, also $x + z < y + z$.

(b) Gilt wieder $y-x \in P$ und $z \in P$, so folgt aus Definition 4.13 (c) auch $(y-x)z = yz - xz \in P$, also $xz < yz$. Ist hingegen $z < 0$, also $-z \in P$, so gilt diesmal $(y-x)(-z) = xz - yz \in P$, also $yz < xz$.

(c) Ist $x \in P$, so ist natürlich auch $x^2 \in P$ nach Definition 4.13 (c). Ist $-x \in P$, so folgt genauso $x^2 = (-x)^2 \in P$. Also ist für $x \neq 0$ in jedem Fall $x^2 > 0$. Insbesondere ist damit $1 = 1 \cdot 1 > 0$.

(d) Gilt $x \in P$, so folgt aus (c) und Definition 4.13 (c) zunächst $x^{-1} = x \cdot (x^{-1})^2 \in P$, also $x^{-1} > 0$. Genauso ergibt sich $y^{-1} > 0$. Ist nun $x < y$, so folgt aus (b) durch Multiplikation mit der positiven Zahl $x^{-1}y^{-1}$ die Ungleichung $xx^{-1}y^{-1} < yx^{-1}y^{-1}$, also $y^{-1} < x^{-1}$, was zu zeigen war. □

06

Notation 4.17 (Intervalle und Betrag)**.** Die folgenden Notationen verwendet man häufig in einem geordneten Körper K.

(a) Sind $a, b \in K$ mit $a \leq b$, so definiert man die folgenden Teilmengen von K:

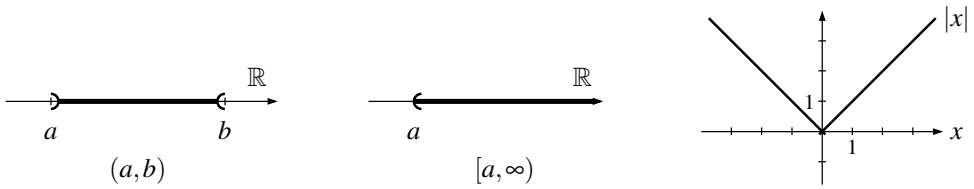
- $[a,b] := \{x \in K : a \leq x \leq b\}$ (**abgeschlossene** bzw. **kompakte Intervalle**);
- $(a,b) := \{x \in K : a < x < b\}$ (**offene Intervalle**);
- $[a,b) := \{x \in K : a \leq x < b\}$ (**halboffene Intervalle**);
- $[a,\infty) := \{x \in K : a \leq x\}$ (**uneigentliche Intervalle**);

und analog natürlich $(a,b]$, (a,∞), $(-\infty,b]$ und $(-\infty,b)$. Wenn wir derartige Intervalle im Fall $K = \mathbb{R}$ graphisch darstellen, deuten wir wie im Bild unten meistens durch Rundungen an den Intervallgrenzen an, ob die Randpunkte mit dazugehören sollen oder nicht.

(b) Für $x \in K$ definieren wir den **Betrag** von x als

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

Insbesondere gilt also immer $|x| \geq 0$. Im Fall $K = \mathbb{R}$ sieht die Betragsfunktion natürlich wie im folgenden Bild rechts aus.



Die wichtigsten beiden Eigenschaften der Betragsfunktion sind ihre „Verträglichkeit“ mit Addition und Multiplikation:

Lemma 4.18 (Eigenschaften der Betragsfunktion)**.** *Für alle x, y in einem geordneten Körper K gilt:*

(a) $|xy| = |x| \cdot |y|$. *Insbesondere ergibt sich daraus für $y = -1$, dass $|-x| = |x|$.*

(b) $x \leq |x|$.

(c) $|x+y| \leq |x| + |y|$. *Diese Ungleichung bezeichnet man als **Dreiecksungleichung** – wir werden in Bemerkung 6.10 (a) sehen, warum.*

Beweis.

(a) Wir machen eine Fallunterscheidung je nach Vorzeichen von x und y. Ist z. B. $x \geq 0$ und $y \leq 0$, so ist $xy \leq 0$ und damit nach Definition des Betrages $|x| = x$, $|y| = -y$ und $|xy| = -xy$. Zusammensetzen dieser Gleichungen ergibt die Behauptung $|xy| = -xy = x \cdot (-y) = |x| \cdot |y|$. Die anderen Fälle der möglichen Vorzeichenverteilungen beweist man genauso.

(b) Für $x \leq 0$ ist $x \leq 0 \leq |x|$; für $x \geq 0$ ist $x = |x|$.

(c) Nach (b), angewendet auf x und y, gilt (mit Lemma 4.16 (a))

$$x+y \leq |x|+|y|. \tag{1}$$

Wenden wir (b) hingegen auf $-x$ und $-y$ an, so folgt auch

$$-x-y \leq |-x|+|-y| = |x|+|y|. \tag{2}$$

Aber $|x+y|$ ist in jedem Fall eine der beiden Zahlen $x+y$ oder $-x-y$. Damit folgt die Behauptung $|x+y| \leq |x|+|y|$ aus (1) und (2). □

Bemerkung 4.19 (Dreiecksungleichung nach oben und unten)**.** Die Dreiecksungleichung aus Lemma 4.18 (c) schätzt den Betrag $|x+y|$ einer Summe nach oben ab. Offensichtlich gilt im Allgemeinen keine Gleichheit, wie das Beispiel $x=1$, $y=-1$ zeigt: Hier ist $|x+y| = 0 < 2 = |x|+|y|$.

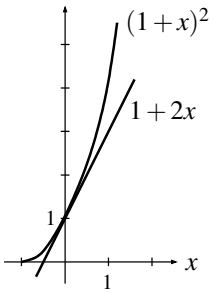
Eine Abschätzung nach unten kann man erhalten, indem man Lemma 4.18 (c) auf die Zahlen $x+y$ und $-y$ anwendet: Man erhält dann nämlich $|x| = |(x+y)-y| \leq |x+y|+|-y| = |x+y|+|y|$ und damit $|x+y| \geq |x|-|y|$. Insgesamt haben wir also für alle $x,y \in K$

$$|x|-|y| \leq |x+y| \leq |x|+|y|.$$

Eine weitere Anwendung der Eigenschaften eines geordneten Körpers ist die folgende Ungleichung, die oftmals dann nützlich ist, wenn die Größe von Potenzen x^n mit der von Produkten $n \cdot x$ verglichen werden soll.

Satz 4.20 (**Bernoullische Ungleichung**)**.** *Es seien K ein geordneter Körper, $x \in K$ mit $x \geq -1$, und $n \in \mathbb{N}$. Dann gilt $(1+x)^n \geq 1+nx$.*

Beweis. Wir zeigen die Aussage mit Induktion über n. Das Bild rechts unten veranschaulicht die Ungleichung im Fall $K = \mathbb{R}$ und $n = 2$.



Der Induktionsanfang für $n = 0$ ist klar: dann sind beide Seiten gleich 1, die Ungleichung ist also erfüllt.

Für den Induktionsschritt nehmen wir nun an, dass

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

für alle $x \geq -1$ und ein gegebenes $n \in \mathbb{N}$ gilt. Mit Lemma 4.16 (b) können wir diese Ungleichung mit der nach Voraussetzung nicht-negativen Zahl $1+x$ multiplizieren und erhalten

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

Nach Lemma 4.16 (c) ist nun $nx^2 \geq 0$ und damit $(1+x)^{n+1} \geq 1+(n+1)x$, was zu zeigen war. □

Aufgabe 4.21. Für welche $x,y \in \mathbb{R}$ bzw. $n \in \mathbb{N}$ gelten die folgenden Ungleichungen?

(a) $\dfrac{2xy}{x+y} \leq \dfrac{x+y}{2}$ (b) $\dfrac{4^n}{2n+1} < \dbinom{2n}{n} < 4^n$ (c) $n! < \left(\dfrac{n}{2}\right)^n$

4.C Supremum und Infimum

Wie bereits angekündigt wollen wir in diesem Abschnitt nun endlich eine eindeutige axiomatische Charakterisierung der reellen Zahlen angeben. Bisher haben wir nur gesehen, dass $\mathbb{R}$ ein geordneter Körper ist. Aber auch $\mathbb{Q}$ ist ein geordneter Körper, und daher müssen wir noch untersuchen, wie sich $\mathbb{R}$ von $\mathbb{Q}$ unterscheidet. Anschaulich würden wir dies wohl so formulieren, dass die Menge $\mathbb{Q}$ der rationalen Zahlen „Löcher“ auf der Zahlengeraden hat, also dass es Punkte wie z. B. $\sqrt{2}$ auf der Zahlengeraden gibt, die keiner rationalen Zahl entsprechen (siehe Lemma 4.35). Aber wie formuliert man so etwas exakt?

Um dies zu sehen, müssen wir genauer untersuchen, wann Teilmengen eines geordneten Körpers K größte bzw. kleinste Elemente haben. Natürlich können wir zunächst auf die offensichtliche Art das **Maximum** und **Minimum** zweier Elemente $x, y \in K$ definieren, nämlich durch

$$\max(x,y) := \begin{cases} x & \text{falls } x \geq y, \\ y & \text{falls } y \geq x \end{cases} \qquad \text{und} \qquad \min(x,y) := \begin{cases} x & \text{falls } x \leq y, \\ y & \text{falls } y \leq x \end{cases}.$$

Analog kann man auch das Maximum und Minimum von mehr als zwei Zahlen definieren, *sofern es nur endlich viele sind.* Betrachten wir dagegen eine (unendliche) Teilmenge $M \subset K$, so können wir in der Regel nicht mehr erwarten, dass M ein kleinstes bzw. größtes Element hat, allein schon weil die Menge dann in folgendem Sinne unbeschränkt sein kann:

Definition 4.22 (Beschränkte Mengen)**.** Es sei M eine Teilmenge eines geordneten Körpers K.

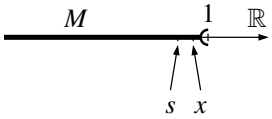
(a) Ein Element $s \in K$ heißt **obere Schranke** für M, wenn $x \leq s$ für alle $x \in M$. Existiert eine solche obere Schranke für M, so nennt man M **nach oben beschränkt**. Analog definiert man den Begriff einer **unteren Schranke** bzw. einer **nach unten beschränkten** Menge.

(b) Die Menge M heißt **beschränkt**, wenn sie nach oben und unten beschränkt ist, oder – äquivalent dazu – wenn die Menge ihrer Beträge beschränkt ist, also wenn es ein $s \in K$ gibt mit $|x| \leq s$ für alle $x \in M$.

Beispiel 4.23. Es sei $M = (-\infty, 1) = \{x \in \mathbb{R} : x < 1\} \subset \mathbb{R}$.

(a) Die Menge M ist nach oben (aber nicht nach unten) beschränkt, denn $s = 1$ ist eine obere Schranke für M. Genauso ist auch $s = 2$ eine obere Schranke für M, auch wenn man anschaulich vielleicht sagen würde, dass diese Schranke „nicht so gut“ ist, weil $x \leq 2$ für alle $x \in M$ eine schwächere Aussage ist als $x \leq 1$ für alle $x \in M$.

(b) Ist $s \in M$, also $s < 1$, so gilt für den Mittelpunkt $x := \frac{s+1}{2}$ zwischen s und 1 wie im Bild rechts natürlich



$$x = \frac{s+1}{2} > \frac{s+s}{2} = s \quad \text{und} \quad x = \frac{s+1}{2} < \frac{1+1}{2} = 1.$$

Hieraus ergeben sich sofort zwei einfache Folgerungen:

- Es gibt keine größte Zahl in M (denn zu $s \in M$ liegt die größere Zahl x wegen $x < 1$ ebenfalls noch in M).
- Die Menge M ist aber nach (a) nach oben beschränkt, und die kleinste mögliche obere Schranke für M ist 1 (denn $s < 1$ kann keine obere Schranke mehr sein, da $x > s$ ebenfalls noch in M liegt).

Die Zahl 1 kann damit als „Obergrenze der Zahlen in M“ angesehen werden, auch wenn sie kein Element von M und daher keine größte Zahl in M ist. Dieses Konzept wollen wir jetzt exakt definieren.

Definition 4.24 (Supremum und Infimum)**.** Es sei M eine Teilmenge eines geordneten Körpers K.

(a) Eine Zahl $s \in K$ heißt ein **Supremum** von M, wenn s eine „kleinste obere Schranke für M“ ist, d. h. wenn gilt:

- s ist eine obere Schranke für M; und
- für jede obere Schranke s' für M gilt $s \leq s'$.

(b) Eine Zahl $s \in K$ heißt ein **Maximum** von M, wenn s eine „obere Schranke für M in M“ ist, d. h. wenn gilt:

- s ist eine obere Schranke für M; und
- $s \in M$.

Analog heißt $s \in K$ ein **Infimum** von M, wenn s eine „größte untere Schranke für M“ ist, und ein **Minimum**, wenn s eine untere Schranke für M in M ist.

Bemerkung und Notation 4.25 ($\sup M$, $\max M$, $\inf M$, $\min M$)**.**

(a) Jedes Maximum s einer Menge M ist auch ein Supremum von M: Ist dann nämlich $s' \in K$ eine weitere obere Schranke, so folgt wegen $s \in M$ natürlich sofort $s \leq s'$.

(b) Wenn die Menge M ein Supremum besitzt, dann ist es eindeutig bestimmt: Sind nämlich s_1 und s_2 zwei kleinste obere Schranken, so folgt nach Definition 4.24 (a) sofort $s_1 \leq s_2$ (weil s_1 eine kleinste obere Schranke und s_2 auch eine obere Schranke ist) und $s_2 \leq s_1$ (weil s_2 Supremum und s_1 auch eine obere Schranke ist), also $s_1 = s_2$. Nach (a) ist damit auch ein Maximum von M eindeutig bestimmt, falls es existiert.

Wenn ein Supremum oder Maximum von M existiert, können wir also von *dem* Supremum bzw. *dem* Maximum von M reden. Wir bezeichnen es dann mit $\sup M$ bzw. $\max M$.

(c) Analog sind natürlich auch Infimum und Minimum von M eindeutig bestimmt, sofern sie existieren; wir bezeichnen sie dann mit $\inf M$ bzw. $\min M$.

Beispiel 4.26.

(a) Die Menge $M = (-\infty, 1] \subset \mathbb{R}$ hat offensichtlich das Maximum 1. Nach Bemerkung 4.25 (a) ist 1 damit auch das Supremum von M, d. h. es ist $\sup M = \max M = 1$.

(b) Das Intervall $M = (-\infty, 1) \subset \mathbb{R}$ hat nach Beispiel 4.23 (b) kein Maximum, aber es gilt $\sup M = 1$.

(c) Das Intervall $M = (1, \infty)$ besitzt kein Supremum (und damit auch kein Maximum), da M nach oben unbeschränkt ist, also nicht einmal irgendeine obere Schranke für M existiert – insbesondere also keine kleinste. Auch die leere Menge hat kein Supremum, weil für sie jede reelle Zahl eine obere Schranke ist und damit keine kleinste obere Schranke existiert.

Analog ergeben sich natürlich Supremum, Maximum, Infimum und Minimum von allen Intervallen aus Beispiel 4.17 (a).

Aufgabe 4.27. Es seien A und B zwei Teilmengen eines geordneten Körpers K, die ein Supremum $\sup A$ bzw. $\sup B$ besitzen. Setzen wir $A + B := \{x + y : x \in A, y \in B\}$ und $-A := \{-x : x \in A\}$, so zeige man, dass auch die folgenden Suprema und Infima existieren und die behaupteten Werte haben:

(a) $\sup(A \cup B) = \max(\sup A, \sup B)$.

(b) $\inf(-A) = -\sup A$.

(c) $\sup(A + B) = \sup A + \sup B$.

Unsere bisherigen Fälle von Mengen ohne Supremum in Beispiel 4.26 (c) waren letztlich trivial – also die leere Menge sowie Mengen, die überhaupt keine obere Schranke besitzen. Ist es auch möglich, dass eine Menge zwar nicht leer und nach oben beschränkt ist, aber trotzdem keine *kleinste* obere Schranke hat? In der Tat ist dies in $\mathbb{R}$ im Gegensatz zu $\mathbb{Q}$ nicht möglich, und wie wir sehen werden ist genau dies der wesentliche Unterschied zwischen diesen beiden Körpern. Wir definieren daher zunächst:

Definition 4.28 (Supremumsaxiom)**.** Wir sagen, dass ein geordneter Körper das **Supremumsaxiom** erfüllt, wenn in ihm jede nicht leere, nach oben beschränkte Teilmenge ein Supremum besitzt. (Natürlich besitzt dann nach Aufgabe 4.27 (b) auch jede nicht leere, nach unten beschränkte Teilmenge ein Infimum.)

Die reellen Zahlen erfüllen also dieses Supremumsaxiom – das werden wir in dieser Vorlesung axiomatisch voraussetzen, und das ist nun endlich auch die letzte Eigenschaft der reellen Zahlen, die wir benötigen. Wenn wir dieses und das vorangegangene Kapitel zusammenfassen, setzen wir nun also insgesamt über die reellen Zahlen voraus:

$\boxed{\mathbb{R} \text{ ist ein geordneter Körper, der das Supremumsaxiom erfüllt.}}$

Wie schon in Notation 1.15 gesagt, kann man die Existenz der reellen Zahlen auch aus den Axiomen der Logik und Mengenlehre herleiten – und dann ist es natürlich ein beweisbarer Satz, dass $\mathbb{R}$ ein geordneter Körper ist, der das Supremumsaxiom erfüllt. Man kann sogar noch mehr zeigen, nämlich dass diese Eigenschaften die reellen Zahlen auch vollständig charakterisieren: $\mathbb{R}$ ist in der Tat der *einzige* geordnete Körper, der das Supremumsaxiom erfüllt. Der Beweis dieser Aussage ist jedoch sehr schwierig und soll hier nicht gegeben werden, zumal wir diese Aussage auch nicht benötigen werden. Wir werden lediglich in Bemerkung 4.36 noch sehen, dass $\mathbb{Q}$ das Supremumsaxiom in der Tat nicht erfüllt.

Für uns bedeutet diese Tatsache letztlich nur, dass wir ab jetzt alles, was wir mit den reellen Zahlen tun möchten, ausschließlich auf den Axiomen eines geordneten Körpers und dem Supremumsaxiom aufbauen können und werden.

Wir wollen nun ein paar erste elementare Folgerungen aus dem Supremumsaxiom ziehen. Seine wahre Stärke wird dieses Axiom jedoch erst im nächsten Kapitel bei der Untersuchung von Grenzwerten von Folgen zeigen.

Lemma 4.29 ($\mathbb{R}$ ist **archimedisch geordnet**)**.** *Die Teilmenge $\mathbb{N} = \{0,1,2,\dots\} \subset \mathbb{R}$ ist nach oben unbeschränkt.*

Beweis. Angenommen, $\mathbb{N}$ wäre nach oben beschränkt. Dann würde nach dem Supremumsaxiom $s := \sup \mathbb{N} \in \mathbb{R}$ existieren. Da s die *kleinste* obere Schranke ist, ist $s-1$ keine obere Schranke; es gibt also ein $n \in \mathbb{N}$ mit $n > s-1$. Dann ist aber $n+1 \in \mathbb{N}$ mit $n+1 > s$, im Widerspruch dazu, dass s eine obere Schranke für $\mathbb{N}$ ist. □

Bemerkung 4.30. Eine einfache, aber oft verwendete Folgerung aus Lemma 4.29 ist, dass es zu jeder positiven Zahl $x \in \mathbb{R}_{>0}$ ein $n \in \mathbb{N}_{>0}$ gibt mit $\frac{1}{n} < x$: Da $\mathbb{N}$ nach oben unbeschränkt ist, ist insbesondere $\frac{1}{x}$ keine obere Schranke für $\mathbb{N}$; also gibt es ein $n \in \mathbb{N}$ mit $n > \frac{1}{x}$ und damit mit $\frac{1}{n} < x$.

Aufgabe 4.31. Bestimme Supremum, Infimum, Maximum und Minimum (sofern sie existieren) der Menge

$$M = \left\{ \frac{m+n}{mn} : m,n \in \mathbb{N}_{>0} \right\} \subset \mathbb{R}.$$

07

Folgerung 4.32. *Jede nicht-leere, nach oben beschränkte Teilmenge von $\mathbb{Z}$ hat ein Maximum. Entsprechend hat jede nicht-leere, nach unten beschränkte Teilmenge von $\mathbb{Z}$ ein Minimum.*

*Insbesondere hat also jede nicht-leere Teilmenge von $\mathbb{N}$ ein Minimum. (Man sagt dafür auch: $\mathbb{N}$ ist **wohlgeordnet**.)*

Beweis. Es sei $M \subset \mathbb{Z}$ nicht-leer und nach oben beschränkt, mit oberer Schranke s. Nach Lemma 4.29 gibt es weiterhin eine natürliche Zahl $b > s$, die dann natürlich ebenfalls eine obere Schranke für M ist.

Ist $a \in M$ nun ein beliebiges Element, so ist $M_{\geq a} = M \cap \{a, a+1, \dots, b\}$ eine nicht-leere endliche Menge, die demzufolge natürlich ein Maximum besitzt. Alle anderen Zahlen von M sind aber noch kleiner als a, so dass dieses Maximum also auch das Maximum von M ist. □

Bemerkung 4.33 (*Gaußklammer*)**.** Nach Folgerung 4.32 kann man für alle $x \in \mathbb{R}$ die Zahl

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\}$$

definieren, da die Menge aller $k \in \mathbb{Z}$ mit $k \leq x$ nach Lemma 4.29 nicht leer ist (es gibt ja ein $n \in \mathbb{N}$ mit $n \geq -x$, wir können dann $k = -n$ setzen), und natürlich durch x nach oben beschränkt. Als die größte ganze Zahl kleiner oder gleich x kann man sie sich als „Abrundung“ von x vorstellen; es ist also z. B.

$$\left\lfloor \frac{7}{2} \right\rfloor = 3 \quad \text{und} \quad \left\lfloor -\frac{7}{2} \right\rfloor = -4.$$

Folgerung 4.34 ($\mathbb{Q}$ liegt **dicht** in $\mathbb{R}$)**.** *Jedes nicht-leere offene Intervall (a,b) in $\mathbb{R}$ enthält eine rationale Zahl.*

Beweis. Nach Bemerkung 4.30 gibt es ein $n \in \mathbb{N}_{>0}$ mit $\frac{1}{n} < b - a$. Weiterhin ist die Menge

$$M := \left\{ k \in \mathbb{Z} : \frac{k}{n} > a \right\} = \{k \in \mathbb{Z} : k > na\}$$

nach Lemma 4.29 nicht-leer und durch na nach unten beschränkt, und besitzt damit nach Folgerung 4.32 ein Minimum k. Für dieses Minimum gilt also:

(a) $k \in M$ und damit $\frac{k}{n} > a$;

(b) $k - 1 \notin M$ und damit $\frac{k-1}{n} \leq a$, d. h.

$$\frac{k}{n} = \frac{k-1}{n} + \frac{1}{n} < a + (b - a) = b.$$

Also ist $\frac{k}{n}$ eine rationale Zahl im offenen Intervall (a, b). □

Als eine Konsequenz aus dieser Folgerung wollen wir wie bereits angekündigt nun sehen, dass die rationalen Zahlen das Supremumsaxiom nicht erfüllen, dass hier also der entscheidende Unterschied zwischen $\mathbb{Q}$ und $\mathbb{R}$ liegt. Wir untersuchen dazu zunächst, ob es eine Quadratwurzel aus 2 gibt, also eine (positive) Zahl x mit $x^2 = 2$.

Lemma 4.35 (Irrationalität von $\sqrt{2}$)**.** *Es gibt keine rationale Zahl $x \in \mathbb{Q}$ mit $x^2 = 2$.*

Beweis. Angenommen, es gäbe eine rationale Zahl $x \in \mathbb{Q}$ mit $x^2 = 2$. Wir können x als gekürzten Bruch $x = \frac{p}{q}$ (mit $p, q \in \mathbb{Z}$ und $q \neq 0$) schreiben und erhalten

$$x^2 = \frac{p^2}{q^2} = 2, \quad \text{also} \quad p^2 = 2q^2. \tag{$*$}$$

Also muss p^2 und damit auch p selbst eine gerade Zahl, d. h. durch 2 teilbar sein. Wir können daher $p = 2r$ für eine ganze Zahl $r \in \mathbb{Z}$ setzen. Einsetzen in $(*)$ liefert

$$(2r)^2 = 2q^2, \quad \text{also} \quad q^2 = 2r^2.$$

Aber dann muss auch q^2 und damit q eine gerade Zahl sein – was ein Widerspruch dazu ist, dass die Darstellung von x als Bruch $\frac{p}{q}$ als gekürzt vorausgesetzt worden ist. □

Wie ihr aus der Schule wisst, gibt es aber in den reellen Zahlen $\mathbb{R}$ eine Wurzel $\sqrt{2}$ aus 2. Da unsere Axiome, dass $\mathbb{R}$ ein geordneter Körper ist, der das Supremumsaxiom erfüllt, die reellen Zahlen ja vollständig beschreiben, könnten wir die Existenz dieser Wurzel jetzt sogar aus unseren Axiomen beweisen (die Menge $\{x \in \mathbb{R} : x^2 < 2\}$ ist offensichtlich eine nicht-leere, nach oben beschränkte Menge, die demzufolge ein Supremum s besitzt – und für dieses Supremum kann man zeigen, dass $s^2 = 2$ gilt, also dass s eine Wurzel aus 2 ist). Dieser Beweis ist jedoch recht technisch, und da wir in Folgerung 5.29 ohnehin die Existenz reeller Quadratwurzeln aus beliebigen nicht-negativen Zahlen aus unseren Axiomen beweisen werden, wollen wir hier darauf verzichten und für die folgende Bemerkung der Einfachheit halber annehmen, dass $\sqrt{2}$ in $\mathbb{R}$ existiert, zumal wir diese Bemerkung im Folgenden nicht verwenden werden.

Bemerkung 4.36 ($\mathbb{Q}$ erfüllt das Supremumsaxiom nicht)**.** Es sei $M := \{x \in \mathbb{Q} : x < \sqrt{2}\}$. Diese Menge ist offensichtlich nicht leer (denn $0 \in M$) und nach oben beschränkt (z. B. mit oberer Schranke 2). Würde $\mathbb{Q}$ das Supremumsaxiom erfüllen, müsste sie also ein Supremum $s := \sup M \in \mathbb{Q}$ besitzen. Nach Lemma 4.35 ist damit $s = \sqrt{2}$ ausgeschlossen, also kann nur $s < \sqrt{2}$ oder $s > \sqrt{2}$ gelten. Aber beides führt sofort zum Widerspruch:

- $s < \sqrt{2}$ kann keine obere Schranke für M sein, denn nach Folgerung 4.34 gäbe es dann eine rationale Zahl $q \in (s, \sqrt{2})$, also mit $q \in M$, aber $q > s$.
- $s > \sqrt{2}$ kann keine *kleinste* obere Schranke für M sein, denn wieder nach Folgerung 4.34 gäbe es nun eine rationale Zahl $q \in (\sqrt{2}, s)$, die kleiner ist als s, aber immer noch eine obere Schranke für M (da aus $x \in M$, also $x < \sqrt{2}$, ja sofort auch $x < q$ folgt).

Also erfüllt $\mathbb{Q}$ nicht das Supremumsaxiom.

Bemerkung 4.37 (Uneigentliche Suprema)**.** Nach dem Supremumsaxiom existiert das Supremum $\sup M$ für jede nicht leere, nach oben beschränkte Teilmenge $M \subset \mathbb{R}$. Oft ist es praktisch, diese Notation wie folgt auf beliebige Teilmengen von $\mathbb{R}$ zu erweitern:

(a) Ist $M = \emptyset$, so schreiben wir formal $\sup M = -\infty$. (Anschaulich: Jede reelle Zahl ist eine obere Schranke der leeren Menge, daher ist „die kleinste" davon $-\infty$.)

(b) Ist $M \neq \emptyset$ nach oben unbeschränkt, so schreiben wir formal $\sup M = \infty$. (Anschaulich: Ist M nach oben unbeschränkt, so ist keine reelle Zahl eine obere Schranke für M, die einzige und damit kleinste „obere Schranke" für M ist also ∞.)

Man spricht in diesem Fall von *uneigentlichen Suprema*. Analog setzt man natürlich $\inf \emptyset = \infty$ und $\inf M = -\infty$ für jede nach unten unbeschränkte Menge $M \subset \mathbb{R}$.

Mit „wir schreiben formal" ist dabei oben gemeint, dass $-\infty$ und ∞ natürlich keine Zahlen sind, mit denen man uneingeschränkt wie gewohnt rechnen kann. Manche Rechenoperationen wie z. B. $\infty + x := \infty$ für alle $x \in \mathbb{R}$ lassen sich zwar noch intuitiv sinnvoll definieren (siehe Bemerkung 5.41), aber andere wie z. B. $\infty - \infty$ nicht. Mit anderen Worten ist $\mathbb{R} \cup \{\pm\infty\}$ kein Körper. Trotzdem hat die Einführung uneigentlicher Suprema und Infima den Vorteil, dass $\sup M$ und $\inf M$ für *jede* Teilmenge M von $\mathbb{R}$ definiert sind, und Aussagen darüber oft auch in den neuen Fällen gültig bleiben. So gelten z. B. die Aussagen aus Aufgabe 4.27 sogar für beliebige Teilmengen von $\mathbb{R}$ (sofern in (c) rechts nicht der unbestimmte Ausdruck $\infty - \infty$ auftritt) – allerdings brauchen wir dafür natürlich einen neuen Beweis, da ja schon die Definition eines uneigentlichen Supremums eine andere als üblich ist.

5. Folgen und Grenzwerte

Nachdem wir die reellen Zahlen vollständig charakterisiert haben, wollen wir jetzt zur eigentlichen Analysis kommen. Der zentrale Begriff ist dabei der des Grenzwerts, den ihr ja sicher in der Schule schon in der einen oder anderen Form kennengelernt habt und den wir jetzt exakt einführen wollen. Wir beginnen dabei mit Grenzwerten von Folgen, da sie für den Anfang einfacher sind als die später auch noch wichtigen Grenzwerte von Funktionen.

5.A Grenzwerte von Folgen

Zur Untersuchung des Grenzwertbegriffs müssen wir als Erstes exakt definieren, was wir damit meinen, dass sich eine (unendlich lange) Folge reeller Zahlen einem Wert beliebig genau annähert.

Definition 5.1 (Folgen und Grenzwerte)**.**

(a) Eine **Folge** in einer Menge M ist eine Abbildung

$$\mathbb{N} \to M, \;\; n \mapsto a_n.$$

Man schreibt eine solche Folge als $(a_n)_{n\in\mathbb{N}}$, einfach nur als $(a_n)_n$, oder durch Aufzählen der Folgenglieder als $(a_0, a_1, a_2, \dots)$. Hin und wieder ist in der Literatur auch die noch weiter verkürzte Schreibweise (a_n) zu finden, die wir hier allerdings nicht verwenden wollen, um Verwechslungen der Folge $(a_n)_n$ mit einem zufällig eingeklammerten Folgenglied a_n zu vermeiden.

Manchmal ist es bequem, Folgen nicht beim Index 0, sondern bei einem anderen Startindex $n_0 \in \mathbb{Z}$ beginnen zu lassen – wenn man dies in der Notation deutlich machen möchte, schreibt man derartige Folgen als $(a_n)_{n\geq n_0}$.

In diesem Kapitel werden wir nur den Fall $M = \mathbb{R}$, also sogenannte *reelle Folgen* betrachten. Wir werden daher oft nur von einer Folge sprechen und damit dann immer eine reelle Folge meinen. Später werden wir auch noch andere Folgen kennenlernen, z. B. Folgen komplexer Zahlen in Abschnitt 6.C oder Folgen von Funktionen in Abschnitt 8.C.

(b) Eine Zahl $a \in \mathbb{R}$ heißt **Grenzwert** einer Folge $(a_n)_n$, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0} \; \exists n_0 \in \mathbb{N} \; \forall n \geq n_0 : |a_n - a| < \varepsilon.$$

Wir werden gleich in Lemma 5.5 sehen, dass eine Folge höchstens einen solchen Grenzwert besitzen kann. Wenn ein solches a existiert, können wir also sagen, dass a *der* Grenzwert der Folge $(a_n)_n$ ist. Man nennt die Folge in diesem Fall **konvergent** (gegen a) und schreibt dies als

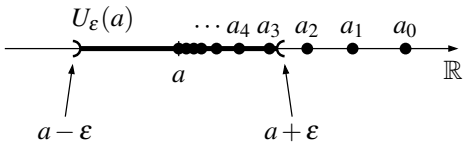
$$\lim_{n\to\infty} a_n = a$$

oder manchmal auch kurz als $a_n \to a$ (die Bezeichnung kommt vom englischen Wort „limit“ bzw. dem lateinischen „limes“). Existiert ein solcher Grenzwert nicht, so heißt die Folge **divergent**.

Bemerkung 5.2 (Anschauliche Deutung des Grenzwertbegriffs)**.** Um die Definition des Grenzwertes in leicht verständliche Worte zu fassen, führen wir ein paar intuitive Notationen ein. Für $a \in \mathbb{R}$ und $\varepsilon \in \mathbb{R}_{>0}$ heißt das offene Intervall

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

die **ε-Umgebung** von a. Die Grenzwertbedingung besagt nun genau, dass in jeder solchen ε-Umgebung von a – egal wie klein das ε gewählt ist – alle Folgenglieder ab einem gewissen n_0 liegen, wobei dieses n_0 natürlich von dem gewählten ε abhängen darf. Im Beispielbild unten wäre das z. B. für $n_0 = 3$ der Fall, denn $a_3, a_4, a_5, \dots$ liegen alle in $U_\varepsilon(a)$.



Man kann diese Tatsache auch so ausdrücken, dass in jeder ε-Umgebung *alle bis auf endlich viele* Folgenglieder liegen müssen (nämlich alle bis auf evtl. $a_0,\dots,a_{n_0-1}$). In der Analysis verwendet man gerne den Buchstaben ε für eine kleine positive Zahl und die Sprechweise „**fast alle**" für „alle bis auf endlich viele", und kann damit die Grenzwertbedingung auch in Worten formulieren:

> *Eine Zahl a ist genau dann Grenzwert einer Folge, wenn in jeder ε-Umgebung von a fast alle Folgenglieder liegen.*

Anschaulich bedeutet das natürlich einfach, dass sich die Folgenglieder immer mehr dem Grenzwert annähern. Beachte auch, dass dies insbesondere bedeutet, dass das Abändern oder Weglassen endlich vieler Folgenglieder nichts daran ändert, ob und gegen welchen Grenzwert eine Folge konvergiert. Der Startindex einer Folge wie in Definition 5.1 (a) ist für ihre Konvergenz also irrelevant.

Beispiel 5.3. Hier sind ein paar sehr wichtige Beispiele von Grenzwerten:

(a) Es ist offensichtlich, dass eine konstante Folge, in der alle Folgenglieder den gleichen Wert $a \in \mathbb{R}$ haben, gegen eben dieses a konvergiert, d. h. dass $\lim_{n\to\infty} a = a$ gilt: Hier liegen ja sogar *alle* Folgenglieder in jeder beliebigen ε-Umgebung von a.

(b) Wir behaupten, dass $\lim_{n\to\infty} \frac{1}{n} = 0$ gilt.

Um dies mit Hilfe der Definition 5.1 (b) zu beweisen, sei zunächst ein $\varepsilon > 0$ beliebig vorgegeben; wir müssen zeigen, dass fast alle Glieder der Folge $\left(\frac{1}{n}\right)_{n\geq 1}$ in der ε-Umgebung von 0 liegen. Dies ist aber sehr einfach: Nach der archimedischen Ordnung von $\mathbb{R}$ wie in Bemerkung 4.30 gibt es ein $n_0 \in \mathbb{N}$ mit $\frac{1}{n_0} < \varepsilon$. Mit einem solchen n_0 gilt dann für alle $n \geq n_0$

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

wobei wir die Rechenregeln für Ungleichungen aus Lemma 4.16 verwendet haben. Fast alle Folgenglieder, nämlich alle $\frac{1}{n}$ für $n \geq n_0$, liegen also in der ε-Umgebung von 0. Damit gilt nach Definition $\lim_{n\to\infty} \frac{1}{n} = 0$.

Beachte, dass wir hierbei zu unserem ε gar kein *konkretes* n_0 angegeben haben, das die Grenzwertbedingung erfüllt. Wir hätten dies hier leicht tun können, z. B.

$$n_0 := \left\lfloor \frac{1}{\varepsilon} + 1 \right\rfloor$$

mit der Gaußklammer aus Bemerkung 4.33, denn dann ist n_0 eine natürliche Zahl größer als $\frac{1}{\varepsilon}$, und damit wie oben $\frac{1}{n_0} < \varepsilon$. Aber zur Überprüfung der Grenzwertbedingung ist es nicht nötig, ein konkretes n_0 anzugeben – erst recht nicht das *kleinste* (also „beste") mögliche n_0. In der Tat wäre eine solche Bestimmung des kleinstmöglichen n_0 für die allermeisten Folgen auch sehr aufwändig oder sogar gar nicht praktisch durchführbar.

(c) (**Geometrische Folge**) Es sei $q \in \mathbb{R}$ mit $|q| < 1$; wir behaupten, dass dann $\lim_{n\to\infty} q^n = 0$ gilt.

Für $q = 0$ ist dies natürlich klar, da wir dann eine konstante Folge haben. Ansonsten sei wie in (b) wieder $\varepsilon > 0$ beliebig vorgegeben. Wir setzen $x := \frac{1}{|q|} - 1$, also $|q| = \frac{1}{1+x}$; wegen $|q| < 1$ ist natürlich $x > 0$. Nach Bemerkung 4.30 gibt es nun ein $n_0 \in \mathbb{N}$ mit $\frac{1}{n_0} < \varepsilon x$. Es gilt

dann für alle $n \geq n_0$

$$\begin{aligned}
|q^n - 0| = |q|^n &\overset{(1)}{=} \frac{1}{(1+x)^n} \\
&\overset{(2)}{\leq} \frac{1}{1+nx} && \text{(mit } x > 0 \text{ nach der Bernoulli-Ungleichung aus Satz 4.20)} \\
&\overset{(3)}{<} \frac{1}{nx} && \text{(wegen } 1 > 0\text{)} \\
&\overset{(4)}{\leq} \frac{1}{n_0 x} && \text{(wegen } n \geq n_0\text{)} \\
&\overset{(5)}{<} \varepsilon && \left(\text{wegen } \frac{1}{n_0} < \varepsilon x\right),
\end{aligned}$$

woraus die Behauptung folgt.

Bemerkung 5.4 (Rückwärtsrechnen)**.** Wenn ihr euch die Rechnung in Beispiel 5.3 (c) angeschaut habt, werdet ihr vermutlich keine Probleme haben, sie nachzuvollziehen – aber euch sicher auch fragen, wie ihr darauf jemals selbst hättet kommen sollen. Insbesondere die Festlegungen von x und n_0 vor Beginn der Rechnung fallen ja doch sehr vom Himmel.

Die Antwort hierauf ist einfach, dass ich den Beweis zunächst „rückwärts“ durchgeführt habe, bevor ich angefangen habe, ihn aufzuschreiben. Ich habe also mit der Rechnung oben begonnen, bevor ich wusste, was z. B. n_0 später einmal sein würde, und mir etwa folgendes gedacht:

> *Okay, wir müssen sehen, dass $|q^n|$ für $n \to \infty$ kleiner als das gegebene ε wird. Nehmen wir der Einfachheit halber erst einmal $q > 0$ an, dann müssen wir also eine Ungleichungskette $q^n < \cdots < \varepsilon$ finden. Bisher wissen wir nichts darüber, wie sich Potenzen mit wachsendem Exponenten verhalten … aber wir hatten die Bernoulli-Ungleichung $(1+x)^n \geq 1+nx$ gezeigt, die Potenzen durch lineare Funktionen abschätzen kann. Um die anwenden zu können, könnten wir vielleicht $q = 1+x$ setzen? Nein, das hilft nicht, denn dann würde die Ungleichung $q^n = (1+x)^n \geq 1+nx$ ja in die falsche Richtung gehen. Also versuchen wir lieber $q = \frac{1}{1+x}$, das dreht „$\geq$“ zu „$\leq$“ um. Moment, gibt es so ein x überhaupt und erfüllt das die Voraussetzungen der Bernoulli-Ungleichung? Ja, die Gleichung ist ja äquivalent zu $x = \frac{1}{q} - 1$, und es ist $q < 1$, also $x > 0$, das passt. Damit haben wir die Schritte (1) und (2) oben.*
>
> *Jetzt müssen wir also $\frac{1}{1+nx}$ weiter abschätzen und sehen, warum dieser Term gegen 0 geht. Die 1 im Nenner stört. Also, wir könnten sicher auch mit ihr weiter rechnen, aber einfacher wäre der Ausdruck ohne sie. Es ist ja auch $\frac{1}{1+nx} < \frac{1}{nx}$, d. h. die Abschätzung geht in die richtige Richtung, und der neue Ausdruck $\frac{1}{nx}$ geht immer noch gegen 0. Also lassen wir die 1 in (3) einfach weg. Wie wir jetzt weiter machen können, wissen wir aus Beispiel 5.3 (b): Ist nun $n \geq n_0$ und $\frac{1}{n_0} < \varepsilon x$, so erhalten wir in (4) und (5) die gewünschte Abschätzung.*

Nachdem wir diese Überlegungen durchgeführt haben, können wir schließlich noch die Beträge wieder einarbeiten und den Beweis dann so aufschreiben wie oben.

Beachte, dass es natürlich viele verschiedene Arten gibt, derartige Abschätzungen durchzuführen. Aber nicht jede Abschätzung, die richtig ist, ist auch zielführend: So hätten wir z. B. in (3) oben auch versuchen können, den Term nx wegzulassen und die Abschätzung mit $\frac{1}{1+nx} < \frac{1}{1}$ fortzusetzen. Diese Ungleichung ist genauso richtig wie (3), aber der neue Ausdruck $\frac{1}{1} = 1$ geht offensichtlich mit $n \to \infty$ nicht mehr gegen 0, so dass wir die gewünschte Folgerung $\cdots < \varepsilon$ jetzt nicht mehr erreichen können. Man muss beim Abschätzen also stets einen geeigneten Mittelweg finden und aufpassen, dass man weder zu wenig noch zu viel abschätzt. Dadurch erfordern derartige Rechnungen oft eine geschickte und vielleicht nicht ganz offensichtliche Idee. Am Anfang ist das sicher ungewohnt, aber im Laufe der Zeit werdet ihr ein gewisses Gefühl dafür entwickeln, welche Art von Abschätzung in welchen Fällen sinnvoll sein könnte. Aber so oder so – für das reine *Nachvollziehen* einer Abschätzung, die jemand anders gefunden hat (wie z. B. wenn ihr den Beweis in Beispiel 5.3 (c) lest und verstehen wollt), sind solche Ideen natürlich nicht notwendig.

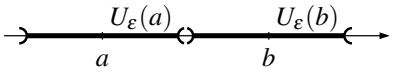
08

Wir wollen nun die bereits in Definition 5.1 versprochene Aussage beweisen, dass der Grenzwert einer Folge (sofern er existiert) immer eindeutig ist. Anschaulich ist diese Aussage natürlich sofort einleuchtend: Es können nicht fast alle Folgenglieder beliebig nahe an zwei verschiedenen Zahlen liegen. Denn wenn wir disjunkte ε-Umgebungen der beiden Grenzwerte wählen, kann jedes Folgenglied natürlich immer nur in einer der beiden Umgebungen liegen – und somit können nicht fast alle in *beiden* Umgebungen liegen. Formal aufgeschrieben sieht diese Beweisidee so aus:

Lemma 5.5 (Eindeutigkeit des Grenzwerts)**.** *Jede Folge hat höchstens einen Grenzwert.*

Beweis. Angenommen, die Aussage wäre falsch, d. h. es gäbe eine Folge $(a_n)_n$ mit $a_n \to a$ und $a_n \to b$ für gewisse $a,b \in \mathbb{R}$ mit $a \neq b$.

Wählen wir dann $\varepsilon := \frac{|a-b|}{2}$ als den halben Abstand zwischen a und b, so gilt wie im Bild rechts $U_\varepsilon(a) \cap U_\varepsilon(b) = \emptyset$, denn aus $x \in U_\varepsilon(a) \cap U_\varepsilon(b)$ würde mit der Dreiecksungleichung der Widerspruch



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

folgen.

Aber wegen $a_n \to a$ gilt $a_n \in U_\varepsilon(a)$ für fast alle n (also für alle $n \geq n_1$ mit einem gewissen $n_1 \in \mathbb{N}$), und wegen $a_n \to b$ genauso $a_n \in U_\varepsilon(b)$ für fast alle n (also für alle $n \geq n_2$ mit einem gewissen $n_2 \in \mathbb{N}$). Damit folgt auch $a_n \in U_\varepsilon(a) \cap U_\varepsilon(b)$ für fast alle n (nämlich für alle n, bei denen beide Aussagen gelten, also für $n \geq \max(n_1, n_2)$), was ein Widerspruch zu $U_\varepsilon(a) \cap U_\varepsilon(b) = \emptyset$ ist und somit das Lemma beweist. □

Bemerkung 5.6. Wir sehen im Beweis von Lemma 5.5, dass die „fast alle"-Notation den Vorteil hat, dass wir uns oft das explizite Arbeiten mit dem n_0 aus Definition 5.1 (b) (von dem wir ja meistens ohnehin nicht wirklich wissen müssen, welchen Wert es genau hat) sparen können. Die einzige Eigenschaft, die wir hier wirklich gebraucht haben, ist die: Wenn eine Aussage $A(n)$ für fast alle n gilt, und eine weitere Aussage $B(n)$ ebenfalls für fast alle (aber nicht notwendig für die gleichen), dann gelten auch $A(n)$ und $B(n)$ zusammen für fast alle n – nämlich für alle bis auf die endlich vielen Ausnahmen für $A(n)$ und $B(n)$.

Natürlich gibt es auch Folgen ohne Grenzwert. Die einfachste Möglichkeit dafür ist, dass ihre Glieder unbeschränkt wachsen und sich somit keiner Zahl annähern können. Dies wollen wir jetzt formal untersuchen.

Definition 5.7 (Beschränkte Folgen)**.** Eine Folge $(a_n)_n$ heißt **beschränkt**, wenn die Menge ihrer Folgenglieder beschränkt ist, also wenn es ein $s \in \mathbb{R}$ gibt mit $|a_n| \leq s$ für alle $n \in \mathbb{N}$.

Lemma 5.8. *Jede konvergente Folge ist beschränkt.*

Beweis. Es sei $(a_n)_n$ eine konvergente Folge mit Grenzwert a. Dann gibt es zu $\varepsilon = 1$ ein n_0, so dass $|a_n - a| < \varepsilon = 1$ und damit nach der Dreiecksungleichung auch

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

für alle $n \geq n_0$ gilt. Damit ist dann aber $|a_n| \leq s$ für alle $n \in \mathbb{N}$, wenn wir

$$s := \max(|a_0|, |a_1|, \dots, |a_{n_0-1}|, 1 + |a|)$$

setzen. Also ist $(a_n)_n$ beschränkt. □

Beispiel 5.9.

(a) Die Folge

$$(a_n)_n = (1, 0, 2, 0, 3, 0, \dots)$$

ist unbeschränkt (da die Menge $\mathbb{N}$ ihrer Folgenglieder nach Lemma 4.29 unbeschränkt ist) und damit nach Lemma 5.8 divergent.

(b) Auch die geometrische Folge $(q^n)_{n\in\mathbb{N}}$ aus Beispiel 5.3 (c) ist für $|q| > 1$ unbeschränkt und damit divergent: Ist nämlich $s \in \mathbb{R}_{>0}$ beliebig, so können wir ein $n \in \mathbb{N}$ wählen mit $n > \frac{s}{|q|-1}$ und erhalten mit der Bernoulli-Ungleichung

$$|q|^n = (1+|q|-1)^n \overset{4.20}{\geq} 1+n(|q|-1) > n(|q|-1) > s.$$

Wie rechnet man nun aber Grenzwerte konkret aus, wenn man nicht jedesmal wieder auf die Definition zurückgehen möchte? Glücklicherweise gibt es dafür die Grenzwertsätze, die besagen, dass man Grenzwerte mit Summen, Differenzen, Produkten und Quotienten vertauschen kann, und die damit oft eine konkrete Berechnung ermöglichen. Zum Beweis dieser Aussage benötigen wir zunächst ein Lemma.

Definition 5.10 (Nullfolgen)**.** Eine Folge heißt **Nullfolge**, wenn sie gegen 0 konvergiert. Offensichtlich konvergiert eine Folge $(a_n)_n$ damit nach Definition genau dann gegen $a \in \mathbb{R}$, wenn $(a_n - a)_n$ eine Nullfolge ist.

Lemma 5.11. *Ist $(a_n)_n$ eine beschränkte Folge und $(b_n)_n$ eine Nullfolge, so ist auch $(a_n b_n)_n$ eine Nullfolge.*

Beweis. Es sei $\varepsilon > 0$ beliebig. Da $(a_n)_n$ beschränkt ist, gilt $|a_n| \leq s$ für ein $s \in \mathbb{R}_{>0}$ und alle $n \in \mathbb{N}$. Da $(b_n)_n$ eine Nullfolge ist, ist weiterhin $|b_n| < \frac{\varepsilon}{s}$ für fast alle n. Also gilt für fast alle n auch die Abschätzung $|a_n b_n| = |a_n| \cdot |b_n| < s \cdot \frac{\varepsilon}{s} = \varepsilon$, d. h. $(a_n b_n)_n$ ist eine Nullfolge. □

Bemerkung 5.12 (Folgen mit Grenzwert ungleich 0)**.** Es sei $(a_n)_n$ eine Folge, die gegen einen Grenzwert $a \neq 0$ konvergiert. Eine unmittelbare, aber dennoch oft nützliche Folgerung aus der Grenzwertdefinition 5.1 ergibt sich, wenn wir dort $\varepsilon = \frac{|a|}{2} > 0$ setzen: Für fast alle n ist dann $a_n \in U_\varepsilon(a)$, mit der Dreiecksungleichung nach unten aus Bemerkung 4.19 also

$$|a_n| \geq |a| - |a - a_n| > |a| - \varepsilon = \frac{|a|}{2}.$$

Hat eine Folge also einen Grenzwert $a \neq 0$, so sind insbesondere auch fast alle Folgenglieder ungleich 0 (und betragsmäßig sogar größer als $\frac{|a|}{2}$).

Satz 5.13 (**Grenzwertsätze** für Folgen)**.** *Es seien $(a_n)_n$ und $(b_n)_n$ zwei konvergente Folgen mit $a_n \to a$ und $b_n \to b$. Dann gilt:*

(a) *$a_n + b_n \to a + b$ und $a_n - b_n \to a - b$.*

(b) *$a_n b_n \to ab$.*

(c) *Ist $b \neq 0$, so sind auch fast alle $b_n \neq 0$, und es gilt $\frac{a_n}{b_n} \to \frac{a}{b}$.*

Beweis.

(a) Es sei $\varepsilon > 0$ beliebig. Wegen $a_n \to a$ und $b_n \to b$ gilt $|a_n - a| < \frac{\varepsilon}{2}$ und $|b_n - b| < \frac{\varepsilon}{2}$ für fast alle n. Damit folgt für fast alle n (siehe Bemerkung 5.6) mit der Dreiecksungleichung

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

also wie behauptet $a_n + b_n \to a + b$. Die Aussage über die Differenz der Grenzwerte folgt natürlich genauso.

(b) Für alle $n \in \mathbb{N}$ gilt zunächst

$$a_n b_n - ab = a_n b_n - a_n b + a_n b - ab = a_n(b_n - b) + b(a_n - a). \tag{1}$$

Die Folge $(a_n)_n$ ist nach Voraussetzung konvergent und damit beschränkt nach Lemma 5.8. Weiterhin ist $b_n - b$ eine Nullfolge wegen $b_n \to b$. Also ist auch $(a_n(b_n - b))_n$, d. h. der erste Summand rechts in (1), nach Lemma 5.11 eine Nullfolge. Genauso ergibt sich, dass auch der zweite Summand $(b(a_n - a))_n$ eine Nullfolge ist. Damit ist (1) die Summe zweier Nullfolgen, nach (a) also ebenfalls eine Nullfolge. Dies zeigt $a_n b_n - ab \to 0$ und damit $a_n b_n \to ab$.

(c) Nach Bemerkung 5.12 sind mit $b \neq 0$ auch fast alle b_n ungleich 0, so dass wir (nach evtl. Weglassen endlich vieler Glieder) die Quotientenfolge $\left(\frac{a_n}{b_n}\right)_n$ betrachten können. Weil nach derselben Bemerkung dann sogar $|b_n| > \frac{|b|}{2}$ und damit $\left|\frac{1}{b_n}\right| < \frac{2}{|b|}$ gilt, ist die Folge $\left(\frac{1}{b_n}\right)_n$ außerdem beschränkt. Schreiben wir also

$$\frac{a_n}{b_n} - \frac{a}{b} = \frac{a_n b - a b_n}{b b_n} = \frac{a_n b - ab + ab - a b_n}{b b_n} = \frac{1}{b_n}(a_n - a) + \frac{a}{b b_n}(b - b_n), \qquad (2)$$

so ergibt sich die Behauptung genauso wie in (b): $\left(\frac{1}{b_n}(a_n - a)\right)_n$ ist eine Nullfolge (nach Lemma 5.11 als Produkt der beschränkten Folge $\left(\frac{1}{b_n}\right)_n$ mit der Nullfolge $(a_n - a)_n$), analog ist auch $\left(\frac{a}{b b_n}(b - b_n)\right)_n$ eine Nullfolge. Damit ist (2) wieder die Summe zweier Nullfolgen, nach (a) also ebenfalls eine Nullfolge – woraus $\frac{a_n}{b_n} \to \frac{a}{b}$ folgt. □

Beispiel 5.14 (Grenzwerte von Quotienten von Polynomen)**.** Wollen wir den Grenzwert der Folge $\left(\frac{2n^2}{n^2+1}\right)_n$ bestimmen, so können wir nicht direkt die Grenzwertsätze anwenden, da Zähler und Nenner natürlich unbeschränkt sind und damit nach Lemma 5.8 divergieren. Durch Kürzen mit n^2 können wir die Folgenglieder aber umschreiben, so dass wir den Grenzwert dann mit Satz 5.13 in den Quotienten, die Summe und das Produkt hineinziehen können und (mit Beispiel 5.3)

$$\frac{2n^2}{n^2+1} = \frac{2}{1+\frac{1}{n^2}} = \frac{2}{1+\frac{1}{n}\cdot\frac{1}{n}} \to \frac{2}{1+0\cdot 0} = 2 \quad \text{für } n \to \infty$$

erhalten. Auf die gleiche Art kann man offensichtlich den Grenzwert jeder Folge berechnen, die ein Quotient von zwei Polynomfunktionen in n ist, indem man zuerst mit der höchsten auftretenden Potenz von n kürzt.

Bemerkung 5.15. Beachte, dass Satz 5.13 nur angewendet werden kann, wenn beide Grenzwerte $\lim_{n\to\infty} a_n$ und $\lim_{n\to\infty} b_n$ existieren – ansonsten macht der Satz keine Aussage. Eine Rechnung hinzuschreiben wie z. B.

$$\lim_{n\to\infty} \frac{n}{n+1}\cdot\frac{n+2}{n+3} \overset{(*)}{=} \lim_{n\to\infty}\frac{n}{n+1}\cdot\lim_{n\to\infty}\frac{n+2}{n+3} = \lim_{n\to\infty}\frac{1}{1+\frac{1}{n}}\cdot\lim_{n\to\infty}\frac{1+\frac{2}{n}}{1+\frac{3}{n}} = 1\cdot 1 = 1$$

(mit Verweis an der Stelle $(*)$ auf den Grenzwertsatz 5.13 (b)) ist daher eigentlich *nicht korrekt*, da wir bei $(*)$ ja noch nicht überprüft haben, ob die Grenzwerte der beiden einzelnen Brüche auch wirklich existieren. Man müsste also theoretisch zuerst die Grenzwerte von $\frac{n}{n+1}$ und $\frac{n+2}{n+3}$ separat berechnen (bzw. ihre Existenz zeigen), und könnte dann erst die obige Rechnung $(*)$ hinschreiben. Da dies aber deutlich mehr Schreibaufwand wäre und die Darstellung auch unübersichtlicher machen würde, wollen wir vereinbaren, dass wir die Grenzwertsätze in einer Rechnung wie oben auch schon benutzen dürfen, wenn wir erst *nachträglich* überprüfen, dass die Einzelgrenzwerte existieren.

Aufgabe 5.16. Bestimme die Grenzwerte (sofern sie existieren)

$$\text{(a)} \lim_{n\to\infty}\frac{n^2+2}{n^2+n}, \qquad \text{(b)} \lim_{n\to\infty}\frac{n^3}{\binom{n}{3}}, \qquad \text{(c)} \lim_{n\to\infty}\frac{3^n}{n^3}.$$

Für (a) beweise man diesen Grenzwert zusätzlich direkt nach Definition, d. h. man gebe zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ an, so dass die Grenzwertbedingung aus Definition 5.1 (b) gilt.

Als Beispiele für divergente Folgen haben wir bisher nur unbeschränkte Folgen gesehen. Aber auch beschränkte Folgen können natürlich divergent sein, wie z. B. die Folge

$$\left((-1)^n\right)_{n\in\mathbb{N}} = (1,-1,1,-1,\dots),$$

in der alle geraden Folgenglieder gleich 1 und alle ungeraden gleich -1 sind, so dass für die gesamte Folge kein Grenzwert existieren kann. Formal können wir dies mit dem Begriff der Teilfolgen und Häufungspunkte ausdrücken.

Definition 5.17 (Umordnungen, Teilfolgen und Häufungspunkte)**.** Es sei $(a_n)_{n\in\mathbb{N}}$ eine Folge.

(a) Eine **Umordnung** von $(a_n)_{n\in\mathbb{N}}$ ist eine Folge der Form $(a_{\sigma(0)}, a_{\sigma(1)}, a_{\sigma(2)}, \dots) = (a_{\sigma(n)})_{n\in\mathbb{N}}$ für eine bijektive Abbildung $\sigma\colon \mathbb{N} \to \mathbb{N}$. Sie entsteht also einfach durch eine beliebige Permutation aller Folgenglieder.

(b) Eine **Teilfolge** von $(a_n)_{n\in\mathbb{N}}$ ist eine Folge der Form $(a_{n_0}, a_{n_1}, a_{n_2}, \dots) = (a_{n_k})_{k\in\mathbb{N}}$ für gewisse $n_0 < n_1 < n_2 < \cdots$, also eine Folge, die aus $(a_n)_{n\in\mathbb{N}}$ durch Auswählen bestimmter Folgenglieder unter Beibehaltung ihrer Reihenfolge entsteht.

(c) Eine Zahl $a \in \mathbb{R}$ heißt **Häufungspunkt** von $(a_n)_{n\in\mathbb{N}}$, wenn es eine Teilfolge von $(a_n)_{n\in\mathbb{N}}$ gibt, die gegen a konvergiert.

Lemma 5.18 (Grenzwerte von Umordnungen und Teilfolgen)**.** *Konvergiert eine Folge $(a_n)_n$ gegen einen Grenzwert a, so konvergiert auch jede Umordnung und jede Teilfolge von $(a_n)_n$ gegen a.*

Insbesondere hat eine konvergente Folge also genau einen Häufungspunkt, nämlich ihren Grenzwert.

Beweis. Es sei $\varepsilon > 0$ beliebig. Da die Folge $(a_n)_n$ gegen a konvergiert, hat sie nur endlich viele Glieder, die außerhalb von $U_\varepsilon(a)$ liegen. Jede Umordnung oder Teilfolge von $(a_n)_n$ hat damit aber ebenfalls nur endlich viele Glieder außerhalb von $U_\varepsilon(a)$, und somit konvergiert eine solche Umordnung oder Teilfolge ebenfalls gegen a. □

Beispiel 5.19.

(a) Die oben betrachtete Folge $(a_n)_{n\in\mathbb{N}} = ((-1)^n)_{n\in\mathbb{N}} = (1, -1, 1, -1, \dots)$ besitzt die beiden konstanten Teilfolgen

$$(a_{2n})_{n\in\mathbb{N}} = (1, 1, 1, \dots)$$
$$\text{und} \qquad (a_{2n+1})_{n\in\mathbb{N}} = (-1, -1, -1, \dots)$$

und damit die beiden Häufungspunkte 1 und -1. Sie ist also nach Lemma 5.18 divergent. In der Tat werden wir in Beispiel 5.22 (a) sehen, dass 1 und -1 auch die einzigen Häufungspunkte von $(a_n)_n$ sind.

(b) Für die Folge $(a_n)_n = (n+1)_n = (1, 2, 3, \dots)$ ist jede ihrer Teilfolgen unbeschränkt und damit nach Lemma 5.8 divergent. Also besitzt $(a_n)_n$ keine Häufungspunkte.

09

Zur konkreten Berechnung von Häufungspunkten sind oft die folgenden beiden Lemmata nützlich.

Lemma 5.20 (Äquivalente Charakterisierung von Häufungspunkten)**.** *Eine Zahl $a \in \mathbb{R}$ ist genau dann ein Häufungspunkt einer Folge $(a_n)_n$, wenn in jeder ε-Umgebung von a unendlich viele Folgenglieder a_n liegen.*

Beweis.

„$\Rightarrow$“: Konvergiert eine Teilfolge von $(a_n)_n$ gegen a, so liegen in jeder ε-Umgebung von a fast alle Glieder der Teilfolge und somit insbesondere auch unendlich viele Glieder von $(a_n)_n$.

„$\Leftarrow$“: Wir konstruieren eine Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ der gewünschten Art wie folgt: Als Startindex nehmen wir $n_0 = 0$. Ist nun für ein $k \in \mathbb{N}_{>0}$ der Index n_{k-1} bereits konstruiert, so wählen wir ein $n_k > n_{k-1}$ mit $|a_{n_k} - a| < \frac{1}{k}$ (dies ist möglich, da in der $\frac{1}{k}$-Umgebung von a nach Voraussetzung unendlich viele Folgenglieder liegen, also auch eines hinter $a_{n_{k-1}}$).

Die so konstruierte Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ konvergiert dann gegen a: Ist $\varepsilon > 0$ gegeben und $k_0 \in \mathbb{N}_{>0}$ mit $\frac{1}{k_0} < \varepsilon$, so ist $|a_{n_k} - a| < \frac{1}{k} \leq \frac{1}{k_0} < \varepsilon$ für alle $k \geq k_0$. □

Lemma 5.21 (Mischfolgen)**.** *Es seien $(a_n)_{n\in\mathbb{N}}$ eine Folge sowie $(a_{n_k})_{k\in\mathbb{N}}$ und $(a_{m_k})_{k\in\mathbb{N}}$ zwei Teilfolgen, die zusammen die gesamte Folge ergeben, also so dass $\{n_k : k \in \mathbb{N}\} \cup \{m_k : k \in \mathbb{N}\} = \mathbb{N}$ gilt. Man sagt in diesem Fall auch, dass $(a_n)_{n\in\mathbb{N}}$ eine **Mischfolge** von $(a_{n_k})_{k\in\mathbb{N}}$ und $(a_{m_k})_{k\in\mathbb{N}}$ ist.*

Dann ist eine Zahl $a \in \mathbb{R}$ genau dann ein Häufungspunkt von $(a_n)_{n\in\mathbb{N}}$, wenn a ein Häufungspunkt von $(a_{n_k})_{k\in\mathbb{N}}$ oder $(a_{m_k})_{k\in\mathbb{N}}$ ist.

Beweis. Eine Zahl a ist nach Lemma 5.20 genau dann ein Häufungspunkt von $(a_n)_n$, wenn in jeder ε-Umgebung von a unendlich viele Folgenglieder von $(a_n)_n$ liegen. Da $(a_n)_n$ eine Mischfolge von $(a_{n_k})_k$ und $(a_{m_k})_k$ ist, ist dies natürlich äquivalent dazu, dass in jeder solchen ε-Umgebung unendlich viele Folgenglieder von $(a_{n_k})_k$ oder $(a_{m_k})_k$ liegen, also dass a ein Häufungspunkt (mindestens) einer dieser beiden Teilfolgen ist. □

Beispiel 5.22.

(a) Die Folge $(a_n)_n = ((-1)^n)_{n\in\mathbb{N}}$ aus Beispiel 5.19 (a) ist eine Mischfolge ihrer positiven und negativen Glieder, die konstant gleich 1 bzw. -1 sind. Aus Lemma 5.21 folgt also, dass $(a_n)_n$ genau die beiden Häufungspunkte 1 und -1 hat.

(b) Die divergente Folge $(a_n)_n = (1,0,2,0,3,0,\dots)$ aus Beispiel 5.9 (a) ist eine Mischfolge der Folge $(1,2,3,\dots)$ (die nach Beispiel 5.19 (b) keine Häufungspunkte hat) und der konstanten Folge 0 (die natürlich 0 als einzigen Häufungspunkt hat). Damit hat $(a_n)_n$ nach Lemma 5.21 den einzigen Häufungspunkt 0. Wir sehen also (im Gegensatz zu Lemma 5.18), dass eine Folge mit genau einem Häufungspunkt nicht notwendig konvergiert.

5.B Konvergenzkriterien für Folgen

Nicht in allen Fällen lässt sich die Berechnung von Grenzwerten mit Hilfe der Grenzwertsätze auf bereits bekannte zurückführen. Wir benötigen daher noch weitere Techniken zur Grenzwertbestimmung und beginnen mit einem einfachen Vergleichskriterium.

Satz 5.23 (Verträglichkeit des Grenzwerts mit Ungleichungen)**.** *Es seien $(a_n)_n$ und $(b_n)_n$ konvergente Folgen mit $a_n \to a$ und $b_n \to b$. Dann gilt:*

(a) *Ist $a_n \le b_n$ für fast alle n, so auch $a \le b$.*

(b) *(**Einschachtelungssatz**) Ist $a = b$, konvergieren also beide Folgen gegen denselben Grenzwert, und ist $(c_n)_n$ eine weitere reelle Folge mit $a_n \le c_n \le b_n$ für fast alle n, so konvergiert auch $(c_n)_n$ gegen diesen Grenzwert.*

Beweis.

(a) Angenommen, es wäre $a > b$. Wir setzen $\varepsilon = \frac{a-b}{2}$. Wegen $a_n \to a$ und $b_n \to b$ wäre dann (nach Bemerkung 5.6) für fast alle n

$$a_n \in (a-\varepsilon, a+\varepsilon) \quad \text{und} \quad b_n \in (b-\varepsilon, b+\varepsilon).$$

Zusammensetzen liefert $a-\varepsilon < a_n \le b_n < b+\varepsilon$ für fast alle n, und damit $a-b < 2\varepsilon$ im Widerspruch zu $\varepsilon = \frac{a-b}{2}$.

(b) Es sei $\varepsilon > 0$ beliebig. Diesmal gilt wegen $a_n \to a$ und $b_n \to a$ für fast alle n

$$a_n \in (a-\varepsilon, a+\varepsilon) \quad \text{und} \quad b_n \in (a-\varepsilon, a+\varepsilon),$$

und damit $a-\varepsilon < a_n \le c_n \le b_n < a+\varepsilon$, also $c_n \in (a-\varepsilon, a+\varepsilon)$. Da $\varepsilon > 0$ beliebig war, folgt daraus wie behauptet $c_n \to a$. □

Bemerkung 5.24. Beachte, dass Satz 5.23 (a) *nicht* auch analog für die echte Ungleichung „$<$“ gilt: Ist z. B. $a_n = 0$ und $b_n = \frac{1}{n}$ für alle $n \ge 1$, so gilt zwar $a_n < b_n$ für alle n, aber die Grenzwerte beider Folgen sind natürlich gleich 0, d. h. es gilt nur $\lim_{n\to\infty} a_n \le \lim_{n\to\infty} b_n$ gemäß Satz 5.23 (a), aber nicht $\lim_{n\to\infty} a_n < \lim_{n\to\infty} b_n$.

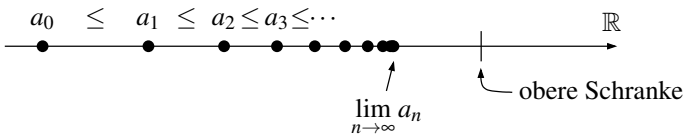
Aufgabe 5.25. Bestimme den Grenzwert

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{n}{n^2+k}.$$

Alle unsere bisherigen Kriterien haben den entscheidenden Nachteil, dass sie Grenzwerte nur auf andere bereits bekannte zurückführen können. In der Praxis werden aber viele Größen wie z. B. Quadratwurzeln, π, e oder der Sinus und Kosinus einer gegebenen Zahl überhaupt erst als Grenzwerte geeigneter Folgen konstruiert. Die Konvergenz solcher Folgen werden wir also mit unseren bisherigen Methoden nie nachweisen können.

Wir benötigen daher auch Kriterien, mit denen man die Konvergenz einer Folge selbst dann nachweisen kann, wenn man ihren Grenzwert noch nicht vorher kennt oder gleichzeitig aus bereits bekannten anderen Grenzwerten berechnen kann. Im Gegensatz zu unseren Ergebnissen aus Abschnitt 5.A, die unverändert auch in $\mathbb{Q}$ gelten würden, handelt es sich hierbei nun um Resultate, die ganz zentral das Supremumsaxiom verwenden und daher nur in $\mathbb{R}$ gelten.

Das erste Kriterium dieser Art, das wir jetzt behandeln wollen, ist für Folgen anwendbar, deren Folgenglieder mit wachsendem n immer größer werden. Ist eine solche Folge nach oben beschränkt, so ist anschaulich klar, dass die Folgenglieder wie im Bild unten für wachsendes n „immer näher zusammen rücken“ müssen, was letztlich zur Konvergenz der Folge führen sollte. Dies ist der Inhalt des folgenden Satzes.



Definition 5.26 (Monotone und beschränkte Folgen)**.** Es sei $(a_n)_n$ eine Folge.

(a) Die Folge $(a_n)_n$ heißt **monoton wachsend** oder **steigend**, wenn $a_n \leq a_{n+1}$ für alle $n \in \mathbb{N}$, also $a_0 \leq a_1 \leq a_2 \leq \cdots$ und damit $a_m \leq a_n$ für alle $m \leq n$ gilt. Gilt sogar $a_n < a_{n+1}$ für alle $n \in \mathbb{N}$, so heißt $(a_n)_n$ **streng monoton wachsend** oder **steigend**.

Analog heißt $(a_n)_n$ **(streng) monoton fallend**, wenn $a_n \geq a_{n+1}$ (bzw. $a_n > a_{n+1}$) für alle $n \in \mathbb{N}$ gilt.

(b) Analog zu Definition 5.7 heißt $(a_n)_n$ **nach oben beschränkt**, wenn die Menge ihrer Folgenglieder nach oben beschränkt ist, also wenn es ein $s \in \mathbb{R}$ gibt mit $a_n \leq s$ für alle n.

Analog heißt $(a_n)_n$ **nach unten beschränkt**, wenn es ein $s \in \mathbb{R}$ gibt mit $a_n \geq s$ für alle n; die Folge ist also genau dann beschränkt, wenn sie nach oben und unten beschränkt ist.

Satz 5.27 (**Monotoniekriterium**)**.** *Jede monoton wachsende, nach oben beschränkte Folge $(a_n)_n$ in $\mathbb{R}$ ist konvergent. (Analog ist dann natürlich auch jede monoton fallende, nach unten beschränkte Folge konvergent.)*

Beweis. Da die Menge $M := \{a_n : n \in \mathbb{N}\} \subset \mathbb{R}$ aller Folgenglieder nach oben beschränkt ist, existiert $a := \sup M$ nach dem Supremumsaxiom. Wir behaupten, dass $a_n \to a$.

Es sei dazu $\varepsilon > 0$ beliebig. Da a die kleinste obere Schranke für M ist, ist $a - \varepsilon$ keine obere Schranke mehr. Es gibt also ein $n_0 \in \mathbb{N}$ mit $a_{n_0} > a - \varepsilon$. Für alle $n \geq n_0$ folgt dann

$$\begin{aligned} a - \varepsilon < a_{n_0} &\leq a_n && \text{(Monotonie)} \\ &\leq a && (a \text{ ist obere Schranke der Folgenglieder}) \\ &< a + \varepsilon, \end{aligned}$$

also $|a_n - a| < \varepsilon$. Damit konvergiert $(a_n)_n$ gegen a. □

Als Beispiel für die Anwendung des Monotoniekriteriums wollen wir nun als Erstes zeigen, dass jede nicht-negative reelle Zahl eine Quadratwurzel besitzt. Auch wenn euch diese Tatsache aus der Schule vielleicht „offensichtlich“ erscheint, folgt sie dennoch nicht unmittelbar aus den Axiomen für $\mathbb{R}$ und muss damit bewiesen werden. Eine Möglichkeit dafür ist, eine Folge $(a_n)_n$ zu konstruieren, deren Konvergenz wir beweisen können, und deren Grenzwert nur die gewünschte Wurzel sein kann. Die Konstruktion von $(a_n)_n$ ist dabei *rekursiv*, d. h. wir können (analog zur vollständigen Induktion) nicht direkt a_n für alle $n \in \mathbb{N}$ angeben, sondern legen nur das erste Folgenglied a_0 fest und geben

dann eine Formel an, mit der für alle $n \in \mathbb{N}$ aus a_n das nächste Folgenglied a_{n+1} berechnet werden kann.

Lemma 5.28 (Existenz von Wurzeln)**.** *Es seien $c \in \mathbb{R}_{>0}$ und $a_0 \in \mathbb{R}_{>0}$ gegeben. Dann konvergiert die mit diesem Startwert a_0 rekursiv definierte Folge $(a_n)_n$ mit*

$$a_{n+1} := \frac{1}{2}\left(a_n + \frac{c}{a_n}\right) \quad \text{für alle } n \in \mathbb{N} \tag{$*$}$$

gegen eine positive reelle Zahl a mit $a^2 = c$.

Beweis. Wir zeigen die Konvergenz von $(a_n)_n$ mit dem Monotoniekriterium, indem wir nachweisen, dass die Folge nach unten beschränkt und monoton fallend ist.

Aus der Rekursionsvorschrift $(*)$ ist offensichtlich, dass mit c und a_0 auch alle Folgenglieder positiv sind. Die Folge ist also sicher durch 0 nach unten beschränkt. In der Tat gilt für alle $n \in \mathbb{N}$ sogar

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

und somit $a_n^2 \geq c$ für alle $n \geq 1$. Daraus folgt für alle $n \geq 1$ aber auch

$$\frac{a_{n+1}}{a_n} = \frac{1}{2}\left(1 + \frac{c}{a_n^2}\right) \leq \frac{1}{2}\left(1 + \frac{c}{c}\right) = 1,$$

und damit $a_{n+1} \leq a_n$, die Folge ist also (mit Ausnahme evtl. des ersten Folgengliedes) monoton fallend. Damit konvergiert $(a_n)_n$ nach Satz 5.27, d. h. der Grenzwert $a := \lim_{n\to\infty} a_n$ existiert.

Um Informationen über den Grenzwert a zu bekommen, multiplizieren wir die Rekursionsgleichung $(*)$ zunächst mit a_n und erhalten $a_{n+1}a_n = \frac{1}{2}(a_n^2 + c)$ für alle $n \in \mathbb{N}$. Gehen wir in dieser Gleichung nun zum Grenzwert über, so ergibt sich

$$\lim_{n\to\infty} a_{n+1}a_n = \lim_{n\to\infty} \frac{1}{2}\left(a_n^2 + c\right), \qquad \text{und damit} \qquad a^2 = \frac{1}{2}\left(a^2 + c\right),$$

da die Folge $(a_{n+1})_n = (a_1, a_2, a_3, \dots)$ gegenüber $(a_n)_n = (a_0, a_1, a_2, \dots)$ ja nur um ein Folgenglied verschoben ist und somit als Teilfolge von $(a_n)_n$ ebenfalls gegen a konvergiert. Auflösen dieser Gleichung liefert nun sofort wie behauptet $a^2 = c$. Da alle a_n positiv sind, ergibt Satz 5.23 (a) außerdem auch $a \geq 0$; in der Tat ist wegen $a^2 = c > 0$ dann sogar $a > 0$. □

Folgerung und Definition 5.29. *Zu jedem $c \in \mathbb{R}_{\geq 0}$ gibt es eine eindeutig bestimmte nicht-negative reelle Zahl a mit $a^2 = c$. Wir nennen sie die **(Quadrat-)Wurzel** aus c und schreiben sie als $\sqrt{c}$.*

Beweis. Für $c = 0$ ist die Aussage mit $a = 0$ klar, daher können wir im Folgenden $c > 0$ annehmen. Die Existenz einer Wurzel a folgt dann direkt aus Lemma 5.28. Das Polynom $x \mapsto x^2 - c$ hat nun die positive Nullstelle a und die negative Nullstelle $-a$, und kann als Polynom vom Grad 2 nach Satz 3.18 (b) keine weiteren Nullstellen haben. Also ist die Wurzel a auch eindeutig bestimmt. □

Bemerkung 5.30 (Grenzwert rekursiver Folgen)**.** Das Monotoniekriterium bietet sich oft für rekursiv definierte Folgen $(a_n)_n$ wie in Lemma 5.28 an, da die Monotonie ja durch den Vergleich von a_{n+1} und a_n nachgewiesen werden kann. Sehr nützlich ist dabei auch der Trick, wie im Beweis des Lemmas in der Rekursionsgleichung zum Grenzwert überzugehen, um eine bestimmende Gleichung für den Grenzwert zu finden.

Beispiel 5.31. Die folgende Tabelle zeigt den Anfang der Folge aus Lemma 5.28 im Fall $c = 2$ und $a_0 = 1$. Beachte, dass die Folge „extrem schnell“ konvergiert und daher sehr gut zur näherungsweisen Berechnung von Wurzeln geeignet ist – z. B. wenn man einem Computer, der bisher nur weiß, wie man die vier Grundrechenarten ausführt, das Wurzelziehen beibringen möchte. In der Tat kann man zeigen, dass jeder Schritt die Anzahl der korrekten Dezimalstellen (die in der Tabelle unten fett gedruckt sind) näherungsweise verdoppelt. Wir werden uns in dieser Vorlesung jedoch nicht weiter mit der „Geschwindigkeit“ der Konvergenz von Folgen beschäftigen; derartige Fragestellungen werdet ihr später in den Vorlesungen zur Praktischen Mathematik untersuchen.

n	a_n
0	**1,**0000000000000000000000000000000000000000000000000...
1	**1,**5000000000000000000000000000000000000000000000000...
2	**1,41**6666666666666666666666666666666666666666666666666...
3	**1,4142**1568627450980392156862745098039215686274509 8...
4	**1,414213562374**68991062629557889013491011655962211 5...
5	**1,41421356237309504880168**9623502530243614981925776...
6	**1,4142135623730950488016887242096980785696718753 77**...

Aufgabe 5.32. Untersuche, ob die folgenden rekursiv definierten Folgen $(a_n)_n$ konvergieren, und bestimme im Fall der Konvergenz den Grenzwert.

(a) $a_0 = 1$, $a_n = \sqrt{a_{n-1}+6}$ für alle $n \geq 1$;

(b) $a_0 = 1$, $a_1 = 2$, $a_n = \frac{1}{2}(a_{n-1}+a_{n-2})$ für alle $n \geq 2$.

(Hinweis: Leite zunächst eine Formel für die Differenz $a_n - a_{n-1}$ zweier aufeinander folgender Glieder her.)

Aufgabe 5.33. Es seien $a, b \in \mathbb{R}_{\geq 0}$. Zeige, dass die Quadratwurzel die folgenden Eigenschaften hat:

(a) Gilt $a \leq b$, so auch $\sqrt{a} \leq \sqrt{b}$ (Monotonie).

(b) $\sqrt{ab} = \sqrt{a} \cdot \sqrt{b}$.

(c) $\sqrt{ab} \leq \frac{a+b}{2}$.

(d) Konvergiert eine Folge $(a_n)_n$ in $\mathbb{R}_{\geq 0}$ gegen a, so gilt auch $\lim_{n\to\infty} \sqrt{a_n} = \sqrt{a}$.

Aufgabe 5.34.

(a) Man zeige: Für alle $a, b \in \mathbb{Z}$ und $k, n \in \mathbb{N}$ ist $(a+b\sqrt{k})^n + (a-b\sqrt{k})^n \in \mathbb{Z}$.

(b) Bestimme die 100. Nachkommastelle (im Dezimalsystem) von $(2+\sqrt{5})^{2015}$.

Aufgabe 5.35. Zeige, dass jedes (nicht-leere) offene Intervall in $\mathbb{R}$ unendlich viele rationale und unendlich viele irrationale Zahlen enthält.

Aufgabe 5.36 (Existenz höherer Wurzeln in $\mathbb{R}$)**.** In dieser Aufgabe wollen wir analog zu Lemma 5.28 und Folgerung 5.29 beweisen, dass jede nicht-negative reelle Zahl c für alle $k \in \mathbb{N}_{>0}$ eine eindeutige k-te Wurzel besitzt. Wir definieren dazu für ein gegebenes $c \in \mathbb{R}_{>0}$, $k \in \mathbb{N}_{>0}$ und einen beliebigen positiven Startwert a_0 die Folge $(a_n)_n$ rekursiv durch

$$a_{n+1} := \frac{1}{k}\left((k-1)\,a_n + \frac{c}{a_n^{k-1}}\right) \quad \text{für alle } n \in \mathbb{N}.$$

Man beweise nun:

(a) Für alle $n \in \mathbb{N}$ ist $a_{n+1}^k \geq c$.

(b) Die Folge $(a_n)_n$ ist ab dem zweiten Folgenglied monoton fallend.

(c) Zu jeder Zahl $c \in \mathbb{R}_{\geq 0}$ gibt es ein eindeutiges $a \in \mathbb{R}_{\geq 0}$ mit $a^k = c$. Wir nennen dieses a die k-te **Wurzel** aus c und schreiben sie als $\sqrt[k]{c}$.

Aufgabe 5.37. Es sei $M \subset \mathbb{R}$ nicht leer und nach oben beschränkt. Zeige, dass für $s \in \mathbb{R}$ die folgenden Aussagen äquivalent sind:

(a) $s = \sup M$.

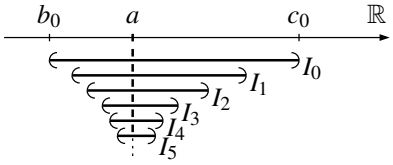
(b) s ist eine obere Schranke für M, und es gibt eine Folge $(a_n)_n$ von Elementen aus M mit $\lim\limits_{n\to\infty} a_n = s$.

10

Kombiniert man die Monotoniekriterien für wachsende und fallende Folgen miteinander, kann man einen Grenzwert wie folgt von beiden Seiten einschachteln.

Satz 5.38 (**Intervallschachtelung**). *Gegeben sei für alle $n \in \mathbb{N}$ ein abgeschlossenes Intervall $I_n = [b_n, c_n]$ in $\mathbb{R}$, so dass $I_0 \supset I_1 \supset I_2 \supset \cdots$ (also die Intervalle ineinander liegen) und $\lim_{n\to\infty}(c_n - b_n) = 0$ (also die Längen der Intervalle gegen 0 konvergieren).*

Dann gibt es genau ein $a \in \mathbb{R}$, das in allen diesen Intervallen liegt, und es gilt $b_n \to a$ und $c_n \to a$.



Beweis. Die Folge $(b_n)_n$ der unteren Intervallgrenzen ist monoton wachsend und nach oben beschränkt (z. B. durch c_0), nach dem Monotoniekriterium aus Satz 5.27 also konvergent gegen ein $b \in \mathbb{R}$. Genauso ist $(c_n)_n$ monoton fallend und nach unten beschränkt, und somit konvergent gegen ein $c \in \mathbb{R}$. Da die Längen der Intervalle gegen 0 konvergieren, folgt nach Satz 5.13 aber

$$0 = \lim_{n\to\infty}(c_n - b_n) = c - b,$$

und damit $b = c$. Wir bezeichnen den gemeinsamen Grenzwert dieser beiden Folgen mit $a := b = c$.

Nach dem Beweis von Satz 5.27 ist a eine obere Schranke für alle b_n und eine untere Schranke für alle c_n. Es gilt also $a \in [b_n, c_n] = I_n$ für alle n. Ist umgekehrt $a' \in \mathbb{R}$ mit $a' \in I_n$ und damit $b_n \le a' \le c_n$ für alle n, so folgt daraus durch Grenzwertbildung mit Satz 5.23 auch $a \le a' \le a$, also $a' = a$. Somit gibt es genau eine Zahl in allen gegebenen Intervallen, nämlich a. □

Zum Abschluss dieses Abschnitts wollen wir analog zu den uneigentlichen Suprema in Bemerkung 4.37 auch *uneigentliche Grenzwerte* definieren, also festlegen, was es heißt, dass eine Folge „den Grenzwert ∞ besitzt". Dies hat den Vorteil, dass viele Aussagen über konvergente Folgen mit gewöhnlichen Grenzwerten in $\mathbb{R}$ auf diesen Fall verallgemeinert werden können.

Definition 5.39 (Uneigentliche Grenzwerte von Folgen). Für eine Folge $(a_n)_n$ in $\mathbb{R}$ schreiben wir $a_n \to \infty$ bzw. $\lim_{n\to\infty} a_n = \infty$, wenn

$$\forall s \in \mathbb{R}\ \exists n_0 \in \mathbb{N}\ \forall n \ge n_0 : a_n > s,$$

also wenn zu jeder vorgegebenen Schranke s fast alle Folgenglieder größer als s sind. Analog definiert man die Eigenschaft $\lim_{n\to\infty} a_n = -\infty$.

Beachte, dass derartige Folgen natürlich insbesondere unbeschränkt und damit nach Lemma 5.8 divergent sind. Man bezeichnet sie als **bestimmt divergent** und nennt ∞ bzw. $-\infty$ einen **uneigentlichen Grenzwert**. Ist $(a_n)_n$ divergent und besitzt nicht in obigem Sinne den Grenzwert ∞ oder $-\infty$, so nennt man $(a_n)_n$ **unbestimmt divergent**.

Beispiel 5.40. Die Folge $(a_n)_n = (n)_n = (0, 1, 2, 3, \dots)$ ist bestimmt divergent mit uneigentlichem Grenzwert $\lim_{n\to\infty} a_n = \infty$. Die Folge $(b_n)_n = (1, 0, 2, 0, 3, 0, \dots)$ aus Beispiel 5.9 (a) ist dagegen unbestimmt divergent (da z. B. nicht fast alle Folgenglieder größer als 1 sind).

Bemerkung 5.41 (Grenzwertsätze für uneigentliche Grenzwerte). Die Grenzwertsätze aus Satz 5.13 gelten auch für uneigentliche Grenzwerte wie in Definition 5.39, wenn man die formalen Rechenregeln für ∞

$$\begin{aligned} a + \infty &= \infty && \text{für } a \in \mathbb{R}, \\ \infty + \infty &= \infty, \\ a \cdot \infty &= \infty && \text{für } a \in \mathbb{R}_{>0}, \\ \infty \cdot \infty &= \infty, \\ \frac{a}{\infty} &= 0 && \text{für } a \in \mathbb{R} \end{aligned}$$

und analog für $-\infty$ bzw. $a \in \mathbb{R}_{<0}$ definiert. Die Beweise dieser Aussagen sind letztlich analog zu denen von Satz 5.13, jedoch in den einzelnen Fällen immer etwas unterschiedlich, da die Bedingung für den Grenzwert ∞ aus Definition 5.39 ja formal anders aussieht als die eines endlichen Grenzwerts in Definition 5.1. Wir werden die Beweise hier nur exemplarisch in Aufgabe 5.42 betrachten.

Beachte aber, dass die Grenzwertsätze auch weiterhin keine Aussage liefern, wenn eine der betrachteten Folgen unbestimmt divergent ist oder sich Ausdrücke der Form $\infty - \infty$, $0 \cdot \infty$ oder $\frac{\infty}{\infty}$ ergeben, die sich nicht sinnvoll definieren lassen.

Aufgabe 5.42. Es seien $(a_n)_n$ und $(b_n)_n$ zwei reelle Zahlenfolgen.

(a) Man zeige: Gilt $a_n \to a \in \mathbb{R}_{>0}$ und $b_n \to \infty$, so ist auch $a_n b_n \to \infty$.

(b) Man zeige: Gilt $a_n \to \infty$ und $b_n \to \infty$, so ist auch $a_n + b_n \to \infty$.

(c) Kann man in (b) die Bedingung des Grenzwerts ∞ auch durch „nach oben unbeschränkt" ersetzen, d. h. gilt für nach oben unbeschränkte Folgen $(a_n)_n$ und $(b_n)_n$ auch, dass $(a_n + b_n)_n$ nach oben unbeschränkt ist?

Bemerkung 5.43 (Monotoniekriterium mit uneigentlichem Grenzwert)**.** Auch das Monotoniekriterium lässt sich auf eine Variante mit uneigentlichen Grenzwerten erweitern: Ist eine reelle Folge $(a_n)_n$ zwar wie in Satz 5.27 monoton wachsend, aber nicht nach oben beschränkt, so gibt es zu jedem $s \in \mathbb{R}$ zunächst ein $n_0 \in \mathbb{N}$ mit $a_{n_0} > s$, und wegen der Monotonie dann auch mit $a_n > s$ für alle $n \geq n_0$. Nach Definition 5.39 ist damit dann also $\lim_{n\to\infty} a_n = \infty$. Wir können Satz 5.27 also dahingehend verallgemeinern, dass jede monotone reelle Folge einen Grenzwert in $\mathbb{R} \cup \{\pm\infty\}$ hat, also konvergent oder bestimmt divergent ist.

5.C Limes superior und inferior

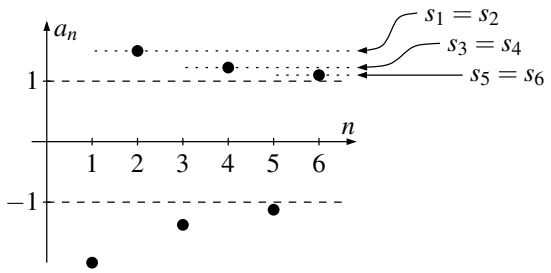
Bisher haben wir das Verhalten von Folgen „im Unendlichen" in der Regel durch ihren Grenzwert beschrieben. Selbst wenn wir hierbei uneigentliche Grenzwerte wie in Definition 5.39 zulassen, funktioniert dies aber natürlich nicht bei unbestimmt divergenten Folgen, die keinen solchen Grenzwert besitzen.

Um solche Folgen zu untersuchen, können wir wie in Definition 5.17 (c) versuchen, auf Häufungspunkte auszuweichen. In der Tat wollen wir jetzt zeigen, dass *jede* Folge (zumindest im uneigentlichen Sinne) auch wirklich mindestens einen Häufungspunkt besitzt. Der Einfachheit halber betrachten wir hierfür zunächst nur beschränkte reelle Folgen, deren Häufungspunkte dann also in $\mathbb{R}$ liegen müssen. In diesem Fall werden wir in Folgerung 5.47 sehen, dass die folgende Konstruktion stets einen Häufungspunkt liefert – und zwar sogar einen ganz bestimmten, nämlich den größten.

Konstruktion 5.44. Es sei $(a_n)_n$ eine beschränkte reelle Folge. Wie im Bild unten für die Folge $(a_n)_n$ mit $a_n = (-1)^n (1 + \frac{1}{n})$ dargestellt konstruieren wir nun die Hilfsfolge $(s_n)_n$ durch

$$s_n := \sup\{a_k : k \geq n\},$$

d. h. wir betrachten das Supremum aller Folgenglieder, wobei wir aber für s_n erst beim n-ten Folgenglied anfangen.



Beachte dabei, dass die Mengen $\{a_k : k \geq n\}$ nach Voraussetzung beschränkt sind und die Suprema s_n damit nach dem Supremumsaxiom in $\mathbb{R}$ existieren. Außerdem folgt aus der Beschränktheit der Folgenglieder, dass auch $(s_n)_n$ eine beschränkte Folge ist.

Darüber hinaus ist die Folge $(s_n)_n$ monoton fallend: Für $m \geq n$ ist die obere Schranke s_n der Menge $\{a_k : k \geq n\}$ ja auch eine obere Schranke der Teilmenge $\{a_k : k \geq m\} \subset \{a_k : k \geq n\}$ und muss damit größer oder gleich der kleinsten oberen Schranke s_m von $\{a_k : k \geq m\}$ sein – d. h. es ist $s_m \leq s_n$.

Wir haben also gesehen, dass $(s_n)_n$ eine monoton fallende und (nach unten) beschränkte reelle Folge ist. Nach dem Monotoniekriterium aus Satz 5.27 besitzt sie also einen Grenzwert. Im oben dargestellten Beispiel ist dieser Grenzwert offensichtlich 1. Da die Häufungspunkte unserer Beispielfolge nach dem Mischfolgenlemma 5.21 genau ± 1 sind, ist der Grenzwert von $(s_n)_n$ hier also gerade der größte Häufungspunkt von $(a_n)_n$. Bevor wir zeigen, dass dies immer der Fall ist, geben wir der so konstruierten Zahl noch einen Namen.

Definition 5.45 (Limes superior und Limes inferior)**.** Für eine beschränkte reelle Folge $(a_n)_n$ definieren wir

$$\text{den } \textbf{Limes superior} \quad \limsup_{n\to\infty} a_n := \lim_{n\to\infty} \big(\sup\{a_k : k \geq n\}\big)$$

$$\text{und} \quad \text{den } \textbf{Limes inferior} \quad \liminf_{n\to\infty} a_n := \lim_{n\to\infty} \big(\inf\{a_k : k \geq n\}\big).$$

In der Literatur sind hierfür auch die Schreibweisen $\overline{\lim}\, a_n$ bzw. $\underline{\lim}\, a_n$ üblich.

Das folgende Lemma zeigt, dass sich der Limes superior (und analog der Limes inferior) in gewissem Sinne wie eine „Mischung“ aus einem Grenzwert und einem Häufungspunkt verhält: Während für einen Grenzwert a ja in jeder ε-Umgebung $U_\varepsilon(a)$ fast alle Folgenglieder, für einen Häufungspunkt nach Lemma 5.20 aber nur unendlich viele Glieder liegen müssen, ist der Limes superior die (eindeutig bestimmte) Zahl a, für die für alle ε fast alle Folgenglieder kleiner als $a+\varepsilon$, und unendlich viele größer als $a-\varepsilon$ sind.

Lemma 5.46. *Es seien $(a_n)_n$ eine beschränkte reelle Folge und $a \in \mathbb{R}$. Dann gilt $a = \limsup_{n\to\infty} a_n$ genau dann, wenn für alle $\varepsilon > 0$ die folgenden beiden Bedingungen erfüllt sind:*

(a) *Für fast alle n ist $a_n < a+\varepsilon$.*

(b) *Für unendlich viele n ist $a_n > a-\varepsilon$.*

(Eine analoge Aussage gilt natürlich auch für den Limes inferior.)

Beweis. Wie in Konstruktion 5.44 sei $s_n = \sup\{a_k : k \geq n\}$, so dass also $\limsup_{n\to\infty} a_n = \lim_{n\to\infty} s_n$. Weiterhin sei $\varepsilon > 0$ beliebig.

„$\Rightarrow$“: Es gelte $s_n \to a$, also $s_n \in (a-\varepsilon, a+\varepsilon)$ für fast alle n. Wir müssen (a) und (b) zeigen.

Da s_n eine obere Schranke für $\{a_k : k \geq n\}$ (also insbesondere für a_n) ist, gilt $a_n \leq s_n < a+\varepsilon$ für fast alle n. Dies zeigt (a).

Um (b) zu zeigen, nehmen wir an, es gäbe nur endlich viele n mit $a_n > a-\varepsilon$. Dann gäbe es also ein $n_0 \in \mathbb{N}$ mit $a_n \leq a-\varepsilon$ für alle $n \geq n_0$, d. h. $a-\varepsilon$ wäre eine obere Schranke für alle diese Folgenglieder. Daraus folgt dann aber auch $s_n \leq a-\varepsilon$ für alle $n \geq n_0$, im Widerspruch zu $s_n \to a$.

„$\Leftarrow$“: Wir setzen nun (a) und (b) voraus und müssen $s_n \in (a-\varepsilon, a+\varepsilon)$ für fast alle n zeigen.

Nach (a) gibt es ein $n_0 \in \mathbb{N}$ mit $a_n < a + \frac{\varepsilon}{2}$ für alle $n \geq n_0$. Damit ist für diese n auch $s_n = \sup\{a_k : k \geq n\} \leq a + \frac{\varepsilon}{2} < a+\varepsilon$.

Weiterhin gibt es nach (b) zu jedem $n \in \mathbb{N}$ ein $k \geq n$ mit $a_k > a-\varepsilon$, woraus natürlich auch $s_n = \sup\{a_k : k \geq n\} > a-\varepsilon$ folgt.

Insgesamt gilt also $s_n \in (a-\varepsilon, a+\varepsilon)$ für fast alle n. □

Aus diesem Lemma ergibt sich nun die folgende wichtige Charakterisierung des Limes superior (und inferior), mit der man diese Zahl oft sehr einfach bestimmen kann.

Folgerung 5.47. *Für jede beschränkte reelle Folge $(a_n)_n$ ist $\limsup_{n\to\infty} a_n$ der größte Häufungspunkt von $(a_n)_n$. (Analog ist dann natürlich $\liminf_{n\to\infty} a_n$ der kleinste Häufungspunkt von $(a_n)_n$.)*

*Insbesondere besitzt also jede beschränkte reelle Folge einen Häufungspunkt (**Satz von Bolzano-Weierstraß** für $\mathbb{R}$).*

Beweis. Es seien $a = \limsup_{n\to\infty} a_n$ und $\varepsilon > 0$ beliebig. Aus (a) und (b) von Lemma 5.46 folgt dann $a - \varepsilon < a_n < a + \varepsilon$ für unendlich viele n, d. h. nach Lemma 5.20 ist a ein Häufungspunkt von $(a_n)_n$.

Andererseits ist aber kein $b > a$ ein Häufungspunkt von $(a_n)_n$: Setzen wir nämlich $\varepsilon = \frac{b-a}{2} > 0$, so ist $a + \varepsilon = b - \varepsilon$, und somit gilt nach Lemma 5.46 (a) für fast alle n

$$a_n < a + \varepsilon = b - \varepsilon \quad \Rightarrow \quad a_n \notin (b - \varepsilon, b + \varepsilon).$$

Damit kann b kein Häufungspunkt von $(a_n)_n$ sein. □

Beispiel 5.48.

(a) Die Folge $(a_n)_{n\in\mathbb{N}_{>0}}$ mit $a_n = (-1)^n(1 + \frac{1}{n})$ ist eine Mischfolge aus den geraden Gliedern $a_{2n} = 1 + \frac{1}{2n} \to 1$ und den ungeraden Gliedern $a_{2n+1} = -(1 + \frac{1}{2n+1}) \to -1$, sie hat nach Lemma 5.21 also die einzigen Häufungspunkte 1 und -1. Aus Folgerung 5.47 ergibt sich damit sofort $\limsup_{n\to\infty} a_n = 1$ und $\liminf_{n\to\infty} a_n = -1$.

(b) Ist $(a_n)_n$ eine konvergente reelle Folge, so ist ihr Grenzwert a nach Lemma 5.18 der einzige Häufungspunkt. Also ist dann $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} = a$ nach Folgerung 5.47.

Ist umgekehrt $(a_n)_n$ eine beschränkte reelle Folge mit $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} =: a$, so folgt aus Lemma 5.46 (a) für alle $\varepsilon > 0$, dass $a - \varepsilon < a_n < a + \varepsilon$ für fast alle n ist – wobei sich die erste Ungleichung aus $\liminf_{n\to\infty} a_n = a$ und die zweite aus $\limsup_{n\to\infty} a_n = a$ ergibt. Also ist $(a_n)_n$ dann konvergent mit Grenzwert a.

Aufgabe 5.49. Berechne $\limsup_{n\to\infty} a_n$ und $\liminf_{n\to\infty} a_n$ für $a_n = \frac{1}{\sqrt{n^2+n+(-1)^n\cdot n}}$.

Aufgabe 5.50.

(a) Es seien $(a_n)_n$ und $(b_n)_n$ zwei beschränkte Folgen positiver Zahlen. Zeige, dass

$$\limsup_{n\to\infty}(a_n b_n) \le \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n,$$

und gib ein Beispiel dafür an, dass hier im Allgemeinen keine Gleichheit gilt.

(b) Beweise, dass in (a) jedoch stets die Gleichheit gilt, wenn $(a_n)_n$ oder $(b_n)_n$ konvergent ist.

Bemerkung 5.51 (Uneigentliche Werte für Limes superior und Limes inferior)**.** Lässt man für Supremum, Infimum und Grenzwerte wie in Bemerkung 4.37 und Definition 5.39 formal auch $\pm\infty$ zu, so kann man den Limes superior und Limes inferior nach Bemerkung 5.43 genau wie in Definition 5.45 auch für beliebige reelle Folgen konstruieren und erhält dann Werte in $\mathbb{R} \cup \{\pm\infty\}$.

Aufgabe 5.52. Zeige, dass die Aussage von Folgerung 5.47 auch für beliebige reelle Folgen richtig ist, wenn man wie in Bemerkung 5.51 die uneigentlichen Werte $\pm\infty$ für Häufungspunkte sowie den Limes superior und inferior zulässt.

Insbesondere gibt es also auch vom Satz von Bolzano-Weierstraß die erweiterte Form, dass jede (nicht notwendig beschränkte) Folge einen (evtl. uneigentlichen) Häufungspunkt hat.

11

5.D Mächtigkeiten von Mengen

Zum Abschluss dieses Kapitels wollen wir noch ein Thema behandeln, das wir auch schon deutlich früher hätten ansprechen können, aber das durch unsere Ergebnisse zu Folgen nun einfacher zu untersuchen ist: die Frage nach der „Größe“ von Mengen. Gibt es zu zwei Mengen M und N eine bijektive Abbildung $f\colon M \to N$ und damit wie im rechten Bild von Definition 2.8 eine 1:1-Beziehung zwischen ihren Elementen, so können wir uns diese beiden Mengen anschaulich als „gleich groß“ vorstellen. Ist z. B. $M = \{x_1, \dots, x_n\}$ eine endliche Menge mit n Elementen, so ist $N = \{f(x_1), \dots, f(x_n)\}$ (da f surjektiv ist und somit alle Elemente von N trifft) – und in dieser Aufzählung der Elemente von N steht auch kein Element doppelt, da f injektiv ist. Also hat N dann ebenfalls n Elemente, d. h. genauso viele Elemente wie M.

Wir wollen dieses Konzept nun für unendliche Mengen untersuchen. Ist es auch in diesem Fall noch sinnvoll, sich zwei Mengen M und N als „gleich groß“ vorzustellen, wenn eine bijektive Abbildung

$f\colon M \to N$ zwischen ihnen existiert? Gibt es überhaupt „verschieden große unendliche Mengen“? Da einen die Intuition bei derartigen Fragen schnell in die Irre leitet, sollten wir zunächst erst einmal exakt definieren, worüber wir reden wollen.

Definition 5.53 (Gleichmächtige und abzählbare Mengen)**.**

(a) Zwei Mengen M und N heißen **gleichmächtig**, wenn es zwischen ihnen eine bijektive Abbildung $f\colon M \to N$ gibt.

(b) Eine Menge M heißt ...

- **abzählbar unendlich**, wenn sie gleichmächtig zu $\mathbb{N}$ ist, also wenn es eine bijektive Abbildung $f\colon \mathbb{N} \to M$ gibt.
- **abzählbar**, wenn sie endlich oder abzählbar unendlich ist.
- **überabzählbar**, wenn sie nicht abzählbar ist.

Bemerkung 5.54. Die Gleichmächtigkeit erfüllt formal die Eigenschaften einer Äquivalenzrelation:

(a) Jede Menge M ist gleichmächtig zu sich selbst (mit der Identität $\mathrm{id}_M\colon M \to M$).

(b) Ist M gleichmächtig zu N, so ist auch N gleichmächtig zu M (denn mit einer bijektiven Abbildung $f\colon M \to N$ ist nach Aufgabe 2.23 auch ihre Umkehrung $f^{-1}\colon N \to M$ bijektiv).

(c) Ist M gleichmächtig zu N und N gleichmächtig zu R, so ist auch M gleichmächtig zu R (denn die Verkettung bijektiver Abbildungen ist nach Aufgabe 2.23 wieder bijektiv).

Man könnte daher versucht sein zu sagen, dass die Gleichmächtigkeit eine Äquivalenzrelation auf der Menge aller Mengen ist. Dies ist jedoch nicht ganz korrekt, da wir wie in Bemerkung 1.14 erläutert keine „Menge aller Mengen“ bilden können.

Beispiel 5.55.

(a) Wie wir am Anfang dieses Abschnitts gesehen haben, sind zwei endliche Mengen M und N genau dann gleichmächtig, wenn sie gleich viele Elemente haben, also wenn $|M| = |N|$ gilt. Insbesondere ist eine endliche Menge also nie gleichmächtig zu einer echten Teilmenge von ihr: Wenn wir von einer endlichen Menge Elemente entfernen, wird sie in diesem Sinne „kleiner“ – was natürlich nicht allzu überraschend sein sollte.

(b) Für unendliche Mengen ist dies jedoch falsch: Die Menge $\mathbb{N} = \{0,1,2,3,\dots\}$ ist gleichmächtig zu ihrer echten Teilmenge $\mathbb{N}\backslash\{0\} = \{1,2,3,\dots\}$, z. B. durch die bijektive Abbildung

$$f\colon \mathbb{N} \to \mathbb{N}\backslash\{0\},\ n \mapsto n+1,$$

die jede Zahl um 1 erhöht.

Bemerkung 5.56.

(a) Die abzählbar unendlichen Mengen sind genau diejenigen, die sich als Aufzählung in der Form $M = \{x_0, x_1, x_2, \dots\}$ mit $x_m \neq x_n$ für alle $m,n \in \mathbb{N}$ mit $m \neq n$ schreiben lassen: Die Funktion $f\colon \mathbb{N} \to M$, $n \mapsto x_n$ ist dann die geforderte bijektive Abbildung. Dies erklärt auch den Begriff „*abzählbar* unendlich“. Wir sehen so auch z. B. schon, dass auch die Menge $\mathbb{Z}$ der ganzen Zahlen abzählbar ist, da wir sie z. B. als $\mathbb{Z} = \{0,-1,1,-2,2,-3,3,\dots\}$ schreiben können.

(b) Aus (a) ergibt sich direkt, dass jede Teilmenge M einer abzählbaren Menge N wieder abzählbar ist: Ist $N = \{x_0, x_1, x_2, \dots\}$ abzählbar (mit einer evtl. abbrechenden Aufzählung), so lässt sich jede Teilmenge $M \subset N$ durch Weglassen gewisser Elemente als $M = \{x_{n_0}, x_{n_1}, x_{n_2}, \dots\}$ für geeignete (evtl. endlich viele) $n_0 < n_1 < n_2 < \cdots$ schreiben, ist damit also ebenfalls abzählbar.

Abzählbare Mengen bleiben aber nicht nur abzählbar, wenn man Elemente von ihnen entfernt. Man kann sie umgekehrt auch noch um „sehr viele“ Elemente vergrößern, ohne dass sie dadurch überabzählbar werden. Konkret wollen wir jetzt zeigen, dass wir sogar abzählbar viele abzählbare Mengen vereinigen können und dabei immer noch eine abzählbare Menge erhalten. Im folgenden Satz sind

diese abzählbaren Mengen dazu mit M_i bezeichnet, wobei i die Mengen durchnummeriert und damit selbst abzählbar viele Werte (in der sogenannten Indexmenge) annehmen kann.

Satz 5.57 (Abzählbare Vereinigungen abzählbarer Mengen sind abzählbar)**.** *Es seien I eine abzählbare Indexmenge sowie M_i für alle $i \in I$ eine abzählbare Menge. Dann ist die Vereinigung aller dieser Mengen M_i, geschrieben $\bigcup_{i\in I} M_i$, ebenfalls abzählbar.*

Beweis. Nach Bemerkung 5.56 (a) können wir die Elemente von I sowie allen M_i mit $i \in I$ in der Form

$$I = \{i_0, i_1, i_2, \dots\} \qquad \text{und} \qquad M_{i_k} = \{x_{k,0}, x_{k,1}, x_{k,2}, \dots\} \quad \text{für } k \in \mathbb{N}$$

aufzählen (wobei einige dieser Mengen auch endlich sein können, so dass die Aufzählungen dann also irgendwo abbrechen). Wir können die Elemente aller M_i damit in der folgenden Form auflisten und abzählen:

$$\begin{array}{lllllll}
M_{i_0}: & x_{0,0} & x_{0,1} & x_{0,2} & x_{0,3} & x_{0,4} & \cdots \\
M_{i_1}: & x_{1,0} & x_{1,1} & x_{1,2} & x_{1,3} & & \\
M_{i_2}: & x_{2,0} & x_{2,1} & x_{2,2} & & & \\
M_{i_3}: & x_{3,0} & & & & &
\end{array}$$

Wir haben also

$$\bigcup_{i\in I} M_i = \{x_{0,0},\ x_{0,1}, x_{1,0},\ x_{0,2}, x_{1,1}, x_{2,0},\ x_{0,3}, x_{1,2}, x_{2,1}, x_{3,0},\ x_{0,4}, x_{1,3}, x_{2,2}, \dots\}.$$

Dabei müssen wir in dieser Aufzählung alle nicht vorhandenen Positionen (wenn einige der Mengen I oder M_i endlich sind) und bereits vorher vorgekommene Elemente (wenn die M_i nicht disjunkt sind) weglassen. Auf diese Art sehen wir also, dass $\bigcup_{i\in I} M_i$ endlich oder abzählbar unendlich sein muss. Man bezeichnet die obige Abzählart auch als das **Cantorsche Diagonalverfahren**. □

Beispiel 5.58.

(a) Sind M und N abzählbare Mengen, so ist nach Satz 5.57 auch ihr Produkt $M \times N$ abzählbar, da man es als abzählbare Vereinigung $\bigcup_{m\in M}(\{m\} \times N)$ abzählbarer Mengen schreiben kann.

(b) Für ein festes $q \in \mathbb{N}_{>0}$ ist die Menge $M_q = \left\{\frac{p}{q} : p \in \mathbb{Z}\right\}$ aller rationalen Zahlen, die sich als Bruch mit Nenner q schreiben lassen, bijektiv zu $\mathbb{Z}$ und damit nach Bemerkung 5.56 (a) abzählbar. Damit ist nach Satz 5.57 auch die Menge $\mathbb{Q} = \bigcup_{q\in\mathbb{N}_{>0}} M_q$ aller rationalen Zahlen abzählbar. Auch wenn es der ersten Intuition vermutlich widerspricht, gibt es in diesem Sinne also „genauso viele“ rationale wie natürliche Zahlen.

Auch wenn wir mit Satz 5.57 jetzt von sehr vielen Mengen sehen können, dass sie abzählbar sind, gibt es dennoch unendliche Mengen, die „zu groß“ sind, um eine bijektive Abbildung nach $\mathbb{N}$ zuzulassen. Das einfachste Beispiel hierfür ist die Menge der reellen Zahlen.

Satz 5.59. *Die Menge $\mathbb{R}$ ist überabzählbar.*

Beweis. Dies ist eine unmittelbare Folge der Intervallschachtelung aus Satz 5.38: Angenommen, wir hätten eine Aufzählung $\mathbb{R} = \{x_0, x_1, x_2, \dots\}$ der reellen Zahlen. Wir konstruieren dann rekursiv eine Intervallschachtelung $I_0 \supset I_1 \supset I_2 \supset \cdots$ aus Intervallen I_n der Länge $\frac{1}{3^n}$, von der wir lediglich verlangen, dass für alle n die Zahl x_n nicht in I_n liegt. Dies ist natürlich sehr einfach: Wir wählen für I_{n+1} immer das linke Drittel von I_n – es sei denn, die Zahl x_{n+1} liegt in diesem Drittel; dann wählen wir für I_{n+1} das rechte Drittel von I_n.

Nach Satz 5.38 gibt es nun aber ein $a \in \mathbb{R}$, das in allen I_n liegt und damit keine der Zahlen $x_n \in \mathbb{R}\backslash I_n$ sein kann, also im Widerspruch zur Annahme in unserer Aufzählung von $\mathbb{R}$ nicht enthalten ist. □

Zusammenfassend können wir also sagen, dass es im Sinne der Gleichmächtigkeit zwar „genauso viele“ natürliche wie ganze oder rationale, aber „deutlich mehr“ reelle Zahlen gibt.

Beispiel 5.60. Nach Beispiel 5.58 (b) gibt es eine Aufzählung $\mathbb{Q} = \{x_0, x_1, x_2, \dots\}$ der rationalen Zahlen. Fassen wir diese als Folge $(x_n)_{n\in\mathbb{N}}$ auf, so liegen in jeder ε-Umgebung jeder reellen Zahl nach Aufgabe 5.35 unendlich viele Folgenglieder. Also ist jede reelle Zahl ein Häufungspunkt dieser Folge: Eine (abzählbare) Folge kann durchaus überabzählbar viele Häufungspunkte haben!

Aufgabe 5.61. Untersuche die folgenden Mengen auf Abzählbarkeit:

(a) die Menge aller zweielementigen Teilmengen von $\mathbb{N}$;

(b) die Menge aller endlichen Teilmengen von $\mathbb{N}$;

(c) die Menge aller Teilmengen von $\mathbb{N}$ (also die Potenzmenge $\mathscr{P}(\mathbb{N})$ von $\mathbb{N}$).

6. Komplexe Zahlen

Bisher haben wir uns nahezu ausschließlich mit dem euch aus der Schule bekannten Körper $\mathbb{R}$ der reellen Zahlen befasst. In der Mathematik – und zwar sowohl in der Analysis als auch in der Algebra – ist jedoch noch ein weiterer damit eng zusammenhängender Körper sehr wichtig: der Körper der komplexen Zahlen. Die Idee dabei ist, einen Körper zu konstruieren, der die reellen Zahlen als Teilmenge enthält (man nennt so etwas auch eine *Körpererweiterung* von $\mathbb{R}$), und in dem jede nicht konstante Polynomfunktion (wie z. B. $x \mapsto x^2+1$) eine Nullstelle besitzt. Außerdem werden die komplexen Zahlen zur Einführung und Untersuchung der Winkelfunktionen in Kapitel 9.B sehr nützlich sein.

6.A Die Konstruktion der komplexen Zahlen

Im Gegensatz zu den reellen Zahlen brauchen wir die komplexen nicht mehr axiomatisch vorauszusetzen – wir können sie explizit aus den reellen konstruieren.

Definition 6.1 (Komplexe Zahlen)**.** Die Menge der **komplexen Zahlen** ist definiert als $\mathbb{C} := \mathbb{R}^2$. Wir betrachten auf dieser Menge die beiden Verknüpfungen

$$\begin{aligned} &(x_1,y_1)+(x_2,y_2) := (x_1+x_2, y_1+y_2) && \text{(Addition)} \\ \text{und} \quad &(x_1,y_1)\cdot(x_2,y_2) := (x_1x_2-y_1y_2, x_1y_2+y_1x_2) && \text{(Multiplikation).} \end{aligned}$$

Notation 6.2 (Komplexe Zahlen in der Form $x+\mathrm{i}y$)**.** Beachte, dass für komplexe Zahlen, deren zweite Komponente 0 ist, die Addition und Multiplikation

$$(x_1,0)+(x_2,0) = (x_1+x_2,0) \qquad \text{bzw.} \qquad (x_1,0)\cdot(x_2,0) = (x_1x_2,0)$$

genauso definiert ist wie für reelle Zahlen. Wir schreiben die komplexe Zahl $(x,0) \in \mathbb{C}$ daher in der Regel auch einfach als x und fassen auf diese Art die reellen Zahlen als Teilmenge der komplexen auf. Setzen wir weiterhin $\mathrm{i} := (0,1)$, so gilt

$$\mathrm{i}^2 = (0,1)\cdot(0,1) = (-1,0) = -1$$

sowie für alle $x,y \in \mathbb{R}$

$$x+\mathrm{i}y = (x,0)+(0,1)(y,0) = (x,0)+(0,y) = (x,y).$$

Diese Darstellung als $x+\mathrm{i}y$ ist in der Tat auch die übliche Schreibweise für eine komplexe Zahl – wir werden Elemente von $\mathbb{C}$ ab jetzt immer in dieser Form schreiben. Diese Notation hat den Vorteil, dass sich die Rechenregeln für die Addition und Multiplikation aus Definition 6.1 ganz von selbst ergeben, wenn man i als Variable auffasst und die Gleichung $\mathrm{i}^2 = -1$ berücksichtigt: Es ist dann nämlich

$$\begin{aligned} &(x_1+\mathrm{i}y_1)+(x_2+\mathrm{i}y_2) = (x_1+x_2)+\mathrm{i}\,(y_1+y_2) \\ \text{und} \quad &(x_1+\mathrm{i}y_1)\cdot(x_2+\mathrm{i}y_2) = x_1x_2+\mathrm{i}x_1y_2+\mathrm{i}x_2y_1+\mathrm{i}^2y_1y_2 = (x_1x_2-y_1y_2)+\mathrm{i}\,(x_1y_2+x_2y_1). \end{aligned}$$

Wenn man in ingenieurswissenschaftliche Bücher schaut, werden die komplexen Zahlen dort in der Tat sogar oft so eingeführt: Man nehme einfach an, dass es eine Zahl i mit $\mathrm{i}^2 = -1$ gibt, und rechne damit dann ganz normal weiter, als wäre nichts Besonderes passiert. Es sollte aber hoffentlich klar sein, dass eine solche „Definition“ aus mathematischer Sicht unsinnig ist: Wenn wir bisher nur die reellen Zahlen kennen, gibt es nach Lemma 4.16 (c) einfach keine Zahl, deren Quadrat gleich -1 ist – und diese Situation wird natürlich auch nicht dadurch besser, dass wir diesem nicht existierenden Objekt einen Namen i geben. Stattdessen müssen wir den Umweg über die korrekte Konstruktion aus Definition 6.1 gehen, die uns garantiert, dass $\mathbb{C}$ erst einmal widerspruchsfrei definiert ist, und können dann erst im Nachhinein untersuchen, welche Eigenschaften der reellen Zahlen sich tatsächlich auf

die komplexen übertragen. Dies sind nämlich auch nicht alle – so werden wir z. B. in Lemma 6.6 und Bemerkung 6.8 sehen, dass $\mathbb{C}$ zwar ein Körper, aber kein geordneter Körper ist.

Definition 6.3 (Real- und Imaginärteil, Konjugation und Betrag)**.** Es sei $z = x + \mathrm{i}y \in \mathbb{C}$ mit $x, y \in \mathbb{R}$ wie in Notation 6.2.

(a) Man nennt x den **Realteil** und y den **Imaginärteil** von z; die Notation hierfür ist $x = \operatorname{Re} z$ und $y = \operatorname{Im} z$.

(b) Man nennt

$$\bar{z} := x - \mathrm{i}y \qquad \text{die zu } z \textbf{ komplex konjugierte Zahl}$$

$$\text{und} \qquad |z| := \sqrt{x^2 + y^2} \in \mathbb{R}_{\geq 0} \qquad \text{den } \textbf{Betrag} \text{ von } z$$

(mit der reellen Wurzel aus Definition 5.29).

Bemerkung 6.4. Offensichtlich lassen sich der Real- und Imaginärteil von $z = x + \mathrm{i}y \in \mathbb{C}$ ausdrücken als

$$\operatorname{Re} z = \frac{1}{2}(z + \bar{z}) \qquad \text{und} \qquad \operatorname{Im} z = -\frac{\mathrm{i}}{2}(z - \bar{z}),$$

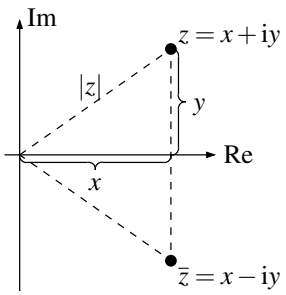
während der Betrag wegen $z\bar{z} = (x + \mathrm{i}y)(x - \mathrm{i}y) = x^2 - (\mathrm{i}y)^2 = x^2 + y^2$ auch als

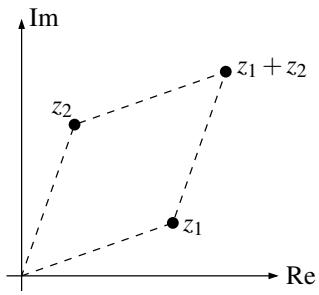
$$|z| = \sqrt{z\bar{z}}$$

geschrieben werden kann.

Bemerkung 6.5 (Geometrische Interpretation von $\mathbb{C}$)**.** Geometrisch können wir die Elemente von $\mathbb{C} = \mathbb{R}^2$ natürlich als Punkte der Ebene, der sogenannten **komplexen Zahlenebene**, zeichnen. Wir wollen jetzt sehen, wie man die oben eingeführten Operationen für komplexe Zahlen in dieser Zahlenebene grafisch veranschaulichen kann. Da diese Interpretation zwar für die Vorstellung sehr wichtig ist, aber nicht für unsere späteren exakten Rechnungen benötigt wird, wollen wir dabei ein paar einfache und sicherlich bekannte Prinzipien der Schulgeometrie ohne Beweis verwenden.

Zunächst einmal ist klar, dass die *reellen* Zahlen in $\mathbb{C}$, also diejenigen der Form $x + \mathrm{i} \cdot 0$, genau die auf der horizontalen Achse sind. Der Betrag $|z|$ einer komplexen Zahl ist nach Definition genau der Abstand des Punktes z vom Ursprung, und die komplexe Konjugation entspricht einer Spiegelung an der reellen Achse (wie im Bild unten links). Ebenso offensichtlich ist, dass zwei komplexe Zahlen genau so addiert werden, wie ihr in der Schule Vektoren im $\mathbb{R}^2$ addiert habt, also indem man die Verbindungsstrecken vom Ursprung zu z_1 und z_2 wie im folgenden Bild rechts zu einem Parallelogramm zusammensetzt.



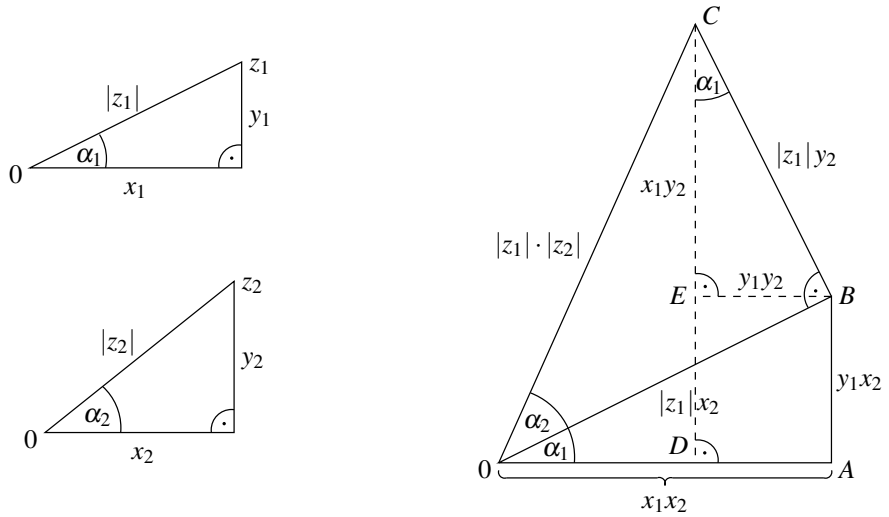


Die Multiplikation dagegen ist schon interessanter. Der Einfachheit halber beschränken wir uns im Bild unten auf den Fall, in dem Real- und Imaginärteil beider Zahlen positiv sind – die anderen Fälle lassen sich analog behandeln. Wir haben dort links zwei komplexe Zahlen z_1 und z_2 wie oben dargestellt (und zusätzlich die Winkel eingezeichnet, die die Verbindungsstrecken zum Ursprung mit der positiven reellen Achse einschließen), und die zugehörigen rechtwinkligen Dreiecke rechts wie folgt zusammengesetzt:

(a) Das Dreieck für z_1 haben wir um den Faktor x_2 zum Dreieck $0AB$ gestreckt.

(b) Das Dreieck für z_2 haben wir um den Faktor $|z_1|$ gestreckt und um den Winkel α_1 gedreht, so dass das Dreieck $0BC$ mit Seitenlängen $|z_1|x_2$, $|z_1|y_2$ und $|z_1|\cdot|z_2|$ entstanden ist (insbesondere hat dieses Dreieck mit dem aus (a) also eine gemeinsame Kante $0B$ mit der Seitenlänge $|z_1|x_2$).

(c) CD ist die Senkrechte auf $0A$, und BE die Senkrechte auf CD. Damit ist das Dreieck CEB ähnlich zu $0AB$, es ist daher die Streckung des Dreiecks für z_1 um den Faktor y_2 und hat Seitenlängen x_1y_2, y_1y_2 und $|z_1|y_2$.



Aus diesem Bild lesen wir nun direkt ab, dass C die Koordinaten $(x_1x_2 - y_1y_2, x_1y_2 + y_1x_2)$ hat, also genau der Punkt $z_1 \cdot z_2$ ist. Da dieser Punkt den Betrag $|z_1|\cdot|z_2|$ hat und den Winkel $\alpha_1 + \alpha_2$ mit der positiven reellen Achse einschließt, sehen wir:

Komplexe Zahlen werden multipliziert, indem man ihre Beträge *multipliziert* und ihre Winkel *addiert*.

12

Wir wollen nun sehen, dass die Addition und Multiplikation auf $\mathbb{C}$ die erwarteten Eigenschaften haben, also die Struktur eines Körpers bilden. Unsere Ergebnisse aus Kapitel 3 und Abschnitt 4.A gelten somit unverändert auch für die komplexen Zahlen.

Lemma 6.6. *$\mathbb{C}$ ist ein Körper.*

Beweis. Die Kommutativität der Addition und Multiplikation ist aus der Definition offensichtlich. Die Assoziativität der Addition und Multiplikation sowie die Distributivität rechnet man einfach nach; wir zeigen hier exemplarisch die Distributivität: Für $z_1 = x_1 + \mathrm{i}y_1$, $z_2 = x_2 + \mathrm{i}y_2$, $z_3 = x_3 + \mathrm{i}y_3$ folgt (letztlich wegen der Distributivität in $\mathbb{R}$)

$$\begin{aligned}(z_1+z_2)\,z_3 &= ((x_1+x_2)+\mathrm{i}\,(y_1+y_2))\cdot(x_3+\mathrm{i}y_3)\\ &= (x_1+x_2)x_3-(y_1+y_2)y_3+\mathrm{i}\,((x_1+x_2)y_3+(y_1+y_2)x_3)\\ &= \big(x_1x_3-y_1y_3+\mathrm{i}\,(x_1y_3+y_1x_3)\big)+\big(x_2x_3-y_2y_3+\mathrm{i}\,(x_2y_3+y_2x_3)\big)\\ &= z_1z_3+z_2z_3.\end{aligned}$$

Das additive neutrale Element ist 0, das additive Inverse zu $z = x + \mathrm{i}y$ natürlich $-z = -x - \mathrm{i}y$. Das multiplikative neutrale Element ist 1, das multiplikative Inverse zu $z = x + \mathrm{i}y \neq 0$ ist

$$\frac{x}{x^2+y^2}+\mathrm{i}\,\frac{-y}{x^2+y^2}, \quad \text{denn} \quad \left(\frac{x}{x^2+y^2}+\mathrm{i}\,\frac{-y}{x^2+y^2}\right)\cdot(x+\mathrm{i}y) = \frac{(x-\mathrm{i}y)(x+\mathrm{i}y)}{x^2+y^2} = \frac{x^2+y^2}{x^2+y^2} = 1. \qquad \square$$

Beispiel 6.7 (Division komplexer Zahlen)**.** Erwähnenswert ist an Lemma 6.6 wohl vor allem die Existenz einer Division, da ja zunächst einmal nicht offensichtlich ist, wie man für eine komplexe Zahl $z = x + \mathrm{i}y$ das multiplikative Inverse $\frac{1}{z} = \frac{1}{x+\mathrm{i}y}$ wieder in der Form $x' + \mathrm{i}y'$ schreiben kann. Die Merkregel hierfür ist, dass man den Bruch mit der konjugiert komplexen Zahl $\overline{z} = x - \mathrm{i}y$ erweitert, so dass der Nenner zu der *reellen* Zahl $z\overline{z} = x^2 + y^2$ wird und somit das i aus dem Nenner verschwindet. So ist z. B.

$$\frac{1}{1+2\mathrm{i}} = \frac{1-2\mathrm{i}}{(1+2\mathrm{i})(1-2\mathrm{i})} = \frac{1-2\mathrm{i}}{1^2+2^2} = \frac{1}{5} - \frac{2}{5}\mathrm{i}.$$

Bemerkung 6.8 ($\mathbb{C}$ ist kein geordneter Körper)**.** $\mathbb{C}$ ist zwar ein Körper, kann aber nicht zu einem *geordneten* Körper gemacht werden. Andernfalls müsste nämlich i^2 als Quadrat einer Zahl ungleich 0 nach Lemma 4.16 (c) positiv sein – was aber natürlich ein Widerspruch ist, da andererseits $\mathrm{i}^2 = -1$ nach demselben Lemma auch eine negative Zahl sein müsste.

Es ergibt also keinen Sinn zu fragen, welche von zwei gegebenen komplexen Zahlen größer ist als die andere. Damit sind unsere Ergebnisse aus den Abschnitten 4.B und 4.C auf die komplexen Zahlen nicht anwendbar; z. B. sind die Begriffe von Supremum und Infimum sowie Maximum und Minimum für Teilmengen von komplexen Zahlen nicht definiert.

6.B Eigenschaften der komplexen Zahlen

Auch wenn $\mathbb{C}$ kein geordneter Körper ist, haben wir in Definition 6.3 (b) bereits wie für $\mathbb{R}$ auch für $\mathbb{C}$ eine Betragsfunktion eingeführt, die immer reelle Werte annimmt und es uns somit erlaubt, komplexe Zahlen *betragsmäßig* miteinander zu vergleichen. Wir wollen nun sehen, dass diese komplexe Betragsfunktion in der Tat sogar die gleichen Eigenschaften wie die reelle Betragsfunktion in Lemma 4.18 (a) und (c) hat, auch wenn der Beweis dafür in $\mathbb{C}$ ganz anders ist als in $\mathbb{R}$.

Lemma 6.9 (Eigenschaften der komplexen Konjugation und Betragsfunktion)**.** *Für alle $z_1, z_2 \in \mathbb{C}$ gilt*

(a) $\overline{\overline{z_1}} = z_1$, $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ *und* $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$*;*

(b) $|z_1 z_2| = |z_1| \cdot |z_2|$*;*

(c) $|z_1 + z_2| \le |z_1| + |z_2|$ ***(Dreiecksungleichung)****.*

Beweis. Wie üblich sei $z_1 = x_1 + \mathrm{i}y_1$ und $z_2 = x_2 + \mathrm{i}y_2$.

(a) Dies rechnet man einfach nach: Es ist $\overline{\overline{z_1}} = \overline{x_1 - \mathrm{i}y_1} = x_1 + \mathrm{i}y_1 = z_1$, sowie

$$\overline{z_1 + z_2} = \overline{x_1 + x_2 + \mathrm{i}(y_1 + y_2)} = x_1 + x_2 - \mathrm{i}(y_1 + y_2) = \overline{z_1} + \overline{z_2}$$

und

$$\begin{aligned}\overline{z_1 z_2} &= \overline{x_1 x_2 - y_1 y_2 + \mathrm{i}(x_1 y_2 + y_1 x_2)} = x_1 x_2 - y_1 y_2 - \mathrm{i}(x_1 y_2 + y_1 x_2) = (x_1 - \mathrm{i}y_1)(x_2 - \mathrm{i}y_2) \\ &= \overline{z_1} \cdot \overline{z_2}.\end{aligned}$$

(b) Bei der geometrischen Deutung der komplexen Multiplikation in Bemerkung 6.5 haben wir dies bereits gesehen; man rechnet es aber auch mit (a) sofort nach: Nach Bemerkung 6.4 ist

$$|z_1 z_2| = \sqrt{z_1 z_2 \cdot \overline{z_1 z_2}} = \sqrt{z_1 \overline{z_1} z_2 \overline{z_2}} \overset{5.33\,(b)}{=} \sqrt{z_1 \overline{z_1}} \cdot \sqrt{z_2 \overline{z_2}} = |z_1| \cdot |z_2|.$$

(c) Zunächst ist nach (a) und Bemerkung 6.4

$$\begin{aligned}|z_1 + z_2|^2 &= (z_1 + z_2)\overline{(z_1 + z_2)} = z_1 \overline{z_1} + z_2 \overline{z_2} + z_1 \overline{z_2} + z_2 \overline{z_1} = |z_1|^2 + |z_2|^2 + z_1 \overline{z_2} + \overline{z_1 \overline{z_2}} \\ &= |z_1|^2 + |z_2|^2 + 2\,\mathrm{Re}(z_1 \overline{z_2}).\end{aligned}$$

Da nun für jede komplexe Zahl $z = x + \mathrm{i}y$ die Ungleichung

$$\mathrm{Re}\, z = x \overset{4.18\,(b)}{\le} |x| = \sqrt{x^2} \overset{5.33\,(a)}{\le} \sqrt{x^2 + y^2} = |z|$$

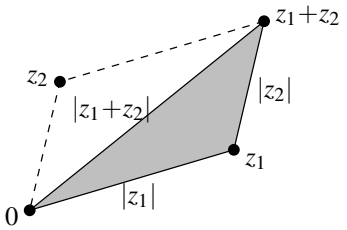
gilt, folgt durch Anwendung auf $z = z_1\,\overline{z_2}$ also $\mathrm{Re}(z_1\,\overline{z_2}) \leq |z_1\,\overline{z_2}| \overset{\text{(b)}}{=} |z_1| \cdot |\overline{z_2}| = |z_1| \cdot |z_2|$ und damit

$$|z_1 + z_2|^2 \leq |z_1|^2 + |z_2|^2 + 2\,|z_1| \cdot |z_2| = (|z_1| + |z_2|)^2.$$

Wurzelziehen liefert nun die Behauptung. □

Bemerkung 6.10.

(a) Die Dreiecksungleichung hat eine sehr anschauliche Bedeutung, die auch ihren Namen erklärt: Nach der geometrischen Interpretation der Addition komplexer Zahlen aus Bemerkung 6.5 besagt sie einfach, dass eine Seite in einem Dreieck (wie $|z_1 + z_2|$ im Bild rechts) höchstens so lang ist wie die Summe der beiden anderen (hier $|z_1|$ und $|z_2|$).



(b) Wenn ihr gleichzeitig die Parallelvorlesung „Algebraische Strukturen“ hört, werdet ihr sicher sehen, dass Lemma 6.9 (a) gerade besagt, dass die komplexe Konjugation $z \mapsto \overline{z}$ ein Gruppenhomomorphismus von $(\mathbb{C}, +)$ nach $(\mathbb{C}, +)$ und von $(\mathbb{C}\backslash\{0\}, \cdot)$ nach $(\mathbb{C}\backslash\{0\}, \cdot)$ ist. Zusammen macht dies die komplexe Konjugation zu einem Körperhomomorphismus von $\mathbb{C}$ nach $\mathbb{C}$ (in der Tat sogar zu einem Körperisomorphismus, da die Abbildung $\mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$ natürlich bijektiv ist).

Wie schon am Anfang dieses Kapitels erwähnt, besteht aber die wesentliche Eigenschaft der komplexen Zahlen darin, dass in $\mathbb{C}$ jede Polynomfunktion eine Nullstelle besitzt. Beachte, dass dies ganz und gar nicht offensichtlich ist – da man jede komplexe Zahl als $x + \mathrm{i}y$ mit $x, y \in \mathbb{R}$ und $\mathrm{i}^2 = -1$ schreiben kann, sieht es ja eher so aus, als ob wir durch den Übergang von $\mathbb{R}$ nach $\mathbb{C}$ nur eine „Quadratwurzel aus -1“ hinzugefügt haben, also nur der Polynomfunktion $z^2 + 1 = 0$ (oder bestenfalls noch anderen quadratischen Polynomfunktionen) eine Nullstelle gegeben haben. Dass dies in der Tat auch für Polynomfunktionen beliebigen Grades gilt, und zwar sogar noch, wenn sie auch komplexe Koeffizienten haben dürfen, ist der Inhalt des sogenannten Fundamentalsatzes der Algebra:

Satz 6.11 (Fundamentalsatz der Algebra). *Jede nicht konstante komplexe Polynomfunktion hat eine Nullstelle in $\mathbb{C}$.*

Beweisidee. Es gibt mehrere (völlig) verschiedene Möglichkeiten, den Fundamentalsatz der Algebra zu beweisen. Leider sind alle diese Beweise für uns aber momentan noch zu schwierig, und so muss ich euch für einen exakten Beweis dieses Satzes auf weiterführende Vorlesungen vertrösten – in den Vorlesungen „Einführung in die Funktionentheorie“, „Einführung in die Algebra“ und „Einführung in die Topologie“ könnt ihr z. B. drei ganz verschiedene Beweise dieses Satzes sehen. Wir können aber auch jetzt zumindest schon eine *Beweisidee* angeben, die hoffentlich dafür ausreicht, dass ihr den Satz glaubt und ein Gefühl dafür bekommt, warum er richtig ist.

Es sei dazu f eine komplexe, nicht konstante Polynomfunktion, die wir der Einfachheit halber natürlich als normiert annehmen können. Es ist also

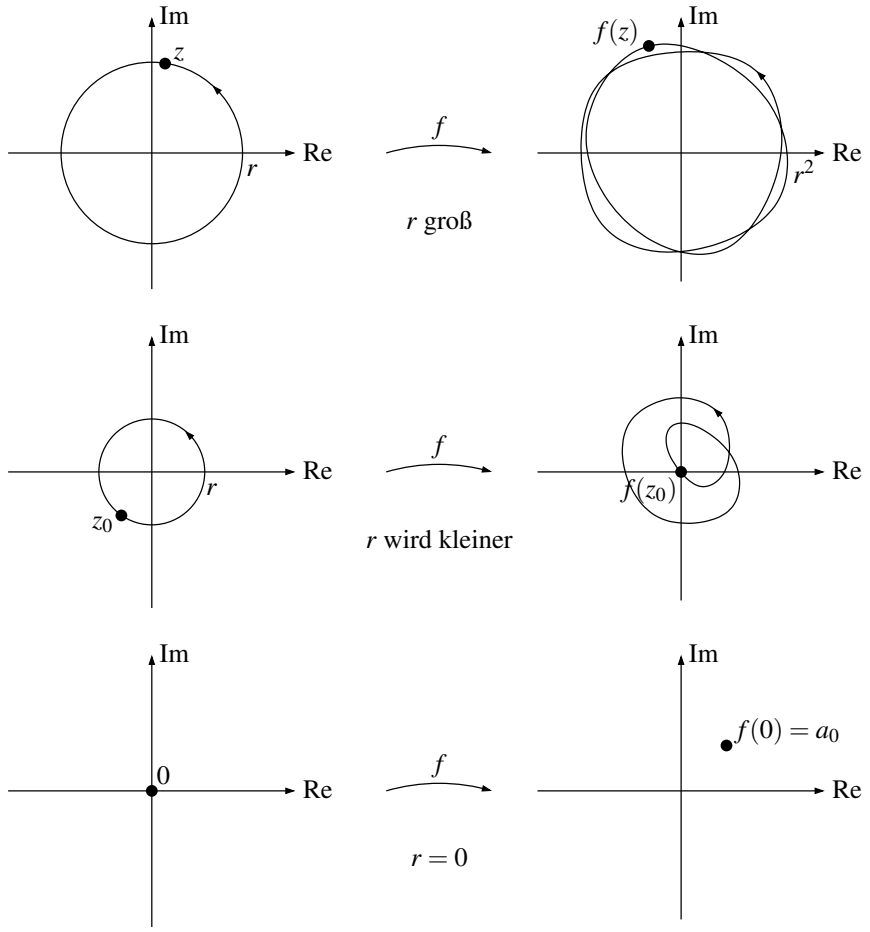
$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

für gewisse $n \in \mathbb{N}_{>0}$ und $a_0, \ldots, a_{n-1} \in \mathbb{C}$. Wie können wir uns eine solche Funktion grafisch vorstellen? Da ihre Start- und Zielmenge $\mathbb{C}$ ist, können wir ihren Graphen, der ja dann in $\mathbb{C} \times \mathbb{C} = \mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4$ liegt, nicht mehr wirklich zeichnen. In den Bildern unten haben wir daher den Startraum $\mathbb{C}$ links und den Zielraum $\mathbb{C}$ rechts dargestellt, und für einige Punkte im Startraum die zugehörigen Bildpunkte im Zielraum eingezeichnet.

Als Erstes wählen wir uns einmal eine feste, sehr große Zahl $r \in \mathbb{R}_{>0}$ und schauen, was passiert, wenn wir mit z den Kreis um 0 mit Radius r durchlaufen. Wenn unsere Funktion einfach $z \mapsto z^n$ wäre, dann wüssten wir genau, wie $f(z)$ auf dieser Kreislinie aussehen würde: Da bei der komplexen Multiplikation nach Bemerkung 6.5 ja gerade Beträge multipliziert und Winkel addiert werden, ist die n-te Potenz einer komplexen Zahl mit Betrag r und Winkel α genau die Zahl mit Betrag r^n und

Winkel $n\alpha$. Läuft also z einmal beim Radius r im Kreis herum, d. h. α von 0 bis 2π, so läuft z^n beim Radius r^n genau n-mal im Kreis herum, nämlich mit Winkel $n\alpha$ von 0 bis $2n\pi$.

Nun ist unsere Polynomfunktion zwar nicht wirklich genau $z \mapsto z^n$, aber für sehr große Beträge von z ist der Term z^n in $f(z)$ mit der höchsten z-Potenz natürlich betragsmäßig viel größer als die anderen Terme $a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$. Anschaulich bedeutet das, dass $f(z)$ immer „in der Nähe" von z^n ist. Wenn also z^n beim Radius r^n insgesamt n-mal auf einer exakten Kreislinie herumläuft, wird $f(z)$ ein klein wenig von diesem Weg abweichen, aber letztlich immer noch n-mal um den Ursprung herumlaufen. Das Bild unten zeigt in der ersten Zeile einen solchen möglichen Weg für $n = 2$, bei dem also $f(z)$ in einem ungefähren Abstand von r^2 zweimal um den Ursprung läuft, während z einmal auf dem Kreis mit Radius r entlang läuft.



Was passiert nun, wenn wir den Radius r des Kreises für z langsam kleiner machen und zu schließlich 0 werden lassen, so wie im Bild von oben nach unten dargestellt? Natürlich wird sich dann auch der von $f(z)$ durchlaufene Weg in irgendeiner Form langsam ändern. Wir können nicht viel darüber aussagen, wie diese Änderung genau aussieht – klar ist die Situation aber natürlich, wenn der Radius wie in der unteren Zeile des Bildes gleich 0 geworden ist: Dann ist der Kreis für z zu einem Punkt zusammengeschrumpft, und folglich muss natürlich auch der Weg von $f(z)$ von der ursprünglichen Schleife zu einem Punkt (nämlich zum Punkt $f(0) = a_0$) zusammenschrumpfen. Aber es ist anschaulich klar, dass man einen geschlossenen Weg, der ursprünglich n-mal um den Ursprung herumgelaufen ist, nicht auf einen Punkt zusammenziehen kann, ohne ihn dabei mindestens einmal über den Nullpunkt zu ziehen. Und genau an so einer Stelle, wo der Weg für $f(z)$ den Nullpunkt trifft, haben wir natürlich, was wir wollen: eine Nullstelle z_0 von f, so wie in der mittleren Zeile oben im Bild. $\square$

Auch wenn diese Beweisidee jetzt hoffentlich sehr anschaulich war, wäre es doch noch ein sehr weiter Weg für uns, diese Argumente zu einem exakten Beweis zu machen. Ein wichtiger fehlender Punkt ist z. B., dass wir irgendwie formalisieren müssten, was es genau heißt, dass „sich $f(z)$ langsam ändert, wenn sich z langsam ändert". Denn nur wenn sich der Weg für $f(z)$ oben langsam und kontinuierlich ändert, können wir schließen, dass wir ihn irgendwann einmal über den Nullpunkt ziehen müssen.

Bemerkung 6.12.

(a) Nach Satz 3.18 (a) folgt durch wiederholte Anwendung des Fundamentalsatzes der Algebra sofort, dass jede komplexe Polynomfunktion komplett in Linearfaktoren zerfällt, dass sich also jede solche Polynomfunktion f mit $\deg f = n \in \mathbb{N}_{>0}$ als

$$f(z) = c\,(z - z_1) \cdot \cdots \cdot (z - z_n)$$

für gewisse $c, z_1, \ldots, z_n \in \mathbb{C}$ mit $c \neq 0$ schreiben lässt. Manchmal wird in der Literatur auch diese Aussage als Fundamentalsatz der Algebra bezeichnet.

(b) Der Fundamentalsatz der Algebra garantiert uns zwar die Existenz einer Nullstelle einer nicht konstanten komplexen Polynomfunktion, er sagt uns aber nicht, wie wir eine solche Nullstelle konkret finden können. In der Tat haben wir ja schon in Bemerkung 3.20 erwähnt, dass es zum exakten Auffinden von Nullstellen von Polynomfunktionen im Allgemeinen nur für kleine Grade explizite Formeln gibt. Einen sehr einfachen und oft vorkommenden Fall, in dem sich die Nullstellen jedoch schnell finden lassen, wollen wir hier kurz erwähnen:

Beispiel 6.13. Es sei $f\colon \mathbb{C} \to \mathbb{C}$ eine Polynomfunktion mit $\deg f = 2$ und *reellen* Koeffizienten, das der Einfachheit halber wieder normiert sei, d. h. es sei $f(z) = z^2 + pz + q$ für gewisse $p, q \in \mathbb{R}$. In diesem Fall lassen sich die (komplexen) Nullstellen von f schnell berechnen: Aus $z^2 + pz + q = 0$ folgt durch quadratische Ergänzung

$$\left(z + \frac{p}{2}\right)^2 = \frac{p^2}{4} - q =: D.$$

Für $D \geq 0$ ergeben sich durch Wurzelziehen natürlich die (reellen) Nullstellen $-\frac{p}{2} \pm \sqrt{D}$. Für $D < 0$ gibt es keine reellen Lösungen, aber wegen $\mathrm{i}^2 = -1$ erhalten wir stattdessen die beiden komplexen Lösungen $-\frac{p}{2} \pm \mathrm{i}\sqrt{-D}$.

Aufgabe 6.14. Für $n = 1, 2, 3$ bestimme und skizziere man die Menge aller $z \in \mathbb{C}$, für die die Gleichung $2 \operatorname{Im} z \cdot \operatorname{Im} \frac{1}{\bar{z}} = n$ gilt.

Aufgabe 6.15.

(a) Zeige (ohne Verwendung des Fundamentalsatzes der Algebra), dass auch in $\mathbb{C}$ Quadratwurzeln existieren, also dass es zu jedem $w \in \mathbb{C}$ ein $z \in \mathbb{C}$ gibt mit $z^2 = w$.

(b) Beweise den Fundamentalsatz der Algebra für Polynome vom Grad 2.

Aufgabe 6.16. Stelle die folgenden Zahlen in der Form $x + \mathrm{i}y$ mit $x, y \in \mathbb{R}$ dar:

(a) $z = \frac{2+\mathrm{i}}{1-\mathrm{i}}$;

(b) $z = \left(\frac{1+\mathrm{i}}{\sqrt{2}}\right)^{-2022}$;

(c) alle Lösungen der Gleichung $z^4 + z^2 + 1 = 0$;

(d) alle $z \in \mathbb{C}$ mit $\left|\frac{z-1}{z-\mathrm{i}}\right| < 1$.

6.C Reelle und komplexe Folgen

Auch wenn $\mathbb{C}$ nach Bemerkung 6.8 kein geordneter Körper ist, können wir mit Hilfe der Betragsfunktion aus Definition 6.3 sagen, was es bedeutet, dass sich eine Folge komplexer Zahlen einem Grenzwert annähert. In der Tat können wir die reelle Grenzwertdefinition 5.1 (b) wörtlich auf den reellen Fall übertragen:

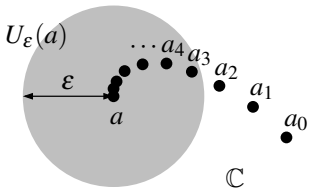
Definition 6.17 (Grenzwerte komplexer Folgen)**.** Eine komplexe Zahl a heißt **Grenzwert** einer Folge $(a_n)_n$ in $\mathbb{C}$, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : |a_n - a| < \varepsilon.$$

Bemerkung 6.18 (Anschauliche Deutung des Grenzwertbegriffs in $\mathbb{C}$)**.** Definieren wir für ein *komplexes* $a \in \mathbb{C}$ und *reelles* $\varepsilon \in \mathbb{R}_{>0}$ analog zu Bemerkung 5.2 wieder die **ε-Umgebung** von a als

$$U_\varepsilon(a) := \{x \in \mathbb{C} : |x - a| < \varepsilon\},$$

so ist dies wie im Bild rechts nun ein Kreis in der komplexen Ebene mit Mittelpunkt a und Radius ε. Die Grenzwertbedingung besagt weiterhin, dass in jeder ε-Umgebung fast alle Folgenglieder liegen, und kann damit wieder so interpretiert werden, dass sich die Folgenglieder immer mehr dem Grenzwert nähern.



Bemerkung 6.19 (Übertragung der Grenzwerteigenschaften von $\mathbb{R}$ auf $\mathbb{C}$)**.** Aufgrund der gleichen Grenzwertdefinition 5.1 bzw. 6.17 sowie der gleichen Eigenschaften der Betragsfunktion aus Lemma 4.18 bzw. 6.9 (insbesondere der Dreiecksungleichung) gelten sehr viele Resultate über reelle Grenzwerte genauso auch für komplexe. *In der Tat übertragen sich alle Definitionen und Sätze aus Abschnitt 5.A mit wörtlich den gleichen Beweisen unmittelbar auf komplexe Folgen $(a_n)_n$:*

(a) die Definitionen von Nullfolgen, Häufungspunkten und beschränkten Folgen (wobei die Schranke s mit $|a_n| \leq s$ für alle n natürlich weiterhin reell bleibt);

(b) die Eindeutigkeit des Grenzwerts (und somit die Schreibweise $\lim_{n\to\infty} a_n$), der Grenzwert $\lim_{n\to\infty} q^n = 0$ für $|q| < 1$, die Beschränktheit konvergenter Folgen, die Grenzwertsätze und die äquivalenten Charakterisierungen von Häufungspunkten.

Ihr könnt euch gerne selbst davon überzeugen und Abschnitt 5.A noch einmal unter der Voraussetzung durchlesen, dass alle Folgen nun komplex sind – es werden keinerlei Änderungen erforderlich sein. Wir werden die Ergebnisse dieses Abschnitts daher im Folgenden auch im Komplexen verwenden, ohne jedesmal wieder darauf hinzuweisen. Um solche Aussagen in Zukunft für den reellen und komplexen Fall gleichzeitig aufschreiben zu können, vereinbaren wir:

$$\boxed{\text{Im Folgenden steht } \mathbb{K} \text{ immer für einen der Körper } \mathbb{R} \text{ oder } \mathbb{C}.}$$

Die Inhalte der Abschnitte 5.B und 5.C benötigen jedoch wirklich einen geordneten Körper und nicht nur das Konzept des Abstandes zweier Zahlen. Ergebnisse wie das Monotoniekriterium, die Intervallschachtelung oder die Existenz eines Limes superior haben daher keine Entsprechung im Komplexen.

Aufgabe 6.20. Es sei $(a_n)_n$ eine Folge in $\mathbb{C}$. Beweise, dass $(a_n)_n$ genau dann gegen die komplexe Zahl a konvergiert, wenn die Folgen $(\operatorname{Re} a_n)_n$ und $(\operatorname{Im} a_n)_n$ ihrer Real- und Imaginärteile gegen $\operatorname{Re} a$ bzw. $\operatorname{Im} a$ konvergieren.

Da wir die Konvergenzkriterien aus Abschnitt 5.B, mit denen wir die Konvergenz einer Folge auch ohne Kenntnis oder gleichzeitige Berechnung des Grenzwerts beweisen konnten, in $\mathbb{C}$ nicht mehr zur Verfügung haben, wollen wir nun noch zwei sehr wichtige Konvergenzkriterien behandeln, die sowohl in $\mathbb{R}$ als auch in $\mathbb{C}$ gelten.

Satz 6.21 (**Satz von Bolzano-Weierstraß**)**.** *Jede beschränkte Folge $(a_n)_n$ in $\mathbb{K}$ besitzt einen Häufungspunkt.*

Beweis. Für den Fall $\mathbb{K} = \mathbb{R}$ haben wir dies bereits in Folgerung 5.47 gesehen: Der Limes superior von $(a_n)_n$ ist ein Häufungspunkt. Im Fall $\mathbb{K} = \mathbb{C}$ stellen wir zunächst fest, dass wegen

$$|\operatorname{Re} a_n| = \sqrt{(\operatorname{Re} a_n)^2} \leq \sqrt{(\operatorname{Re} a_n)^2 + (\operatorname{Im} a_n)^2} = |a_n| \qquad (*)$$

mit $(a_n)_n$ auch die reelle Folge $(\operatorname{Re} a_n)_n$ beschränkt ist. Nach dem Satz von Bolzano-Weierstraß für $\mathbb{R}$ (den wir ja schon bewiesen haben) gibt es also eine Teilfolge von $(a_n)_n$, in der die Realteile gegen ein $a \in \mathbb{R}$ konvergieren. Nun ist analog zu $(*)$ aber auch $(\operatorname{Im} a_n)_n$ beschränkt. Wir können also wiederum den Satz von Bolzano-Weierstraß für $\mathbb{R}$ *auf die gerade gefundene Teilfolge* anwenden und erhalten so innerhalb dieser Teilfolge eine weitere Teilfolge, in der auch die Imaginärteile gegen ein $b \in \mathbb{R}$ konvergieren. Nach Aufgabe 6.20 konvergiert diese Teilfolge dann gegen $a + \mathrm{i}b$, d. h. $a + \mathrm{i}b$ ist ein Häufungspunkt von $(a_n)_n$. □

Das letzte wichtige Konvergenzkriterium, das wir hier beweisen wollen – das sogenannte Cauchy-Kriterium – sieht fast so aus wie die Definition der Konvergenz. Der Unterschied besteht lediglich darin, dass wir nicht verlangen, dass sich die Folgenglieder *einem gegebenen Grenzwert* immer weiter annähern, sondern nur, dass sie sich *untereinander* beliebig nahe kommen. Auf diese Art müssen wir den Grenzwert der Folge also wiederum nicht vorher kennen, um das Kriterium anwenden zu können. Im Gegensatz zu unseren bisherigen Kriterien hat das Cauchy-Kriterium aber auch noch den weiteren entscheidenden Vorteil, dass es *äquivalent* zur Konvergenz ist und somit auch zum Beweis der Divergenz einer Folge verwendet werden kann.

Die Eigenschaft, dass sich die Folgenglieder untereinander immer näher kommen, sieht formal wie folgt aus.

Definition 6.22 (Cauchyfolgen)**.** Eine Folge $(a_n)_n$ in $\mathbb{K}$ heißt **Cauchyfolge**, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists n_0 \in \mathbb{N}\ \forall m,n \geq n_0 : |a_m - a_n| < \varepsilon.$$

Bemerkung 6.23. Jede konvergente Folge ist eine Cauchyfolge: Ist $(a_n)_n$ konvergent mit Grenzwert $a \in \mathbb{K}$, so gibt es zu jedem $\varepsilon \in \mathbb{R}_{>0}$ ein $n_0 \in \mathbb{N}$ mit $|a_n - a| < \frac{\varepsilon}{2}$ für alle $n \geq n_0$. Dann gilt nach der Dreiecksungleichung aber auch für alle $m,n \geq n_0$

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

d. h. $(a_n)_n$ ist eine Cauchyfolge.

13

Diese Tatsache, dass eine konvergente Folge immer eine Cauchyfolge ist, ist also sehr einfach zu zeigen und wäre z. B. auch in $\mathbb{Q}$ richtig: Wenn die Folgenglieder immer mehr gegen einen Grenzwert streben, müssen sie sich natürlich auch untereinander immer näher kommen. Die Umkehrung dagegen ist weit weniger klar: Da $\mathbb{Q}$ ja „Löcher“ auf der Zahlengeraden hat, könnte es ja sein, dass sich die Glieder einer rationalen Folge zwar immer näher kommen, aber sich an einem solchen Loch häufen und daher kein Grenzwert der Folge in $\mathbb{Q}$ existiert. Dass so etwas in $\mathbb{R}$ oder $\mathbb{C}$ nicht passieren kann, weil es dort keine solchen Löcher gibt, wird als *Vollständigkeit* dieser Körper bezeichnet (siehe auch Definition 23.28). Um dies zu zeigen, benötigen wir zunächst ein kleines Lemma analog zu Lemma 5.8:

Lemma 6.24. *Jede Cauchyfolge in $\mathbb{K}$ ist beschränkt.*

Beweis. Nicht nur die Aussage, sondern auch ihr Beweis ist völlig analog zu Lemma 5.8: Es sei $(a_n)_n$ eine Cauchyfolge in $\mathbb{K}$. Dann gibt es zu $\varepsilon = 1$ ein n_0, so dass $|a_m - a_n| < \varepsilon = 1$ für alle $m,n \geq n_0$ ist. Insbesondere gilt dies also für $m = n_0$, und damit erhalten wir nach der Dreiecksungleichung für alle $n \geq n_0$

$$|a_n| = |a_n - a_{n_0} + a_{n_0}| \leq |a_n - a_{n_0}| + |a_{n_0}| < 1 + |a_{n_0}|.$$

Damit folgt nun aber $|a_n| \leq s$ für alle $n \in \mathbb{N}$, wenn wir

$$s := \max(|a_0|, |a_1|, \ldots, |a_{n_0-1}|, 1 + |a_{n_0}|)$$

setzen. Also ist $(a_n)_n$ beschränkt. □

Satz 6.25 (**Cauchy-Kriterium** für Folgen, **Vollständigkeit** von $\mathbb{K}$)**.** *Jede Cauchyfolge in $\mathbb{K}$ konvergiert.*

Nach Bemerkung 6.23 konvergiert eine Folge in $\mathbb{K}$ also genau dann, wenn sie eine Cauchyfolge ist.

Beweis. Es sei $(a_n)_n$ eine Cauchyfolge in $\mathbb{K}$. Dann ist $(a_n)_n$ nach Lemma 6.24 beschränkt und besitzt damit nach dem Satz 6.21 von Bolzano-Weierstraß einen Häufungspunkt a. Wir behaupten, dass $(a_n)_n$ sogar schon gegen a konvergiert.

Um dies zu zeigen, sei $\varepsilon \in \mathbb{R}_{>0}$ beliebig. Da $(a_n)_n$ eine Cauchyfolge ist, gibt es ein $n_0 \in \mathbb{N}$, so dass

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

Weil a ein Häufungspunkt von $(a_n)_n$ ist, gilt nach Lemma 5.20 weiterhin

$$|a_m - a| < \frac{\varepsilon}{2} \quad \text{für unendlich viele } m,$$

und damit insbesondere für ein $m \geq n_0$. Wir können diese beiden Ungleichungen also miteinander kombinieren und erhalten für ein solches m nach der Dreiecksungleichung

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

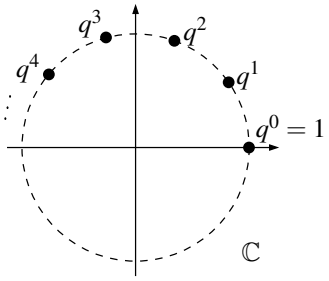
für alle $n \geq n_0$. Damit ist $(a_n)_n$ konvergent gegen a. □

Beispiel 6.26 (Noch einmal die geometrische Folge)**.** Wir betrachten noch einmal die geometrische Folge $(q^n)_n$ für ein $q \in \mathbb{K}$. Aus Beispiel 5.3 (c) und 5.9 (b) wissen wir bereits, dass $(q^n)_n$ für $|q| < 1$ gegen 0 konvergiert und für $|q| > 1$ divergiert. Außerdem ist klar, dass die Folge für $q = 1$ konstant ist und damit konvergiert. Wir zeigen nun mit dem Cauchy-Kriterium in den übrigen Fällen, also wenn $|q| = 1$ und $q \neq 1$, dass die Folge divergiert. Dazu müssen wir also beweisen, dass $(q^n)_n$ keine Cauchyfolge ist, d. h. (nach den Regeln der Negation aus Bemerkung 1.8)

$$\exists \varepsilon > 0 \ \forall n_0 \in \mathbb{N} \ \exists n,m \geq n_0 : |q^n - q^m| \geq \varepsilon.$$

Um dies zu zeigen, setzen wir $\varepsilon := |q-1| > 0$. Nun sei $n_0 \in \mathbb{N}$ beliebig; wir setzen dann $n = n_0 + 1$ und $m = n_0$. Mit diesen Werten folgt

$$|q^n - q^m| = |q^{n_0+1} - q^{n_0}| = |q^{n_0}(q-1)| = \underbrace{|q|^{n_0}}_{=1} \cdot \underbrace{|q-1|}_{=\varepsilon} = \varepsilon.$$



Also ist $(q^n)_n$ keine Cauchyfolge und damit nach Satz 6.25 nicht konvergent. Das Bild rechts illustriert dies: Nach der geometrischen Interpretation der komplexen Multiplikation aus Bemerkung 6.5 läuft die Folge für $|q| = 1$ und $q \neq 1$ „mit konstanter Geschwindigkeit“ auf dem Einheitskreis herum und nähert sich somit keinem Grenzwert beliebig an.

Aufgabe 6.27. Es sei $(a_n)_n$ eine Folge in $\mathbb{K}$. Man zeige: Gibt es ein $q \in \mathbb{R}_{\geq 0}$ mit $q < 1$, so dass

$$|a_{n+1} - a_n| \leq q^n \quad \text{für alle } n \in \mathbb{N},$$

so ist $(a_n)_n$ eine Cauchyfolge.

Aufgabe 6.28. Für ein fest gegebenes $c \in \mathbb{C}$ mit $|c| < \frac{1}{4}$ definieren wir eine komplexe Folge $(a_n)_n$ rekursiv durch

$$a_0 = 0 \quad \text{und} \quad a_{n+1} = a_n^2 + c \quad \text{für alle } n \in \mathbb{N}.$$

Zeige, dass $(a_n)_n$ konvergiert.

(Hinweis: Zeige zunächst, dass $\frac{1}{4} + |c|$ eine obere Schranke für die Beträge aller Folgenglieder ist.)

7. Reihen

Wir wollen uns nun mit einem speziellen Typ von Folgen beschäftigen, der in der Praxis sehr häufig vorkommt: nämlich Folgen, die in der Form

$$(a_0,\ a_0+a_1,\ a_0+a_1+a_2,\ \dots)$$

gegeben sind, deren Grenzwert wir also anschaulich als die „unendliche Summe“ $a_0+a_1+a_2+\cdots$ auffassen können. Derartige Folgen bezeichnet man als Reihen. Wir werden solche Reihen im reellen und komplexen Fall gleichzeitig betrachten und arbeiten daher im Folgenden in der Regel über dem Körper $\mathbb{K}$ wie in Bemerkung 6.19.

7.A Grenzwerte von Reihen

Da Reihen letztlich nichts anderes als spezielle Folgen sind, können wir die Definition und die ersten Eigenschaften von Folgen und Grenzwerten natürlich unmittelbar auf unsere neue Situation übertragen. Dies wollen wir nun im ersten Abschnitt dieses Kapitels tun.

Definition 7.1 (Reihen)**.** Es sei $(a_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{K}$. Dann heißt die Folge $(s_N)_{N\in\mathbb{N}}$ mit

$$s_N=\sum_{n=0}^{N}a_n=a_0+a_1+\cdots+a_N$$

die Folge der **Partialsummen** von $(a_n)_n$ bzw. die zu $(a_n)_n$ gehörige **Reihe**. Wir bezeichnen sowohl diese Reihe als auch ihren Grenzwert $\lim\limits_{N\to\infty}s_N$ (sofern er existiert) mit

$$\sum_{n=0}^{\infty}a_n \quad \text{bzw.} \quad a_0+a_1+a_2+\cdots.$$

Genau wie bei Folgen kann auch eine Reihe bei einem anderen Startindex $n_0\in\mathbb{Z}$ als bei 0 anfangen; in diesem Fall schreiben wir sie natürlich als

$$\sum_{n=n_0}^{\infty}a_n \quad \text{bzw.} \quad a_{n_0}+a_{n_0+1}+a_{n_0+2}+\cdots.$$

Bemerkung 7.2.

(a) Da jede Reihe nach Definition eine Folge ist, übertragen sich die Begriffe Konvergenz und Divergenz, Beschränktheit usw. aus Kapitel 5 direkt auf Reihen.

(b) Die Doppelbelegung des Symbols $\sum_{n=0}^{\infty}a_n$ sowohl für die Reihe (also die Folge ihrer Partialsummen) als auch für ihren Grenzwert ist zwar mathematisch unschön, aber in der Literatur so fest verankert, dass wir hier nicht davon abweichen wollen. Es sollte dadurch keine Verwirrung entstehen: Wenn wir von Eigenschaften einer Folge reden, also z. B. sagen, dass $\sum_{n=0}^{\infty}a_n$ konvergiert oder divergiert, so meinen wir natürlich die Partialsummenfolge – während z. B. in Gleichungen der Form $\sum_{n=0}^{\infty}a_n=a$ der Grenzwert der Reihe gemeint ist. Wenn Verwechslungen zu befürchten sind, können wir aber natürlich auch immer die eindeutige Schreibweise $(\sum_{n=0}^{N}a_n)_{N\in\mathbb{N}}$ für die Reihe und $\lim_{N\to\infty}\sum_{n=0}^{N}a_n$ für ihren Grenzwert benutzen.

Beispiel 7.3.

(a) (**Unendliche geometrische Reihe**) Wir betrachten die Reihe $\sum_{n=0}^{\infty}q^n$ für ein $q\in\mathbb{K}$. Für $q=1$ ist diese Reihe $1+1+1+\dots$ natürlich unbeschränkt und damit divergent. Ansonsten haben wir in Satz 4.1 gesehen, dass

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

Der Grenzwert für $N \to \infty$ ergibt sich nun sofort aus Beispiel 6.26: Da $\lim_{N\to\infty} q^{N+1}$ nur für $|q| < 1$ existiert und dann gleich 0 ist, erhalten wir also

$$\sum_{n=0}^{\infty} q^n = \lim_{N\to\infty} \sum_{n=0}^{N} q^n = \frac{1}{1-q} \qquad \text{für } |q| < 1, \tag{$*$}$$

während die Reihe in allen anderen Fällen divergiert.

Ein interessanter konkreter Fall dieser Reihe ist die Frage, ob die Dezimalzahl $0,9999\ldots$ gleich 1 oder „etwas kleiner" als 1 ist. Dies können wir nun beantworten, denn die einzig mögliche mathematisch korrekte Definition dieser Zahl ist natürlich die geometrische Reihe

$$0,9999\ldots = \sum_{n=1}^{\infty} 9 \cdot 10^{-n} = \frac{9}{10} \cdot \sum_{n=0}^{\infty} \left(\frac{1}{10}\right)^n \overset{(*)}{=} \frac{9}{10} \cdot \frac{1}{1-\frac{1}{10}} = 1.$$

Die Zahl $0,9999\ldots$ ist daher wirklich *gleich* 1 – in diesem Fall ist die Dezimaldarstellung einer reellen Zahl also nicht eindeutig.

(b) (Teleskopreihen) Wir wollen den Grenzwert der Reihe $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ bestimmen. Normalerweise lassen sich derartige Reihen nicht ohne weiteres berechnen, aber in diesem ganz speziellen Fall können wir einen Trick anwenden: Wegen $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$ können wir die Partialsummen der Reihe schreiben als

$$\begin{aligned}
\sum_{n=1}^{N} \frac{1}{n(n+1)} &= \sum_{n=1}^{N} \left(\frac{1}{n} - \frac{1}{n+1}\right) \\
&= \left(\frac{1}{1} - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{N} - \frac{1}{N+1}\right) \\
&= \frac{1}{1} - \frac{1}{N+1}.
\end{aligned}$$

Derartige Reihen, bei denen sich in den Partialsummen durch geeignete Differenzen alle Terme bis auf einen Start- und Endterm wegheben, bezeichnet man als **Teleskopreihen** (weil die Summe sozusagen wie ein Teleskop „zusammengeschoben" werden kann). Der Grenzwert der Reihe lässt sich dann natürlich einfach berechnen; in diesem Fall ist er

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{N\to\infty} \sum_{n=1}^{N} \frac{1}{n(n+1)} = \lim_{N\to\infty} \left(1 - \frac{1}{N+1}\right) = 1.$$

(c) (**Harmonische Reihe**) Die Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ divergiert: Für die Partialsummen mit Index $N = 2^k$ gilt

$$\begin{aligned}
\sum_{n=1}^{2^k} \frac{1}{n} &= 1 + \tfrac{1}{2} + \left(\tfrac{1}{3} + \tfrac{1}{4}\right) + \left(\tfrac{1}{5} + \tfrac{1}{6} + \tfrac{1}{7} + \tfrac{1}{8}\right) + \cdots + \left(\tfrac{1}{2^{k-1}+1} + \cdots + \tfrac{1}{2^k}\right) \\
&\geq 1 + \tfrac{1}{2} + \left(\tfrac{1}{4} + \tfrac{1}{4}\right) + \left(\tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8}\right) + \cdots + \left(\tfrac{1}{2^k} + \cdots + \tfrac{1}{2^k}\right) \\
&= 1 + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2} + \cdots + \tfrac{1}{2} \\
&= 1 + \tfrac{k}{2}.
\end{aligned}$$

Da $1 + \frac{k}{2}$ mit k unbeschränkt wächst, ist die gegebene Reihe also unbeschränkt und damit nach Lemma 5.8 divergent.

Die folgenden einfachen Rechenregeln für Reihen – die Verträglichkeit mit Summen, Differenzen, Multiplikation mit Konstanten sowie im Fall des Körpers $\mathbb{R}$ mit Ungleichungen – ergeben sich sofort aus denen für Folgen in Kapitel 5.

Lemma 7.4 (Rechenregeln für Reihen)**.** *Es seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ konvergente Reihen in $\mathbb{K}$. Dann gilt:*

(a) $\displaystyle\sum_{n=0}^{\infty} (a_n + b_n) = \sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} b_n$ *und* $\displaystyle\sum_{n=0}^{\infty} (a_n - b_n) = \sum_{n=0}^{\infty} a_n - \sum_{n=0}^{\infty} b_n$.

(b) *Für* $c \in \mathbb{K}$ *ist* $\sum_{n=0}^{\infty} ca_n = c \cdot \sum_{n=0}^{\infty} a_n$.

(c) *Ist* $\mathbb{K} = \mathbb{R}$ *und* $a_n \leq b_n$ *für alle* n, *so ist* $\sum_{n=0}^{\infty} a_n \leq \sum_{n=0}^{\infty} b_n$.

Beweis. Alle behaupteten Aussagen gelten trivialerweise für die Partialsummen der Reihen (also wenn die Summen bis zu einem festen $N \in \mathbb{N}$ laufen). Übergang zum Grenzwert liefert dann mit den Sätzen 5.13 und 5.23 die Behauptungen. □

Eine analoge direkte Verträglichkeit mit der Multiplikation ist natürlich nicht zu erwarten, weil ja schon für die Partialsummen $\left(\sum_{n=0}^{N} a_n\right) \cdot \left(\sum_{n=0}^{N} b_n\right)$ nicht dasselbe ist wie $\sum_{n=0}^{N} a_n b_n$. Wir werden aber später in Satz 7.34 noch eine Formel für das Produkt von Reihen finden.

Bevor wir nun mit der Herleitung allgemeiner Konvergenzkriterien für Reihen beginnen, wollen wir noch zwei sehr einfache Hilfsaussagen festhalten, die aber dennoch oft nützlich sind. Die erste von ihnen ist so einfach, dass sie üblicherweise als *Trivialkriterium* bezeichnet wird: Eine Reihe kann höchstens dann konvergieren, wenn die aufsummierten Zahlen zumindest gegen 0 konvergieren.

Lemma 7.5 (**Trivialkriterium**). *Ist die Reihe* $\sum_{n=0}^{\infty} a_n$ *konvergent, so ist* $(a_n)_n$ *eine Nullfolge.*

Beweis. Existiert der Grenzwert $\sum_{n=0}^{\infty} a_n$, so folgt aus den Grenzwertsätzen

$$a_N = \sum_{n=0}^{N} a_n - \sum_{n=0}^{N-1} a_n \to \sum_{n=0}^{\infty} a_n - \sum_{n=0}^{\infty} a_n = 0 \quad \text{für } N \to \infty. \qquad \square$$

Beispiel 7.6.

(a) Die Reihe $\sum_{n=0}^{\infty} \frac{2n^2}{n^2+1}$ ist divergent, denn nach Beispiel 5.14 ist $\lim_{n\to\infty} \frac{2n^2}{n^2+1} = 2 \neq 0$.

(b) Das Trivialkriterium ist nicht umkehrbar: So ist z. B. zwar $(\frac{1}{n})_n$ eine Nullfolge, aber die harmonische Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ nach Beispiel 7.3 (c) trotzdem divergent. Man kann mit diesem Kriterium also immer nur die Divergenz einer Reihe nachweisen, aber nie die Konvergenz.

Dieses Beispiel zeigt auch noch etwas anderes: Bezeichnen wir mit $a_n = \sum_{k=1}^{n} \frac{1}{k}$ die Partialsummen der harmonischen Reihe, so ist die Folge $(a_n)_n$ zwar divergent, aber die Folge $(a_n - a_{n-1})_n = (\frac{1}{n})_n$ konvergiert trotzdem gegen 0, d. h. es gilt

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : |a_n - a_{n-1}| < \varepsilon.$$

Um die Äquivalenz zwischen konvergenten Folgen und Cauchyfolgen zu erhalten, genügt es in der Definition 6.22 einer Cauchyfolge also nicht, zwei *benachbarte* Folgenglieder a_n und a_{n-1} miteinander zu vergleichen, sondern wir müssen zwei *beliebige* Folgenglieder a_m und a_n (mit $m, n \geq n_0$) nehmen!

Lemma 7.7. *Eine Reihe* $\sum_{n=0}^{\infty} a_n$ *mit* $a_n \in \mathbb{R}_{\geq 0}$ *für alle* $n \in \mathbb{N}$ *ist genau dann konvergent, wenn sie beschränkt ist.*

Beweis. Da alle aufsummierten Zahlen reell und nicht-negativ sind, ist die Folge ihrer Partialsummen monoton wachsend. Für eine reelle, monoton wachsende Folge ist die Konvergenz nach Lemma 5.8 und dem Monotoniekriterium aus Satz 5.27 aber äquivalent zur Beschränktheit. □

7.B Konvergenzkriterien für Reihen

Wie im Fall von Folgen im letzten Kapitel wollen wir nun einige Kriterien herleiten, mit denen man die Konvergenz einer Reihe beweisen kann, ohne ihren Grenzwert zu kennen. Dabei bleiben natürlich alle Ergebnisse aus Abschnitt 5.B unverändert anwendbar, da Reihen ja letztlich auch nur Folgen sind. Es gibt aber einige zusätzliche Kriterien, die speziell auf den Fall von Reihen zugeschnitten und meistens einfacher zu überprüfen sind. Wir beginnen dabei mit einem Kriterium für reelle Reihen, in denen abwechselnd positive und negative Glieder aufsummiert werden.

Satz 7.8 (**Leibniz-Kriterium**). *Ist $(a_n)_n$ eine monoton fallende Nullfolge in $\mathbb{R}_{\geq 0}$, so ist die Reihe*

$$\sum_{n=0}^{\infty} (-1)^n a_n = a_0 - a_1 + a_2 - a_3 \pm \cdots$$

*konvergent, und ihre Partialsummen sind abwechselnd obere und untere Schranken für ihren Grenzwert. (Derartige reelle Reihen, bei denen sich das Vorzeichen in der Summe immer abwechselt, nennt man **alternierend**.)*

Beweis. Es sei $s_N = \sum_{n=0}^{N} (-1)^n a_n$, also $(s_N)_N$ die Folge der Partialsummen der betrachteten Reihe. Da $(a_n)_n$ monoton fallend und die Differenz zweier aufeinander folgender Glieder von $(a_n)_n$ damit nicht negativ ist, ist die Folge $(s_{2N})_{N\in\mathbb{N}}$ der *geraden* Partialsummen monoton fallend: Es gilt

$$s_{2N+2} = s_{2N} - \underbrace{(a_{2N+1} - a_{2N+2})}_{\geq 0} \leq s_{2N}.$$

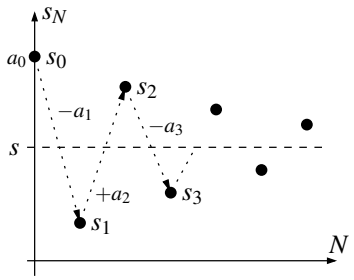
Analog ist die Folge $(s_{2N+1})_{N\in\mathbb{N}}$ der ungeraden Partialsummen monoton wachsend, wie auch das Bild unten rechts zeigt. Damit haben wir ineinander liegende Intervalle

$$[s_1, s_2] \supset [s_3, s_4] \supset [s_5, s_6] \supset \cdots,$$

die eine Intervallschachtelung definieren, da die Länge

$$s_{2N} - s_{2N-1} = a_{2N}$$

dieser Intervalle mit $N \to \infty$ gegen 0 konvergiert. Nach Satz 5.38 konvergieren also die geraden und ungeraden Partialsummen monoton fallend bzw. wachsend gegen den gleichen Grenzwert s. Insbesondere sind die geraden und ungeraden Partialsummen also obere bzw. untere Schranken für s.



Außerdem liegen damit in jeder ε-Umgebung von s fast alle geraden und fast alle ungeraden Partialsummen, und somit konvergiert auch die gesamte Folge der Partialsummen gegen s. □

14

Beispiel 7.9 (**Alternierende harmonische Reihe**). Die Reihe

$$\sum_{n=1}^{\infty} (-1)^n \frac{1}{n} = -\frac{1}{1} + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} \mp \cdots$$

ist nach dem Leibniz-Kriterium konvergent, denn $\frac{1}{n}$ ist eine monoton fallende Nullfolge. Ihren Grenzwert können wir momentan noch nicht berechnen (in der Tat ist er gleich $-\log 2$, wie wir in Beispiel 11.16 (a) sehen werden), aber nach Satz 7.8 liegt er sicher zwischen den ersten beiden Partialsummen $-\frac{1}{1} = -1$ und $-\frac{1}{1} + \frac{1}{2} = -\frac{1}{2}$.

Übrigens ist diese Reihe (ganz im Gegensatz z. B. zur Folge aus Beispiel 5.36) eine, die „extrem langsam“ konvergiert: Um hier den Grenzwert auf k Nachkommastellen genau zu berechnen, müssen wir natürlich mindestens die ersten 10^k Summanden mitnehmen, denn der 10^k-te Summand ist ja 10^{-k} und ändert somit in jedem Fall noch die k-te Nachkommastelle.

Wir können an dieser alternierenden harmonischen Reihe aber noch eine weitere überraschende Eigenschaft sehen. Dazu sortieren wir die aufzusummierenden Zahlen mal etwas um und schreiben unsere Reihe als

$$\left(-\frac{1}{1} + \frac{1}{2}\right) + \frac{1}{4} + \left(-\frac{1}{3} + \frac{1}{6}\right) + \frac{1}{8} + \left(-\frac{1}{5} + \frac{1}{10}\right) + \frac{1}{12} + \left(-\frac{1}{7} + \frac{1}{14}\right) + \frac{1}{16} + \cdots.$$

Das Prinzip hierbei ist, dass die Terme $(-1)^n \frac{1}{n} \ldots$

- für ungerade n der Reihe nach als erste Summanden in den Klammern stehen,
- für gerade, aber nicht durch 4 teilbare n der Reihe nach als zweite Summanden in den Klammern stehen,
- für durch 4 teilbare n der Reihe nach außerhalb der Klammern stehen.

Es ist klar, dass wir hier wirklich nur die Summanden umsortiert, also keinen vergessen oder doppelt hingeschrieben haben. Rechnen wir jetzt aber mal die Klammern aus, so erhalten wir

$$-\frac{1}{2}+\frac{1}{4}-\frac{1}{6}+\frac{1}{8}-\frac{1}{10}+\frac{1}{12}-\frac{1}{14}+\frac{1}{16}\mp\cdots$$

und damit genau die Hälfte der ursprünglichen Reihe! Da die Reihe nicht den Wert 0 hat (wie wir oben schon gesehen haben, liegt ihr Wert ja zwischen -1 und $-\frac{1}{2}$), haben wir ihren Wert durch das Umsortieren also tatsächlich geändert und müssen damit wohl oder übel feststellen:

Das Umordnen der Summanden in einer konvergenten Reihe kann ihren Grenzwert ändern.

Das ist natürlich extrem lästig, weil uns das sozusagen die Kommutativität der Addition im Fall von unendlichen Summen kaputt macht – was völlig der Intuition widerspricht und natürlich auch beim Rechnen mit solchen Reihen große Probleme bereitet. Glücklicherweise gibt es einen relativ eleganten Ausweg aus dieser Situation: Es gibt eine Eigenschaft von Reihen, die etwas stärker als die normale Konvergenz ist, in vielen Fällen aber dennoch erfüllt ist und die Umsortierbarkeit ohne Änderung des Grenzwerts garantiert. Diese wollen wir jetzt einführen.

Definition 7.10 (Absolute Konvergenz)**.** Eine Reihe $\sum_{n=0}^{\infty} a_n$ in $\mathbb{K}$ heißt **absolut konvergent**, wenn die Reihe $\sum_{n=0}^{\infty} |a_n|$ ihrer Beträge konvergiert, also nach Lemma 7.7 wenn diese Reihe $\sum_{n=0}^{\infty} |a_n|$ beschränkt ist.

(Der Name kommt einfach daher, dass man den Betrag einer Zahl oft auch als *Absolutbetrag* bezeichnet.)

Für Reihen, in denen nur nicht-negative reelle Zahlen aufsummiert werden, stimmen die Begriffe „konvergent“ und „absolut konvergent“ offensichtlich überein. Wir wollen nun sehen, dass der Begriff der absoluten Konvergenz für allgemeine Reihen wirklich „stärker“ als die gewöhnliche Konvergenz ist, also dass aus der absoluten Konvergenz einer Reihe auch die Konvergenz folgt. Dazu müssen wir zunächst das Cauchy-Kriterium aus Satz 6.25 auf Reihen übertragen. Auch hier ist dieses Kriterium wieder besonders deswegen wichtig, weil es zum einen zur Konvergenz *äquivalent* ist (man mit ihm also Konvergenz genauso wie Divergenz nachweisen kann) und es außerdem in $\mathbb{R}$ und $\mathbb{C}$ gleichermaßen funktioniert.

Folgerung 7.11 (**Cauchy-Kriterium** für Reihen)**.** *Eine Reihe $\sum_{n=0}^{\infty} a_n$ in $\mathbb{K}$ ist genau dann konvergent, wenn*

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall m \geq n \geq n_0 : \left| \sum_{k=n+1}^{m} a_k \right| < \varepsilon.$$

Beweis. Nach Definition ist die Reihe $\sum_{n=0}^{\infty} a_n$ genau dann konvergent, wenn die Folge $(s_N)_N$ der Partialsummen mit $s_N = \sum_{n=0}^{N} a_n$ konvergiert. Wenden wir das Cauchy-Kriterium für Folgen aus Satz 6.25 auf $(s_N)_N$ an, sehen wir, dass dies genau dann der Fall ist, wenn

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall m, n \geq n_0 : |s_n - s_m| < \varepsilon.$$

Natürlich können wir hier aus Symmetriegründen $m \geq n$ annehmen, und aus $s_m - s_n = \sum_{k=n+1}^{m} a_k$ folgt dann sofort die Behauptung. □

Lemma 7.12. *Jede absolut konvergente Reihe in $\mathbb{K}$ ist konvergent.*

Beweis. Es sei $\sum_{n=0}^{\infty} a_n$ eine absolut konvergente Reihe, d. h. die Reihe $\sum_{n=0}^{\infty} |a_n|$ sei konvergent. Nach dem Cauchy-Kriterium aus Folgerung 7.11 gibt es also zu jedem $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ mit

$$\left| \sum_{k=n+1}^{m} |a_k| \right| = \sum_{k=n+1}^{m} |a_k| < \varepsilon$$

für alle $m \geq n \geq n_0$. Dann ist nach der Dreiecksungleichung aber erst recht

$$\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k| < \varepsilon,$$

und damit ist wiederum nach dem Cauchy-Kriterium auch die Reihe $\sum_{n=0}^{\infty} a_n$ konvergent. □

Beispiel 7.13. Die alternierende harmonische Reihe $\sum_{n=1}^{\infty}(-1)^n \frac{1}{n}$ ist nach Beispiel 7.9 konvergent. Sie ist aber nicht absolut konvergent, da die harmonische Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ nach Beispiel 7.3 (c) divergiert. Die Umkehrung von Lemma 7.12 gilt also nicht.

Als Nächstes hatten wir behauptet, dass die absolute Konvergenz einer Reihe sicher stellt, dass man die Summanden ohne Änderung des Grenzwerts umordnen kann. Dies wollen wir jetzt zeigen.

Definition 7.14 (Umordnungen einer Reihe)**.** Es sei $\sum_{n=0}^{\infty} a_n$ eine Reihe in $\mathbb{K}$ und $\sigma : \mathbb{N} \to \mathbb{N}$ eine bijektive Abbildung. Dann heißt die Reihe $\sum_{n=0}^{\infty} a_{\sigma(n)}$ (die offensichtlich aus den gleichen Summanden besteht, nur evtl. in anderer Reihenfolge) eine **Umordnung** von $\sum_{n=0}^{\infty} a_n$.

Satz 7.15 (**Umordnungssatz**)**.** *Jede Umordnung einer absolut konvergenten Reihe ist ebenfalls absolut konvergent und konvergiert gegen denselben Grenzwert.*

Beweis. Es seien $\sum_{n=0}^{\infty} a_n$ eine absolut konvergente Reihe und $\sigma : \mathbb{N} \to \mathbb{N}$ eine bijektive Abbildung. Ferner sei $\varepsilon > 0$ beliebig. Da $\sum_{n=0}^{\infty} |a_n|$ nach Voraussetzung konvergiert, gibt es nach dem Cauchy-Kriterium aus Folgerung 7.11 ein $n_0 \in \mathbb{N}$ mit $\sum_{k=n+1}^{m} |a_k| < \frac{\varepsilon}{2}$ für alle $m \geq n \geq n_0$. Insbesondere haben wir für $n = n_0$ und $m \to \infty$ also

$$\sum_{k=n_0+1}^{\infty} |a_k| \leq \frac{\varepsilon}{2} < \varepsilon. \tag{1}$$

Da σ surjektiv ist, können wir nun ein $n_0' \geq n_0$ wählen, so dass

$$\{0, 1, \ldots, n_0\} \subset \{\sigma(0), \sigma(1), \ldots, \sigma(n_0')\}. \tag{2}$$

Wir betrachten nun für beliebiges $n \geq n_0'$ die Summe

$$\sum_{k=0}^{n} (a_{\sigma(k)} - a_k) = (a_{\sigma(0)} - a_0) + (a_{\sigma(1)} - a_1) + \cdots + (a_{\sigma(n)} - a_n).$$

Wegen (2) und $n \geq n_0' \geq n_0$ treten in dieser Summe alle Glieder $a_0, \ldots, a_{n_0}$ sowohl einmal mit positivem als auch einmal mit negativem Vorzeichen auf, heben sich also heraus. Die übrigen a_n mit $n > n_0$ können sich ebenfalls herausheben, oder mit einem positiven oder negativen Vorzeichen auftreten. Wir können dies symbolisch schreiben als

$$\sum_{k=0}^{n} (a_{\sigma(k)} - a_k) = \sum_k \pm a_k,$$

wobei die Summe hier über gewisse (endlich viele) $k > n_0$ läuft und für jedes solche k das Vorzeichen von a_k positiv oder negativ sein kann. Damit können wir diesen Ausdruck mit der Dreiecksungleichung betragsmäßig abschätzen durch

$$\left| \sum_{k=0}^{n} (a_{\sigma(k)} - a_k) \right| = \left| \sum_k \pm a_k \right| \leq \sum_k |a_k| \leq \sum_{k=n_0+1}^{\infty} |a_k| \overset{(1)}{<} \varepsilon.$$

Daraus ergibt sich $\sum_{n=0}^{\infty}(a_{\sigma(n)} - a_n) = 0$, und damit nach den üblichen Rechenregeln aus Lemma 7.4

$$\sum_{n=0}^{\infty} a_{\sigma(n)} = \sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} (a_{\sigma(n)} - a_n) = \sum_{n=0}^{\infty} a_n.$$

Die Umordnung $\sum_{n=0}^{\infty} a_{\sigma(n)}$ konvergiert also gegen den gleichen Grenzwert wie die ursprüngliche Reihe. Wenden wir dieses Ergebnis nun auch noch auf die Reihe $\sum_{n=0}^{\infty} |a_n|$ an, so erhalten wir genauso $\sum_{n=0}^{\infty} |a_{\sigma(n)}| = \sum_{n=0}^{\infty} |a_n|$, woraus die absolute Konvergenz der Umordnung folgt. □

Bemerkung 7.16 (Summen mit abzählbar unendlicher Indexmenge)**.** Da es bei „unendlichen Summen“ im Fall der absoluten Konvergenz also nicht auf die Reihenfolge der Summanden ankommt, können wir damit auch derartige Summen definieren, bei denen die Summanden zunächst einmal überhaupt keine vorgegebene Reihenfolge haben, sondern durch eine beliebige abzählbar unendliche Menge I indiziert werden: Ist $a_i \in \mathbb{K}$ für alle $i \in I$, so wählen wir eine bijektive Abbildung

$\sigma : \mathbb{N} \to I$. Ist dann die Reihe $\sum_{n=0}^{\infty} a_{\sigma(n)}$ absolut konvergent, so schreiben wir den Wert dieser Reihe als

$$\sum_{i \in I} a_i := \sum_{n=0}^{\infty} a_{\sigma(n)} \quad \in \mathbb{K}.$$

Dies hängt dann nach dem Umordnungssatz 7.15 nicht von der Wahl von σ ab, da sich die durch eine andere Bijektion entstehende Reihe nur durch eine Umordnung unterscheidet und somit nichts an der absoluten Konvergenz bzw. dem Grenzwert der Reihe ändert.

Ist die Reihe $\sum_{n=0}^{\infty} a_{\sigma(n)}$ hingegen nicht absolut konvergent, so können wir $\sum_{i \in I} a_i$ nicht sinnvoll definieren.

Aufgabe 7.17. Zeige die folgende Verallgemeinerung des Leibniz-Kriteriums ins Komplexe: Ist $(a_n)_n$ eine reelle, monoton fallende Nullfolge, so konvergiert die Reihe $\sum_{n=0}^{\infty} a_n x^n$ für alle $x \in \mathbb{K}$ mit $|x| = 1$ und $x \neq 1$.

(Hinweis: Untersuche die Reihe $\displaystyle\sum_{n=0}^{\infty} a_n (x-1) x^n$.)

Aufgrund der schönen Eigenschaften absolut konvergenter Reihen werden wir uns im Folgenden oftmals eher für die absolute als für die „gewöhnliche“ Konvergenz von Reihen interessieren. Wir wollen nun ein paar Kriterien zusammentragen, mit denen man die absolute Konvergenz von Reihen in vielen Fällen einfach nachprüfen kann. Das erste von ihnen ist eigentlich sehr offensichtlich:

Satz 7.18 (Majoranten-/Minorantenkriterium). *Es seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ zwei Reihen in $\mathbb{K}$ mit $|a_n| \leq |b_n|$ für fast alle n.*

(a) *Ist $\sum_{n=0}^{\infty} b_n$ absolut konvergent, so auch $\sum_{n=0}^{\infty} a_n$.*
*(Man nennt $\sum_{n=0}^{\infty} b_n$ in diesem Fall eine konvergente **Majorante** von $\sum_{n=0}^{\infty} a_n$.)*

(b) *Ist $\sum_{n=0}^{\infty} a_n$ divergent, so auch $\sum_{n=0}^{\infty} |b_n|$.*
*(Man nennt $\sum_{n=0}^{\infty} a_n$ in diesem Fall eine divergente **Minorante** von $\sum_{n=0}^{\infty} |b_n|$.)*

Beweis.

(a) Ist $\sum_{n=0}^{\infty} b_n$ absolut konvergent, also $\sum_{n=0}^{\infty} |b_n|$ beschränkt, so ist wegen $|a_n| \leq |b_n|$ für fast alle n auch $\sum_{n=0}^{\infty} |a_n|$ beschränkt, und damit $\sum_{n=0}^{\infty} a_n$ nach Lemma 7.7 absolut konvergent.

(b) Ist die Reihe $\sum_{n=0}^{\infty} a_n$ nicht konvergent, so ist sie nach Lemma 7.12 insbesondere auch nicht absolut konvergent, und daher kann nach (a) auch $\sum_{n=0}^{\infty} b_n$ nicht absolut konvergent sein, d. h. die Reihe $\sum_{n=0}^{\infty} |b_n|$ divergiert. □

Beispiel 7.19.

(a) Die Reihe $\sum_{n=1}^{\infty} \frac{1}{n^2}$ ist konvergent: Wegen $\frac{1}{(n+1)^2} \leq \frac{1}{n(n+1)}$ für $n \geq 1$ ist $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ nach Beispiel 7.3 (b) eine (absolut) konvergente Majorante von $\sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$. Damit konvergiert die Reihe

$$\frac{1}{1^2} + \sum_{n=1}^{\infty} \frac{1}{(n+1)^2} = \frac{1}{1^2} + \sum_{n=2}^{\infty} \frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Beachte, dass man auf diese Art mit Hilfe des Majorantenkriteriums zwar die Konvergenz der Reihe $\sum_{n=1}^{\infty} \frac{1}{n^2}$ beweisen, aber nicht ihren Grenzwert bestimmen kann (in der Tat kann man zeigen, dass der Wert dieser Reihe gleich $\frac{\pi^2}{6}$ ist).

(b) Für $k \geq 2$ ist die Reihe $\sum_{n=1}^{\infty} \frac{1}{n^k}$ konvergent, denn wegen $\frac{1}{n^k} \leq \frac{1}{n^2}$ für alle n ist $\sum_{n=1}^{\infty} \frac{1}{n^2}$ nach (a) eine konvergente Majorante.

(c) Die Reihe $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ dagegen ist divergent, denn wegen $\frac{1}{\sqrt{n}} \geq \frac{1}{n}$ ist die harmonische Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ aus Beispiel 7.3 (c) eine divergente Minorante.

Aufgabe 7.20. Es sei $(a_n)_n$ eine Folge in $\mathbb{K}$ mit $a_n \neq -1$ für alle $n \in \mathbb{N}$. Man zeige:

$$\sum_{n=0}^{\infty} a_n \text{ ist absolut konvergent} \quad \Leftrightarrow \quad \sum_{n=0}^{\infty} \frac{a_n}{1+a_n} \text{ ist absolut konvergent.}$$

Wenn man mit dem Majorantenkriterium die (absolute) Konvergenz einer Reihe nachweisen möchte, stellt sich natürlich die Frage, wo man eine konvergente Majorante herbekommt. Sehr oft kann man hierfür einfach eine geometrische Reihe $\sum_{n=0}^{\infty} q^n$ für ein $q \in \mathbb{R}_{>0}$ mit $q < 1$ wie in Beispiel 7.3 (a) verwenden. Aus diesem Ansatz ergeben sich in der Tat die folgenden beiden allgemeinen Kriterien, die sehr oft anwendbar sind:

Satz 7.21 (Quotientenkriterium). *Es sei $\sum\limits_{n=0}^{\infty} a_n$ eine Reihe in $\mathbb{K}$ mit $a_n \neq 0$ für fast alle n. Dann gilt:*

(a) *Ist $\lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1$, so ist die Reihe $\sum\limits_{n=0}^{\infty} a_n$ absolut konvergent.*

(b) *Ist $\lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| > 1$, so ist die Reihe $\sum\limits_{n=0}^{\infty} a_n$ divergent.*

Der Fall $\lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \infty$ ist dabei in (b) zugelassen. Ist die Folge $\left(\left|\frac{a_{n+1}}{a_n}\right|\right)_n$ jedoch unbestimmt divergent oder konvergiert sie gegen 1, so macht das Quotientenkriterium keine Aussage.

15

Beweis. Es sei $a := \lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| \in \mathbb{R}_{\geq 0} \cup \{\infty\}$.

(a) Ist $a < 1$, so können wir ein $\varepsilon > 0$ wählen, so dass auch noch $q := a + \varepsilon < 1$ gilt. Wegen $\lim\limits_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = a$ gibt es dann ein $n_0 \in \mathbb{N}$, so dass

$$\left|\frac{a_{n+1}}{a_n}\right| < a + \varepsilon = q \quad \text{für alle } n \geq n_0,$$

und damit $|a_{n+1}| < q\,|a_n|$. Daraus ergibt sich für alle $n \geq n_0$

$$|a_n| < q\,|a_{n-1}| < q^2\,|a_{n-2}| < \dots < q^{n-n_0}\,|a_{n_0}|.$$

Also ist die Reihe $\sum_{n=0}^{\infty} q^{n-n_0}\,|a_{n_0}|$ eine Majorante der gegebenen Reihe $\sum_{n=0}^{\infty} a_n$. Wegen $q < 1$ konvergiert sie nach Beispiel 7.3 (a) absolut, denn es ist

$$\sum_{n=0}^{\infty} q^{n-n_0}\,|a_{n_0}| = q^{-n_0}\,|a_{n_0}| \sum_{n=0}^{\infty} q^n = q^{-n_0}\,|a_{n_0}| \cdot \frac{1}{1-q}.$$

Die zu beweisende Aussage folgt damit aus dem Majorantenkriterium von Satz 7.18.

(b) Ist $a \in \mathbb{R}$ mit $a > 1$, so können wir ein $\varepsilon > 0$ finden mit $a - \varepsilon > 1$. In diesem Fall gibt es nach der Grenzwertbedingung ein $n_0 \in \mathbb{N}$, so dass

$$\left|\frac{a_{n+1}}{a_n}\right| > a - \varepsilon > 1 \quad \text{für alle } n \geq n_0.$$

Beachte, dass dies auch im Fall $a = \infty$ gilt, denn auch dann sind ja insbesondere fast alle $\left|\frac{a_{n+1}}{a_n}\right|$ größer als 1.

Also gilt $|a_{n+1}| > |a_n|$ für alle $n \geq n_0$. Damit ist $(|a_n|)_n$ ab n_0 aber eine monoton wachsende Folge positiver Zahlen, und kann somit keine Nullfolge sein. Die gegebene Reihe divergiert also nach dem Trivialkriterium aus Lemma 7.5. □

Das zweite, recht ähnliche Kriterium, das auf dem Vergleich mit der geometrischen Reihe beruht, benutzt die höheren Wurzeln aus Aufgabe 5.36.

Satz 7.22 (Wurzelkriterium). *Für jede Reihe $\sum\limits_{n=0}^{\infty} a_n$ in $\mathbb{K}$ gilt:*

(a) *Ist $\limsup\limits_{n\to\infty} \sqrt[n]{|a_n|} < 1$, so ist die Reihe $\sum\limits_{n=0}^{\infty} a_n$ absolut konvergent.*

(b) *Ist* $\limsup\limits_{n\to\infty} \sqrt[n]{|a_n|} > 1$, *so ist die Reihe* $\sum\limits_{n=0}^{\infty} a_n$ *divergent.*

Der Fall $\limsup\limits_{n\to\infty} \sqrt[n]{|a_n|} = \infty$ *ist dabei in (b) wieder zugelassen. Ist jedoch* $\limsup\limits_{n\to\infty} \sqrt[n]{|a_n|} = 1$, *so macht das Wurzelkriterium keine Aussage.*

Beweis. Es sei $a := \limsup_{n\to\infty} \sqrt[n]{|a_n|} \in \mathbb{R}_{\geq 0} \cup \{\infty\}$.

(a) Für $a < 1$ sei wieder $\varepsilon > 0$ mit $q := a + \varepsilon < 1$. Nach Lemma 5.46 (a) gilt dann

$$\sqrt[n]{|a_n|} < a + \varepsilon = q, \qquad \text{also} \quad |a_n| < q^n$$

für fast alle n. Also ist $\sum_{n=0}^{\infty} q^n$ eine Majorante der gegebenen Reihe. Da diese wegen $q < 1$ nach Beispiel 7.3 (a) (absolut) konvergiert, konvergiert auch $\sum_{n=0}^{\infty} a_n$ nach dem Majorantenkriterium aus Satz 7.18 absolut.

(b) Ist $a \in \mathbb{R}$ mit $a > 1$, so wählen wir ein $\varepsilon > 0$ mit $a - \varepsilon > 1$. Diesmal folgt dann aus Lemma 5.46 (b), dass

$$\sqrt[n]{|a_n|} > a - \varepsilon > 1$$

für unendlich viele n. Beachte, dass dies auch im Fall $a = \infty$ gilt, weil die Folge $(\sqrt[n]{|a_n|})_n$ dann eine Teilfolge mit uneigentlichem Grenzwert ∞ hat.

Damit ist aber auch $|a_n| > 1$ für unendlich viele n. Also ist $(a_n)_n$ keine Nullfolge, und die gegebene Reihe divergiert nach dem Trivialkriterium aus Lemma 7.5. □

Bemerkung 7.23 (Vergleich von Quotienten- und Wurzelkriterium)**.** Das Quotientenkriterium hat gegenüber dem Wurzelkriterium den Vorteil, dass sich der Quotient $\left|\frac{a_{n+1}}{a_n}\right|$ oft einfacher berechnen lässt als die Wurzel $\sqrt[n]{|a_n|}$. Allerdings benötigen wir im Quotientenkriterium einen Grenzwert der Quotientenfolge, während im Wurzelkriterium der Limes superior der Wurzelfolge genügt, der ja nach Bemerkung 5.51 zumindest im uneigentlichen Sinne stets existiert.

Dies liegt daran, dass wir für die Induktion im Beweis von Satz 7.21 brauchten, dass fast alle Quotienten in (a) kleiner als $a + \varepsilon$ und in (b) größer als $a - \varepsilon$ sind, so dass a dort der Grenzwert der Quotientenfolge sein musste. Im Beweis von Satz 7.22 brauchten wir dagegen in (a) zwar auch, dass fast alle Wurzeln kleiner als $a + \varepsilon$ sind, aber in (b) reichten unendlich viele Wurzeln größer als $a - \varepsilon$.

Mit dieser Beobachtung sieht man allerdings mit Hilfe von Lemma 5.46 auch, dass wir den Grenzwert im Quotientenkriterium von Satz 7.21 in (a) durch den Limes superior und in (b) durch den Limes inferior ersetzen könnten, um so noch allgemeinere Aussagen zu erhalten. Hat die Quotientenfolge jedoch mehrere Häufungspunkte, von denen einer größer als 1 und einer kleiner als 1 ist, so lässt sich aus der Idee des Quotientenkriteriums aber endgültig keine Aussage über die Konvergenz der Reihe mehr herleiten.

Beispiel 7.24.

(a) Betrachten wir für ein $q \in \mathbb{K}$ mit $q \neq 0$ die geometrische Reihe $\sum_{n=0}^{\infty} q^n$ selbst, ist also $a_n = q^n$ in der Notation von Satz 7.21 und 7.22, so ist offensichtlich $\left|\frac{a_{n+1}}{a_n}\right| = \sqrt[n]{|a_n|} = |q|$ unabhängig von n, und sowohl Quotienten- als auch Wurzelkriterium reproduzieren einfach das Ergebnis aus Beispiel 7.3 (a) in den Fällen mit $|q| \neq 1$.

(b) Für die alternierende harmonische Reihe $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} (-1)^n \frac{1}{n}$ aus Beispiel 7.9 macht das Quotientenkriterium keine Aussage, denn dort gilt

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{(-1)^{n+1}/(n+1)}{(-1)^n/n}\right| = \frac{n}{n+1} = \frac{1}{1+\frac{1}{n}} \to 1$$

für $n \to \infty$. Dies war natürlich zu erwarten, da diese Reihe ja auch weder absolut konvergent noch divergent ist.

(c) Die Reihe $\sum_{n=0}^{\infty}\left(\frac{n}{2n+1}\right)^n$ ist nach dem Wurzelkriterium (absolut) konvergent, denn es ist

$$\sqrt[n]{\left(\frac{n}{2n+1}\right)^n} = \frac{n}{2n+1} = \frac{1}{2+\frac{1}{n}} \to \frac{1}{2}$$

für $n \to \infty$.

(d) Für alle $x \in \mathbb{K}$ ist die Reihe

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots$$

nach dem Quotientenkriterium absolut konvergent, denn es gilt

$$\left|\frac{x^{n+1}/(n+1)!}{x^n/n!}\right| = \frac{|x|}{n+1} \to 0$$

für $n \to \infty$. Auf diese Art haben wir also letztlich eine Funktion von $\mathbb{K}$ nach $\mathbb{K}$ definiert, die jedem x den Wert der Reihe $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ zuordnet.

Es handelt sich bei diesem letzten Beispiel genau um die Exponentialfunktion, die ihr zumindest im reellen Fall bereits aus der Schule kennt. Sie ist aber letztlich nur ein spezielles Beispiel für eine sehr große Klasse von Funktionen, die sich in der Form $x \mapsto \sum_{n=0}^{\infty} a_n x^n$ für gewisse $a_n \in \mathbb{K}$ schreiben lassen. Wir wollen derartige Funktionen, die in dieser Vorlesung immer wieder vorkommen werden, daher jetzt einführen und etwas genauer untersuchen.

7.C Potenzreihen

Potenzreihen kann man sich in gewissem Sinne als Verallgemeinerung der Polynome aus Abschnitt 3.C vorstellen: statt endlicher Summen $a_0 + a_1 x + \cdots + a_n x^n$ in einer Variablen x betrachten wir nun unendliche Reihen der Form

$$\sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots .$$

Wir beginnen mit der formalen Definition solcher Reihen, zusammen mit dem wohl wichtigsten Beispiel: der Exponentialfunktion.

Definition 7.25 (Potenzreihen und die Exponentialfunktion)**.**

(a) Ist $(a_n)_n$ eine Folge in $\mathbb{K}$ und $x \in \mathbb{K}$, so heißt die Reihe $\sum_{n=0}^{\infty} a_n x^n$ die **Potenzreihe** in x mit Koeffizienten $(a_n)_n$. Ist $D \subset \mathbb{K}$ die Menge aller x, für die diese Reihe konvergiert, so können wir die Potenzreihe offensichtlich als Funktion von D nach $\mathbb{K}$ auffassen.

Der Startindex einer Potenzreihe darf auch größer als 0 sein, aber nie kleiner als 0 – eine Potenzreihe in x enthält nach Definition keine negativen Potenzen von x.

(b) Die **Exponentialfunktion** ist die Potenzreihenfunktion

$$\exp\colon \mathbb{K} \to \mathbb{K} \qquad \text{mit} \qquad \exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

(die nach Beispiel 7.24 (d) für alle $x \in \mathbb{K}$ absolut konvergiert).

Aus dem Wurzelkriterium können wir sofort eine allgemeine Aussage ableiten, auf welchen Gebieten derartige Potenzreihen konvergieren: nämlich auf um 0 zentrierten Intervallen (im Fall $\mathbb{K} = \mathbb{R}$) bzw. auf Kreisen um 0 (im Fall $\mathbb{K} = \mathbb{C}$).

Satz und Definition 7.26 (Konvergenzgebiete von Potenzreihen, Formel von **Cauchy-Hadamard**)**.** *Es sei $\sum_{n=0}^{\infty} a_n x^n$ eine Potenzreihe über $\mathbb{K}$ und*

$$r = \frac{1}{\limsup\limits_{n\to\infty} \sqrt[n]{|a_n|}} \qquad \in \mathbb{R}_{\geq 0} \cup \{\infty\}$$

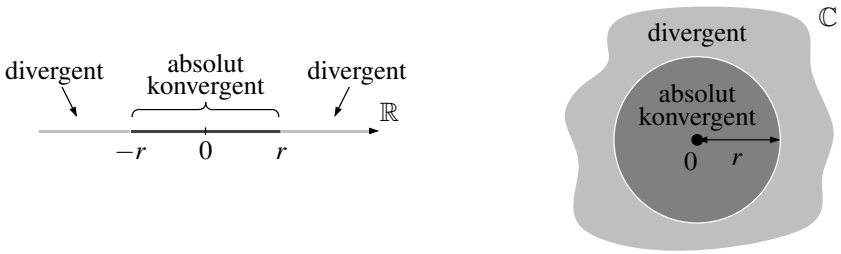
(beachte, dass dieser Wert nach Bemerkung 5.51 in jedem Fall existiert). Dann gilt:

(a) *für alle $x \in \mathbb{K}$ mit $|x| < r$ ist die Potenzreihe absolut konvergent;*

(b) *für alle $x \in \mathbb{K}$ mit $|x| > r$ ist die Potenzreihe divergent.*

Im Fall $|x| = r$ kann keine allgemeine Aussage über die Konvergenz der Reihe getroffen werden.

Die geometrische Deutung dieser Konvergenzaussagen im reellen bzw. komplexen Fall zeigt das folgende Bild. Man nennt r den ***Konvergenzradius*** *und $\{x \in \mathbb{K} : |x| < r\}$ das* ***Konvergenzgebiet*** *der Potenzreihe.*



Beweis. Wir wenden das Wurzelkriterium aus Satz 7.22 auf die Potenzreihe $\sum_{n=0}^{\infty} a_n x^n$ an: Nach Aufgabe 5.50 (b) ist für alle $x \in \mathbb{K}$

$$\limsup_{n\to\infty} \sqrt[n]{|a_n x^n|} = \limsup_{n\to\infty} \left(|x| \cdot \sqrt[n]{|a_n|}\right) = |x| \cdot \limsup_{n\to\infty} \sqrt[n]{|a_n|} = \frac{|x|}{r},$$

also konvergiert die Reihe für $|x| < r$ absolut und divergiert für $|x| > r$. □

Bemerkung 7.27. Beachte, dass die Eigenschaften (a) und (b) aus Satz 7.26 den Konvergenzradius eindeutig charakterisieren als

$$r = \sup \left\{ |x| : x \in \mathbb{K} \text{ mit } \sum_{n=0}^{\infty} a_n x^n \text{ konvergent} \right\}. \qquad (*)$$

So ist z. B. auch ohne Berechnung des Ausdrucks für r in Satz 7.26 klar, dass die Exponentialreihe aus Definition 7.25 (b) den Konvergenzradius ∞ hat, da sie ja auf ganz $\mathbb{K}$ konvergiert. In der Tat wird die Gleichung $(*)$ in der Literatur auch oft als Definition des Konvergenzradius einer Potenzreihe benutzt.

Es sollte nicht überraschen, dass man nicht nur mit dem Wurzelkriterium, sondern auch mit dem Quotientenkriterium eine Aussage über den Konvergenzradius einer Potenzreihe treffen kann. Allerdings ist diese nicht ganz so universell, da sie wie in Satz 7.21 die Existenz des Grenzwerts der Quotientenfolge der Koeffizienten voraussetzt.

Satz 7.28 (Alternative Formel für den Konvergenzradius einer Potenzreihe)**.** *Es sei $\sum_{n=0}^{\infty} a_n x^n$ eine Potenzreihe in $\mathbb{K}$ mit $a_n \neq 0$ für fast alle n. Existiert dann der Grenzwert*

$$r = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right| \quad \in \mathbb{R}_{\geq 0} \cup \{\infty\},$$

so ist dies der Konvergenzradius der Potenzreihe.

Beweis. Wegen

$$\lim_{n\to\infty} \left| \frac{a_{n+1} x^{n+1}}{a_n x^n} \right| = \lim_{n\to\infty} \left(|x| \cdot \left| \frac{a_{n+1}}{a_n} \right| \right) = |x| \cdot \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| = \frac{|x|}{r}$$

konvergiert die Potenzreihe nach Satz 7.21 absolut für $|x| < r$ und divergiert für $|x| > r$. Nach Bemerkung 7.27 ist r damit der Konvergenzradius der Reihe. □

Beispiel 7.29.

(a) Die Potenzreihe $\sum_{n=1}^{\infty} \frac{x^n}{n}$ hat nach Satz 7.28 den Konvergenzradius

$$r = \lim_{n\to\infty} \left| \frac{1/n}{1/(n+1)} \right| = \lim_{n\to\infty} \frac{n+1}{n} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right) = 1,$$

konvergiert also für alle $x \in \mathbb{K}$ mit $|x| < 1$ absolut und divergiert für alle x mit $|x| > 1$. Für $|x| = 1$ treten in der Tat verschiedene Fälle auf: Im Fall $x = 1$ erhalten wir die harmonische Reihe, die nach Beispiel 7.3 (c) divergiert, während wir für $x = -1$ die alternierende harmonische Reihe haben, die nach Beispiel 7.9 konvergiert. Dies zeigt noch einmal, dass unsere obigen nur von $|x|$ abhängigen Kriterien auf dem Rand des Konvergenzgebiets wirklich keine allgemeine Aussage machen können.

(b) Für den Konvergenzradius der Potenzreihe $\sum_{n=0}^{\infty} n x^n$ gilt nach Satz 7.26 und 7.28

$$r = \frac{1}{\limsup\limits_{n\to\infty} \sqrt[n]{n}} \qquad \text{sowie} \qquad r = \lim_{n\to\infty} \left| \frac{n}{n+1} \right| = 1.$$

Vergleich dieser beiden Ergebnisse liefert also $\limsup\limits_{n\to\infty} \sqrt[n]{n} = 1$. Für alle $\varepsilon > 0$ gilt damit nach Lemma 5.46 (a), dass $\sqrt[n]{n} < 1 + \varepsilon$ für fast alle n. Natürlich ist aber auch $\sqrt[n]{n} \geq 1$ für alle n, und damit ergibt sich zusammen

$$\lim_{n\to\infty} \sqrt[n]{n} = 1$$

(was gar nicht so offensichtlich ist, da der Ausdruck $\sqrt[n]{n}$ für wachsendes n ja durch das n unter der Wurzel größer, durch das Ziehen der n-ten Wurzel aber kleiner wird).

(c) Da die Exponentialreihe $\exp(x)$ für alle $x \in \mathbb{K}$ konvergiert, ist ihr Konvergenzradius $r = \infty$.

Aufgabe 7.30. Untersuche die folgenden Reihen auf Konvergenz (im Fall (c) in Abhängigkeit von $x \in \mathbb{R}$):

$$\text{(a)} \sum_{n=0}^{\infty} \frac{n}{(3+(-1)^n)^n}, \qquad \text{(b)} \sum_{n=0}^{\infty} \frac{n!}{n^n}, \qquad \text{(c)} \sum_{n=0}^{\infty} \frac{n}{n^2+1} x^n.$$

Eine Berechnung des Grenzwerts im Fall der Konvergenz ist nicht erforderlich.

Aufgabe 7.31. Zeige, dass jede Potenzreihe $\sum_{n=0}^{\infty} a_n x^n$ in $\mathbb{K}$ denselben Konvergenzradius wie ihre „formale Ableitung“ $\sum_{n=1}^{\infty} n a_n x^{n-1}$ hat, aber nicht notwendig für die gleichen x konvergiert.

(Wir werden in Folgerung 10.28 sehen, dass dies dann auch genau die „gewöhnliche“ Ableitung der ursprünglichen Reihe ist.)

Aufgabe 7.32. Es sei $(a_n)_n$ eine Folge in $\mathbb{K}$. Man zeige:

(a) Ist a_n der Quotient von zwei Polynomen in n, so machen weder das Quotienten- noch das Wurzelkriterium eine Aussage über die Konvergenz der Reihe $\sum_{n=0}^{\infty} a_n$.

(b) Macht das Wurzelkriterium keine Aussage über die Konvergenz der Reihe $\sum_{n=0}^{\infty} a_n$, so macht auch das Quotientenkriterium keine Aussage darüber.

Aufgabe 7.33 (Alternative Darstellung der Exponentialfunktion)**.** Zeige, dass für alle $x \in \mathbb{K}$

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

gilt. (Hinweis: Eine Möglichkeit besteht darin, durch eine geeignete Abschätzung zu zeigen, dass

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

mit $n \to \infty$ gegen 0 konvergiert.)

Der Ausdruck $\lim_{n\to\infty}(1+\frac{x}{n})^n$ in dieser Darstellung der Exponentialfunktion hat übrigens eine sehr anschauliche Interpretation: Wenn ihr 1 Euro zu einem Zinssatz x ein Jahr lang anlegt und sich die Bank bereit erklärt, nicht einmal am Ende des Jahres den Betrag x an Zinsen auszuzahlen, sondern stattdessen n-mal einen Zinssatz von $\frac{x}{n}$ bezahlt, so habt ihr dadurch am Ende des Jahres aufgrund des Zinseszinses natürlich mehr Geld als bei einer einmaligen Zinszahlung, nämlich genau $(1+\frac{x}{n})^n$. Die gerade gezeigte Formel besagt, dass ihr selbst für $n \to \infty$, also wenn ihr die Bank zu einer unendlichen Aufteilung der Zinsen auf diese Art überreden könntet, dadurch kein unendliches Vermögen aufbauen könntet, sondern am Ende des Jahres lediglich den Betrag von $\exp(x)$ hättet.

Zum Schluss dieses Kapitels wollen wir nun noch Produkte von Reihen untersuchen. Haben wir z. B. zwei Potenzreihen $\sum_{k=0}^{\infty} a_k x^k$ und $\sum_{l=0}^{\infty} b_l x^l$, so würden wir erwarten, dass wir ihr Produkt für alle x im Durchschnitt der Konvergenzgebiete der beiden Reihen wie folgt durch „unendliches Ausmultiplizieren“ berechnen und wieder zu einer neuen Potenzreihe zusammenfassen können:

$$\begin{aligned}\left(\sum_{k=0}^{\infty} a_k x^k\right)\cdot\left(\sum_{l=0}^{\infty} b_l x^l\right) &= (a_0 + a_1 x + a_2 x^2 + \cdots)\cdot(b_0 + b_1 x + b_2 x^2 + \cdots)\\ &\overset{?}{=} a_0 b_0 + (a_0 b_1 + a_1 b_0)x + (a_0 b_2 + a_1 b_1 + a_2 b_0)x^2 + \cdots\\ &= \sum_{n=0}^{\infty} c_n x^n\end{aligned}$$

mit $c_n = \sum_{k=0}^{n} a_k b_{n-k}$. Wir wollen nun zeigen, dass dies in der Tat erlaubt ist – und zwar auch für allgemeine Reihen, nicht nur für Potenzreihen (also wenn wir uns das x in der obigen Rechnung wegdenken). Da die einzelnen ausmultiplizierten Summanden in der entstehenden Reihe dabei aber irgendwie sortiert werden müssen, sollte es in Anbetracht des Umordnungssatzes 7.15 nicht überraschen, dass wir für die Gültigkeit dieser Rechnung die *absolute* Konvergenz der Reihen benötigen.

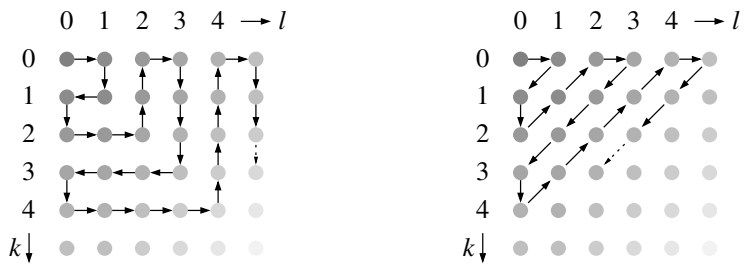
Satz 7.34 (**Cauchy-Produkt** von Reihen)**.** *Es seien $\sum_{k=0}^{\infty} a_k$ und $\sum_{l=0}^{\infty} b_l$ zwei absolut konvergente Reihen in $\mathbb{K}$. Setzen wir dann*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

für alle $n \in \mathbb{N}$, so ist auch die Reihe $\sum_{n=0}^{\infty} c_n$ absolut konvergent, und es gilt

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

Beweis. Nach Voraussetzung existieren die Grenzwerte $A := \sum_{k=0}^{\infty} |a_k|$ und $B := \sum_{l=0}^{\infty} |b_l|$ in $\mathbb{R}$. Wir zeigen, dass die Summe $\sum_{(k,l)\in\mathbb{N}^2} a_k b_l$ über die nach Beispiel 5.58 (a) abzählbare Indexmenge $\mathbb{N}\times\mathbb{N}$ im Sinne von Bemerkung 7.16 existiert und mit dem Wert beider Seiten der Gleichung $(*)$ übereinstimmt. Dazu betrachten wir die beiden im folgenden Bild dargestellten Aufzählungen von $\mathbb{N}^2$.



(a) Summieren wir zunächst die Beträge $|a_k b_l|$ in der Reihenfolge dieser „quadratischen“ Aufzählung, so erhalten wir nach der Summation von höchstens n^2 Termen, also denen im Quadrat links oben mit n Zeilen und Spalten, maximal den Wert

$$\sum_{k=0}^{n-1}\sum_{l=0}^{n-1} |a_k b_l| = \left(\sum_{k=0}^{n-1} |a_k|\right)\cdot\left(\sum_{l=0}^{n-1} |b_l|\right) \le A\cdot B.$$

Die Reihe über alle $|a_k b_l|$ ist (in dieser Aufzählungsreihenfolge) also beschränkt. Damit ist die Reihe über $a_k b_l$ nach Definition 7.10 absolut konvergent, und nach Bemerkung 7.16 existiert somit die Reihe $\sum_{(k,l)\in\mathbb{N}^2} a_k b_l$.

Um den Wert dieser Reihe zu bestimmen, können wir nach Lemma 5.18 auch nur eine Teilfolge der Partialsummenfolge betrachten. Nehmen wir hierzu die Partialsummen, bei denen

wir die ersten n^2 Terme aufaddieren, und lassen dort n gegen ∞ gehen, so erhalten wir

$$\sum_{(k,l)\in\mathbb{N}^2} a_k b_l = \lim_{n\to\infty} \sum_{k=0}^{n-1}\sum_{l=0}^{n-1} a_k b_l = \lim_{n\to\infty} \left(\sum_{k=0}^{n-1} a_k\right)\cdot\left(\sum_{l=0}^{n-1} b_l\right) \overset{5.13}{=} \left(\sum_{k=0}^{\infty} a_k\right)\cdot\left(\sum_{l=0}^{\infty} b_l\right),$$

also die linke Seite von $(*)$.

(b) In dieser „schrägen" Aufzählungsreihenfolge können wir den Wert von $\sum_{(k,l)\in\mathbb{N}^2} a_k b_l$ analog als Grenzwert für $N\to\infty$ der Partialsummen bestimmen, bei denen wir die ersten N Diagonalen aufsummieren. Da die Summe der $a_k b_l$ entlang der Diagonalen mit $k+l=n$ gerade c_n ist, erhalten wir so

$$\sum_{(k,l)\in\mathbb{N}^2} a_k b_l = \lim_{N\to\infty} \sum_{n=0}^{N-1} c_n = \sum_{n=0}^{\infty} c_n$$

und damit durch Vergleich mit (a) die behauptete Gleichheit $(*)$. Weil die ersten N Diagonalen außerdem im Quadrat links oben mit N Zeilen und Spalten enthalten sind, haben wir weiterhin mit der Dreiecksungleichung

$$\sum_{n=0}^{N-1} |c_n| = \sum_{n=0}^{N-1}\left|\sum_{k=0}^{n} a_k b_{n-k}\right| \le \sum_{n=0}^{N-1}\sum_{k=0}^{n} |a_k|\cdot|b_{n-k}| \le \left(\sum_{k=0}^{N-1}|a_k|\right)\cdot\left(\sum_{l=0}^{N-1}|b_l|\right) \le A\cdot B$$

16 und damit auch die absolute Konvergenz von $\sum_{n=0}^{\infty} c_n$ gezeigt. □

Die wohl wichtigste Anwendung des Cauchy-Produkts erhalten wir im Fall der Exponentialfunktion.

Folgerung 7.35 (Funktionalgleichung der Exponentialfunktion)**.** *Für alle $x,y\in\mathbb{K}$ gilt*

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

Beweis. Nach Beispiel 7.24 (d) und Definition 7.25 (b) konvergiert die Exponentialreihe auf ganz $\mathbb{K}$ absolut. Also gilt für alle $x,y\in\mathbb{K}$

$$\begin{aligned}
\exp(x)\cdot\exp(y) &= \left(\sum_{k=0}^{\infty}\frac{x^k}{k!}\right)\cdot\left(\sum_{l=0}^{\infty}\frac{y^l}{l!}\right)\\
&= \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n}\frac{x^k}{k!}\cdot\frac{y^{n-k}}{(n-k)!}\right) && \text{(nach Satz 7.34)}\\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\cdot\left(\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}\right)\\
&= \sum_{n=0}^{\infty}\frac{(x+y)^n}{n!} && \text{(nach Satz 4.7)}\\
&= \exp(x+y),
\end{aligned}$$

was zu zeigen war. □

Bemerkung 7.36 (Produkte von Potenzreihen)**.** Sind $\sum_{k=0}^{\infty} a_k x^k$ und $\sum_{l=0}^{\infty} b_l x^l$ zwei Potenzreihen in x mit Konvergenzradien r_1 bzw. r_2, so gilt nach Satz 7.34 für alle $x\in\mathbb{K}$ mit $|x|<\min(r_1,r_2)$ (wo also nach Satz 7.26 beide Potenzreihen absolut konvergieren)

$$\left(\sum_{k=0}^{\infty} a_k x^k\right)\cdot\left(\sum_{l=0}^{\infty} b_l x^l\right) = \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right)x^n,$$

d. h. das Produkt zweier Potenzreihen mit Konvergenzradien r_1 und r_2 ist wieder eine Potenzreihe, deren Konvergenzradius mindestens $\min(r_1,r_2)$ beträgt und die durch „unendliches Ausmultiplizieren" berechnet werden kann.

Aufgabe 7.37. Es sei $q\in\mathbb{C}$ mit $|q|<1$. Berechne das Cauchy-Produkt $\left(\sum_{n=0}^{\infty} q^n\right)^2$ und damit den Wert der Reihe $\sum_{n=0}^{\infty} n q^n$.

Aufgabe 7.38. Zeige, dass die Reihe $\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}$ konvergiert, aber dass die Formel

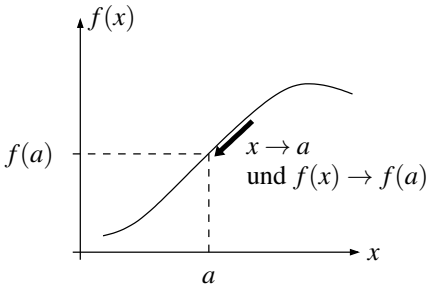
$$\left(\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}\right) \cdot \left(\sum_{n=0}^{\infty} \frac{(-1)^n}{\sqrt{n+1}}\right) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{(-1)^k}{\sqrt{k+1}} \frac{(-1)^{n-k}}{\sqrt{n-k+1}}$$

des Cauchy-Produkts aus Satz 7.34 nicht gilt.

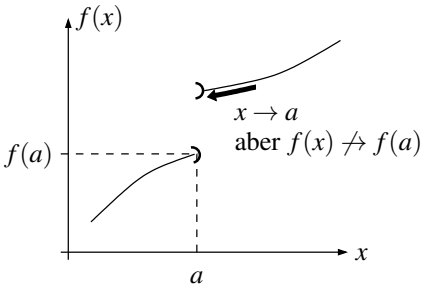
8. Stetigkeit

Nachdem wir uns gerade ausführlich mit Grenzwerten von Folgen und Reihen befasst haben, wollen wir den Grenzwertbegriff nun auf Funktionen einer reellen (oder evtl. komplexen) Variablen ausdehnen, also auf Funktionen $f\colon D \to \mathbb{K}$ mit einer Definitionsmenge $D \subset \mathbb{K}$. Dies führt zum zentralen Begriff der Stetigkeit solcher Funktionen.

Anschaulich ist die Frage dabei: Wenn eine Stelle $a \in D$ mit Funktionswert $f(a)$ gegeben ist, und wir nun andere Punkte $x \in D$ in der Nähe von a betrachten, liegt dann auch $f(x)$ in der Nähe von $f(a)$? Im Bild unten links ist dies der Fall: Laufen wir entlang des dick eingezeichneten Pfeils mit x auf den Punkt a zu, so nähern sich auch die Funktionswerte $f(x)$ dem Punkt $f(a)$. In diesem Fall werden wir sagen, dass f im Punkt a stetig ist.



f ist stetig in a



f ist nicht stetig in a

Im Bild oben rechts dagegen führt der Sprung im Funktionsgraphen zu einem anderen Verhalten: Nähern wir uns hier entlang des dick eingezeichneten Pfeils mit x dem Punkt a, so nähert sich $f(x)$ nicht dem Wert $f(a)$, sondern dem oberen Punkt der Sprungstelle. In diesem Fall ist f im Punkt a unstetig.

Um dies mathematisch exakt zu formulieren, wollen wir jetzt den Begriff von Funktionsgrenzwerten einführen. Im linken Fall können wir dann sagen, dass $f(x)$ mit $x \to a$ gegen $f(a)$ konvergiert, während dies im Fall rechts nicht so ist.

8.A Grenzwerte von Funktionen

Wie eben erläutert wollen wir das Verhalten von Funktionen $f\colon D \to \mathbb{K}$ mit $D \subset \mathbb{K}$ untersuchen, wenn wir uns einem vorgegebenen Punkt a nähern. Dieser Wert a kann dabei, muss aber nicht unbedingt selbst Element von D sein. Wir sollten aber natürlich sicherstellen, dass wir uns zumindest innerhalb von D dem Punkt a beliebig annähern können – also anschaulich gesprochen, dass a entweder in D oder am Rand von D liegt. Formal bedeutet dies, dass a im Sinne der folgenden Definition ein Berührpunkt von D sein muss.

Definition 8.1 (Berührpunkte)**.** Es sei $D \subset \mathbb{K}$ eine Menge. Eine Zahl $a \in \mathbb{K}$ heißt **Berührpunkt** von D, wenn jede ε-Umgebung von a (siehe Bemerkung 5.2) mindestens einen Punkt aus D enthält, also wenn es zu jedem $\varepsilon > 0$ ein $x \in D$ gibt mit $|x-a| < \varepsilon$. Die Menge aller Berührpunkte von D wird mit $\overline{D}$ bezeichnet und heißt der **Abschluss** von D.

Beispiel 8.2.

(a) Für ein $D \subset \mathbb{K}$ ist jedes $a \in D$ Berührpunkt von D: Wir können in diesem Fall in der Definition 8.1 einfach $x = a$ für jedes ε wählen. Es gilt also stets $D \subset \overline{D}$.

(b) Für ein offenes reelles Intervall $D = (a,b) \subset \mathbb{R}$ ist $\overline{D} = [a,b]$ das zugehörige abgeschlossene Intervall. Die Randpunkte a und b sind also Berührpunkte von D, die nicht selbst zu D gehören.

Für derartige Berührpunkte können wir nun Grenzwerte von Funktionen definieren. Die Konstruktion ist völlig analog zur Definition 5.1 (b) des Grenzwerts von Folgen:

Definition 8.3 (Grenzwerte von Funktionen)**.** Es seien $D \subset \mathbb{K}$ eine Menge und $f\colon D \to \mathbb{K}$ eine Funktion. Ferner sei $a \in \overline{D}$ ein Berührpunkt von D.

Dann heißt eine Zahl $c \in \mathbb{K}$ **Grenzwert** von f in a, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0}\ \forall x \in D : |x-a| < \delta \Rightarrow |f(x)-c| < \varepsilon.$$

Wie schon im Fall von Folgen werden wir sehen (siehe Bemerkung 8.12), dass ein solcher Grenzwert im Fall der Existenz eindeutig ist, so dass wir also von *dem* Grenzwert von f in a sprechen können. Wir schreiben dies dann als

$$\lim_{x\to a} f(x) = c \qquad \text{bzw.} \qquad \lim_{\substack{x\to a\\ x\in D}} f(x) = c$$

oder auch als „$f(x) \to c$ für $x \to a$“, und sagen, dass f in a gegen c **konvergiert**. Existiert ein solcher Grenzwert nicht, so heißt f **divergent** in a.

Bemerkung 8.4. Gilt in Definition 8.3 sogar $a \in D$, so kommt als Grenzwert c nur $f(a)$ in Frage: Wir können dann nämlich $x = a$ in der Grenzwertbedingung von Definition 8.3 setzen (so dass $|x-a| < \delta$ in jedem Fall erfüllt ist) und erhalten damit $|f(a)-c| < \varepsilon$ für alle ε – was nur möglich ist, wenn $c = f(a)$ ist.

Definition 8.5 (Stetigkeit)**.** Es seien $D \subset \mathbb{K}$ eine Menge und $f\colon D \to \mathbb{K}$ eine Funktion.

(a) Ist $a \in D$, so heißt f **stetig** in a, wenn der Grenzwert $\lim_{x\to a} f(x)$ existiert (und nach Bemerkung 8.4 damit zwangsläufig gleich $f(a)$ ist), d. h. wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0}\ \forall x \in D : |x-a| < \delta \Rightarrow |f(x)-f(a)| < \varepsilon.$$

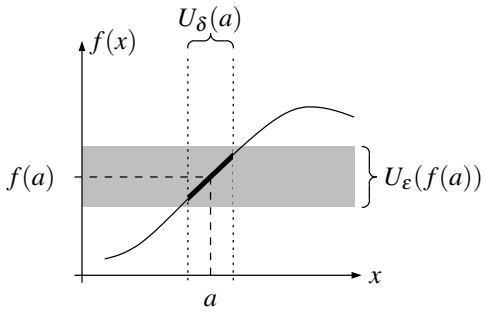
Die Funktion f heißt **stetig** (auf D), wenn sie in jedem Punkt $a \in D$ stetig ist.

(b) Ist $a \in \overline{D}\backslash D$, so heißt f **stetig fortsetzbar** nach a, wenn der Grenzwert $c = \lim_{x\to a} f(x)$ existiert. (In diesem Fall erhält man nämlich eine in a stetige Funktion

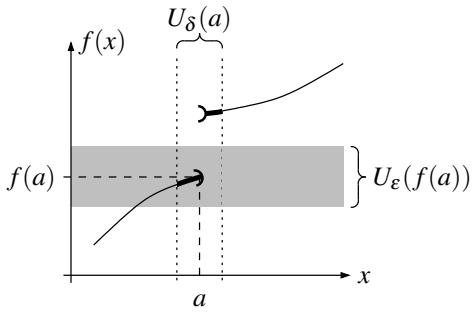
$$\tilde{f}\colon D \cup \{a\} \to \mathbb{K},\ x \mapsto \begin{cases} f(x) & \text{für } x \in D, \\ c & \text{für } x = a, \end{cases}$$

die man als stetige Fortsetzung von f nach a bezeichnet.)

Bemerkung 8.6. Das Bild unten zeigt noch einmal das Beispiel vom Anfang dieses Kapitels mit den eben eingeführten Notationen: Nach Definition ist eine Funktion $f\colon D \to \mathbb{K}$ genau dann stetig in einem Punkt $a \in D$, wenn es zu jeder (beliebig kleinen) ε-Umgebung $U_\varepsilon(f(a))$ von $f(a)$ eine δ-Umgebung $U_\delta(a)$ von a gibt, in der alle Punkte von D nach $U_\varepsilon(f(a))$ abgebildet werden. Im Bild unten bedeutet dies, dass zu jedem auch noch so schmal gewählten grauen horizontalen Streifen um $f(a)$ eine Einschränkung von f auf eine hinreichend kleine Umgebung von a dazu führt, dass alle Funktionswerte dort (im Bild unten dick eingezeichnet) in dem gewählten Streifen liegen. Dies entspricht genau der ursprünglichen Motivation, dass eine kleine Änderung von x um a herum auch nur zu einer kleinen Änderung der Funktionswerte $f(x)$ um $f(a)$ führen darf. Bei der linken Funktion ist dies also der Fall, bei der rechten aufgrund der Sprungstelle jedoch nicht.



f ist stetig in a



f ist nicht stetig in a

Die ebenfalls oft gehörte geometrische Interpretation, dass eine reelle Funktion stetig ist, wenn man „ihren Graphen zeichnen kann, ohne den Stift abzusetzen", ist übrigens etwas mit Vorsicht zu genießen, wie die Beispiele 8.7 (e) und (f) unten zeigen.

Beispiel 8.7.

(a) Die Identität $f\colon \mathbb{K} \to \mathbb{K}$, $x \mapsto x$ ist stetig: Sind $a \in \mathbb{K}$ und $\varepsilon > 0$ gegeben, so setze man $\delta := \varepsilon$. Dann gilt natürlich für alle $x \in \mathbb{K}$ mit $|x-a| < \delta$, dass $|f(x) - f(a)| = |x-a| < \delta = \varepsilon$. Genauso zeigt man, dass konstante Funktionen stetig sind.

(b) Wir zeigen, dass die Betragsfunktion $f\colon \mathbb{K} \to \mathbb{K}$, $x \mapsto |x|$ stetig ist. Es seien dazu $a \in \mathbb{K}$ und $\varepsilon > 0$ gegeben. Wir setzen wieder $\delta := \varepsilon$. Dann folgt für alle $x \in \mathbb{K}$ mit $|x-a| < \delta$ mit Hilfe der Dreiecksungleichung nach unten

$$|x-a| \geq |x| - |a| \quad \text{und} \quad |x-a| \geq |a| - |x|.$$

Da $|f(x) - f(a)| = \big||x| - |a|\big|$ aber eine der beiden Zahlen $|x| - |a|$ und $|a| - |x|$ sein muss, ergibt sich in jedem Fall

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

Damit ist f stetig.

(c) Analog ist die komplexe Konjugation $f\colon \mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$ (siehe Notation 6.2) stetig: Sind $a \in \mathbb{C}$ und $\varepsilon > 0$ gegeben, so setzen wir $\delta := \varepsilon$ und erhalten für alle $z \in \mathbb{C}$ mit $|z-a| < \delta$

$$|f(z) - f(a)| = |\overline{z} - \overline{a}| = |\overline{z-a}| = |z-a| < \delta = \varepsilon.$$

(d) Die Funktion

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} 0 & \text{für } x \neq 0, \\ 1 & \text{für } x = 0 \end{cases}$$

(siehe Bild unten) ist in $a = 0$ nicht stetig. Wollen wir dies formal zeigen, müssen wir die Negation der Bedingung aus Definition 8.5 (a) beweisen, d. h.

$$\exists \varepsilon > 0\ \forall \delta > 0\ \exists x \in \mathbb{R} : |x-a| < \delta \text{ und } |f(x) - f(a)| \geq \varepsilon.$$

(Beachte dabei, dass die Negation der Aussage „$|x-a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon$" nach Beispiel 1.9 die angegebene Bedingung „$|x-a| < \delta$ *und* $|f(x) - f(a)| \geq \varepsilon$" ist, und nicht etwa eine Folgerung „$|x-a| < \delta \Rightarrow |f(x) - f(a)| \geq \varepsilon$"!)

Dies zu zeigen ist hier aber sehr einfach: Setzen wir $\varepsilon = \frac{1}{2}$ und ist $\delta > 0$ beliebig, so können wir $x = \frac{\delta}{2}$ setzen und erhalten $|x-a| = \frac{\delta}{2} < \delta$ und $|f(x) - f(a)| = |0-1| = 1 \geq \varepsilon$.

Anders ausgedrückt existiert in diesem Fall der Grenzwert $\lim_{x\to 0} f(x)$ nicht. Falls ihr jetzt gedacht hättet, dass dieser Grenzwert doch existiert und gleich 0 ist, so habt ihr damit sicher gemeint, dass sich $f(x)$ dem Wert 0 nähert, wenn x in der Nähe von 0, *aber nicht gleich* 0 ist. Der Fall $x = 0$ (bzw. $x = a$) ist in Definition 8.3 aber nicht ausgeschlossen! Wenn wir dies

ausschließen wollten, so müssten wir die Definitionsmenge von f auf $\mathbb{R}\backslash\{0\}$ einschränken und würden dann in der Tat

$$\lim_{\substack{x\to 0\\ x\neq 0}} f(x) = 0$$

erhalten.

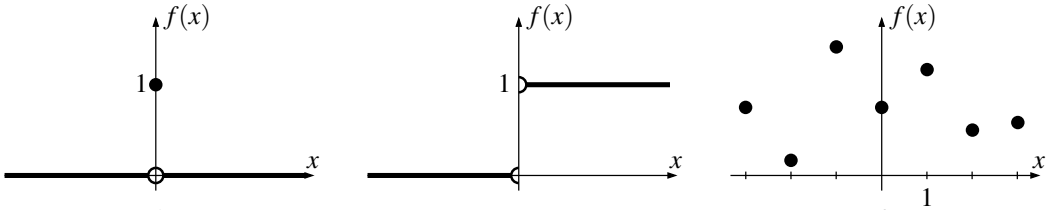
Beachte jedoch, dass es hier in der Literatur zwei verschiedene Konventionen gibt: In manchen Büchern werden Funktionsgrenzwerte so definiert, dass $\lim_{x\to a}$ immer für $\lim_{x\to a, x\neq a}$ steht.

(e) Die unten im Bild dargestellte Funktion

$$f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R},\ x \mapsto \begin{cases} 0 & \text{für } x < 0, \\ 1 & \text{für } x > 0 \end{cases}$$

ist stetig – ja, wirklich! Sie ist nämlich an jedem Punkt *der Definitionsmenge*, also an jedem $a \neq 0$ stetig, weil sie in der Nähe eines jeden solchen Punktes (genauer: in der $|a|$-Umgebung von a) konstant ist. Die Funktion f ist aber natürlich *nicht stetig fortsetzbar* nach 0: Der Grenzwert $\lim_{x\to 0} f(x)$ existiert nicht.

(f) *Jede* Funktion $f\colon \mathbb{Z} \to \mathbb{R}$ ist stetig (siehe unten). Das liegt anschaulich einfach daran, dass wir hier gar keine Möglichkeit haben, ein gegebenes $a \in \mathbb{Z}$ ein wenig so zu verändern, dass es immer noch in der Definitionsmenge liegt. Formal können wir in der Bedingung aus Definition 8.5 (a) für jedes gegebene $\varepsilon > 0$ immer $\delta = \frac{1}{2}$ setzen und haben damit sicher gestellt, dass $|x-a| < \delta$ mit $x \in \mathbb{Z}$ nur für $x = a$ erfüllt ist, womit dann natürlich auch $|f(x) - f(a)| = 0 < \varepsilon$ ist.



(d) (e) (f)

Bemerkung 8.8 (Funktionen mit Grenzwert ungleich 0)**.** Es seien $D \subset \mathbb{K}$, $f\colon D \to \mathbb{K}$ eine Funktion und $a \in \overline{D}$ mit $c := \lim_{x\to a} f(x) \neq 0$. Aus Definition 8.3 für $\varepsilon = \frac{|c|}{2}$ erhalten wir dann wie in Bemerkung 5.12 ein $\delta > 0$, so dass $|f(x) - c| < \varepsilon$ und damit

$$|f(x)| = |f(x) - c + c| \geq |c| - |f(x) - c| > |c| - \varepsilon = \frac{|c|}{2} > 0$$

für alle $x \in D$ mit $|x - a| < \delta$ gilt.

Insbesondere ergibt sich im Fall $a \in D$ also, dass eine in a stetige Funktion f mit $f(a) \neq 0$ auch in einer δ-Umgebung von a ungleich 0 ist. Beispiel 8.7 (d) zeigt (bei $a = 0$), dass dies für unstetige Funktionen im Allgemeinen natürlich falsch ist.

Eine analoge Aussage gilt im Fall $\mathbb{K} = \mathbb{R}$ auch ohne Beträge: Eine in a stetige reelle Funktion f mit $f(a) > 0$ ist auch in einer δ-Umgebung von a positiv.

Aufgabe 8.9. Es seien $m, n \in \mathbb{N}_{>0}$. Berechne den Grenzwert $\lim_{x\to 1} \frac{x^m - 1}{x^n - 1}$.

Aufgabe 8.10. Zeige durch Rückgang auf die ε-δ-Definition der Stetigkeit, dass die Funktion $f\colon [-1,1] \to \mathbb{R}$, $x \mapsto \sqrt{1-x^2}$ stetig ist.

Wie im Fall von Folgengrenzwerten wollen wir nun natürlich auch für Grenzwerte von Funktionen ein paar einfache Rechenregeln zeigen, z. B. dass solche Grenzwerte wie in Satz 5.13 mit Summen und Produkten vertauschen. Glücklicherweise lassen sich Grenzwerte von Funktionen mit Hilfe des folgenden Satzes immer auf Grenzwerte von Folgen zurückführen, so dass wir viele unserer Ergebnisse dann sofort von Folgen auf Funktionen übertragen können:

Satz 8.11 (**Folgenkriterium**). *Es seien $D \subset \mathbb{K}$ und $f\colon D \to \mathbb{K}$ eine Funktion.*

(a) ***(Folgenkriterium für Funktionsgrenzwerte)*** *Für $a \in \overline{D}$ und $c \in \mathbb{K}$ gilt*

$$\lim_{x\to a} f(x) = c \quad \Leftrightarrow \quad \text{Für jede Folge } (x_n)_n \text{ in } D \text{ mit } x_n \to a \text{ gilt } f(x_n) \to c.$$

(b) ***(Folgenkriterium für Stetigkeit)*** *Für $a \in D$ gilt*

$$f \text{ ist stetig in } a \quad \Leftrightarrow \quad \text{Für jede Folge } (x_n)_n \text{ in } D \text{ mit } x_n \to a \text{ gilt } f(x_n) \to f(a).$$

Es ist dann also

$$\lim_{n\to\infty} f(x_n) = f\big(\lim_{n\to\infty} x_n\big),$$

d. h. „eine stetige Funktion f vertauscht mit der Grenzwertbildung von Folgen“.

Beweis. Wir beweisen zunächst Teil (a).

„$\Rightarrow$“ Es seien $\lim_{x\to a} f(x) = c$ und $(x_n)_n$ eine Folge in D mit $x_n \to a$; wir müssen $f(x_n) \to c$ zeigen. Dazu sei $\varepsilon > 0$ beliebig. Wegen $\lim_{x\to a} f(x) = c$ gibt es ein $\delta > 0$, so dass $|f(x) - c| < \varepsilon$ für alle $x \in D$ mit $|x - a| < \delta$ gilt. Wegen $x_n \to a$ ist aber $|x_n - a| < \delta$ für fast alle n, und damit dann auch $|f(x_n) - c| < \varepsilon$ für diese n. Damit gilt $f(x_n) \to c$.

„$\Leftarrow$“ Wir zeigen diese Richtung durch einen Widerspruchsbeweis und nehmen also an, dass c kein Grenzwert von $f(x)$ in a ist, d. h. (durch Negation der Definition 8.3)

$$\exists\, \varepsilon > 0\ \forall\, \delta > 0\ \exists\, x \in D : |x - a| < \delta \text{ und } |f(x) - c| \geq \varepsilon.$$

Wir wählen nun ein solches ε. Indem wir $\delta = \frac{1}{n}$ für $n \in \mathbb{N}_{>0}$ setzen, erhalten wir für alle n ein $x_n \in D$ mit $|x_n - a| < \frac{1}{n}$ und $|f(x_n) - c| \geq \varepsilon$. Für diese Folge gilt dann aber $x_n \to a$ und $f(x_n) \not\to c$ im Widerspruch zur Annahme.

17 Teil (b) folgt nun mit Definition 8.5 (a) sofort aus (a). □

Bemerkung 8.12. Mit Hilfe des Folgenkriteriums können wir nun sehr schnell viele Resultate über Grenzwerte von Folgen auf Funktionen übertragen. So folgt z. B. sofort, dass Grenzwerte von Funktionen immer eindeutig sind, sofern sie existieren: Sind $D \subset \mathbb{K}$, $f\colon D \to \mathbb{K}$, $a \in \overline{D}$ und $f(x) \to c$ für $x \to a$, so können wir wegen $a \in \overline{D}$ eine Folge $(x_n)_n$ in D mit $x_n \to a$ wählen, und erhalten mit Satz 8.11 (a) dann auch $f(x_n) \to c$. Da Folgengrenzwerte nach Lemma 5.5 aber eindeutig sind, ist dies für höchstens ein c möglich.

Die folgenden Rechenregeln ergeben sich ebenfalls sofort aus dem Folgenkriterium und sorgen auch dafür, dass wir für sehr viele Funktionen ohne weitere Rechnung direkt die Stetigkeit nachweisen können:

Satz 8.13 (**Grenzwertsätze** für Funktionen). *Es seien $D \subset \mathbb{K}$ und $f, g\colon D \to \mathbb{K}$ zwei Funktionen. Weiterhin sei $a \in \overline{D}$, so dass beide Grenzwerte $\lim_{x\to a} f(x)$ und $\lim_{x\to a} g(x)$ existieren. Dann existiert auch der Grenzwert $\lim_{x\to a}(f(x) + g(x))$, und es gilt*

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

Eine analoge Aussage gilt auch für $f(x) - g(x)$, $f(x) \cdot g(x)$ und $\frac{f(x)}{g(x)}$ (letzteres natürlich nur, falls $\lim_{x\to a} g(x) \neq 0$).

Insbesondere sind für $a \in D$ also mit f und g auch $f + g$, $f - g$, $f \cdot g$ und $\frac{f}{g}$ in a stetig (letzteres wiederum nur, falls $g(a) \neq 0$).

Beweis. Beachte im Fall $\frac{f}{g}$ zunächst, dass die Definitionsmenge dieses Quotienten nicht ganz D, sondern die evtl. kleinere Menge $D' = \{x \in D : g(x) \neq 0\}$ ist. Um überhaupt über den Grenzwert von $\frac{f(x)}{g(x)}$ für $x \to a$ sprechen zu können, müssen wir also zuerst überprüfen, dass a ein Berührpunkt von D' ist. Dies folgt aber aus Bemerkung 8.8, die besagt, dass g wegen $\lim_{x\to a} g(x) \neq 0$ in einer ε-Umgebung von a nirgends 0 wird, so dass D und D' dort also übereinstimmen.

Die eigentliche Behauptung des Lemmas ist nun eine direkte Übertragung der Grenzwertsätze für Folgen aus Satz 5.13. Wir betrachten hier nur den Fall der Addition, da die anderen drei Fälle wörtlich genauso bewiesen werden. Dazu berechnen wir den Grenzwert von $f(x)+g(x)$ mit dem Folgenkriterium aus Satz 8.11 (a): Es sei $(x_n)_n$ eine beliebige Folge in D mit $x_n \to a$. Dann gilt nach dem Folgenkriterium für f und g

$$\lim_{n\to\infty} f(x_n) = \lim_{x\to a} f(x) \quad \text{und} \quad \lim_{n\to\infty} g(x_n) = \lim_{x\to a} g(x)$$

und damit nach Satz 5.13 (a)

$$\lim_{n\to\infty}(f(x_n)+g(x_n)) = \lim_{n\to\infty} f(x_n) + \lim_{n\to\infty} g(x_n) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x),$$

d. h. $f(x_n)+g(x_n)$ konvergiert für jede solche Folge gegen $\lim_{x\to a} f(x) + \lim_{x\to a} g(x)$, also immer gegen dieselbe Zahl. Wiederum nach dem Folgenkriterium – diesmal für $f+g$ – ergibt sich damit also wie gewünscht auch

$$\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x). \qquad \square$$

Beispiel 8.14. Jede *rationale Funktion*, also jede Funktion der Form $f(x) = \frac{p(x)}{q(x)}$ mit Polynomfunktionen $p(x)$ und $q(x)$, lässt sich natürlich mit den vier Grundrechenarten aus der Identität und den konstanten Funktionen zusammensetzen. Damit folgt aus Satz 8.13, dass jede solche Funktion auf jeder Definitionsmenge $D \subset \{x \in \mathbb{K} : q(x) \neq 0\}$ – also überall dort, wo f überhaupt definiert werden kann – stetig ist.

Als Nächstes wollen wir zeigen, dass auch Verkettungen stetiger Funktionen wieder stetig sind. Dazu beweisen wir zuerst ein etwas allgemeineres Lemma, das analog zur Vertauschbarkeit stetiger Funktionen mit Folgengrenzwerten in Satz 8.11 (b) ist und das auch oft zur Berechnung von Grenzwerten nützlich ist.

Satz 8.15 (Grenzwert einer Verkettung)**.** *Es seien $f\colon D \to \mathbb{K}$ mit $D \subset \mathbb{K}$ sowie $g\colon D' \to \mathbb{K}$ mit $D' \subset \mathbb{K}$ zwei Funktionen mit $f(D) \subset D'$. Ferner sei $a \in \overline{D}$, so dass $\lim_{x\to a} f(x)$ existiert, in D' liegt, und g in diesem Punkt stetig ist. Dann gilt*

$$\lim_{x\to a}(g\circ f)(x) = g\big(\lim_{x\to a} f(x)\big),$$

d. h. „für stetige g kann man die Anwendung von g mit der Grenzwertbildung vertauschen“.

Beweis. Wir zeigen das Lemma wieder mit dem Folgenkriterium aus Satz 8.11 (a). Es sei also $(x_n)_n$ eine beliebige Folge in D mit $x_n \to a$. Weil der Grenzwert $c := \lim_{x\to a} f(x)$ nach Voraussetzung existiert, gilt $f(x_n) \to c$ nach dem Folgenkriterium für f. Da weiterhin g in c stetig ist, gilt nach dem Folgenkriterium für g auch $(g\circ f)(x_n) = g(f(x_n)) \to g(c)$. Die Aussage des Lemmas ergibt sich damit aus dem Folgenkriterium für $g\circ f$. $\square$

Bemerkung 8.16 (Verkettungen stetiger Funktionen sind stetig)**.** Ist in Satz 8.15 sogar $a \in D$, so dass also f in a stetig ist und nach Bemerkung 8.4 dann $\lim_{x\to a} f(x) = f(a)$ gilt, so besagt das Lemma folgendes: Ist f stetig in a und g stetig in $f(a)$, so ist $\lim_{x\to a}(g\circ f)(x) = (g\circ f)(a)$, d. h. auch $g\circ f$ stetig in a.

Aufgabe 8.17. Zeige, dass die Funktion

$$f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} x & \text{für } x \in \mathbb{Q}, \\ 1-x & \text{für } x \notin \mathbb{Q} \end{cases}$$

genau im Punkt $a = \frac{1}{2}$ stetig ist.

Aufgabe 8.18. Man beweise: Ist $f\colon \mathbb{R} \to \mathbb{R}$ eine stetige Funktion, für die die Funktionalgleichung

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

gilt, so gibt es ein $a \in \mathbb{R}$ mit $f(x) = ax$ für alle $x \in \mathbb{R}$, d. h. f ist eine lineare Funktion.

Bleibt die Aussage richtig, wenn man überall $\mathbb{R}$ durch $\mathbb{C}$ ersetzt?

Genau wie bei Folgen wollen wir nun auch für Funktionen im reellen Fall uneigentliche Grenzwerte einführen, und zwar sowohl in der Start- als auch in der Zielmenge: Wir wollen sowohl Grenzwerte der Form $\lim_{x\to\infty} f(x)$ definieren als auch sagen, was es bedeutet, dass der Grenzwert einer Funktion gleich ∞ ist. Die folgende Definition ist völlig analog zu Definition 5.39.

Definition 8.19 (Uneigentliche Grenzwerte von Funktionen)**.** Es seien $D \subset \mathbb{R}$ und $f\colon D \to \mathbb{R}$ eine Funktion.

(a) Für $a \in \overline{D}$ schreiben wir $\lim_{x\to a} f(x) = \infty$, wenn

$$\forall s \in \mathbb{R}\ \exists \delta > 0\ \forall x \in D : |x-a| < \delta \Rightarrow f(x) > s.$$

Wie im Fall von Folgen in Definition 5.39 spricht man in diesem Fall von einem **uneigentlichen Grenzwert** bzw. sagt, dass f für $x \to a$ **bestimmt divergiert**.

(b) Ist D nach oben unbeschränkt (so dass man x in $f(x)$ überhaupt beliebig groß werden lassen kann), so schreibt man $\lim_{x\to\infty} f(x) = c$ für ein $c \in \mathbb{R}$, wenn

$$\forall \varepsilon > 0\ \exists r \in \mathbb{R}\ \forall x \in D : x \geq r \Rightarrow |f(x) - c| < \varepsilon.$$

Kombiniert man dies nun noch mit (a), so erhält man die Schreibweise $\lim_{x\to\infty} f(x) = \infty$ für

$$\forall s \in \mathbb{R}\ \exists r \in \mathbb{R}\ \forall x \in D : x \geq r \Rightarrow f(x) > s.$$

Beachte, dass diese letzten beiden Notationen sogar exakt mit der Definition von Folgengrenzwerten aus Definition 5.1 (b) und 5.39 übereinstimmen, wenn man sie auf eine Folge (a_n) als Funktion mit Definitionsmenge $\mathbb{N}$ anwendet.

Analog definiert man derartige Grenzwerte mit $-\infty$ statt ∞.

Bemerkung 8.20. Man prüft leicht nach, dass mit Definition 8.19 sowohl das Folgenkriterium für Funktionsgrenzwerte aus Satz 8.11 (a) als auch die Grenzwertsätze aus Satz 8.13 auch für diese uneigentlichen Grenzwerte gelten, wenn man die üblichen Rechenregeln für $\pm\infty$ verwendet.

8.B Eigenschaften stetiger Funktionen

Nachdem wir nun von vielen Funktionen gesehen haben, wie man ihre Stetigkeit nachweisen kann, wollen wir jetzt untersuchen, was wir davon haben, wenn wir wissen, dass eine gegebene Funktion stetig ist. Dazu wollen wir einige sehr anschauliche Aussagen formal beweisen, die für *reelle* stetige Funktionen auf einem abgeschlossenen Intervall $[a,b]$ mit $a < b$ gelten. Die erste von ihnen besagt, dass eine solche Funktion mit je zwei Funktionswerten auch jeden Wert dazwischen annehmen muss – was bei einer kontinuierlichen Änderung der Funktionswerte natürlich zu erwarten ist.

Satz 8.21 (**Zwischenwertsatz**)**.** *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion. Dann gibt es zu jedem c zwischen $f(a)$ und $f(b)$ ein $x \in [a,b]$ mit $f(x) = c$.*

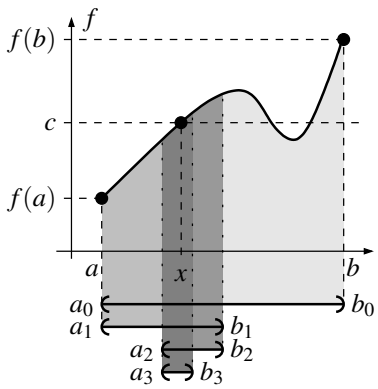
Beweis. Wir können ohne Einschränkung annehmen, dass wie im Bild unten rechts $f(a) \leq f(b)$ und damit $f(a) \leq c \leq f(b)$ gilt. Ausgehend von $[a_0,b_0] := [a,b]$ konstruieren wir nun rekursiv eine Intervallschachtelung

$$[a,b] = [a_0,b_0] \supset [a_1,b_1] \supset [a_2,b_2] \supset \cdots$$

mit in jedem Schritt halbierter Länge der Intervalle, so dass $f(a_n) \leq c \leq f(b_n)$ für alle $n \in \mathbb{N}$ gilt:

Ist $[a_n, b_n]$ für ein $n \in \mathbb{N}$ bereits konstruiert, so betrachten wir den Funktionswert $f(\frac{a_n+b_n}{2})$ in der Intervallmitte.

- Ist $f(\frac{a_n+b_n}{2}) \geq c$ (wie im Bild im Fall $n = 0$), so ersetzen wir die rechte Intervallgrenze durch den Mittelpunkt, setzen also $a_{n+1} := a_n$ und $b_{n+1} := \frac{a_n+b_n}{2}$.
- Ist dagegen $f(\frac{a_n+b_n}{2}) < c$ (wie im Fall $n = 1$ im Bild rechts), so ersetzen wir die linke Intervallgrenze durch den Mittelpunkt, setzen also $a_{n+1} := \frac{a_n+b_n}{2}$ und $b_{n+1} := b_n$.



Für den nach Satz 5.38 durch diese Intervallschachtelung definierten Punkt $x \in [a,b]$ gilt dann $a_n \to x$ und $b_n \to x$, nach dem Folgenkriterium aus Satz 8.11 also

$$f(x) = \lim_{n\to\infty} f(a_n) \leq c \leq \lim_{n\to\infty} f(b_n) = f(x)$$

und damit $f(x) = c$. □

Als Nächstes wollen wir zeigen, dass eine stetige Funktion auf einem abgeschlossenen Intervall wie in Satz 8.21 immer beschränkt ist.

Definition 8.22 (Beschränkte und monotone Funktionen)**.** Es seien $D \subset \mathbb{R}$ und $f\colon D \to \mathbb{R}$ eine Funktion. Dann heißt f auf D ...

(a) (nach oben bzw. unten) **beschränkt**, wenn die Menge $f(D) \subset \mathbb{R}$ aller Bildpunkte von f (nach oben bzw. unten) beschränkt ist.

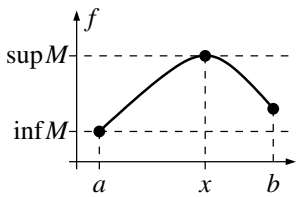
(b) **monoton wachsend** oder **steigend** (bzw. **streng monoton wachsend** oder **steigend**), wenn für alle $x, y \in D$ mit $x < y$ gilt, dass $f(x) \leq f(y)$ (bzw. $f(x) < f(y)$). Analog definiert man **(streng) monoton fallend**.

Satz 8.23. *Jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$ ist beschränkt.*

Beweis. Angenommen, f wäre unbeschränkt. Dann gäbe es zu jedem $n \in \mathbb{N}$ ein $x_n \in [a,b]$ mit $|f(x_n)| > n$. Beachte, dass die Folge $(f(x_n))_n$ dann natürlich unbeschränkt, die Folge $(x_n)_n$ aber beschränkt ist (weil ja stets $x_n \in [a,b]$ gilt). Nach dem Satz 6.21 von Bolzano-Weierstraß besitzt $(x_n)_n$ also eine konvergente Teilfolge $(x_{n_k})_k$. Der Grenzwert x dieser Teilfolge liegt nach Satz 5.23 (a) ebenfalls in $[a,b]$ und damit in der Definitionsmenge von f. Nach dem Folgenkriterium aus Satz 8.11 (b) müsste dann aber die Folge $(f(x_{n_k}))_k$ gegen $f(x)$ konvergieren – was ein Widerspruch dazu ist, dass diese Folge nach Konstruktion unbeschränkt und damit divergent ist. □

Bemerkung 8.24. Für nicht abgeschlossene Intervalle ist Satz 8.23 natürlich im Allgemeinen falsch, wie das Beispiel $f(x) = \frac{1}{x}$ auf dem offenen Intervall $(0,1)$ zeigt.

Wir haben gerade gesehen, dass das Bild $M = f([a,b])$ einer stetigen Funktion auf einem abgeschlossenen Intervall $[a,b]$ immer beschränkt ist und damit also stets zwischen $\inf M$ und $\sup M$ liegt. Wir wollen nun zeigen, dass Infimum und Supremum dieser Menge in der Tat sogar Minimum und Maximum sind, also dass diese beiden Zahlen auch als Werte von f angenommen werden – so wie z. B. im Bild rechts $f(x) = \sup M$ ist.



Satz 8.25 (**Satz vom Maximum und Minimum**)**.** *Jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$ nimmt ein Maximum und Minimum an, d. h. die Menge $M = f([a,b]) = \{f(x) : x \in [a,b]\}$ hat ein Maximum und Minimum.*

Beweis. Wir zeigen die Aussage für das Maximum; das Minimum ergibt sich analog. Die Menge M ist natürlich nicht leer und nach Satz 8.23 beschränkt, also existiert $s := \sup M$. Da s die kleinste

obere Schranke für M ist, ist $s-\frac{1}{n}$ für alle $n \in \mathbb{N}_{>0}$ dann keine obere Schranke mehr. Wir finden also ein $x_n \in [a,b]$ mit

$$s - \frac{1}{n} < f(x_n) \leq s. \qquad (*)$$

Nach dem Einschachtelungssatz 5.23 (b) konvergiert $(f(x_n))_n$ damit gegen s. Nun können wir aber wieder nach dem Satz 6.21 von Bolzano-Weierstraß aus $(x_n)_n$ eine konvergente Teilfolge $(x_{n_k})_k$ auswählen, die gegen ein $x \in [a,b]$ konvergiert. Weil f in x stetig ist, gilt nach dem Folgenkriterium aus Satz 8.11 (b) damit

$$f(x) = f\left(\lim_{k\to\infty} x_{n_k}\right) = \lim_{k\to\infty} f(x_{n_k}) = s. \qquad \square$$

Die Ergebnisse aus den Sätzen Satz 8.21, 8.23 und 8.25 lassen sich übrigens einfach in einer einzigen Aussage zusammenfassen:

Folgerung 8.26. *Ist $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion auf einem abgeschlossenen Intervall $[a,b]$, so ist das Bild von f ebenfalls ein abgeschlossenes Intervall $[c,d]$.*

Beweis. Nach Satz 8.25 existieren $c := \min f([a,b])$ und $d := \max f([a,b])$. Insbesondere gilt damit $f([a,b]) \subset [c,d]$, wobei die Werte c und d von f angenommen werden. Nach dem Zwischenwertsatz werden damit von f aber auch alle Werte zwischen c und d angenommen, d. h. es ist in der Tat $f([a,b]) = [c,d]$. $\square$

Aufgabe 8.27. Man zeige:

(a) Es seien $f\colon \mathbb{R} \to \mathbb{R}$ eine stetige Funktion mit $f(0) = 0$ sowie $g\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$ eine beschränkte stetige Funktion.

Dann ist die Funktion $f \cdot g\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $x \mapsto f(x)\cdot g(x)$ stetig in den Nullpunkt fortsetzbar.

(b) Jede bijektive, monoton wachsende Funktion $f\colon [a,b] \to [c,d]$ zwischen abgeschlossenen reellen Intervallen ist stetig.

18

Eine der wichtigsten Anwendungen dieser Aussage ist die Konstruktion von (stetigen) Umkehrfunktionen für streng monotone Funktionen:

Satz 8.28 (Existenz und Stetigkeit von Umkehrfunktionen)**.** *Es sei $f\colon [a,b] \to [c,d]$ eine streng monoton wachsende, stetige Funktion mit $c = f(a)$ und $d = f(b)$. Dann ist f bijektiv, und ihre Umkehrfunktion $f^{-1}\colon [c,d] \to [a,b]$ ist ebenfalls streng monoton wachsend und stetig.*

Analog gilt dies mit „streng monoton fallend“ statt „streng monoton wachsend“.

Beweis. Die Abbildung f ist surjektiv nach Folgerung 8.26. Sie ist auch injektiv, da sie streng monoton wachsend ist. Also ist f bijektiv, und die Umkehrfunktion $f^{-1}\colon [c,d] \to [a,b]$ existiert. Sie ist notwendigerweise streng monoton wachsend, denn wenn es $x,y \in [c,d]$ mit $x < y$ und $f^{-1}(x) \geq f^{-1}(y)$ gäbe, würde sich daraus durch Anwenden der streng monotonen Funktion f der Widerspruch $f(f^{-1}(x)) \geq f(f^{-1}(y))$, also $x \geq y$ ergeben. Nach Aufgabe 8.27 (b) ist f^{-1} damit auch stetig. $\square$

Beispiel 8.29 (Wurzelfunktionen)**.** Es seien $n \in \mathbb{N}_{>0}$ und $R \in \mathbb{R}_{>0}$ gegeben. Dann ist die Funktion $f\colon [0,R] \to [0,R^n]$, $x \mapsto x^n$ nach Lemma 4.16 streng monoton wachsend und nach Beispiel 8.14 stetig. Also ist die Umkehrfunktion $f^{-1}\colon [0,R^n] \to [0,R]$, $x \mapsto \sqrt[n]{x}$, die wir bereits aus Aufgabe 5.36 kennen, ebenfalls streng monoton wachsend und stetig. Betrachtet man diese Aussage für alle R zusammen, ist damit auch die Wurzelfunktion $f^{-1}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $x \mapsto \sqrt[n]{x}$ streng monoton wachsend und stetig. Ihr Graph ensteht wie im Bild unten durch Spiegelung des Graphen von f an der gestrichelt eingezeichneten Diagonalen.

Aufgabe 8.30. Man beweise:

(a) Jede stetige Funktion $f\colon [a,b] \to [a,b]$ hat einen Fixpunkt, d. h. ein $x \in [a,b]$ mit $f(x) = x$.

Ist f darüber hinaus monoton wachsend, so konvergiert die rekursiv definierte Folge $(x_n)_n$ mit $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$ für ein beliebiges $x_0 \in [a,b]$ gegen einen Fixpunkt von f.

(b) Ist $f\colon \mathbb{R} \to \mathbb{R}$ eine stetige Funktion mit $f(x) = f(x+1)$ für alle $x \in \mathbb{R}$ (d. h. „f ist periodisch mit Periodenlänge 1“), dann gibt es ein $a \in \mathbb{R}$ mit $f(a) = f(a+\frac{1}{2})$. (Anschaulich bedeutet dies z. B., dass es auf dem Äquator der Erde (mit Umfang 1 und Koordinate x) stets zwei gegenüberliegende Punkte gibt, an denen die gleiche Temperatur $f(x)$ herrscht.)

Aufgabe 8.31. Es sei $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ eine stetige Funktion mit $\lim_{x\to\infty} f(x) = \infty$. Zeige, dass f ein Minimum annimmt.

Aufgabe 8.32. Man zeige:

(a) Es gibt keine stetige Funktion $f\colon \mathbb{R} \to \mathbb{R}$, unter der jede reelle Zahl genau zwei Urbilder hat.

(b) Jede stetige Funktion $f\colon \mathbb{R} \to \mathbb{R}$, die offene Intervalle auf offene Intervalle abbildet, ist streng monoton.

(c) Ist $f\colon \mathbb{R} \to \mathbb{R}$ eine beschränkte stetige Funktion, so gibt es eine Gerade in $\mathbb{R}^2$, die mit dem Graphen von f mindestens drei Punkte gemeinsam hat.

8.C Gleichmäßige Konvergenz und Stetigkeit

Wir haben nun einige schöne Eigenschaften stetiger Funktionen gesehen und auch Methoden kennengelernt, mit denen wir von vielen Funktionen ihre Stetigkeit nachweisen können. Allerdings haben wir dabei bisher eine wichtige Klasse von Funktionen ausgelassen – nämlich solche, die durch den Grenzwert einer konvergenten Folge oder Reihe definiert sind, wie z. B. die Exponentialfunktion oder ganz generell allgemeine Potenzreihen wie in Definition 7.25. Zur Untersuchung der Stetigkeit derartiger Funktionen starten wir mit einem einfachen Beispiel.

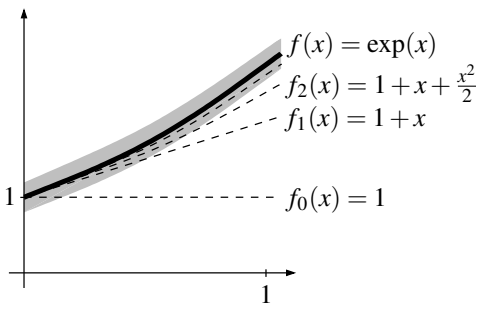
Beispiel 8.33. In Definition 7.25 (b) hatten wir die Exponentialfunktion als $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ für alle $x \in \mathbb{C}$ definiert, also als den Grenzwert

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

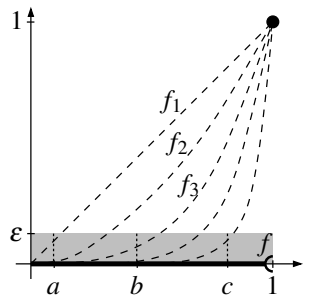
Natürlich ist jede einzelne Partialsumme f_n nach Beispiel 8.14 eine stetige Funktion in x. Da sich diese Partialsummen für $n \to \infty$ immer mehr der Exponentialfunktion annähern, würden wir nun hoffen, dass aus der Stetigkeit aller f_n auch die Stetigkeit der Grenzfunktion, also der Exponentialfunktion folgt. Allgemein fragen wir uns also: Sind $f_n\colon D \to \mathbb{K}$ für $n \in \mathbb{N}$ stetige Funktionen auf einer Menge $D \subset \mathbb{K}$, so dass für alle $x \in D$ der Grenzwert

$$f(x) := \lim_{n\to\infty} f_n(x)$$

existiert (wir sagen in diesem Fall auch, dass die Funktionen f_n *punktweise gegen f konvergieren* – siehe Definition 8.34), ist dann auch diese Grenzfunktion $f\colon D \to \mathbb{K}$ stetig? Der Fall der Exponentialfunktion ist im folgenden Bild links dargestellt, wobei die einzelnen f_n gestrichelt und die Grenzfunktion f dick eingezeichnet ist.



$f_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}, \quad f(x) = \exp(x)$



$f_n(x) = x^n, \quad f(x) = \begin{cases} 0 & \text{für } x < 1 \\ 1 & \text{für } x = 1 \end{cases}$

Es sieht hier also bereits so aus, als ob die Grenzfunktion wie gewünscht stetig ist, und in der Tat werden wir auch sehen, dass dies bei der Exponentialfunktion wirklich der Fall ist. Allerdings ist die Situation im Allgemeinen leider nicht ganz so schön, wie man es sich wünschen würde. Betrachten wir z. B. einmal die Funktionen

$$f_n \colon D = [0,1] \to \mathbb{R},\ x \mapsto x^n$$

wie im Bild oben rechts, so existiert nach Beispiel 5.3 (c) zwar der Grenzwert

$$f(x) := \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n = \begin{cases} 0 & \text{für } x < 1, \\ 1 & \text{für } x = 1 \end{cases}$$

für alle $x \in D$, aber die Grenzfunktion ist hier offensichtlich *nicht* stetig! Wir halten also fest:

> Konvergiert eine Folge stetiger Funktionen $f_n \colon D \to \mathbb{K}$ punktweise gegen eine Grenzfunktion $f \colon D \to \mathbb{K}$, so muss f nicht notwendig stetig sein!

Analog zum Fall der Umordnungen von Reihen in Beispiel 7.9 wird auch hier der Ausweg aus diesem Problem darin bestehen, eine stärkere Form der Konvergenz einer Folge stetiger Funktionen einzuführen, die letztlich die Stetigkeit der Grenzfunktion sicherstellt.

In der Tat können wir an unserem obigen Beispiel $f_n(x) = x^n$ auch schon motivieren, wie dieses stärkere Kriterium aussehen wird, denn man sieht an diesem Bild bereits recht deutlich, wo das Problem liegt: Es ist zwar richtig, dass für jedes $x \in [0,1)$ der Grenzwert $\lim_{n\to\infty} x^n$ gleich 0 ist, d. h. dass $x^n < \varepsilon$ für alle n ab einem gewissen n_0 gilt – aber dieses n_0 hängt extrem vom betrachteten Punkt x ab und wird insbesondere für $x \to 1$ immer größer. So kann man z. B. für den Wert $x = a$ im Bild oben rechts noch $n_0 = 1$ wählen, beim Wert $x = b$ braucht man mindestens $n_0 = 3$, beim Wert $x = c$ schon mindestens $n_0 = 5$. Je weiter sich x dem Wert 1 nähert, um so größer muss man dieses n_0 wählen – bis es im Grenzfall $x = 1$ schließlich gar kein solches n_0 mehr gibt, so dass 0 nicht mehr der Grenzwert der Folge $\lim_{n\to\infty} x^n$ ist. Im Bild oben links hingegen kann man für die dargestellte ε-Umgebung um f z. B. $n_0 = 3$ *für alle* x (in dem dort betrachteten Intervall) wählen, denn $f_3, f_4, f_5, \ldots$ liegen komplett in dem grau eingezeichneten Streifen.

Es kommt bei der Grenzwertdefinition also anscheinend darauf an, ob man das in der Grenzwertdefinition verlangte n_0 unabhängig vom betrachteten Punkt x wählen kann. Dies führt auf die folgenden Definitionen:

Definition 8.34 (Gleichmäßige Konvergenz und Stetigkeit)**.** Es seien $D \subset \mathbb{K}$ und $f \colon D \to \mathbb{K}$ eine Funktion.

(a) Für jedes $n \in \mathbb{N}$ sei eine Funktion $f_n \colon D \to \mathbb{K}$ gegeben – man nennt $(f_n)_n$ dann auch eine **Funktionenfolge** auf D. Ist $\lim_{n\to\infty} f_n(x) = f(x)$ für alle $x \in D$, d. h. gilt

$$\forall x \in D\ \forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : |f_n(x) - f(x)| < \varepsilon,$$

so nennt man $(f_n)_n$ **punktweise konvergent** gegen f. Beachte, dass das n_0 hierbei nicht nur von ε, sondern auch vom betrachteten Punkt x abhängen darf. Kann man n_0 jedoch auch unabhängig von x wählen, d. h. gilt sogar

$$\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N} \; \forall x \in D \; \forall n \geq n_0 : |f_n(x) - f(x)| < \varepsilon,$$

so heißt die Funktionenfolge $(f_n)_n$ auf D **gleichmäßig konvergent** gegen f.

(b) Bekanntlich heißt eine Funktion f nach Definition 8.5 stetig, wenn sie in jedem Punkt $a \in D$ stetig ist, also wenn gilt

$$\forall a \in D \; \forall \varepsilon > 0 \; \exists \delta > 0 \; \forall x \in D : |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

Beachte, dass das δ hierbei analog zu (a) nicht nur von ε, sondern auch vom betrachteten Punkt a abhängen darf. Kann man δ jedoch auch unabhängig von a wählen, gilt also

$$\forall \varepsilon > 0 \; \exists \delta > 0 \; \forall a \in D \; \forall x \in D : |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon,$$

so nennt man f auf D **gleichmäßig stetig**.

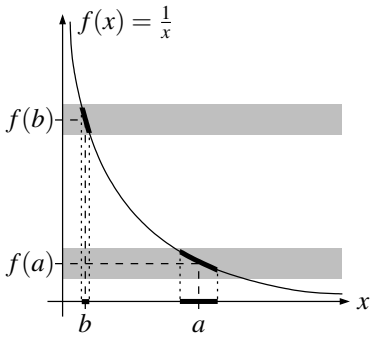
Wir wollen diese Konzepte der gleichmäßigen Konvergenz und Stetigkeit jetzt genauer untersuchen. Natürlich ist jede gleichmäßig konvergente Funktionenfolge auch punktweise konvergent, und jede gleichmäßig stetige Funktion auch stetig. Beachte auch, dass die gleichmäßige Konvergenz und Stetigkeit nach Definition keine „punktweisen“ Konzepte sind, also dass es nicht sinnvoll ist, zu sagen, eine Funktionenfolge sei „in jedem Punkt von D gleichmäßig konvergent“, oder eine Funktion „in jedem Punkt von D gleichmäßig stetig“.

Wir beginnen mit einem Beispiel dafür, dass die gleichmäßige Stetigkeit wirklich eine stärkere Bedingung als die normale Stetigkeit ist.

Beispiel 8.35. Wir behaupten, dass die nach Beispiel 8.14 stetige Funktion

$$f : \mathbb{R}_{>0} \to \mathbb{R}, \; x \mapsto \frac{1}{x}$$

nicht gleichmäßig stetig ist. Anschaulich ist diese Aussage im Bild rechts dargestellt: Die Stetigkeit von f bedeutet ja gerade, dass wir zu jeder ε-Umgebung $U_\varepsilon(f(x))$ eines Bildpunktes $f(x)$ eine δ-Umgebung von x finden, die ganz nach $U_\varepsilon(f(x))$ abgebildet wird. Zum Punkt $x = a$ haben wir in der Skizze eine ε-Umgebung von $f(a)$ grau eingezeichnet, und auf der x-Achse eine dazu passende δ-Umgebung wie in Bemerkung 8.6 dick markiert. Wenn wir nun auf einen (viel) kleineren Wert $x = b$ gehen und das gleiche ε wie oben behalten, so sehen wir, dass wir das zugehörige δ viel kleiner wählen müssen. Wenn sich x der 0 nähert, müssen wir bei gleich bleibendem ε das δ in der Tat sogar gegen 0 gehen lassen. Dies bedeutet, dass wir für festgehaltenes ε kein δ finden können, das in *jedem* Punkt $x > 0$ funktioniert – und das wiederum bedeutet genau, dass f nicht gleichmäßig stetig ist.



Wollen wir diese Aussage formal beweisen, so müssen wir die Negation der Bedingung aus Definition 8.34 (b) nachweisen, d. h.

$$\exists \varepsilon > 0 \; \forall \delta > 0 \; \exists a \in D \; \exists x \in D : |x - a| < \delta \text{ und } |f(x) - f(a)| \geq \varepsilon.$$

Dies ist schnell gezeigt: Wir setzen $\varepsilon := 1$; es sei $\delta > 0$ beliebig. Dann wählen wir $x = \delta$ und $a = \frac{\delta}{1+\delta}$, und es gilt wegen $x > a$

$$|x - a| = \delta - \frac{\delta}{1+\delta} < \delta \qquad \text{und} \qquad |f(x) - f(a)| = \left| \frac{1}{\delta} - \frac{1+\delta}{\delta} \right| = 1 \geq \varepsilon.$$

Wir wollen nun aber zeigen, dass eine stetige Funktion auf einem abgeschlossenen reellen Intervall glücklicherweise immer gleichmäßig stetig ist, so dass wir in diesem Fall die eigentlich stärkere Bedingung der gleichmäßigen Stetigkeit immer umsonst mitgeliefert bekommen.

Satz 8.36. *Jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$ auf einem abgeschlossenen reellen Intervall ist dort auch gleichmäßig stetig.*

Beweis. Angenommen, f wäre nicht gleichmäßig stetig. Dann würde wie in Beispiel 8.35 das Gegenteil der Bedingung aus Definition 8.34 (b) gelten, d. h.

$$\exists \varepsilon > 0\ \forall \delta > 0\ \exists x,y \in [a,b] : |y-x| < \delta \text{ und } |f(y)-f(x)| \geq \varepsilon.$$

Wählen wir ein solches ε, so finden wir also mit $\delta = \frac{1}{n}$ zu jedem $n \in \mathbb{N}_{>0}$ Zahlen $x_n, y_n \in [a,b]$ mit $|y_n - x_n| < \frac{1}{n}$ und $|f(y_n) - f(x_n)| \geq \varepsilon$. Insbesondere ist dann $\lim_{n\to\infty}(y_n - x_n) = 0$. Wählen wir nun nach dem Satz 6.21 von Bolzano-Weierstraß eine gegen ein $x \in [a,b]$ konvergente Teilfolge $(x_{n_k})_k$ von $(x_n)_n$, so folgt

$$\lim_{k\to\infty} x_{n_k} = x \qquad \text{und} \qquad \lim_{k\to\infty} y_{n_k} = \lim_{k\to\infty} x_{n_k} + \lim_{k\to\infty}(y_{n_k} - x_{n_k}) = x + 0 = x,$$

d. h. die entsprechende Teilfolge von $(y_n)_n$ konvergiert ebenfalls gegen x. Wegen der Stetigkeit von f in x ergibt sich dann aber nach dem Folgenkriterium aus Satz 8.11 (b)

$$\lim_{k\to\infty}(f(y_{n_k}) - f(x_{n_k})) = \lim_{k\to\infty} f(y_{n_k}) - \lim_{k\to\infty} f(x_{n_k}) = f\big(\lim_{k\to\infty} y_{n_k}\big) - f\big(\lim_{k\to\infty} x_{n_k}\big) = f(x) - f(x) = 0,$$

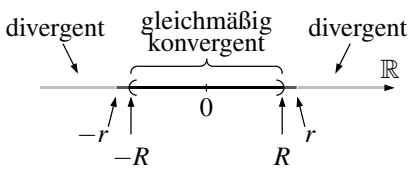
im Widerspruch zu $|f(y_{n_k}) - f(x_{n_k})| \geq \varepsilon$ für alle k. □

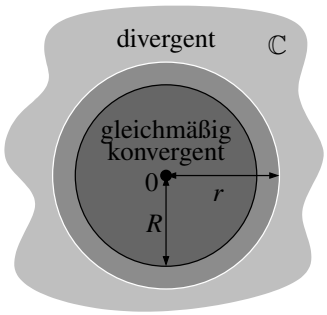
Nach der gleichmäßigen Stetigkeit wollen wir uns nun mit der gleichmäßigen Konvergenz von Funktionenfolgen befassen. Unser wichtigstes Beispiel von Funktionenfolgen sind Potenzreihen, und glücklicherweise sind diese in folgendem Sinne immer gleichmäßig konvergent.

Satz 8.37 (Gleichmäßige Konvergenz von Potenzreihen)**.** *Jede Potenzreihe $\sum_{k=0}^{\infty} a_k x^k$ in $\mathbb{K}$ mit Konvergenzradius r ist gleichmäßig konvergent auf jeder Menge der Form*

$$K_R := \{x \in \mathbb{K} : |x| \leq R\} \qquad \textit{für } 0 \leq R < r$$

(d. h. die Folge $(f_n)_n$ der Partialsummen $f_n(x) = \sum_{k=0}^{n} a_k x^k$ konvergiert gleichmäßig auf jedem K_R gegen die Grenzfunktion f mit $f(x) = \sum_{k=0}^{\infty} a_k x^k$).





Mit anderen Worten konvergieren Potenzreihen also gleichmäßig auf jedem abgeschlossenen Intervall (für $\mathbb{K} = \mathbb{R}$) bzw. Kreis (für $\mathbb{K} = \mathbb{C}$) innerhalb des Konvergenzgebiets.

Beweis. Wir weisen das Kriterium aus Definition 8.34 (a) direkt nach. Es sei dazu $\varepsilon > 0$ beliebig. Wegen $R < r$ konvergiert die Reihe $\sum_{k=0}^{\infty} a_k R^k$ nach Satz 7.26 absolut. Es gibt also ein $n_0 \in \mathbb{N}$, so dass

$$\left| \sum_{k=0}^{\infty} |a_k| \cdot R^k - \sum_{k=0}^{n} |a_k| \cdot R^k \right| = \sum_{k=n+1}^{\infty} |a_k| \cdot R^k < \varepsilon$$

für alle $n \geq n_0$ gilt. Dann folgt für alle $n \geq n_0$ und $x \in \mathbb{K}$ mit $|x| \leq R$ aber auch

$$|f(x) - f_n(x)| = \left| \sum_{k=0}^{\infty} a_k x^k - \sum_{k=0}^{n} a_k x^k \right| = \left| \sum_{k=n+1}^{\infty} a_k x^k \right| \leq \sum_{k=n+1}^{\infty} |a_k| \cdot |x|^k \leq \sum_{k=n+1}^{\infty} |a_k| \cdot R^k < \varepsilon.$$

Da wir unser n_0 hierbei unabhängig von $x \in K_R$ wählen konnten, ist die Potenzreihe auf K_R also gleichmäßig konvergent. □

Beachte, dass man den Wert von R in Satz 8.37 beliebig nahe an r wählen darf. Insbesondere findet man also zu jedem $x \in \mathbb{K}$ im Konvergenzgebiet $D = \{x \in \mathbb{K} : |x| < r\}$ ein R, so dass x in K_R enthalten ist. Dies bedeutet jedoch *nicht*, dass die Potenzreihe auch auf ganz D gleichmäßig konvergiert! Hier ist ein einfaches Gegenbeispiel dafür:

Beispiel 8.38. Die geometrische Reihe $f(x) = \sum_{k=0}^{\infty} x^k$ hat nach Beispiel 7.3 (a) das Konvergenzgebiet $D = \{x \in \mathbb{K} : |x| < 1\}$, also den Konvergenzradius 1. Wir behaupten, dass f auf D nicht gleichmäßig konvergent ist, d. h. dass die Umkehrung der Bedingung aus Definition 8.34 (a)

$$\exists \varepsilon > 0 \; \forall n_0 \in \mathbb{N} \; \exists x \in D \; \exists n \geq n_0 : \left| \sum_{k=0}^{\infty} x^k - \sum_{k=0}^{n} x^k \right| \geq \varepsilon$$

gilt. Dazu wählen wir $\varepsilon := 1$; es sei $n_0 \in \mathbb{N}$ gegeben, und wir setzen $n := n_0$. Für alle $x \in D$ mit $x > 0$ ist nun nach der Formel für die geometrische Reihe aus Beispiel 7.3 (a)

$$\left| \sum_{k=0}^{\infty} x^k - \sum_{k=0}^{n} x^k \right| = \sum_{k=n+1}^{\infty} x^k = x^{n+1} \cdot \sum_{k=0}^{\infty} x^k = \frac{x^{n+1}}{1-x}.$$

Nähert sich x nun innerhalb von D dem Wert 1, so wächst dieser Ausdruck offensichtlich unbeschränkt an. Also gibt es insbesondere ein $x \in D$, für das dieser Ausdruck mindestens gleich $1 = \varepsilon$ ist, was zu zeigen war.

19

Wir kommen nun endlich zu dem zentralen Satz, der die Motivation für die Einführung der gleichmäßigen Konvergenz bzw. Stetigkeit war:

Satz 8.39 (Der gleichmäßige Grenzwert stetiger Funktionen ist stetig)**.** *Es seien $D \subset \mathbb{K}$ und $(f_n)_n$ eine Folge stetiger Funktionen $f_n \colon D \to \mathbb{K}$, die gleichmäßig gegen eine Grenzfunktion $f \colon D \to \mathbb{K}$ konvergiert. Dann ist auch f stetig.*

Beweis. Wir weisen das ε-δ-Kriterium aus Definition 8.5 (a) für f nach. Es seien dazu $a \in D$ und $\varepsilon > 0$ beliebig. Da $(f_n)_n$ gleichmäßig gegen f konvergiert, gibt es ein $n_0 \in \mathbb{N}$, so dass

$$|f_n(x) - f(x)| < \frac{\varepsilon}{3} \quad \text{für alle } x \in D \text{ und alle } n \geq n_0 \tag{1}$$

gilt (insbesondere also auch für $x = a$). Wir können hier also $n := n_0$ setzen. Wegen der Stetigkeit von f_n gibt es dann ein $\delta > 0$ mit

$$|f_n(x) - f_n(a)| < \frac{\varepsilon}{3} \quad \text{für alle } x \in D \text{ mit } |x-a| < \delta. \tag{2}$$

Insgesamt folgt damit für alle $x \in D$ mit $|x-a| < \delta$ nach der Dreiecksungleichung

$$\begin{aligned} |f(x) - f(a)| &= |f(x) - f_n(x) + f_n(x) - f_n(a) + f_n(a) - f(a)| \\ &\leq \underbrace{|f(x) - f_n(x)|}_{< \frac{\varepsilon}{3} \text{ nach (1)}} + \underbrace{|f_n(x) - f_n(a)|}_{< \frac{\varepsilon}{3} \text{ nach (2)}} + \underbrace{|f_n(a) - f(a)|}_{< \frac{\varepsilon}{3} \text{ nach (1)}} \\ &< \varepsilon. \end{aligned}$$

□

Folgerung 8.40 (Stetigkeit von Potenzreihen)**.** *Jede Potenzreihe in $\mathbb{K}$ ist in ihrem Konvergenzgebiet stetig.*

Beweis. Wir betrachten eine Potenzreihenfunktion f in $\mathbb{K}$ mit Konvergenzradius r, also eine Funktion der Form $f(x) = \lim_{n\to\infty} f_n(x)$ mit $f_n(x) = \sum_{k=0}^{n} a_k x^k$ für alle $x \in \mathbb{K}$ mit $|x| < r$.

Es sei nun ein $c \in \mathbb{K}$ mit $|c| < r$ gegeben; wir müssen zeigen, dass f in c stetig ist. Wähle dazu ein $R > 0$ mit $|c| < R < r$. Dann ist f nach Satz 8.37 auf $K_R = \{x \in \mathbb{K} : |x| \leq R\}$ der gleichmäßige Grenzwert der stetigen Partialsummen f_n. Also ist diese Grenzfunktion f nach Satz 8.39 auf K_R und damit insbesondere in c stetig. □

Beispiel 8.41.

(a) Die Exponentialfunktion $\exp\colon \mathbb{K} \to \mathbb{K}$ ist als Potenzreihe mit Konvergenzradius ∞ nach Folgerung 8.40 auf ganz $\mathbb{K}$ stetig.

(b) Die reelle Funktionenfolge $(x^n)_n$ auf $[0,1]$ aus Beispiel 8.33 ist nach Satz 8.39 nicht gleichmäßig konvergent, da ihre Grenzfunktion nicht stetig ist. (Natürlich könnte man dies auch analog zu Beispiel 8.35 oder 8.38 direkt nachrechnen.)

Aufgabe 8.42 (Lipschitz-Stetigkeit)**.** Es sei $D \subset \mathbb{K}$. Eine Funktion $f\colon D \to \mathbb{K}$ heißt **Lipschitz-stetig**, wenn es ein $L \in \mathbb{R}_{\geq 0}$ gibt, so dass

$$|f(x) - f(y)| \leq L \cdot |x - y|$$

für alle $x, y \in D$. Man nennt L in diesem Fall eine **Lipschitz-Konstante** für f. Man zeige:

(a) Ist f Lipschitz-stetig, so ist f auch gleichmäßig stetig.

(b) Die Wurzelfunktion $f\colon [0,1] \to \mathbb{R}$, $x \mapsto \sqrt{x}$ ist gleichmäßig stetig, aber nicht Lipschitz-stetig.

Aufgabe 8.43 (Supremumsnorm)**.** Es seien $D \subset \mathbb{K}$ und $f\colon D \to \mathbb{K}$ eine Funktion. Wir definieren die **Supremumsnorm** (siehe auch Beispiel 23.3 (e)) als

$$\|f\|_{\sup} := \sup\{|f(x)| : x \in D\} \quad \in \mathbb{R}_{\geq 0} \cup \{\infty\}.$$

Zeige, dass eine Funktionenfolge $(f_n)_n$ auf D genau dann gleichmäßig gegen f konvergiert, wenn $\|f_n - f\|_{\sup} \to 0$ für $n \to \infty$.

Aufgabe 8.44.

(a) Zeige, dass die reelle Funktionenfolge $(f_n)_n$ mit $f_n(x) = \frac{nx}{1+nx}$ zwar nicht auf $\mathbb{R}_{>0}$, aber auf jedem Intervall $[a, \infty)$ mit $a > 0$ gleichmäßig konvergiert.

(b) Zeige, dass die Funktion $f\colon [1, \infty) \to \mathbb{R}$, $x \mapsto \frac{1}{x^2}$ gleichmäßig stetig ist.

Aufgabe 8.45 (Koeffizientenvergleich für Potenzreihen)**.** Es sei $f(x) = \sum_{n=0}^{\infty} a_n x^n$ eine Potenzreihe über $\mathbb{K}$ mit Konvergenzradius mindestens $r > 0$. Man zeige:

(a) Ist $f(x) = 0$ für alle $x \in \mathbb{K}$ mit $|x| < r$, so gilt bereits $a_n = 0$ für alle $n \in \mathbb{N}$ (d. h. ist der Wert der Reihe gleich 0 für alle diese x, so sind bereits alle Koeffizienten der Reihe gleich 0).

(b) Ist $g(x) = \sum_{n=0}^{\infty} b_n x^n$ eine weitere Potenzreihe mit Konvergenzradius mindestens r und gilt $f(x) = g(x)$ für alle $x \in \mathbb{K}$ mit $|x| < r$, so ist bereits $a_n = b_n$ für alle $n \in \mathbb{N}$.

9. Spezielle Funktionen

Nachdem wir jetzt schon recht viel allgemeine Theorie kennengelernt haben, wollen wir diese nun anwenden, um einige bekannte spezielle Funktionen zu studieren (oder überhaupt erst einmal exakt zu definieren), die ihr bereits aus der Schule kennt: die Exponential- und Logarithmusfunktion, die allgemeine Potenz x^a für $a \in \mathbb{R}$, die Winkelfunktionen und ihre Umkehrfunktionen. Ausgangspunkt aller dieser Funktionen ist letztlich die in Definition 7.25 (b) bereits eingeführte Exponentialfunktion

$$\exp(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots \quad \text{für } x \in \mathbb{K}.$$

Aus Folgerung 7.35 wissen wir schon, dass diese Funktion die Funktionalgleichung

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

erfüllt. Außerdem ist sie nach Beispiel 8.41 (a) stetig, und aus der Reihendarstellung sieht man sofort, dass $\exp(0) = 1$.

Die weiteren Eigenschaften der Exponentialfunktion sind im reellen und komplexen Fall trotz der gleich lautenden Definition sehr unterschiedlich. Wir werden diese beiden Fälle im Folgenden daher separat untersuchen.

9.A Logarithmen und allgemeine Potenzen

Wir beginnen mit der reellen Exponentialfunktion und zeigen zunächst einige ihrer wichtigen Eigenschaften.

Satz 9.1 (Eigenschaften der reellen Exponentialfunktion)**.**

(a) *Es gilt* $\exp(x) > 0$ *für alle* $x \in \mathbb{R}$.

(b) *Die Funktion* $\exp\colon \mathbb{R} \to \mathbb{R}$ *ist streng monoton wachsend.*

(c) *Für alle* $n \in \mathbb{N}$ *gilt*

$$\lim_{x\to\infty} \frac{\exp(x)}{x^n} = \infty \qquad \textit{und} \qquad \lim_{x\to-\infty} x^n \exp(x) = 0.$$

Insbesondere ist also $\lim_{x\to\infty} \exp(x) = \infty$ *und* $\lim_{x\to-\infty} \exp(x) = 0$.

(d) *Für die Zahl* $\mathrm{e} := \exp(1)$ *gilt* $2 < \mathrm{e} < 3$. *(Man nennt* e *die* ***Eulersche Zahl****; eine explizite näherungsweise Berechnung der Exponentialreihe zeigt, dass* $\mathrm{e} = 2{,}71828\ldots$*.)*

Beweis.

(a) Für $x \geq 0$ ist dies aus der Reihendarstellung offensichtlich. Für $x \leq 0$ folgt aus der Funktionalgleichung

$$\exp(x) \cdot \exp(-x) = \exp(0) = 1 \quad \text{und damit} \quad \exp(x) = \frac{1}{\exp(-x)},$$

was nun wegen $-x \geq 0$ ebenfalls positiv ist.

(b) Es seien $x, y \in \mathbb{R}$ mit $x < y$. Wegen $y - x > 0$ folgt aus der Reihendarstellung der Exponentialfunktion dann $\exp(y-x) > 1$. Da nach (a) außerdem $\exp(x) > 0$ gilt, erhalten wir mit der Funktionalgleichung wie gewünscht

$$\exp(y) = \exp(x) \cdot \exp(y-x) > \exp(x) \cdot 1 = \exp(x).$$

(c) Für $x > 0$ ergibt sich aus der Reihendarstellung der Exponentialfunktion natürlich

$$\exp(x) > \frac{x^{n+1}}{(n+1)!} \qquad \text{und damit} \qquad \frac{\exp(x)}{x^n} > \frac{x}{(n+1)!}.$$

für alle $n \in \mathbb{N}$. Wegen $\lim_{x\to\infty} \frac{x}{(n+1)!} = \infty$ folgt damit auch $\lim_{x\to\infty} \frac{\exp(x)}{x^n} = \infty$.

Die Aussage für $x \to -\infty$ zeigt man analog: Für $x < 0$ ist $-x > 0$, und da wir in (a) schon gesehen haben, dass $\exp(x) = \frac{1}{\exp(-x)}$ gilt, erhalten wir

$$\exp(x) = \frac{1}{\exp(-x)} < \frac{1}{(-x)^{n+1}/(n+1)!} \qquad \text{und damit} \qquad |x^n \exp(x)| < \frac{(n+1)!}{|x|}.$$

Wegen $\lim_{x\to-\infty} \frac{(n+1)!}{|x|} = 0$ ergibt sich damit auch $\lim_{x\to-\infty} x^n \exp(x) = 0$.

(d) Aus der Exponentialreihe erhalten wir sofort

$$\mathrm{e} > \frac{1}{0!} + \frac{1}{1!} = 2,$$

und wegen $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n \geq 1 \cdot 2 \cdot 2 \cdot \cdots \cdot 2 = 2^{n-1}$ mit Hilfe der geometrischen Reihe aus Beispiel 7.3 (a) auch

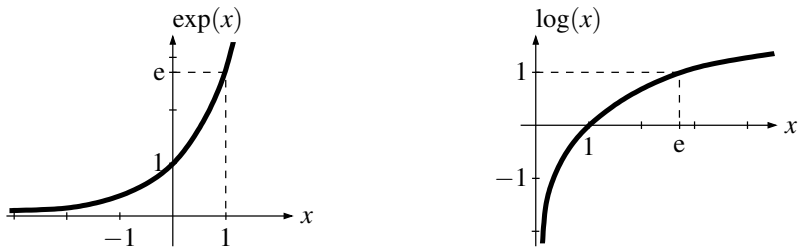
$$\mathrm{e} = \frac{1}{0!} + \sum_{n=1}^{\infty} \frac{1}{n!} < 1 + \sum_{n=1}^{\infty} \frac{1}{2^{n-1}} = 1 + \sum_{n=0}^{\infty} \frac{1}{2^n} = 1 + \frac{1}{1-\frac{1}{2}} = 3. \qquad \square$$

Aufgabe 9.2 (Irrationalität von e)**.** Finde für alle $n \in \mathbb{N}_{>0}$ explizit eine natürliche Zahl $a \in \mathbb{N}$ mit $\frac{a}{n!} < \mathrm{e} < \frac{a+1}{n!}$, und zeige so, dass e irrational ist.

Bemerkung 9.3.

(a) Da die (uneigentlichen) Grenzwerte von $\exp(x)$ für $x \to \pm\infty$ nach Satz 9.1 (b) gleich ∞ bzw. 0 sind, bedeutet die Aussage desselben Satzes für $n > 0$ gerade, dass sich in diesen Grenzwerten, die ja von der Form $\frac{\infty}{\infty}$ bzw. $\pm\infty \cdot 0$ sind, die Exponentialfunktion gegenüber der Potenz x^n durchsetzt. Man sagt auch, „die Exponentialfunktion ist für $x \to \pm\infty$ stärker als jede Potenz“.

(b) Im Bild unten links haben wir den Graphen der reellen Exponentialfunktion gemäß Satz 9.1 skizziert. Da exp nach Beispiel 8.41 (a) stetig und nach Satz 9.1 (b) streng monoton wachsend ist, existiert nach Satz 8.28 eine Umkehrfunktion (wie in Beispiel 8.29 zunächst für Start- und Zielmenge $[-R,R] \to [\exp(-R), \exp(R)]$ für alle $R > 0$, dann durch den Übergang $R \to \infty$ aber auch für $\mathbb{R} \to \mathbb{R}_{>0}$):



Definition 9.4 (Logarithmus)**.** Die Umkehrfunktion der Exponentialfunktion $\exp\colon \mathbb{R} \to \mathbb{R}_{>0}$ wird mit

$$\log\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto \log(x)$$

bezeichnet und heißt die (natürliche) **Logarithmusfunktion**. Sie ist im Bild oben rechts dargestellt. Statt $\log(x)$ ist oft auch die Bezeichnung $\ln(x)$ üblich.

Notation 9.5 (Schreibweise spezieller Funktionen)**.** Bei den speziellen Funktionen, die wir in diesem Kapitel kennenlernen werden, ist es zur Vereinfachung der Notation oft üblich, die Klammern

beim Funktionsargument wegzulassen, wenn es sich nur um eine einfache Zahl oder Variable handelt, also z. B. $\log x$ statt $\log(x)$ zu schreiben. Ist das Funktionsargument jedoch ein zusammengesetzter Ausdruck, sind die Klammern zwingend erforderlich: $\log(x+y)$ kann man nicht als $\log x + y$ schreiben, da $\log x + y$ immer als $(\log x) + y$ zu verstehen ist.

Bemerkung 9.6 (Eigenschaften der Logarithmusfunktion)**.** Unsere bisher gezeigten Eigenschaften der Exponentialfunktion übertragen sich natürlich sofort auf die Logarithmusfunktion:

(a) log ist stetig und streng monoton wachsend nach Satz 8.28.

(b) $\log 1 = 0$ und $\log \mathrm{e} = 1$.

(c) Die Grenzwerte aus Satz 9.1 (b) übertragen sich durch Vertauschen von Start- und Zielraum auf den Logarithmus als $\lim_{x\to\infty} \log x = \infty$ und $\lim_{x\to 0} \log x = -\infty$.

(d) Wenden wir die Funktionalgleichung der Exponentialfunktion auf $\log x$ und $\log y$ für $x, y > 0$ an, so erhalten wir

$$\exp(\log x + \log y) = \exp(\log x) \cdot \exp(\log y) = x \cdot y$$

und damit durch Logarithmieren

$$\log x + \log y = \log(x \cdot y) \qquad \text{für } x, y \in \mathbb{R}_{>0},$$

was die **Funktionalgleichung der Logarithmusfunktion** genannt wird.

Eine der wichtigsten Anwendungen der Logarithmusfunktion ist, dass man mit ihr allgemeine Potenzen definieren kann – also Potenzen der Form x^a, wobei a nun nicht mehr wie bisher in $\mathbb{Z}$ liegen muss, sondern eine beliebige reelle Zahl sein kann:

Definition 9.7 (Allgemeine Potenzen)**.** Für $x \in \mathbb{R}_{>0}$ und $a \in \mathbb{R}$ definieren wir die **Potenz**

$$x^a := \exp(a \log x)$$

(wir werden in Bemerkung 9.9 (a) noch sehen, dass dies für $a \in \mathbb{Z}$ mit unserer alten Definition aus Notation 3.9 (b) übereinstimmt – was dann auch diese neue, allgemeinere Definition motiviert).

Lemma 9.8 (Rechenregeln für allgemeine Potenzen)**.** *Für alle $x, y \in \mathbb{R}_{>0}$ und $a, b \in \mathbb{R}$ gilt*

(a) $x^0 = 1$ *und* $x^1 = x$*;*

(b) $x^{a+b} = x^a \cdot x^b$ *und* $x^{-a} = \frac{1}{x^a}$*;*

(c) $x^{ab} = (x^a)^b$*;*

(d) $(xy)^a = x^a \cdot y^a$.

Beweis. Alle Beweise sind einfaches Nachrechnen mit Hilfe der Funktionalgleichung:

(a) $x^0 = \exp(0 \cdot \log x) = \exp(0) = 1$ und $x^1 = \exp(\log x) = x$.

(b) Es ist

$$x^{a+b} = \exp((a+b)\log x) = \exp(a \log x + b \log x) = \exp(a \log x) \cdot \exp(b \log x) = x^a \cdot x^b.$$

Setzen wir in dieser Gleichung $b = -a$, so ergibt sich ferner $1 = x^a \cdot x^{-a}$ und damit $x^{-a} = \frac{1}{x^a}$.

(c) Es gilt

$$(x^a)^b = \exp(b \log(x^a)) = \exp(b \log(\exp(a \log x))) = \exp(ab \log x) = x^{ab}.$$

(d) Es ist

$$(xy)^a = \exp(a \log(xy)) = \exp(a \log x + a \log y) = \exp(a \log x) \cdot \exp(a \log y) = x^a \cdot y^a. \qquad \square$$

Bemerkung 9.9.

(a) Aus Lemma 9.8 (a) und (b) folgt insbesondere, dass im Fall $a \in \mathbb{N}$ für unsere allgemeine Potenz aus Definition 9.7

$$x^a = x^{1+\cdots+1} = \underbrace{x \cdot \cdots \cdot x}_{a\text{-mal}} \qquad \text{und genauso} \qquad x^{-a} = x^{-1-\cdots-1} = \underbrace{\frac{1}{x} \cdot \cdots \cdot \frac{1}{x}}_{a\text{-mal}}$$

gilt, dass sie dann also mit der alten Definition der Potenz aus Notation 3.9 (b) übereinstimmt.

(b) Nach Lemma 9.8 (c) ist $x \mapsto x^{\frac{1}{a}}$ für $a \neq 0$ eine Umkehrfunktion zu $x \mapsto x^a$, denn es ist

$$(x^a)^{\frac{1}{a}} = x^1 = x \qquad \text{und} \qquad \left(x^{\frac{1}{a}}\right)^a = x^1 = x$$

für alle $x \in \mathbb{R}_{>0}$ und $a \in \mathbb{R}\backslash\{0\}$.

Da Umkehrfunktionen eindeutig sind und wir im Fall $a \in \mathbb{N}_{>0}$ aus Aufgabe 5.36 und Beispiel 8.29 bereits wissen, dass die Wurzelfunktion $x \mapsto \sqrt[a]{x}$ ebenfalls eine solche Umkehrfunktion ist, sehen wir also, dass $x^{\frac{1}{a}} = \sqrt[a]{x}$ für alle $x \in \mathbb{R}_{>0}$ und $a \in \mathbb{N}_{>0}$ gilt.

(c) Mit der Eulerschen Zahl e aus Satz 9.1 (d) ist offensichtlich $\mathrm{e}^a = \exp(a \log \mathrm{e}) = \exp(a)$ für alle $a \in \mathbb{R}$. Man verwendet daher in der Regel die einfachere Potenzschreibweise e^a für $\exp(a)$.

(d) Beachte, dass wir die allgemeine Potenz x^a mit $a \in \mathbb{R}$ nur für positive x definieren konnten, weil für negative Zahlen kein Logarithmus existiert. In der Tat ist es auch einleuchtend, dass ein Ausdruck wie z. B. $(-1)^{\sqrt{2}}$ nicht sinnvoll definiert werden kann, da nicht einmal klar ist, ob diese Zahl positiv oder negativ sein sollte.

9.B Winkelfunktionen

Nach der reellen wollen wir nun die komplexe Exponentialfunktion studieren, die uns schließlich zu den Winkelfunktionen führen wird. Wie in Bemerkung 9.9 (c) werden wir dabei die Exponentialfunktion auch im Komplexen in der Regel mit $z \mapsto \mathrm{e}^z$ bezeichnen (obwohl es keine allgemeine Potenz w^z für $w, z \in \mathbb{C}$ gibt). Ihre wesentlichen Eigenschaften, die wir benötigen, um den Zusammenhang mit Winkelfunktionen herstellen zu können, sind die folgenden:

Satz 9.10 (Eigenschaften der komplexen Exponentialfunktion)**.** *Es gilt:*

(a) $\mathrm{e}^{\overline{z}} = \overline{\mathrm{e}^z}$ *für alle* $z \in \mathbb{C}$*;*

(b) $|\mathrm{e}^{\mathrm{i}x}| = 1$ *für alle* $x \in \mathbb{R}$*.*

Beweis.

(a) Für die Partialsummen $f_n(z) := \sum_{k=0}^n \frac{z^k}{k!}$ der Exponentialfunktion folgt natürlich $f_n(\overline{z}) = \overline{f_n(z)}$ durch fortgesetztes Anwenden von Lemma 6.9 (a). Da die komplexe Konjugation $z \mapsto \overline{z}$ nach Beispiel 8.7 (c) stetig ist, ergibt sich also nach dem Folgenkriterium für Stetigkeit aus Satz 8.11 (b)

$$\mathrm{e}^{\overline{z}} = \lim_{n\to\infty} f_n(\overline{z}) = \lim_{n\to\infty} \overline{f_n(z)} = \overline{\lim_{n\to\infty} f_n(z)} = \overline{\mathrm{e}^z}.$$

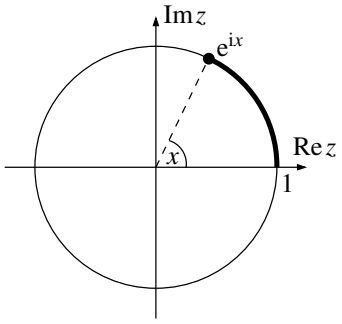
(b) Wegen $|z| = \sqrt{z\overline{z}}$ (siehe Bemerkung 6.4) erhalten wir nun mit (a)

$$|\mathrm{e}^{\mathrm{i}x}| = \sqrt{\mathrm{e}^{\mathrm{i}x} \cdot \overline{\mathrm{e}^{\mathrm{i}x}}} = \sqrt{\mathrm{e}^{\mathrm{i}x} \cdot \mathrm{e}^{\overline{\mathrm{i}x}}} = \sqrt{\mathrm{e}^{\mathrm{i}x} \cdot \mathrm{e}^{-\mathrm{i}x}} = \sqrt{\mathrm{e}^{\mathrm{i}x-\mathrm{i}x}} = \sqrt{\mathrm{e}^0} = 1. \qquad \square$$

Bemerkung 9.11. Satz 9.10 (b) besagt gerade, dass die komplexe Zahl e^{ix} für *reelle* x immer auf dem Rand des Einheitskreises liegt. Multiplizieren wir zwei solche Zahlen e^{ix} und e^{iy} für $x, y \in \mathbb{R}$ miteinander, so erhalten wir einerseits nach der Funktionalgleichung der Exponentialfunktion die Zahl

$$e^{ix} \cdot e^{iy} = e^{i(x+y)}$$

(d. h. die Exponenten addieren sich), andererseits haben wir aber auch schon in Bemerkung 6.5 gesehen, dass sich bei der komplexen Multiplikation die Winkel, die die Zahlen mit der positiven reellen Achse einschließen, ebenfalls addieren. Wir können den Exponenten x der Zahl e^{ix} daher wie im Bild rechts als ein Maß für diesen Winkel auffassen.



In der Tat werden wir in Aufgabe 11.7 sehen, dass dieses x genau die (im Bild oben rechts dick eingezeichnete) Länge des Kreisbogens ist, der von 1 zu der Zahl e^{ix} führt – man sagt auch, dass x der im **Bogenmaß** gemessene Winkel ist. Wir werden diese Aussage im Folgenden nicht benötigen, sondern verwenden sie hier nur als Motivation dafür, dass Real- und Imaginärteil von e^{ix} (also die beiden Koordinaten dieses Punktes in der Ebene) dann wie aus der Schule bekannt der Kosinus bzw. Sinus des Winkels x sein sollten. Diese Idee machen wir nun zu unserer *Definition* von Kosinus und Sinus.

Definition 9.12. Für $x \in \mathbb{R}$ definieren wir **Kosinus** und **Sinus** als die reellen Zahlen

$$\cos x := \operatorname{Re}(e^{ix}) \qquad \text{und} \qquad \sin x := \operatorname{Im}(e^{ix}),$$

so dass also $e^{ix} = \cos x + i \sin x$ die Zerlegung der komplexen Exponentialfunktion in Real- und Imaginärteil ist.

Bevor wir die Eigenschaften dieser beiden Funktionen studieren, wollen wir erst einmal zwei einfache alternative Darstellungsweisen notieren:

Lemma 9.13 (Alternative Darstellung von Kosinus und Sinus)**.**

(a) *Für alle $x \in \mathbb{R}$ ist*

$$\cos x = \frac{1}{2}\left(e^{ix} + e^{-ix}\right) \qquad \textit{und} \qquad \sin x = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right).$$

(b) *Kosinus und Sinus lassen sich darstellen als reelle Potenzreihen mit Konvergenzradius ∞*

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \qquad = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \pm \cdots,$$

$$\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \pm \cdots.$$

Insbesondere sind Kosinus und Sinus nach Folgerung 8.40 also stetige Funktionen auf $\mathbb{R}$.

Beweis.

(a) Dies folgt aus den allgemeinen Formeln $\operatorname{Re} z = \frac{1}{2}(z + \bar{z})$ und $\operatorname{Im} z = \frac{1}{2i}(z - \bar{z})$ (siehe Bemerkung 6.4) zusammen mit Satz 9.10 (a).

(b) Die Potenzreihendarstellungen ergeben sich mit (a) sofort aus

$$e^{ix} = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = 1 + i\frac{x^1}{1!} - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \cdots$$

$$\text{und} \qquad e^{-ix} = \sum_{n=0}^{\infty} \frac{(-ix)^n}{n!} = 1 - i\frac{x^1}{1!} - \frac{x^2}{2!} + i\frac{x^3}{3!} + \frac{x^4}{4!} - i\frac{x^5}{5!} - \frac{x^6}{6!} + i\frac{x^7}{7!} + \cdots,$$

da man konvergente Reihen nach Lemma 7.4 (a) gliedweise addieren kann. Weil diese Reihendarstellung für alle $x \in \mathbb{R}$ gilt, ist der Konvergenzradius dieser Potenzreihen gleich ∞. $\square$

Wir listen nun zunächst die einfachsten Eigenschaften von Kosinus und Sinus auf, die sich direkt aus der Definition ergeben.

Satz 9.14 (Eigenschaften von Kosinus und Sinus)**.** *Für alle $x, y \in \mathbb{R}$ gilt*

(a) $\cos(-x) = \cos x$ *und* $\sin(-x) = -\sin x$. *(Der Graph von* cos *ist also achsensymmetrisch zur vertikalen Achse, der von* sin *punktsymmetrisch zum Ursprung).*

(b) $(\cos x)^2 + (\sin x)^2 = 1$*; insbesondere ist also* $|\cos x| \leq 1$ *und* $|\sin x| \leq 1$.

(c) ***(Additionstheoreme)***

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y$$
$$\textit{und} \qquad \sin(x \pm y) = \sin x \cos y \pm \cos x \sin y$$

(wobei die Gleichungen so zu verstehen sind, dass an beiden Stellen das obere oder an beiden Stellen das untere Vorzeichen zu nehmen ist).

Beweis.

(a) Dies folgt z. B. unmittelbar aus Lemma 9.13 (a).

(b) Nach Satz 9.10 (b) ist $(\cos x)^2 + (\sin x)^2 = (\mathrm{Re}(\mathrm{e}^{\mathrm{i}x}))^2 + (\mathrm{Im}(\mathrm{e}^{\mathrm{i}x}))^2 = |\mathrm{e}^{\mathrm{i}x}|^2 = 1$.

(c) Einerseits gilt nach der Funktionalgleichung der komplexen Exponentialfunktion

$$\begin{aligned} \mathrm{e}^{\mathrm{i}(x+y)} &= \mathrm{e}^{\mathrm{i}x} \cdot \mathrm{e}^{\mathrm{i}y} = (\cos x + \mathrm{i} \sin x)(\cos y + \mathrm{i} \sin y) \\ &= \cos x \cos y - \sin x \sin y + \mathrm{i}\big(\sin x \cos y + \cos x \sin y\big), \end{aligned}$$

andererseits nach Definition aber auch

$$\mathrm{e}^{\mathrm{i}(x+y)} = \cos(x+y) + \mathrm{i} \sin(x+y).$$

Vergleich von Real- und Imaginärteil liefert nun die behaupteten Formeln für $\cos(x+y)$ und $\sin(x+y)$. Die Formeln für $\cos(x-y)$ und $\sin(x-y)$ ergeben sich daraus durch den Übergang $y \to -y$ mit (a). □

Notation 9.15 ($\cos^n x$ und $\sin^n x$)**.** Für $n \in \mathbb{N}$ schreibt man zur Abkürzung oft auch $\cos^n x$ und $\sin^n x$ anstatt $(\cos x)^n$ und $(\sin x)^n$. Die Formel aus Satz 9.14 (b) schreibt sich dann z. B. kürzer als $\cos^2 x + \sin^2 x = 1$. Beachte aber, dass dies leicht zu Verwechslungen führen kann, weil wir die Umkehrfunktion einer Funktion f in Lemma 2.20 (c) ja mit $x \mapsto f^{-1}(x)$ bezeichnet haben, dies aber nach dieser neuen Notation auch als $x \mapsto (f(x))^{-1} = \frac{1}{f(x)}$ interpretiert werden könnte – was natürlich etwas völlig anderes ist. Wir werden daher für Kosinus und Sinus die Schreibweise $\cos^{-1}(x)$ bzw. $\sin^{-1}(x)$ überhaupt nicht verwenden, und den Umkehrfunktionen dieser beiden Funktionen andere Namen geben (siehe Definition 9.24).

Aufgabe 9.16. Es sei $x \in \mathbb{R}$. Finde und beweise eine explizite Formel für die Summen

$$\sum_{k=0}^{n} \cos(kx) \quad \text{und} \quad \sum_{k=0}^{n} \sin(kx).$$

20

Als Nächstes wollen wir die Nullstellen und die Periodizität von Kosinus und Sinus untersuchen. Aus Bemerkung 9.11 (und dem, was ihr aus der Schule wisst) ist z. B. klar, dass diese beiden Funktionen die Periode 2π besitzen sollten. Aber bisher wissen wir überhaupt noch nicht, was π eigentlich genau ist! Wie ihr euch vielleicht schon denken könnt, wird auch hier der Ausweg wieder darin bestehen, die Sache rückwärts anzugehen und die Zahl π über die Eigenschaften der Kosinus- und Sinusfunktion zu *definieren*. Dazu benötigen wir das folgende Lemma.

Lemma 9.17.

(a) *Für alle* $x \in (0,2)$ *ist* $\sin x > 0$.

(b) *Die Kosinusfunktion ist im Intervall* $[0,2]$ *streng monoton fallend, und es gilt* $\cos 0 > 0$ *sowie* $\cos 2 < 0$.

Beweis. Wir bemerken zunächst, dass die Summanden der Exponentialreihe $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ für $0 < x \leq 2$ ab dem x^1-Term betragsmäßig monoton fallend sind, denn für $n \geq 1$ ist

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

Dasselbe gilt dann natürlich auch für die Glieder der Kosinus- und Sinusreihe, die nach Lemma 9.13 (b) ja bis auf das Vorzeichen genau die geraden bzw. ungeraden Terme der Exponentialreihe sind. Da die Kosinus- und Sinusreihe zudem alternierend sind, sind ihre Partialsummen nach Satz 7.8 damit abwechselnd obere und untere Schranken für den Grenzwert (sofern wir mindestens bis zum x^1-Term aufsummieren, ab dem die Summanden betragsmäßig monoton fallen).

(a) Nach unserer Vorbemerkung folgt nun sofort für alle $x \in (0,2)$

$$\sin x \geq x - \frac{x^3}{3!} = x\left(1 - \frac{x^2}{6}\right) > x\left(1 - \frac{2^2}{6}\right) = \frac{x}{3} > 0.$$

(b) Natürlich ist $\cos 0 = 1 > 0$. Für $\cos 2$ gilt wieder nach der Vorbemerkung

$$\cos 2 \leq 1 - \frac{2^2}{2!} + \frac{2^4}{4!} = 1 - \frac{4}{2} + \frac{16}{24} = -\frac{1}{3} < 0.$$

Es bleibt also nur noch die strenge Monotonie zu zeigen. Es seien dazu $x, y \in [0,2]$ mit $x < y$ gegeben. Mit den Additionstheoremen aus Satz 9.14 (c) ergibt sich

$$\cos\left(\frac{y+x}{2} \pm \frac{y-x}{2}\right) = \cos\frac{y+x}{2} \cdot \cos\frac{y-x}{2} \mp \sin\frac{y+x}{2} \cdot \sin\frac{y-x}{2}.$$

Subtraktion dieser beiden Gleichungen voneinander liefert nun

$$\cos\Big(\underbrace{\frac{y+x}{2} + \frac{y-x}{2}}_{=y}\Big) - \cos\Big(\underbrace{\frac{y+x}{2} - \frac{y-x}{2}}_{=x}\Big) = -2\sin\underbrace{\frac{y+x}{2}}_{\in(0,2)} \cdot \sin\underbrace{\frac{y-x}{2}}_{\in(0,2)},$$

mit (a) also $\cos y - \cos x < 0$ und damit wie behauptet $\cos x > \cos y$. □

Da die Kosinusfunktion nach Lemma 9.13 (b) stetig ist, ergibt sich mit dem Zwischenwertsatz 8.21 aus Lemma 9.17 (b) also, dass es *genau ein* $x \in (0,2)$ gibt mit $\cos x = 0$. Aus der Interpretation von Bemerkung 9.11 sehen wir, dass diese Nullstelle des Kosinus genau bei $x = \frac{\pi}{2}$ auftreten sollte, also dort wo e^{ix} auf der imaginären Achse liegt. Wir benutzen dies nun, um die Zahl π zu *definieren*:

Definition 9.18 (Die Zahl π)**.** Wir definieren die Zahl $\pi \in \mathbb{R}$ als das Doppelte der (nach obigen Überlegungen eindeutig bestimmten) Nullstelle der Kosinusfunktion im Intervall $[0,2]$. (Eine näherungsweise Berechnung dieser Nullstelle zeigt, dass $\pi = 3,14159\ldots$.)

Bemerkung 9.19.

(a) Unsere Definition 9.18 ist natürlich nicht die einzig mögliche Art, wie man die Zahl π definieren kann. Man könnte genauso gut auch irgendeine andere charakteristische Eigenschaft dieser Zahl als Definition verwenden, wie z. B. (was ja häufig gesagt wird) den Flächeninhalt des Einheitskreises. Allerdings wissen wir bisher noch gar nicht, wie man derartige Flächeninhalte überhaupt definieren und berechnen kann. Für uns hat die etwas merkwürdig scheinende Definition 9.18 daher einfach den Vorteil, dass sie am schnellsten zu den gewünschten Resultaten führt (insbesondere auch ohne Flächeninhalte berechnen zu können).

(b) Da die Kosinusfunktion nach Lemma 9.17 (b) auf $[0,2]$ streng monoton fallend ist, ist ihre Einschränkung auf das Intervall $[0, \frac{\pi}{2}]$ damit eine bijektive, streng monoton fallende Funktion, die von $\cos 0 = 1$ nach $\cos\frac{\pi}{2} = 0$ läuft. Wir wollen nun sehen, wie wir aus diesem Abschnitt der Kosinusfunktion die Kosinus- und Sinusfunktion auf ganz $\mathbb{R}$ rekonstruieren können.

Satz 9.20 (Periodizität von Kosinus und Sinus)**.**

(a) *An den Stellen* $x \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}, 2\pi\}$ *nehmen* e^{ix}, $\cos x$ *und* $\sin x$ *die folgenden Werte an:*

x	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
e^{ix}	1	i	-1	$-i$	1
$\cos x$	1	0	-1	0	1
$\sin x$	0	1	0	-1	0

(b) *Kosinus und Sinus sind* 2π*-periodisch, d. h. es gilt* $\cos(x+2\pi) = \cos x$ *und* $\sin(x+2\pi) = \sin x$ *für alle* $x \in \mathbb{R}$.

(c) *Für alle* $x \in \mathbb{R}$ *ist* $\cos(\pi - x) = -\cos x$ *und* $\sin(\frac{\pi}{2} \pm x) = \cos x$.

Beweis.

(a) Nach Definition 9.18 ist $\cos\frac{\pi}{2} = 0$. Für den Sinus gilt damit $\sin^2\frac{\pi}{2} = 1 - \cos^2\frac{\pi}{2} = 1 - 0 = 1$ nach Satz 9.14 (b); da außerdem $\sin\frac{\pi}{2} > 0$ nach Lemma 9.17 (a) gilt, muss also $\sin\frac{\pi}{2} = 1$ sein. Damit ist $e^{i\frac{\pi}{2}} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i$.

Die übrigen behaupteten Werte für $e^{in\frac{\pi}{2}}$ mit $n \in \{0,1,2,3,4\}$ folgen damit sofort nach den Rechenregeln für Potenzen aus $e^{in\frac{\pi}{2}} = (e^{i\frac{\pi}{2}})^n = i^n$. Aufteilen dieser Zahlen in Real- und Imaginärteil liefert dann die Werte für Kosinus und Sinus in der Tabelle.

(b) Nach (a) ist $e^{i\cdot 2\pi} = 1$ und damit $e^{i\cdot(x+2\pi)} = e^{ix}\cdot e^{i\cdot 2\pi} = e^{ix}$. Aufteilen in Real- und Imaginärteil liefert wieder die Behauptung.

(c) Wiederum mit (a) ist

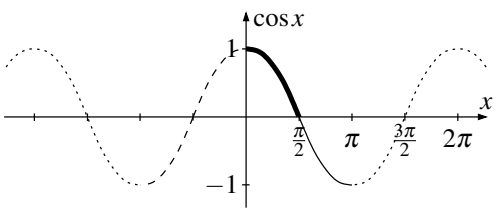
$$\cos(\pi - x) = \cos\pi\cos x + \sin\pi\sin x = (-1)\cdot\cos x + 0\cdot\sin x = -\cos x$$

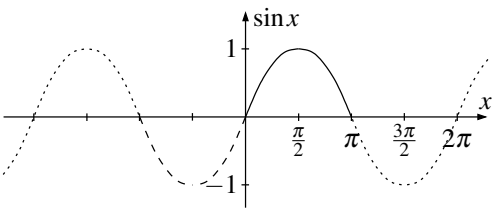
$$\text{und}\quad \sin\left(\frac{\pi}{2} \pm x\right) = \sin\frac{\pi}{2}\cos x \pm \cos\frac{\pi}{2}\sin x = 1\cdot\cos x \pm 0\cdot\sin x = \cos x$$

nach den Additionstheoremen aus Satz 9.14 (c). □

Bemerkung 9.21. Mit Satz 9.20 können wir nun aus dem Verlauf der Kosinusfunktion im Intervall $[0, \frac{\pi}{2}]$ (nach Bemerkung 9.19 (b) streng monoton fallend von 1 nach 0 verlaufend; im Bild unten dick eingezeichnet) die gesamte Kosinus- und Sinusfunktion rekonstruieren:

- Satz 9.20 (c) für $x \in [0, \frac{\pi}{2}]$ legt zunächst Kosinus und Sinus im Intervall $[0, \pi]$ fest: Die Graphen verlaufen hier „genauso“ wie beim Kosinus von 0 bis $\frac{\pi}{2}$, nur gedreht bzw. gespiegelt (im Bild unten als durchgezogene Linie markiert);
- Satz 9.14 (a) bestimmt Kosinus und Sinus damit dann auch im Intervall $[-\pi, \pi]$ (im Bild gestrichelt eingezeichnet);
- Satz 9.20 (b) schließlich besagt dann, dass dieser Verlauf von Kosinus und Sinus in beide Richtungen 2π-periodisch fortgesetzt wird (wie im Bild gepunktet eingezeichnet).





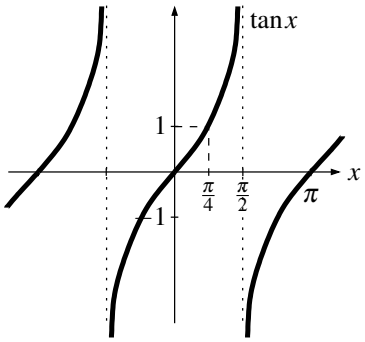
Wie in der Schule definiert man schließlich noch den Tangens:

Definition 9.22 (Tangens)**.** Für alle $x \in \mathbb{R}\backslash\{\frac{\pi}{2} + n\pi : n \in \mathbb{Z}\}$, also nach Bemerkung 9.21 für alle x mit $\cos x \neq 0$, heißt

$$\tan x := \frac{\sin x}{\cos x}$$

der **Tangens** von x.

Bemerkung 9.23 (Eigenschaften des Tangens)**.** Die wesentlichen Eigenschaften des Tangens ergeben sich unmittelbar aus denen des Kosinus und Sinus:



(a) Der Tangens ist auf seiner Definitionsmenge als Quotient stetiger Funktionen stetig.

(b) Es ist

$$\tan(-x) = \frac{\sin(-x)}{\cos(-x)} = \frac{-\sin x}{\cos x} = -\tan x$$

nach Satz 9.14 (a) (d. h. der Graph von tan ist punktsymmetrisch zum Ursprung), und

$$\tan(x+\pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{-\sin x}{-\cos x} = \tan x$$

nach Bemerkung 9.21 (d. h. tan ist periodisch mit Periode π). Der Graph der Tangensfunktion ist wegen dieser Symmetrien also bereits durch den Graphen im Intervall $[0, \frac{\pi}{2})$ bestimmt.

(c) Es ist $\tan 0 = \frac{\sin 0}{\cos 0} = \frac{0}{1} = 0$ sowie $\tan\frac{\pi}{4} = 1$, denn nach Satz 9.20 (c) ist $\sin\frac{\pi}{4} = \cos\frac{\pi}{4}$. Weiterhin ist

$$\lim_{\substack{x\to\pi/2\\ x<\pi/2}} \tan x = \infty,$$

da in diesem Grenzfall $\sin x \to 1$ gilt und $\cos x$ von der positiven Seite gegen 0 konvergiert.

(d) Die Tangensfunktion ist auf $[0, \frac{\pi}{2})$ (und damit nach der Symmetrie aus (b) auch auf $(-\frac{\pi}{2}, \frac{\pi}{2})$) streng monoton wachsend: Für $x, y \in [0, \frac{\pi}{2})$ mit $x < y$ ist $\sin x < \sin y$ und $\cos x > \cos y > 0$ nach Bemerkung 9.21, also

$$\tan x = \frac{\sin x}{\cos x} < \frac{\sin y}{\cos y} = \tan y.$$

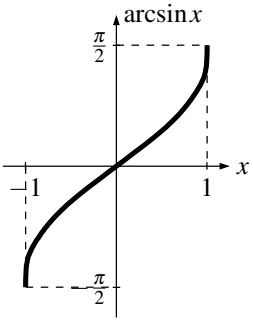
Wir können nun natürlich auf den Intervallen, auf denen die Winkelfunktionen stetig und streng monoton sind, nach Satz 8.28 die Umkehrfunktionen definieren. Aufgrund der Periodizität haben wir dabei in allen drei Fällen eine Wahl, *welches* solche Intervall wir betrachten. Üblicherweise verwendet man die folgenden:

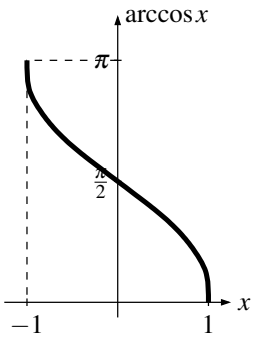
Definition 9.24 (Umkehrfunktionen der Winkelfunktionen)**.** Die Umkehrfunktion von ...

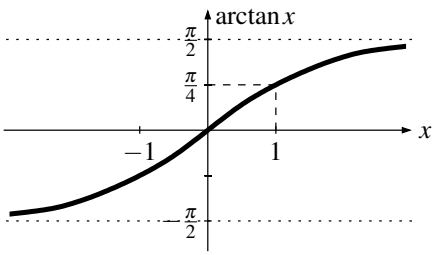
(a) $\sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ heißt **Arkussinus** $\arcsin\colon [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$.

(b) $\cos\colon [0, \pi] \to [-1, 1]$ heißt **Arkuskosinus** $\arccos\colon [-1, 1] \to [0, \pi]$.

(c) $\tan\colon \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ heißt **Arkustangens** $\arctan\colon \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.





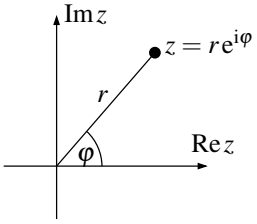


Bemerkung 9.25.

(a) Beachte, dass die Arkusfunktionen *nur in den betrachteten Intervallen* die Umkehrfunktionen der Winkelfunktionen sind; es ist also z. B. $\arcsin(\sin x) = x$ nur für $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, wohingegen z. B. $\arcsin(\sin\pi) = \arcsin 0 = 0$ ist.

(b) Wie bei den anderen bisher betrachteten Umkehrfunktionen ergeben sich die Eigenschaften der Arkusfunktionen natürlich wieder direkt aus denen der Winkelfunktionen. So sind z. B. alle Arkusfunktionen stetig und streng monoton nach Satz 8.28 (wachsend für arcsin und arctan, fallend für arccos), und die wichtigsten speziellen Werte und Symmetrien sind wie im Bild oben dargestellt.

Als Anwendung der Arkusfunktionen wollen wir zum Abschluss dieses Kapitels nun noch die sogenannte Polarkoordinatendarstellung komplexer Zahlen behandeln, mit der sich Rechnungen in $\mathbb{C}$ oft wesentlich vereinfachen lassen. Die Idee dabei ist einfach, dass man durch Multiplikation einer Zahl $\mathrm{e}^{\mathrm{i}\varphi}$ auf dem Einheitskreis (siehe Bemerkung 9.11) mit einer positiven reellen Zahl r jeden Punkt der komplexen Zahlenebene (mit Ausnahme des Nullpunkts) erreichen können sollte, wobei wie im Bild rechts r gerade der Betrag und φ der Winkel des betrachteten Punktes ist. Man kann einen solchen Punkt also auch durch Angabe der Werte von r und φ (anstatt durch Real- und Imaginärteil) charakterisieren:



Satz 9.26 (Polarkoordinatendarstellung)**.**

(a) *Jede komplexe Zahl $z \in \mathbb{C}\backslash\{0\}$ lässt sich in der Form $z = r\,\mathrm{e}^{\mathrm{i}\varphi}$ mit $r \in \mathbb{R}_{>0}$ und $\varphi \in \mathbb{R}$ schreiben.*

(b) *Die Darstellung aus (a) ist eindeutig bis auf ganzzahlige Vielfache von 2π in φ, d. h. für $r, r' \in \mathbb{R}_{>0}$ und $\varphi, \varphi' \in \mathbb{R}$ gilt $r\,\mathrm{e}^{\mathrm{i}\varphi} = r'\mathrm{e}^{\mathrm{i}\varphi'}$ genau dann wenn $r' = r$ und $\varphi' - \varphi = 2\pi n$ für ein $n \in \mathbb{Z}$.*

*Man nennt r und φ die **Polarkoordinaten** von z.*

Beweis.

(a) Wir setzen $r := |z|$. Es bleibt also noch zu zeigen, dass es ein $\varphi \in \mathbb{R}$ gibt mit $z = |z|\,\mathrm{e}^{\mathrm{i}\varphi}$, d. h. dass sich die komplexe Zahl $\frac{z}{|z|}$, die ja jetzt Betrag 1 hat, in der Form $\mathrm{e}^{\mathrm{i}\varphi}$ schreiben lässt. Aufgeteilt in Real- und Imaginärteil bedeutet dies genau, dass wir $\frac{z}{|z|} =: x + \mathrm{i}y$ als $\cos\varphi + \mathrm{i}\sin\varphi$ schreiben wollen, d. h. wir suchen zu $x, y \in \mathbb{R}$ mit $x^2 + y^2 = 1$ eine Zahl $\varphi \in \mathbb{R}$ mit

$$\cos\varphi = x \qquad \text{und} \qquad \sin\varphi = y. \tag{$*$}$$

Wegen $x^2 + y^2 = 1$ ist insbesondere $|x| \le 1$, wir können also $\alpha := \arccos x$ setzen, so dass schon einmal $\cos\alpha = x$ gilt. Nun ist mit Satz 9.14 (b)

$$y^2 = 1 - x^2 = 1 - \cos^2\alpha = \sin^2\alpha$$

und damit $y = \pm\sin\alpha$. Im Fall des Vorzeichens „+“ ergibt nun $\varphi := \alpha$, im Fall des Vorzeichens „−“ hingegen $\varphi := -\alpha$ die gewünschten Relationen $(*)$.

(b) Zunächst ist $r\,\mathrm{e}^{\mathrm{i}\varphi} = r'\,\mathrm{e}^{\mathrm{i}\varphi'}$ genau dann, wenn $r = r'$ und $\mathrm{e}^{\mathrm{i}\varphi} = \mathrm{e}^{\mathrm{i}\varphi'}$ (für die Richtung „$\Leftarrow$“ nehmen wir auf beiden Seiten den Betrag). Weiterhin ist $\mathrm{e}^{\mathrm{i}\varphi} = \mathrm{e}^{\mathrm{i}\varphi'}$, also $\mathrm{e}^{\mathrm{i}(\varphi'-\varphi)} = 1$, äquivalent zu

$$\cos(\varphi' - \varphi) = 1 \qquad \text{und} \qquad \sin(\varphi' - \varphi) = 0.$$

Aus Bemerkung 9.21 ergibt sich nun, dass dies genau dann der Fall ist, wenn $\varphi' - \varphi$ ein ganzzahliges Vielfaches von 2π ist. □

Beispiel 9.27.

(a) In Polarkoordinaten ist insbesondere die Multiplikation zweier komplexer Zahlen sehr einfach: Es ist

$$(r_1\,\mathrm{e}^{\mathrm{i}\varphi_1}) \cdot (r_2\,\mathrm{e}^{\mathrm{i}\varphi_2}) = (r_1 r_2)\,\mathrm{e}^{\mathrm{i}(\varphi_1+\varphi_2)},$$

was die algebraische Version der geometrischen Aussage aus Bemerkung 6.5 ist, dass sich bei der Multiplikation komplexer Zahlen die Beträge multiplizieren und die Winkel addieren.

(b) (Einheitswurzeln) In Polarkoordinaten können wir in $\mathbb{C}$ besonders einfach die Gleichung $z^n = 1$ für $n \in \mathbb{N}_{>0}$ lösen. Schreiben wir nämlich $z = r\,\mathrm{e}^{\mathrm{i}\varphi}$, so wollen wir also

$$z^n = r^n\,\mathrm{e}^{\mathrm{i}n\varphi} = 1 = 1\,\mathrm{e}^{\mathrm{i}\cdot 0},$$

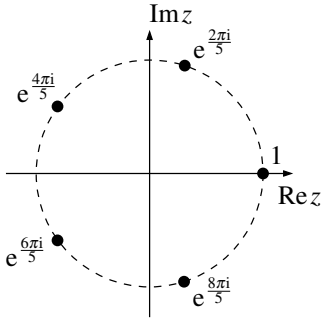
was nach der Eindeutigkeitsaussage aus Satz 9.26 (b) bedeutet, dass

$$r^n = 1 \qquad \text{und} \qquad n\varphi = 2\pi k \text{ für ein } k \in \mathbb{Z},$$

also $r = 1$ und $\varphi = \frac{2\pi k}{n}$ für ein $k \in \mathbb{Z}$. Da man die Polarkoordinaten stets so wählen kann, dass $\varphi \in [0, 2\pi)$ ist, genügt es dabei, sich auf die Werte $k \in \{0, 1, \ldots, n-1\}$ zu beschränken. Die komplexen Lösungen der Gleichung $z^n = 1$ sind also genau die n Zahlen

$$z_k = \mathrm{e}^{\frac{2\pi \mathrm{i} k}{n}} \qquad \text{für } k = 0, \ldots, n-1.$$

Man bezeichnet diese Zahlen als die n-ten **Einheitswurzeln**; sie bilden die Ecken eines regelmäßigen n-Ecks (wie im Bild rechts für das Beispiel $n = 5$ dargestellt).



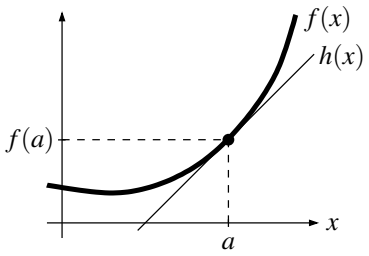
Aufgabe 9.28.

(a) Bestimme $\cos\frac{\pi}{6}$ und $\sin\frac{\pi}{6}$.

(b) Stelle alle komplexen Lösungen der Gleichung $z^6 = -8$ sowohl in Polarkoordinaten $z = r\mathrm{e}^{\mathrm{i}\varphi}$ als auch ohne Verwendung von Winkelfunktionen in der Form $z = x + \mathrm{i}y$ dar. Wo liegen sie in der komplexen Zahlenebene?

10. Differentialrechnung

Wir kommen nun zum wohl wichtigsten Teil der Analysis (in einer Veränderlichen), der sogenannten Differentialrechnung. Ziel der Differentialrechnung ist es, wie im Bild rechts eine Funktion $f\colon D \to \mathbb{K}$ in einem gegebenen Punkt $a \in D$ *linear zu approximieren*, d. h. eine Gerade h zu finden, die f in einer kleinen Umgebung von a möglichst gut annähert. Mit anderen Worten können wir h als *Tangente* an den Graphen von f im Punkt a auffassen.



In der Praxis ist dies natürlich oft wünschenswert, denn immer wenn wir aus irgendwelchen Gründen wissen, dass wir die Funktion f nur in der Nähe von a benötigen werden, dann können wir die womöglich sehr komplizierte Funktion f näherungsweise durch eine Gerade ersetzen, also durch eine viel einfacher zu behandelnde Funktion.

Wie kann man nun diese Tangente h bestimmen? Als Gerade durch den Punkt $(a, f(a))$ muss sie natürlich von der Form $h(x) = f(a) + c\,(x-a)$ für ein $c \in \mathbb{R}$ sein, wobei dieses c die Steigung der Geraden angibt. Wir möchten also erreichen, dass

$$f(x) \approx f(a) + c\,(x-a),$$

wobei das Symbol „$\approx$“ hier nicht exakt definiert ist, sondern nur den anschaulichen Sachverhalt „ist für x in der Nähe von a in etwa gleich“ beschreiben soll. Es müsste dann also

$$c \approx \frac{f(x)-f(a)}{x-a}$$

sein. Aufgrund unserer Vorarbeiten wissen wir aber natürlich nun, wie man dies mathematisch exakt formulieren muss: Die beste Näherung erhalten wir für den Grenzwert

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

(sofern er existiert). Derartige Grenzwerte wollen wir nun also in der Differentialrechnung studieren.

10.A Ableitungen von Funktionen

Bevor wir den obigen Grenzwert von $\frac{f(x)-f(a)}{x-a}$ für $x \to a$ exakt definieren können, müssen wir noch kurz eine (recht schwache) Bedingung an die Definitionsmenge D der betrachteten Funktion stellen: Da dieser Quotient nur für $x \in D \backslash \{a\}$ definiert ist, muss a nach Definition 8.3 ein Berührpunkt von $D \backslash \{a\}$ sein, damit der Grenzwert dieses Ausdrucks für $x \to a$ überhaupt definierbar ist, also damit man sich innerhalb von $D \backslash \{a\}$ dem Punkt a beliebig nähern kann. Wir wollen diese Bedingung nun formalisieren.

Definition 10.1 (Isolierte Punkte)**.** Es sei $D \subset \mathbb{K}$. Ein Punkt $a \in D$ heißt **isolierter Punkt** von D, wenn es eine ε-Umgebung von a gibt, die außer a keinen Punkt von D enthält (also „wenn man sich innerhalb von $D \backslash \{a\}$ dem Punkt a nicht beliebig nähern kann“).

Beispiel 10.2.

(a) Die Menge $\mathbb{Z}$ besteht nur aus isolierten Punkten.

(b) Intervalle in $\mathbb{R}$ – egal ob offene, halboffene, abgeschlossene oder uneigentliche – haben keine isolierten Punkte (solange sie nicht ein einpunktiges Intervall $[a,a]$ sind). Vereinigungen derartiger Intervalle haben ebenfalls keine isolierten Punkte.

Durch Negation der Bedingung aus Definition 10.1 sehen wir, dass ein Punkt $a \in D$ *kein* isolierter Punkt von D ist, wenn in jeder ε-Umgebung von a ein Punkt von $D \backslash \{a\}$ liegt, also wenn a ein Berührpunkt von $D \backslash \{a\}$ ist. Um Grenzwerte für $x \to a$ mit $x \in D$ und $x \neq a$ in jedem Punkt $a \in D$ bilden zu können, machen wir wie oben erläutert jetzt also die

Grundvoraussetzung für dieses Kapitel: Die Definitionsmengen aller betrachteten Funktionen haben keine isolierten Punkte.

Damit können wir nun wie oben motiviert die Steigungen der lokalen linearen Approximationen einer gegebenen Funktion als Grenzwerte berechnen:

Definition 10.3 (Differenzierbarkeit)**.** Es seien $D \subset \mathbb{K}$ und $f\colon D \to \mathbb{K}$ eine Funktion.

(a) Die Funktion f heißt **differenzierbar** in $a \in D$, wenn der Grenzwert

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

in $\mathbb{K}$ existiert (ein uneigentlicher Grenzwert $\pm\infty$ im reellen Fall wie in Definition 8.19 ist hier also nicht zugelassen). Die Zahl $\frac{f(x)-f(a)}{x-a}$ bezeichnet man oft als **Differenzenquotient**, ihren Grenzwert $f'(a)$ – sofern er existiert – als **Differentialquotient** oder **Ableitung** von f in a. Da es offensichtlich ist, dass wir hier bei der Grenzwertbildung $x \neq a$ beachten müssen, werden wir diese Bedingung in der Regel nicht jedesmal wieder explizit hinschreiben.

(b) Man nennt f differenzierbar (auf D), wenn f in jedem Punkt von D differenzierbar ist. Offensichtlich erhält man dann eine Funktion $f'\colon D \to \mathbb{K}$, $x \mapsto f'(x)$, die die Ableitungsfunktion (oder ebenfalls kurz Ableitung) von f genannt wird.

Wie oben erläutert ist $f'(a)$ also die Steigung der Tangenten an den Graphen von f im Punkt $(a, f(a))$.

Beispiel 10.4.

(a) Jede Gerade $f\colon \mathbb{K} \to \mathbb{K}$, $x \mapsto mx + b$ mit $m, b \in \mathbb{K}$ ist differenzierbar mit Ableitung m, denn

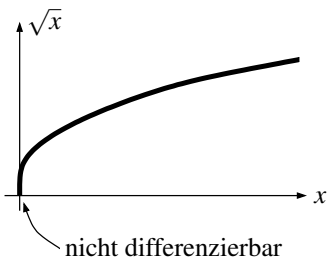
$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{(mx + b) - (ma + b)}{x - a} = m.$$

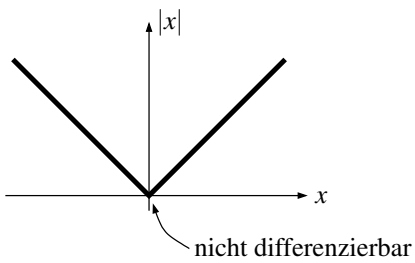
Dies ist geometrisch natürlich klar, denn eine solche Gerade ist ihre eigene Tangente in jedem Punkt und hat damit überall die Steigung m. 21

(b) Es sei $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, $x \mapsto \sqrt{x}$ die Wurzelfunktion. Dann ist für alle $a \in \mathbb{R}_{\geq 0}$

$$\lim_{x \to a} \frac{\sqrt{x} - \sqrt{a}}{x - a} = \lim_{x \to a} \frac{\sqrt{x} - \sqrt{a}}{(\sqrt{x} + \sqrt{a})(\sqrt{x} - \sqrt{a})} = \lim_{x \to a} \frac{1}{\sqrt{x} + \sqrt{a}} = \begin{cases} \frac{1}{2\sqrt{a}} & \text{für } a > 0, \\ \infty & \text{für } a = 0. \end{cases}$$

Also ist f differenzierbar auf $\mathbb{R}_{>0}$ mit Ableitung $f'(x) = \frac{1}{2\sqrt{x}}$, aber nicht differenzierbar auf $\mathbb{R}_{\geq 0}$. Anschaulich ist f nicht differenzierbar in 0, weil f dort „unendliche Steigung“ hat und die lineare Approximation daher eine senkrechte Gerade sein müsste (siehe Bild unten links).





(c) Die Betragsfunktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$ ist nicht differenzierbar in 0 nach dem Folgenkriterium, denn die Folge $(x_n)_n$ mit $x_n = \frac{(-1)^n}{n}$ konvergiert gegen 0, aber der Grenzwert

$$\lim_{n\to\infty} \frac{f(x_n)-f(0)}{x_n - 0} = \lim_{n\to\infty} \frac{|x_n|}{x_n} = \lim_{n\to\infty} \frac{1/n}{(-1)^n/n} = \lim_{n\to\infty} (-1)^n$$

existiert nicht. Anschaulich ist f deswegen nicht differenzierbar in 0, weil der Funktionsgraph dort einen „Knick“ hat und sich die Funktion daher dort nicht durch eine Gerade approximieren lässt (siehe Bild oben rechts).

Wir haben gerade mit Beispiel 10.4 (b) und (c) zwei Beispiele von Funktionen gesehen, die (in einem Punkt) stetig, aber nicht differenzierbar sind. Wir wollen nun zeigen, dass umgekehrt aber jede differenzierbare Funktion stetig ist. Hierfür benötigen wir das folgende Lemma, das wir auch später noch einmal verwenden werden.

Lemma 10.5 (Äquivalentes Kriterium für Differenzierbarkeit)**.** *Es seien $D \subset \mathbb{K}$, $f\colon D \to \mathbb{K}$ eine Funktion und $a \in D$. Dann sind die folgenden beiden Bedingungen äquivalent:*

(a) *f ist differenzierbar in a.*

(b) *Es gibt eine in a stetige Funktion $\varphi\colon D \to \mathbb{K}$ mit $f(x) - f(a) = \varphi(x)\cdot(x-a)$ für alle $x \in D$.*

In diesem Fall ist dann $\varphi(a) = f'(a)$.

Beweis. Die gegebene Bedingung an φ legt diese Funktion für alle $x \neq a$ offensichtlich fest als den Differenzenquotienten

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

Damit besagt (b) also genau, dass diese Funktion stetig nach a fortsetzbar ist. Dies ist gemäß Definition 8.5 (b) exakt dasselbe wie die Aussage, dass der Grenzwert

$$\lim_{\substack{x\to a\\ x\neq a}} \varphi(x) = \lim_{\substack{x\to a\\ x\neq a}} \frac{f(x)-f(a)}{x-a} \qquad (*)$$

existiert, also dass f in a differenzierbar ist. Ist dies der Fall, so ist der Ausdruck $(*)$ dann aber sowohl gleich der stetigen Fortsetzung $\varphi(a)$ von φ in a als auch gleich der Ableitung $f'(a)$. □

Bemerkung 10.6. Anschaulich gibt die Funktion φ aus Lemma 10.5 im Punkt x genau die Steigung der Geraden durch die Punkte $(a, f(a))$ und $(x, f(x))$ an – die im Grenzfall $x \to a$ dann zur Tangentensteigung wird.

Folgerung 10.7. *Es seien $D \subset \mathbb{K}$, $f\colon D \to \mathbb{K}$ eine Funktion und $a \in D$. Ist f differenzierbar in a, so ist f auch stetig in a.*

Beweis. Ist f differenzierbar in a, so gibt es nach Lemma 10.5 eine in a stetige Funktion $\varphi\colon D \to \mathbb{K}$ mit $f(x) = f(a) + \varphi(x)(x-a)$. Insbesondere existiert also der Grenzwert $\lim_{x\to a} \varphi(x) = \varphi(a)$, und damit nach den Grenzwertsätzen aus Satz 8.13 auch

$$\lim_{x\to a} f(x) = \lim_{x\to a} \big(f(a) + \varphi(x)(x-a)\big) = f(a) + \varphi(a)(a-a) = f(a),$$

d. h. f ist stetig in a. □

Genau wie bei unserer Untersuchung der Stetigkeit wollen wir nun zeigen, dass sich die Differenzierbarkeit von Funktionen auf Summen, Differenzen, Produkte, Quotienten, Verkettungen, Umkehrfunktionen und schließlich auch auf Potenzreihen überträgt – und auch wie man dann die Ableitungen dieser neuen Funktionen berechnet. Wir beginnen mit den vier Grundrechenarten.

Satz 10.8 (Rechenregeln für Ableitungen)**.** *Die Funktionen $f, g\colon D \to \mathbb{K}$ seien differenzierbar in $a \in D$. Dann gilt:*

(a) *$f \pm g$ ist differenzierbar in a mit Ableitung $(f \pm g)'(a) = (f' \pm g')(a)$.*

(b) ***(Produktregel)** fg ist differenzierbar in a mit $(fg)'(a) = (f'g + fg')(a)$.*

(c) ***(Quotientenregel)*** *Ist $g(a) \neq 0$, so ist $\frac{f}{g}$ differenzierbar in a mit $\left(\frac{f}{g}\right)'(a) = \left(\frac{f'g - fg'}{g^2}\right)(a)$.*

Beweis.

(a) Wir führen den Beweis hier nur für die Addition, der für die Subtraktion ist analog:

$$\begin{aligned}(f+g)'(a) &= \lim_{x\to a} \frac{f(x)+g(x)-(f(a)+g(a))}{x-a} = \lim_{x\to a}\Big(\underbrace{\frac{f(x)-f(a)}{x-a}}_{\to f'(a)} + \underbrace{\frac{g(x)-g(a)}{x-a}}_{\to g'(a)}\Big)\\ &= (f'+g')(a).\end{aligned}$$

(b) Da g in a differenzierbar, nach Folgerung 10.7 also auch stetig ist, gilt $\lim_{x\to a} g(x) = g(a)$. Damit ergibt sich nach den Rechenregeln für Grenzwerte aus Satz 8.13

$$\begin{aligned}(fg)'(a) &= \lim_{x\to a} \frac{f(x)g(x)-f(a)g(a)}{x-a}\\ &= \lim_{x\to a} \frac{f(x)g(x)-f(a)g(x)+f(a)g(x)-f(a)g(a)}{x-a}\\ &= \lim_{x\to a}\Big(\underbrace{\frac{f(x)-f(a)}{x-a}}_{\to f'(a)} \cdot \underbrace{g(x)}_{\to g(a)} + f(a)\cdot \underbrace{\frac{g(x)-g(a)}{x-a}}_{\to g'(a)}\Big)\\ &= (f'g+fg')(a).\end{aligned}$$

(c) Da g als differenzierbare Funktion nach Folgerung 10.7 auch stetig ist, ist g wegen $g(a) \neq 0$ nach Bemerkung 8.8 in einer ε-Umgebung von a nirgends 0. Damit stimmt die Definitionsmenge von $\frac{f}{g}$ dort mit D überein. Also ist a kein isolierter Punkt dieser Definitionsmenge, und wir können sinnvoll über die Ableitung von $\frac{f}{g}$ in a sprechen.

Die eigentliche Berechnung dieser Ableitung ist nun analog zu (b):

$$\begin{aligned}\left(\frac{f}{g}\right)'(a) &= \lim_{x\to a} \frac{\frac{f(x)}{g(x)} - \frac{f(a)}{g(a)}}{x-a}\\ &= \lim_{x\to a} \frac{1}{g(x)g(a)} \cdot \frac{f(x)g(a)-f(a)g(a)+f(a)g(a)-f(a)g(x)}{x-a}\\ &= \lim_{x\to a} \underbrace{\frac{1}{g(x)g(a)}}_{\to 1/(g(a))^2} \cdot \Big(\underbrace{\frac{f(x)-f(a)}{x-a}}_{\to f'(a)} \cdot g(a) - f(a)\cdot \underbrace{\frac{g(x)-g(a)}{x-a}}_{\to g'(a)}\Big)\\ &= \left(\frac{f'g-fg'}{g^2}\right)(a).\end{aligned}$$

□

Beispiel 10.9.

(a) Wir zeigen mit (auf- und absteigender) Induktion über n, dass die Ableitung der Potenzfunktion $f(x) = x^n$ gleich $f'(x) = nx^{n-1}$ für alle $n \in \mathbb{Z}$ ist. Der Induktionsanfang für $n = 0$ ergibt sich aus Beispiel 10.4 (a). Wissen wir nun für ein festes $n \in \mathbb{Z}$, dass die Ableitung von $x \mapsto x^n$ gleich $x \mapsto nx^{n-1}$ ist, so folgt mit der Produktregel für die Ableitung von $f(x) = x^{n+1} = x^n \cdot x$

$$f'(x) = nx^{n-1}\cdot x + x^n \cdot 1 = (n+1)x^n$$

(da die Ableitung der Funktion $x \mapsto x$ nach Beispiel 10.4 (a) die konstante Funktion 1 ist), und mit der Quotientenregel für die Ableitung von $f(x) = x^{n-1} = \frac{x^n}{x}$ analog

$$f'(x) = \frac{nx^{n-1}\cdot x - x^n \cdot 1}{x^2} = (n-1)x^{n-2}.$$

(b) Aus der Produktregel und Beispiel 10.4 (a) folgt insbesondere für alle $c \in \mathbb{K}$ und jede differenzierbare Funktion $f\colon D \to \mathbb{K}$, dass $(cf)' = c \cdot f'$.

(c) Mit den Regeln aus Satz 10.8 (und Beispiel 10.4 (a)) können wir nun offensichtlich die Ableitung jeder rationalen Funktion, also jeder Funktion der Form $f(x) = \frac{p(x)}{q(x)}$ mit Polynomfunktionen p und q berechnen. Ist z. B. $f(x) = \sum_{k=0}^{n} a_k x^k$ selbst eine Polynomfunktion, so ist $f'(x) = \sum_{k=1}^{n} k a_k x^{k-1}$ nach (a), (b) und Satz 10.8 (a).

Wir kommen jetzt zu Verkettungen und Umkehrfunktionen.

Satz 10.10 (Kettenregel). *Es seien $f\colon D \to \mathbb{K}$ und $g\colon D' \to \mathbb{K}$ zwei Funktionen, so dass $f(D) \subset D'$. Ist dann $a \in D$, so dass f differenzierbar in a und g differenzierbar in $f(a)$ ist, so ist auch die Verkettung $g \circ f$ differenzierbar in a, und es gilt*

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

d. h. „die Ableitung einer Verkettung ist das Produkt der beiden Ableitungen".

Beweis. Da f und g in a bzw. $f(a)$ differenzierbar sind, gibt es nach Lemma 10.5 Funktionen $\varphi\colon D \to \mathbb{K}$ und $\psi\colon D' \to \mathbb{K}$, die in a bzw. $f(a)$ stetig sind, und für die

$$\begin{aligned} f(x) - f(a) &= \varphi(x)\,(x-a) && \text{für alle } x \in D \\ \text{bzw.}\quad g(y) - g(f(a)) &= \psi(y)\,(y - f(a)) && \text{für alle } y \in D' \end{aligned}$$

sowie $\varphi(a) = f'(a)$ und $\psi(f(a)) = g'(f(a))$ gelten. Setzen wir nun $y = f(x)$, so erhalten wir durch Einsetzen der ersten Gleichung in die zweite

$$g(f(x)) - g(f(a)) = \psi(f(x))\,(f(x) - f(a)) = \psi(f(x))\,\varphi(x)\,(x-a)$$

für alle $x \in D$. Da f und φ in a sowie ψ in $f(a)$ stetig sind, ist nun aber auch $x \mapsto \psi(f(x))\,\varphi(x)$ in a stetig, und somit ergibt sich aus der Richtung „(b) $\Rightarrow$ (a)" von Lemma 10.5 angewendet auf $g \circ f$, dass diese Funktion in a differenzierbar ist mit $(g \circ f)'(a) = \psi(f(a))\,\varphi(a) = g'(f(a)) \cdot f'(a)$. □

Satz 10.11 (Ableitung der Umkehrfunktion)**.** *Es seien $D, D' \subset \mathbb{K}$ und $f\colon D \to D'$ eine bijektive Funktion mit Umkehrfunktion $f^{-1}\colon D' \to D$. Ist dann $a \in D$ ein Punkt, so dass f differenzierbar in a ist mit $f'(a) \neq 0$, und so dass f^{-1} stetig in $f(a)$ ist, so ist auch f^{-1} differenzierbar in $f(a)$ mit*

$$(f^{-1})'(f(a)) = \frac{1}{f'(a)}.$$

Beweis. Wir berechnen die Ableitung von f^{-1} in $b := f(a)$ mit dem Folgenkriterium aus Satz 8.11. Es sei dazu $(y_n)_n$ eine beliebige Folge in $D' \backslash \{b\}$ mit $y_n \to b$. Da f^{-1} nach Voraussetzung in b stetig ist, gilt dann $x_n := f^{-1}(y_n) \to f^{-1}(b) = a$. Also ist $\lim_{n\to\infty} \frac{f(x_n)-f(a)}{x_n - a} = f'(a)$, und somit nach dem Grenzwertsatz 5.13 (c) für Quotienten

$$\lim_{n\to\infty} \frac{f^{-1}(y_n) - f^{-1}(b)}{y_n - b} = \lim_{n\to\infty} \frac{x_n - a}{f(x_n) - f(a)} = \frac{1}{f'(a)},$$

woraus die Behauptung folgt. □

Bemerkung 10.12. Im Fall einer reellen, streng monotonen Funktion f benötigen wir die Voraussetzung der Stetigkeit von f^{-1} in $f(a)$ in Satz 10.11 nicht, da dies nach Satz 8.28 automatisch erfüllt ist. Die Bedingung $f'(a) \neq 0$ ist hingegen auch in diesem Fall nicht überflüssig: Das Beispiel der reellen Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$ mit $f'(0) = 0$ zeigt, dass eine differenzierbare, streng monotone Funktion in einem Punkt auch Ableitung Null haben kann.

Beispiel 10.13. In den Sätzen 10.10 und 10.11 werden nicht alle Ableitungen an derselben Stelle a, sondern manche auch an $f(a)$ ausgewertet. Man macht dies „automatisch" richtig, wenn man wie in den folgenden beiden Beispielen für die Definitions- und Wertemengen der beteiligten Funktionen bestimmte Variablennamen festlegt und darauf achtet, dass Funktionen und ihre Ableitungen immer an der entsprechenden Variablen ausgewertet werden.

(a) Die Funktion $h\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto y = \sqrt{x^2+1}$ ist von der Form $h = g \circ f$ mit $f\colon x \mapsto u = x^2+1$ und $g\colon u \mapsto y = \sqrt{u}$. Ihre Ableitung ergibt sich daher nach der Kettenregel aus Satz 10.10 zu

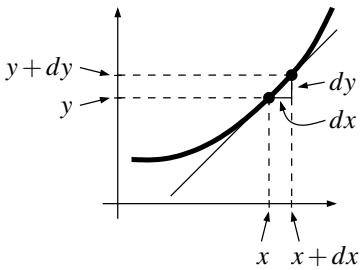
$$h'(x) = g'(u) \cdot f'(x) = \frac{1}{2\sqrt{u}} \cdot 2x = \frac{x}{\sqrt{x^2+1}},$$

denn $f'(x) = 2x$ und $g'(u) = \frac{1}{2\sqrt{u}}$ nach Beispiel 10.9 (c) und 10.4 (b).

(b) Nach Beispiel 10.9 (a) ist die Ableitung der Potenzfunktion $f\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto y = x^n$ für $n \in \mathbb{N}_{>0}$ gleich $f'(x) = nx^{n-1}$. Damit ist die Ableitung ihrer Umkehrfunktion, also der n-ten Wurzelfunktion $f^{-1}\colon y \mapsto x = \sqrt[n]{y} = y^{1/n}$, nach Satz 10.11 gleich

$$(f^{-1})'(y) = \frac{1}{f'(x)} = \frac{1}{nx^{n-1}} = \frac{1}{n\,(\sqrt[n]{y})^{n-1}} = \frac{1}{n} \cdot y^{\frac{1}{n}-1}.$$

Notation 10.14 (Differentialschreibweise)**.** Die Regeln aus den Sätzen 10.10 und 10.11 lassen sich leicht mit Hilfe der sogenannten Differentialschreibweise merken: Man legt hierzu wie in Beispiel 10.13 für eine Funktion $f\colon D \to \mathbb{K}$ bestimmte Variablennamen für Definitions- und Wertemenge fest, etwa $y = f(x)$, und schreibt die Ableitung $f'(x)$ dann als formalen Quotienten $\frac{dy}{dx}$, wobei die „Differentiale“ dx und dy wie im Bild rechts für eine (unendlich kleine) Differenz in den x- und y-Werten stehen sollen.



Wichtig dabei ist, dass dies nur eine formale Schreibweise ist – es gibt nicht wirklich Objekte dx und dy, die hier durcheinander geteilt werden. Dennoch nehmen die Sätze 10.10 und 10.11 in dieser Schreibweise eine sehr natürliche Form an, die so aussieht, als könnte man mit diesen „Brüchen“ wirklich rechnen:

(a) (Verkettung) Ist $h = g \circ f$ eine Verkettung und setzen wir $u = f(x)$, $y = g(u)$ und damit $y = h(x)$, so besagt Satz 10.10 einfach $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$, so als ob man hier mit du erweitern würde.

(b) (Umkehrfunktion) Ist $y = f(x)$, also $x = f^{-1}(y)$, so würden wir die Ableitung von f^{-1} ja als $\frac{dx}{dy}$ schreiben, und damit sagt Satz 10.11 gerade $\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}$, so als ob man hier einfach den Kehrwert des Bruches $\frac{dy}{dx}$ bilden würde.

Aufgabe 10.15. Es sei $f\colon D \to \mathbb{R}$ eine auf einem Intervall D definierte Funktion. Wir setzen voraus, dass es $b, c \in \mathbb{R}_{>0}$ gibt mit $|f(x) - f(y)| \le c\,|x-y|^b$ für alle $x, y \in D$ mit $x \neq y$. Man zeige:

(a) f ist gleichmäßig stetig.

(b) Ist $b > 1$, so ist f konstant.

Aufgabe 10.16. Es seien $f, g\colon \mathbb{R} \to \mathbb{R}$ stetige Funktionen. Man beweise oder widerlege:

(a) Ist f differenzierbar in 0 und $g(0) = 0$, dann ist $f \cdot g$ differenzierbar in 0.

(b) Ist f differenzierbar in 0 und $f(0) = 0$, dann ist $f \cdot g$ differenzierbar in 0.

(c) Sind f und g differenzierbar in 0 mit $f(0) = 0$ und $f(x)g(x) = x$ für alle $x \in \mathbb{R}$, dann gilt $g(0) \neq 0$.

Aufgabe 10.17. Es seien $D \subset \mathbb{R}$ und $f\colon D \to \mathbb{R}$ eine Funktion. Zeige, dass für jedes $a \in D$ und $c \in \mathbb{R}$ die folgenden beiden Bedingungen äquivalent sind:

(a) Die Funktion f ist differenzierbar in a mit $f'(a) = c$.

(b) Für zwei beliebige gegen a konvergente Folgen $(x_n)_n$ und $(y_n)_n$ in D mit $x_n \le a \le y_n$ und $x_n \neq y_n$ für alle $n \in \mathbb{N}$ gilt

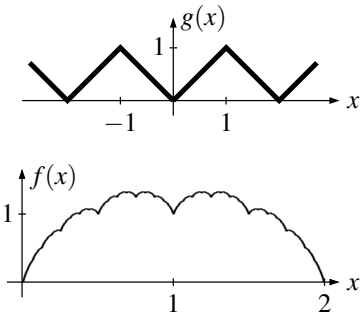
$$\lim_{n\to\infty} \frac{f(y_n) - f(x_n)}{y_n - x_n} = c.$$

Aufgabe 10.18. Wir betrachten wie im Bild rechts dargestellt die „Zickzackfunktion“ $g\colon \mathbb{R} \to \mathbb{R}$, die periodisch mit Periodenlänge 2 ist und auf $[-1,1]$ mit der Betragsfunktion übereinstimmt. Zeige, dass dann die Funktion

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \sum_{n=0}^{\infty} \frac{g(2^n x)}{2^n}$$

in jedem Punkt stetig und in keinem Punkt differenzierbar ist.

(Hinweis: Für die Differenzierbarkeit ist Aufgabe 10.17 nützlich.)



10.B Extremwerte und der Mittelwertsatz

Nachdem wir nun schon einige Ableitungen berechnen können, wollen wir uns als Nächstes anschauen, welche Informationen man über eine differenzierbare Funktion aus ihrer Ableitung erhalten kann. Am wichtigsten ist dabei, dass man mit Hilfe der Ableitung sehr leicht die Stellen finden kann, an denen eine Funktion ihre größten bzw. kleinsten Werte annimmt. Dies zu untersuchen ist natürlich nur für reelle Funktionen sinnvoll, und daher beschränken wir uns im Folgenden auf solche.

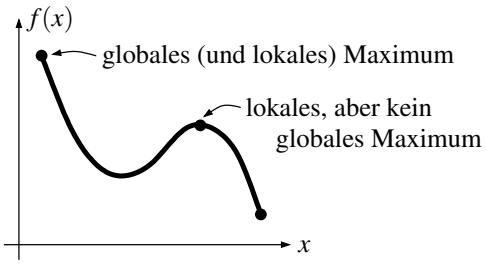
Definition 10.19 (Extrema)**.** Es seien $D \subset \mathbb{R}$, $f\colon D \to \mathbb{R}$ eine Funktion und $a \in D$. Man sagt, ...

(a) f habe in a ein **(globales) Maximum**, wenn $f(a) \geq f(x)$ für alle $x \in D$.

(b) f habe in a ein **lokales Maximum**, wenn es ein $\varepsilon > 0$ gibt, so dass $f(a) \geq f(x)$ für alle $x \in D$ mit $|x-a| < \varepsilon$. Gilt sogar $f(a) > f(x)$ für alle $x \in D$ mit $|x-a| < \varepsilon$ und $x \neq a$, so nennt man das lokale Maximum **isoliert**.

Analog definiert man globale und lokale (isolierte) **Minima**. Hat f in a ein (globales, lokales, isoliertes) Maximum oder Minimum, so sagt man auch, dass f dort ein (globales, lokales, isoliertes) **Extremum** hat.

Bemerkung 10.20. Ein globales Maximum (analog Minimum) bedeutet also gerade, dass f dort den größten aller möglichen Funktionswerte annimmt; ein lokales Maximum dagegen nur, dass f in einer kleinen Umgebung des betrachteten Punktes den größten Wert hat. Offensichtlich ist also jedes globale Maximum auch ein lokales. Das Bild rechts zeigt, dass die Umkehrung nicht notwendig richtig ist.



Wie wir im Bild schon sehen, zeichnet sich ein lokales Extremum, *das nicht am Rand des Definitionsbereichs liegt*, dadurch aus, dass die Ableitung, also die Steigung der Funktion, dort gleich 0 ist:

Lemma 10.21 (Notwendige Bedingung für Extrema)**.** *Hat die Funktion $f\colon (a,b) \to \mathbb{R}$ in einem Punkt $x \in (a,b)$ ein lokales Extremum und ist f in diesem Punkt differenzierbar, so gilt $f'(x) = 0$.*

Beweis. Wir beweisen das Lemma für ein Maximum; der Beweis für ein Minimum ist natürlich analog. Nach eventuellem Verkleinern des Definitionsintervalls (a,b) können wir weiterhin ohne Beschränkung der Allgemeinheit annehmen, dass f in x sogar ein globales Maximum hat. Wähle nun eine Folge $(x_n)_n$ in (a,b) mit $x_n \to x$ und $x_n > x$ für alle $n \in \mathbb{N}$ – also eine Folge in D, die sich dem Punkt x von rechts nähert. Die Ableitung $f'(x)$, die nach Voraussetzung existiert, können wir dann nach dem Folgenkriterium aus Satz 8.11 als

$$f'(x) = \lim_{n\to\infty} \frac{f(x_n) - f(x)}{x_n - x}$$

berechnen. Nun ist der Zähler dieses Bruches immer kleiner oder gleich 0 (weil f in x ein Maximum hat), und der Nenner immer größer als Null – und damit folgt $f'(x) \leq 0$ nach Satz 5.23 (a). Durch eine Folge, die sich von links dem Punkt x nähert, erhält man genauso $f'(x) \geq 0$, und damit letztendlich $f'(x) = 0$. □

22

Bemerkung 10.22. Hat eine reelle Funktion $f\colon [a,b] \to \mathbb{R}$ auf einem abgeschlossenen Intervall in einem Punkt $x \in [a,b]$ ein lokales Extremum, so gibt es also zwei Möglichkeiten:

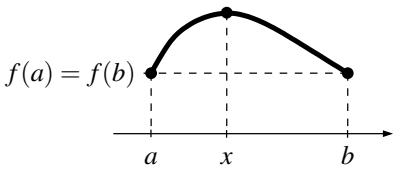
(a) $x \in (a,b)$: Nach Lemma 10.21 muss dann $f'(x) = 0$ sein, falls f dort differenzierbar ist. Beachte aber, dass die Bedingung $f'(x) = 0$ nicht hinreichend dafür ist, dass in x ein lokales Extremum vorliegt – dies zeigt das Beispiel der Funktion $f(x) = x^3$, für die zwar $f'(0) = 0$ gilt, die bei $x = 0$ aber kein Extremum hat. Punkte $x \in (a,b)$, für die $f'(x) = 0$ gilt, die also als lokales Extremum im Inneren des Definitionsintervalls in Frage kommen, werden oft **kritische Punkte** genannt. Wir werden später noch sehen, wie man feststellen kann, ob ein kritischer Punkt wirklich ein lokales Extremum ist oder nicht (siehe Bemerkung 10.26 und Satz 11.19).

(b) $x = a$ oder $x = b$: In diesem Fall muss die Ableitung von f in x nicht notwendig 0 sein (wie z. B. beim globalen Maximum der Funktion in Bemerkung 10.20). Solche Extremwerte am Rand des Definitionsintervalls nennt man **Randextrema**.

Als Nächstes wollen wir mit Hilfe der Ableitung einer reellen Funktion ihre Monotonieeigenschaften untersuchen. Wir beweisen dazu zunächst zwei einfache Resultate, die wir auch noch für spätere Anwendungen benötigen werden.

Satz 10.23 (Satz von Rolle). *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion, die auf (a,b) differenzierbar ist. Gilt dann $f(a) = f(b)$, so gibt es ein $x \in (a,b)$ mit $f'(x) = 0$.*

Beweis. Da f stetig ist, nimmt f nach Satz 8.25 auf dem Intervall $[a,b]$ Maximum und Minimum an. Sind diese beide gleich $f(a) = f(b)$, so ist f offensichtlich konstant und wir können ein beliebiges $x \in (a,b)$ wählen. Andernfalls können wir ohne Beschränkung der Allgemeinheit annehmen, dass das Maximum von f größer als $f(a) = f(b)$ ist, also im Inneren des Definitionsintervalls angenommen wird. Dort gilt dann aber $f'(x) = 0$ nach Lemma 10.21. □

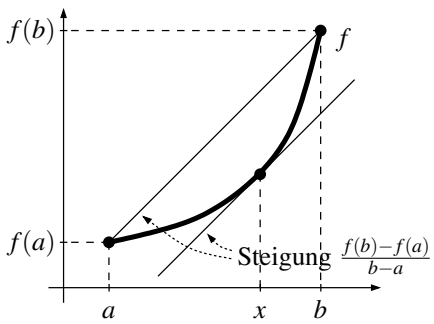


Satz 10.24 (Mittelwertsatz).

(a) *(1. Version) Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion, die auf (a,b) differenzierbar ist. Dann gibt es ein $x \in (a,b)$ mit $f'(x) = \frac{f(b)-f(a)}{b-a}$. Mit anderen Worten wird die Steigung der Geraden zwischen $(a, f(a))$ und $(b, f(b))$ also wie im Bild rechts an einer Stelle $x \in (a,b)$ als Tangentensteigung angenommen.*

(b) *(2. Version) Es seien $f, g\colon [a,b] \to \mathbb{R}$ stetige, auf (a,b) differenzierbare Funktionen. Dann gibt es ein $x \in (a,b)$ mit*

$$f'(x) \cdot (g(b) - g(a)) = g'(x) \cdot (f(b) - f(a)).$$



Beweis. Es genügt, die allgemeinere Aussage (b) zu zeigen, da sich Teil (a) sofort daraus ergibt, wenn man $g(x) = x$ setzt. Wir betrachten dazu die Funktion

$$h\colon [a,b] \to \mathbb{R},\ x \mapsto (f(x) - f(a))\,(g(b) - g(a)) - (g(x) - g(a))\,(f(b) - f(a)).$$

Mit f und g ist auch h auf $[a,b]$ stetig und auf (a,b) differenzierbar; außerdem ist $h(a) = h(b) = 0$. Nach dem Satz 10.23 von Rolle gibt es also ein $x \in (a,b)$ mit $h'(x) = 0$, und wegen

$$h'(x) = f'(x)\,(g(b) - g(a)) - g'(x)\,(f(b) - f(a))$$

ergibt sich daraus genau die Behauptung. □

Der Mittelwertsatz wirkt auf den ersten Blick etwas unscheinbar, ist in der Tat aber sehr wichtig, da er es erlaubt, einen Differenzenquotienten (und damit letztlich die Differenz zweier Funktionswerte) durch einen Differentialquotienten (also eine Ableitung) auszudrücken. Wie bereits angekündigt ist ein erstes Beispiel hierfür, dass man die Monotonie von Funktionen mit Hilfe von Ableitungen untersuchen kann.

Folgerung 10.25 (Monotonie differenzierbarer Funktionen)**.** *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige, auf (a,b) differenzierbare Funktion. Gilt dann für alle $x \in (a,b)$ ...*

(a) $f'(x) \geq 0$ *(bzw. $f'(x) > 0$), so ist f monoton (bzw. streng monoton) wachsend.*

(b) $f'(x) \leq 0$ *(bzw. $f'(x) < 0$), so ist f monoton (bzw. streng monoton) fallend.*

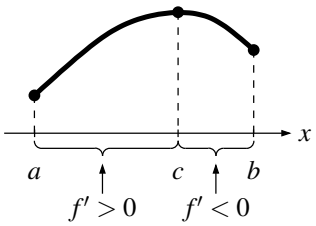
(c) $f'(x) = 0$*, so ist f konstant.*

Beweis. Es seien $x, y \in [a,b]$ mit $x < y$. Nach dem Mittelwertsatz 10.24 (a) angewendet auf das Intervall $[x,y]$ gibt es dann ein $z \in (x,y) \subset (a,b)$ mit $f(y) - f(x) = f'(z)\,(y-x)$. Im Fall (a) ist nun $f'(z) \geq 0$ bzw. $f'(z) > 0$, und damit $f(y) - f(x) \geq 0$ bzw. $f(y) - f(x) > 0$, d. h. f ist monoton (bzw. streng monoton) wachsend. Teil (b) ergibt sich natürlich genauso, und (c) folgt aus der Kombination der beiden Teile. □

Bemerkung 10.26 (Hinreichendes Kriterium für Extrema)**.** Mit Folgerung 10.25 ergibt sich ein einfaches *hinreichendes* Kriterium für ein (lokales) Extremum: Ist $f\colon [a,b] \to \mathbb{R}$ stetig und auf (a,b) differenzierbar, und gibt es ein $c \in (a,b)$ mit

$$f'(x) > 0 \text{ für alle } x < c \text{ und } f'(x) < 0 \text{ für alle } x > c$$

(d. h. hat f' einen Vorzeichenwechsel von $+$ nach $-$ in c), so ist f nach Folgerung 10.25 streng monoton wachsend auf $[a,c]$ und streng monoton fallend auf $[c,b]$, d. h. f hat ein isoliertes lokales Maximum in c. Eine analoge Aussage gilt natürlich auch für ein Minimum.



Wir sehen also, dass man mit Hilfe der Ableitung gut die Extrema von differenzierbaren Funktionen finden kann. Um dies in der Praxis auch anwenden zu können, müssen wir aber auch noch in der Lage sein, von komplizierteren Funktionen – z. B. den „speziellen Funktionen“ aus Kapitel 9 – die Ableitung zu berechnen oder überhaupt erst einmal ihre Differenzierbarkeit nachzuweisen. Da diese Funktionen oftmals über Potenzreihen definiert sind, müssen wir uns also mit der Differenzierbarkeit solcher Potenzreihen (oder allgemeiner von Funktionenfolgen) beschäftigen. Entscheidend hierfür ist die folgende Aussage, die ebenfalls zentral den Mittelwertsatz verwendet.

Satz 10.27 (Vertauschbarkeit von Differentiation und Grenzwertbildung)**.** *Es seien $D \subset \mathbb{R}$ ein Intervall und $(f_n)_n$ mit $f_n\colon D \to \mathbb{R}$ eine Folge differenzierbarer Funktionen. Wir setzen voraus, dass*

- *$(f_n)_n$ punktweise gegen eine Grenzfunktion $f\colon D \to \mathbb{R}$ konvergiert, und*
- *die Ableitungen f_n' stetig sind und gleichmäßig gegen eine Grenzfunktion $g\colon D \to \mathbb{R}$ konvergieren.*

Dann ist f differenzierbar mit $f' = g$ (d. h. „Differentiation und Grenzwertbildung können vertauscht werden“; es ist $(\lim_{n\to\infty} f_n)' = \lim_{n\to\infty} f_n'$).

Beweis. Für alle $a \in D$ zeigen wir direkt mit der Grenzwertdefinition, dass $\lim_{x\to a} \frac{f(x)-f(a)}{x-a} = g(a)$. Es sei also $\varepsilon > 0$ beliebig. Da g nach Satz 8.39 als gleichmäßiger Grenzwert stetiger Funktionen

stetig ist, gibt es zunächst ein $\delta > 0$, so dass

$$|g(x) - g(a)| < \frac{\varepsilon}{3} \qquad \text{für alle } x \in D \text{ mit } |x-a| < \delta. \tag{1}$$

Außerdem konvergiert $(f_n')_n$ nach Voraussetzung gleichmäßig gegen g, d. h. es gibt auch ein (von x unabhängiges) $n_0 \in \mathbb{N}$ mit

$$|f_n'(x) - g(x)| < \frac{\varepsilon}{3} \qquad \text{für alle } x \in D \text{ und } n \geq n_0. \tag{2}$$

Es seien nun $x \in D$ mit $x \neq a$ und $|x-a| < \delta$ sowie $n \geq n_0$ beliebig. Nach dem Mittelwertsatz 10.24 (a) gibt es dann ein z zwischen a und x (für das also insbesondere auch $|z-a| < |x-a| < \delta$ gilt) mit

$$\frac{f_n(x) - f_n(a)}{x-a} = f_n'(z). \tag{3}$$

Setzen wir dies nun alles zusammen, so erhalten wir mit der Dreiecksungleichung

$$\begin{aligned} \left| \frac{f_n(x) - f_n(a)}{x-a} - g(a) \right| &\overset{(3)}{=} |f_n'(z) - g(a)| = |f_n'(z) - g(z) + g(z) - g(a)| \\ &\leq |f_n'(z) - g(z)| + |g(z) - g(a)| \\ &\overset{(1),(2)}{<} \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \frac{2\varepsilon}{3}, \end{aligned}$$

wobei wir (1) und (2) für den Punkt z angewendet haben. Nehmen wir hier nun den Grenzwert für $n \to \infty$, so erhalten wir daraus mit Satz 5.23 (a) für alle x mit $|x-a| < \delta$

$$\left| \frac{f(x) - f(a)}{x-a} - g(a) \right| \leq \frac{2\varepsilon}{3} < \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, bedeutet dies aber genau, dass $\lim_{x \to a} \frac{f(x)-f(a)}{x-a} = g(a)$. □

Auch wenn der Beweis dieses Satzes recht kompliziert war, ist die Aussage doch sehr einfach anzuwenden. So ergibt sich z. B. in dem für uns wichtigsten Fall von Potenzreihen:

Folgerung 10.28 (Differenzierbarkeit von Potenzreihen)**.** *Es sei $f(x) = \sum_{k=0}^{\infty} a_k x^k$ eine reelle Potenzreihe mit Konvergenzradius r. Dann ist f im Konvergenzgebiet $(-r,r)$ differenzierbar mit Ableitung $f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$, d. h. „Potenzreihen können gliedweise differenziert werden“.*

Beweis. Es sei $c \in (-r,r)$; wir wollen zeigen, dass f in c differenzierbar ist mit Ableitung $f'(c) = \sum_{k=1}^{\infty} k a_k c^{k-1}$. Wähle dazu ein R mit $|c| < R < r$. Dann ist die Folge der Partialsummen $f_n(x) = \sum_{k=0}^{n} a_k x^k$ nach Satz 8.37 auf $(-R,R)$ gleichmäßig konvergent gegen f. Die Ableitung dieser Partialsummenfunktionen sind nach Beispiel 10.9 (c) die stetigen Funktionen $f_n'(x) = \sum_{k=1}^{n} k a_k x^{k-1}$. Da die Reihe $g(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}$ nach Aufgabe 7.31 den gleichen Konvergenzradius r wie f hat, konvergieren genauso auch die f_n' auf $(-R,R)$ gleichmäßig gegen g.

Damit haben wir alle Voraussetzungen überprüft, um Satz 10.27 auf die Folge $(f_n)_n$ auf $(-R,R)$ anwenden zu können. Der Satz liefert uns also $f' = g$ auf $(-R,R)$, und damit insbesondere auch im Punkt c. □

Bemerkung 10.29. Man kann zeigen, dass die Aussage von Satz 10.27 (und damit auch von Folgerung 10.28) genauso auch im komplexen Fall gilt. Da wir in unserem Beweis dieser Aussagen den Mittelwertsatz verwendet haben (der nur in $\mathbb{R}$ gilt), benötigt man hierfür jedoch andere Argumente. Wir werden den komplexen Fall im Folgenden in dieser Vorlesung aber nicht benötigen – die Untersuchung komplex differenzierbarer Funktionen bzw. Potenzreihen ist der wesentliche Inhalt der Vorlesung „Einführung in die Funktionentheorie“, die ihr im zweiten Studienjahr hören könnt.

Mit Folgerung 10.28 (und unseren vorherigen Resultaten) können wir jetzt endlich von „praktisch allen“ Funktionen die Ableitungen berechnen:

Beispiel 10.30 (Ableitungen spezieller Funktionen)**.**

(a) Die Ableitung der (reellen) Exponentialfunktion $f(x) = \mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ ergibt sich durch gliedweises Differenzieren zu

$$f'(x) = \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n!} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \mathrm{e}^x;$$

die Exponentialfunktion ist also gleich ihrer eigenen Ableitung.

(b) Die Ableitung der Sinusfunktion $f(x) = \sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$ ist analog

$$f'(x) = \sum_{n=0}^{\infty} (2n+1) \cdot \frac{(-1)^n}{(2n+1)!} x^{2n} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = \cos x.$$

Genauso berechnet man $\cos'(x) = -\sin x$. Aus der Quotientenregel von Satz 10.8 (c) ergibt sich damit

$$\tan'(x) = \left(\frac{\sin}{\cos}\right)'(x) = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = 1 + \tan^2 x.$$

(c) Die Ableitungen der Umkehrfunktionen zu (a) und (b) folgen nun sofort aus Satz 10.11: Mit $f\colon x \mapsto y = \mathrm{e}^x$, also $f'(x) = \mathrm{e}^x$ und $f^{-1}(y) = \log y$ ist z. B.

$$\log'(y) = \frac{1}{f'(x)} = \frac{1}{\mathrm{e}^x} = \frac{1}{y}.$$

Für $f\colon x \mapsto y = \sin x$, also $f'(x) = \cos x$ und $f^{-1}(y) = \arcsin y$ ist analog

$$\arcsin'(y) = \frac{1}{f'(x)} = \frac{1}{\cos x} \overset{(*)}{=} \frac{1}{\sqrt{1-\sin^2 x}} = \frac{1}{\sqrt{1-y^2}},$$

wobei wir in $(*)$ die Gleichung aus Satz 9.14 (b) benutzen (sowie dass der Arkussinus streng monoton wachsend ist, so dass wir hier das positive Vorzeichen der Wurzel nehmen müssen). Genauso zeigt man

$$\arccos'(y) = -\frac{1}{\sqrt{1-y^2}} \qquad \text{und} \qquad \arctan'(y) = \frac{1}{1+y^2}.$$

(d) Die Ableitung der Potenzfunktion $f(x) = x^a = \mathrm{e}^{a \log x}$ (mit festem Exponenten a und variabler Basis x) ist nach der Kettenregel aus Satz 10.10

$$f'(x) = \mathrm{e}^{a \log x} \cdot \frac{a}{x} = x^a \cdot \frac{a}{x} = a x^{a-1},$$

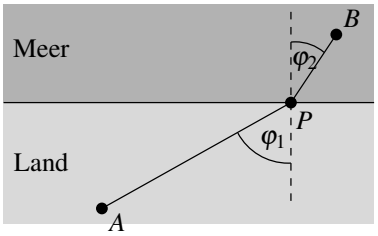
das Ergebnis ist also für alle $a \in \mathbb{R}$ das gleiche wie schon für die speziellen Exponenten in Beispiel 10.9 (a) und 10.13 (b). Die Ableitung der Potenzfunktion $f(x) = a^x = \mathrm{e}^{x \log a}$ mit fester Basis und variablem Exponenten ist hingegen ebenfalls nach der Kettenregel

$$f'(x) = \mathrm{e}^{x \log a} \cdot \log a = \log a \cdot a^x.$$

Bemerkung 10.31. Ohne die Voraussetzung der (gleichmäßigen) Konvergenz der Folge der Ableitungen in Satz 10.27 wäre die Aussage im Allgemeinen falsch: Betrachten wir z. B. die Folge $f_n\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \frac{1}{n}\sin(nx)$, so gilt zwar $\lim_{n\to\infty} f_n(x) = 0$ für alle x, d. h. (f_n) konvergiert punktweise (und in der Tat auch gleichmäßig) gegen die Nullfunktion, die ja Ableitung 0 hat – aber die Folge der Ableitungen $f_n'(x) = \cos(nx)$ konvergiert für $n \to \infty$ noch nicht einmal punktweise! Hier können Differentiation und Grenzwertbildung also nicht vertauscht werden.

Aufgabe 10.32. Bestimme alle lokalen und globalen Extrema der Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \frac{\mathrm{e}^x}{1+2|x|}$.

Aufgabe 10.33. Ein Rettungsschwimmer, der sich an Land am Punkt A befindet, möchte eine im Meer ertrinkende Person am Punkt B retten. Er läuft dazu zunächst entlang einer geraden Linie zu einem Punkt P am Ufer, und schwimmt von dort wieder entlang einer geraden Linie nach B. Wenn er mit der Geschwindigkeit v_1 laufen und mit der Geschwindigkeit v_2 schwimmen kann, wo muss er dann den Punkt P wählen, damit er möglichst schnell bei B ist? Zeige, dass diese minimale Zeit genau dort erreicht wird, wo

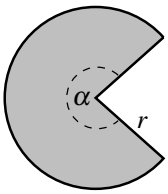


$$\frac{\sin\varphi_1}{\sin\varphi_2} = \frac{v_1}{v_2}.$$

Für die Physiker und physikalisch Interessierten unter euch: Dies ist übrigens genau das Brechungsgesetz für Licht – auch Licht bewegt sich so, dass es schnellstmöglich ans Ziel kommt!

Aufgabe 10.34. Wir betrachten wie im Bild unten rechts Kreissektoren mit variablem Öffnungswinkel $\alpha \in (0, 2\pi)$ und Radius $r \in \mathbb{R}_{>0}$.

Welcher solche Kreissektor hat bei vorgegebenem Flächeninhalt F den kleinstmöglichen Umfang U? Bestimme für diesen Fall r, α und U in Abhängigkeit von F.



(Der Umfang beinhaltet dabei auch die beiden Geradenstücke zum Mittelpunkt. Die Formeln für den Flächeninhalt und Umfang eines Kreissektors können als bekannt vorausgesetzt werden.)

Aufgabe 10.35. Untersuche die Funktionenfolge $(f_n)_n$ mit $f_n\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, $x \mapsto x^{n+1}\mathrm{e}^{-nx}$ auf gleichmäßige Konvergenz.

Aufgabe 10.36. Finde $m \in \mathbb{N}$ und $n \in \mathbb{N}_{>0}$, so dass $f\colon [-1,1] \to \mathbb{R}$, $x \mapsto \begin{cases} x^m \sin\frac{1}{x^n} & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases}$ ...

(a) unstetig ist.

(b) stetig, aber nicht differenzierbar ist.

(c) differenzierbar mit unbeschränkter Ableitung ist.

(d) differenzierbar mit beschränkter, aber unstetiger Ableitung ist.

(e) differenzierbar mit stetiger Ableitung ist.

Skizziere für kleine Werte von m und n auch die Graphen dieser Funktionen!

Aufgabe 10.37. Es sei $f\colon [a,b] \to \mathbb{R}$ eine differenzierbare Funktion. Man zeige:

(a) Ist $f'(a) > 0$ und $f'(b) < 0$, so gibt es ein $x \in (a,b)$ mit $f'(x) = 0$.

(b) Für alle c zwischen $f'(a)$ und $f'(b)$ gibt es ein $x \in [a,b]$ mit $f'(x) = c$. (Ableitungen erfüllen also den Zwischenwertsatz, obwohl sie nach Aufgabe 10.36 (d) nicht stetig sein müssen.)

23

11. Anwendungen der Differentialrechnung

Im letzten Kapitel haben wir gesehen, wie wir von „nahezu allen“ Funktionen ihre Ableitung berechnen können und welche elementaren Eigenschaften der Funktion man daran ablesen kann. In diesem Kapitel wollen wir nun zwei weitere Anwendungen vorstellen, die sich aus der Differentialrechnung ergeben. Der Einfachheit halber beschränken wir uns dabei auf reelle Funktionen.

11.A Die Regel von de l'Hôpital

Als Erstes wollen wir eine einfache Regel vorstellen, mit der man oft Grenzwerte berechnen kann, die anders nur schwer zu bestimmen wären: nämlich Grenzwerte der Form $\lim_{x\to a}\frac{f(x)}{g(x)}$, bei denen die normalen Grenzwertsätze aus Satz 8.13 bzw. Bemerkung 8.20 nicht anwendbar sind, weil sich die unbestimmten Quotienten „$\frac{0}{0}$“ oder „$\frac{\pm\infty}{\pm\infty}$“ ergeben würden.

Es gibt viele Varianten dieser Regel, je nachdem, bei welchen der im folgenden Satz vorkommenden Grenzwerten auch uneigentliche Grenzwerte $\pm\infty$ zugelassen sind. In der Praxis treten alle diese Varianten auch oft auf. Um die Beweisidee des Satzes klar herauszustellen, beschränken wir uns hier aber zunächst (wie auch bei unseren bisherigen Beweisen von Rechenregeln für Grenzwerte) auf den Fall, in dem keine uneigentlichen Grenzwerte vorkommen, und geben die möglichen Verallgemeinerungen dann in der anschließenden Bemerkung 11.2 an.

Satz 11.1 (**Regel von de l'Hôpital**, Grundversion)**.** *Es seien $f,g\colon (a,b)\to\mathbb{R}$ differenzierbare Funktionen mit $g'(x)\neq 0$ für alle $x\in(a,b)$. Wir nehmen ferner an, dass*

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$$

(so dass der Grenzwert $\lim_{x\to a}\frac{f(x)}{g(x)}$ also formal von der unbestimmten Form „$\frac{0}{0}$“ ist). Existiert dann der Grenzwert $\lim_{x\to a}\frac{f'(x)}{g'(x)}$ in $\mathbb{R}$, so auch $\lim_{x\to a}\frac{f(x)}{g(x)}$, und es gilt

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

Beweis. Wegen $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$ können wir f und g durch $f(a) := g(a) := 0$ stetig nach $[a,b)$ fortsetzen. Beachte außerdem, dass g auf (a,b) nirgends gleich 0 sein kann, denn sonst gäbe es im Widerspruch zur Voraussetzung nach dem Satz 10.23 von Rolle zwischen a und dieser Nullstelle von g eine Nullstelle von g'.

Wir zeigen nun den Grenzwert $\lim_{x\to a}\frac{f(x)}{g(x)}$ mit dem Folgenkriterium. Es sei also $(x_n)_n$ eine Folge in (a,b) mit $x_n\to a$. Nach dem Mittelwertsatz 10.24 (b) gibt es dann für alle $n\in\mathbb{N}$ ein $c_n\in(a,x_n)$ mit

$$f'(c_n)\cdot(g(x_n)-g(a)) = g'(c_n)\cdot(f(x_n)-f(a)), \quad \text{also} \quad \frac{f(x_n)}{g(x_n)} = \frac{f'(c_n)}{g'(c_n)}$$

wegen $f(a)=g(a)=0$ und der Nullstellenfreiheit von g und g'. Nun konvergiert wegen $a<c_n<x_n$ mit $(x_n)_n$ aber auch $(c_n)_n$ gegen a, und damit ergibt sich

$$\lim_{n\to\infty}\frac{f(x_n)}{g(x_n)} = \lim_{n\to\infty}\frac{f'(c_n)}{g'(c_n)} = \lim_{x\to a}\frac{f'(x)}{g'(x)},$$

was die Behauptung mit dem Folgenkriterium zeigt. □

Bemerkung 11.2 (**Regel von de l'Hôpital**, Varianten)**.** Der Satz 11.1 von de l'Hôpital hat die folgenden Varianten, die wir im Folgenden ebenfalls verwenden werden. Die Beweise sollen hier nicht gegeben werden – sie lassen sich mit analogen, allerdings oft technisch etwas aufwändigeren Methoden führen.

(a) Die Regel gilt auch, wenn $\lim_{x\to a} f(x)$ und $\lim_{x\to a} g(x)$ beide im uneigentlichen Sinne gleich $\pm\infty$ sind, wir also beim Grenzwert $\lim_{x\to a} \frac{f(x)}{g(x)}$ den unbestimmten Ausdruck „$\frac{\pm\infty}{\pm\infty}$“ haben.

(b) Statt einer Annäherung von rechts an die Grenze der Definitionsmenge (in unserem Fall also an den Punkt a am Rand des Intervalls (a,b)) ist natürlich auch eine Annäherung von links oder eine beidseitige Annäherung möglich. Dabei sind die Fälle $a = -\infty$ und $b = \infty$ zugelassen.

(c) Die Regel gilt auch, wenn der Grenzwert von $\frac{f'(x)}{g'(x)}$ nur im uneigentlichen Sinne existiert, also gleich $\pm\infty$ ist.

Existiert der Grenzwert von $\frac{f'(x)}{g'(x)}$ dagegen auch im uneigentlichen Sinne nicht, so macht die Regel von de l'Hôpital keine Aussage – wir können daraus dann also nicht schließen, dass auch der ursprüngliche Grenzwert von $\frac{f(x)}{g(x)}$ nicht existiert!

Beispiel 11.3.

(a) Für jedes $n \in \mathbb{N}_{>0}$ ist der Grenzwert $\lim_{x\to\infty} \frac{\log x}{x^n}$ von der Form „$\frac{\infty}{\infty}$“. Um ihn mit der Regel von de l'Hôpital (mit $f(x) = \log x$ und $g(x) = x^n$) zu bestimmen, differenzieren wir also Zähler und Nenner separat und erhalten den Bruch $\frac{f'(x)}{g'(x)} = \frac{1/x}{nx^{n-1}}$. Da der Nenner dieses Bruchs für $x > 0$ nirgends gleich 0 ist und der Grenzwert

$$\lim_{x\to\infty} \frac{1/x}{nx^{n-1}} = \lim_{x\to\infty} \frac{1}{nx^n} = 0$$

existiert, folgt mit Satz 11.1 (bzw. der Verallgemeinerung aus Bemerkung 11.2) also auch

$$\lim_{x\to\infty} \frac{\log x}{x^n} \overset{\text{„}\frac{\infty}{\infty}\text{“}}{\overset{\downarrow}{=}} \lim_{x\to\infty} \frac{1/x}{nx^{n-1}} = 0. \qquad (*)$$

Analog zu Bemerkung 5.15 zur Anwendung von Grenzwertsätzen schreibt man dabei die Anwendung der Regel von de l'Hôpital oftmals gleich wie in der oben in $(*)$ mit „$\frac{\infty}{\infty}$“ bezeichneten Gleichung, und überprüft erst nachträglich, dass der neu entstandene Bruch einen (evtl. uneigentlichen) Grenzwert hat und sein Nenner stets ungleich 0 ist.

(b) Analog erhält man

$$\lim_{x\to 0} x^n \log x = \lim_{x\to 0} \frac{\log x}{x^{-n}} \overset{\text{„}\frac{-\infty}{\infty}\text{“}}{\overset{\downarrow}{=}} \lim_{x\to 0} \frac{1/x}{-nx^{-n-1}} = \lim_{x\to 0} -\frac{x^n}{n} = 0,$$

da der nach dem Differenzieren entstandene Nenner $-nx^{-n-1}$ für $x > 0$ ungleich 0 ist. Zusammen mit (a) sehen wir in diesem Sinne also, dass „der Logarithmus für $x \to 0$ oder $x \to \infty$ schwächer ist als jede Potenz“ – in den beiden betrachteten Grenzwerten setzt sich jeweils die Funktion x^n durch. Dies ist natürlich ganz analog zu der Aussage von Bemerkung 9.3 (a), dass die Exponentialfunktion schneller als jede Potenz wächst. Beachte auch, dass wir in der zweiten Rechnung oben gesehen haben, dass es sich auch bei einem ursprünglichen Ausdruck der Form „$0 \cdot (\pm\infty)$“ lohnen kann, ihn künstlich als Bruch umzuschreiben, um dann die Regel von de l'Hôpital anwenden zu können.

(c) Falls sich nach einmaliger Anwendung von Satz 11.1 immer noch ein Bruch der Form „$\frac{0}{0}$“ oder „$\frac{\pm\infty}{\pm\infty}$“ ergibt, kann man den Satz natürlich auch mehrfach hintereinander anwenden, wie z. B. in dem Grenzwert

$$\lim_{x\to\infty} \frac{x^2}{e^x} \overset{\text{„}\frac{\infty}{\infty}\text{“}}{\overset{\downarrow}{=}} \lim_{x\to\infty} \frac{2x}{e^x} \overset{\text{„}\frac{\infty}{\infty}\text{“}}{\overset{\downarrow}{=}} \lim_{x\to\infty} \frac{2}{e^x} = 0$$

(wegen $e^x \neq 0$ für alle x), den wir aber natürlich auch schon aus Satz 9.1 (c) kannten.

Bemerkung 11.4 (Die Regel von de l'Hôpital für Folgengrenzwerte)**.** Manchmal kann man auch den Grenzwert einer reellen Folge $(a_n)_n$ mit der Regel von de l'Hôpital berechnen. Kann man die Folge nämlich – betrachtet als Abbildung $\mathbb{N} \to \mathbb{R}$ – zu einer Abbildung $\mathbb{R} \to \mathbb{R}$ fortsetzen, also auch

für reelles n betrachten, und existiert dann der Grenzwert für *reelle* $n \to \infty$, so existiert er dann natürlich auch für *natürliche* $n \to \infty$, und hat denselben Wert. Für die Berechnung des Grenzwerts für reelle n haben wir dann aber wieder die Regel von de l'Hôpital zur Verfügung.

Betrachten wir als Beispiel hierfür einmal für gegebenes $x \in \mathbb{R}$ den Grenzwert $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$, den wir in Aufgabe 7.33 mit viel Aufwand zu e^x berechnet haben. Da wir Potenzen inzwischen auch für reelle n definiert haben, können wir den Ausdruck $\left(1+\frac{x}{n}\right)^n$ nun aber auch als Funktion einer reellen Variablen n auffassen und seinen Grenzwert für $n \to \infty$ mit der Regel von de l'Hôpital berechnen: Es ist

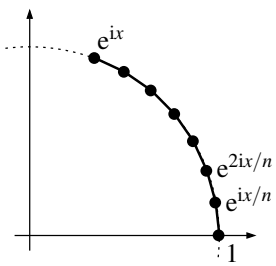
$$\begin{aligned}
\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n &= \lim_{n\to\infty}\exp\left(n\log\left(1+\frac{x}{n}\right)\right) && \text{(Definition 9.7)}\\
&= \exp\left(\lim_{n\to\infty} n\log\left(1+\frac{x}{n}\right)\right) && \text{(Stetigkeit von exp, Satz 8.15)}\\
&= \exp\left(\lim_{n\to\infty}\frac{\log(1+\frac{x}{n})}{n^{-1}}\right)\\
&= \exp\left(\lim_{n\to\infty}\frac{-x/n^2}{1+\frac{x}{n}}\cdot\frac{1}{-1/n^2}\right) && (\text{„}\tfrac{0}{0}\text{“, Differenzieren nach } n\text{, beachte } -\tfrac{1}{n^2}\neq 0)\\
&= \exp(x).
\end{aligned}$$

Aufgabe 11.5. Berechne die folgenden Grenzwerte:

(a) $\displaystyle\lim_{\substack{x\to 0\\ x>0}}\frac{\log(\tan(2x))}{\log(\tan(3x))}$ (b) $\displaystyle\lim_{x\to 1}\left(\frac{1}{x-1}-\frac{1}{\log x}\right)$ (c) $\displaystyle\lim_{\substack{x\to 0\\ x>0}} x^x$

Aufgabe 11.6. Es sei $f\colon [a,b]\to\mathbb{R}$ stetig und auf (a,b) differenzierbar, so dass $\lim_{x\to a} f'(x)$ existiert. Zeige, dass f dann auch in a differenzierbar ist und $f'(a) = \lim_{x\to a} f'(x)$ gilt.

Aufgabe 11.7 (Bogenmaß)**.** Es sei $x \in \mathbb{R}_{\geq 0}$. Wir wollen nun zeigen, dass die „Bogenlänge“ entlang des Einheitskreises von 1 nach $\mathrm{e}^{\mathrm{i}x} \in \mathbb{C}$ gleich x ist und $\mathrm{e}^{\mathrm{i}x}$ damit als der Punkt auf dem Einheitskreis aufgefasst werden kann, der mit der positiven reellen Achse den Winkel x im *Bogenmaß* einschließt (siehe Bemerkung 9.11).



Für alle $n \in \mathbb{N}_{>0}$ unterteilen wir den Kreisbogen dazu wie im Bild durch die Zwischenpunkte $\mathrm{e}^{\mathrm{i}kx/n}$ mit $k = 0,\dots,n$. Die Länge des geraden Streckenzuges, der diese Punkte der Reihe nach miteinander verbindet, ist dann $L_n = \sum_{k=0}^{n-1}\left|\mathrm{e}^{\mathrm{i}(k+1)x/n} - \mathrm{e}^{\mathrm{i}kx/n}\right|$.

Zeige, dass $\lim_{n\to\infty} L_n = x$.

11.B Taylor-Entwicklung

Als weitere Anwendung der Differentialrechnung wollen wir nun unsere ursprüngliche Idee der linearen Approximation einer (reellen) Funktion $f\colon D\to\mathbb{R}$ in einem Punkt $a \in D$ erweitern und uns fragen, ob wir vielleicht noch bessere Näherungen bekommen können, wenn wir als Näherungsfunktion statt einer *linearen* Funktion eine Polynomfunktion von höherem Grad verwenden. Betrachten wir z. B. statt unserer bisherigen Näherung vom Anfang von Kapitel 10

$$f(x) \approx f(a) + c_1(x-a)$$

den Ansatz

$$f(x) \approx f(a) + c_1(x-a) + c_2(x-a)^2,$$

bei dem wir auch einen quadratischen Term zulassen (den wir proportional zu $(x-a)^2$ statt zu x^2 wählen, damit er am Näherungspunkt a selbst verschwindet), so können wir wie im Bild unten erwarten, dass wir eine viel bessere Näherung erhalten, da die Näherungsfunktion ja jetzt eine quadratische Parabel ist und damit auch ein wenig die Krümmung von f an der Stelle a nachbilden kann.

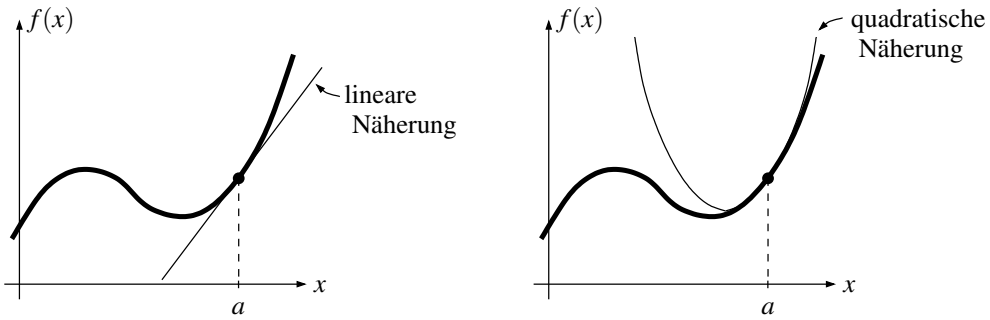
Natürlich können wir dies dann auch noch weiter treiben und Polynomfunktionen höheren Grades zulassen: Wenn wir für ein $n \in \mathbb{N}$ einen Ansatz

$$f(x) \approx \sum_{k=0}^{n} c_k (x-a)^k$$

machen und dabei die Koeffizienten c_k geschickt wählen, sollte die Näherung mit wachsendem n immer besser werden. Wir können sogar versuchen, den Grenzübergang $n \to \infty$ zu machen und uns fragen, ob wir mit einer Potenzreihe

$$\sum_{k=0}^{\infty} c_k (x-a)^k$$

im Grenzfall vielleicht nicht nur eine ganz besonders gute Näherung, sondern sogar *genau* die Funktion f zurück erhalten, also ob wir f letztlich als Potenzreihe in $x-a$ schreiben können – schließlich haben wir ja auch wie z. B. die Exponentialfunktion schon einige Funktionen gesehen, die wir von vornherein bereits als Potenzreihe geschrieben haben.



Um diese Idee zu verfolgen, wollen wir nun als Erstes untersuchen, welche Koeffizienten c_k wir in den obigen Polynomen bzw. Reihen wählen sollten. Da wir bereits wissen, dass der lineare Koeffizient c_1 gerade die Ableitung $f'(a)$ ist, sollte es nicht überraschen, dass wir für die höheren Koeffizienten c_k mit $k > 1$ höhere Ableitungen benötigen. Diese wollen wir daher jetzt einführen.

Definition 11.8 (Höhere Ableitungen)**.** Es seien $f\colon D \to \mathbb{R}$ eine Funktion und $n \in \mathbb{N}$.

(a) Die Funktion f heißt **n-mal differenzierbar** auf D, wenn alle fortgesetzten Ableitungen $f^{(0)} := f$, $f^{(1)} := f'$, $f^{(2)} := f'' := (f')'$, $\ldots$, $f^{(n)} := (f^{(n-1)})'$ existieren.

(b) Die Funktion f heißt **n-mal stetig differenzierbar** auf D, wenn zusätzlich $f^{(n)}$ stetig ist.

(c) Existieren die höheren Ableitungen $f^{(n)}$ für alle n, so heißt f **unendlich oft differenzierbar** auf D.

Die Menge aller n-mal stetig differenzierbaren Funktionen auf D wird mit $C^n(D)$ bezeichnet (der Buchstabe C kommt vom englischen Wort „continuous“ für „stetig“). Inbesondere ist also $C^0(D)$ die Menge aller stetigen und $C^\infty(D)$ die Menge aller unendlich oft differenzierbaren Funktionen auf D.

Wie diese höheren Ableitungen die oben betrachteten Koeffizienten c_k bestimmen, sieht man am besten am Beispiel von Potenzreihen, die ja bereits in einer derartigen Form geschrieben sind:

Satz 11.9 (**Taylor-Formel für Potenzreihen**)**.** *Es seien $a \in \mathbb{R}$ und $f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k$ eine reelle Potenzreihe in $x-a$ mit Konvergenzradius $r > 0$, so dass wir f also als Funktion $f\colon (a-r, a+r) \to \mathbb{R}$ auffassen können.*

Dann ist f auf $(a-r, a+r)$ unendlich oft differenzierbar, und für die Koeffizienten c_k der Potenzreihe gilt

$$c_k = \frac{f^{(k)}(a)}{k!} \qquad \text{für alle } k \in \mathbb{N}.$$

Mit anderen Worten ist also

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$

für alle $x \in (a-r, a+r)$.

Beweis. Nach Folgerung 10.28 ist jede Potenzreihe in ihrem Konvergenzgebiet differenzierbar, und ihre Ableitung ist wieder eine Potenzreihe (mit demselben Konvergenzradius), die sich durch gliedweises Differenzieren berechnen lässt. Insbesondere ist f damit also unendlich oft differenzierbar, und die höheren Ableitungen sind

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)\, c_k\, (x-a)^{k-n}$$

für alle $n \in \mathbb{N}$. Setzen wir hier nun $x = a$ ein, so ist $(x-a)^{k-n}$ gleich 0 für $k > n$ und 1 für $k = n$. In der obigen Summe bleibt dann also nur der Term für $k = n$ übrig, und wir erhalten wie behauptet

$$f^{(n)}(a) = n(n-1)\cdots(n-n+1)\, c_n = n!\, c_n. \qquad \square$$

Beispiel 11.10. Wir betrachten die Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{1+x^2}$ und möchten die 10. Ableitung $f^{(10)}(0)$ im Nullpunkt berechnen. Natürlich könnte man jetzt mit Hilfe der Regeln von Satz 10.8 alle fortgesetzten Ableitungen von f berechnen und schließlich in dem so gefundenen Ausdruck für $f^{(10)}$ den Wert $x = 0$ einsetzen – dies wäre aber sehr zeitaufwändig. Viel schneller geht es mit der Taylor-Formel: Nach der geometrischen Reihe können wir f ja für $|x| < 1$ gemäß

$$f(x) = \frac{1}{1+x^2} = \frac{1}{1-(-x^2)} = \sum_{n=0}^{\infty} (-x^2)^n = 1 - x^2 + x^4 - x^6 + x^8 - x^{10} \pm \cdots$$

als Potenzreihe in x schreiben. Satz 11.9 mit $a = 0$ und $k = 10$ sagt uns also für den Koeffizienten von x^{10} in dieser Reihe, der ja offensichtlich gleich -1 ist, dass

$$-1 = \frac{f^{(10)}(0)}{10!}, \qquad \text{und damit} \qquad f^{(10)}(0) = -10!.$$

Wir sehen an diesem Beispiel schon, dass die Taylor-Formel für Potenzreihen auch dann nützlich ist, wenn die Funktion f ursprünglich gar nicht als Potenzreihe gegeben ist, sondern wir nur wissen, dass es eine solche Darstellung als Potenzreihe gibt. In der Tat benutzt die Taylor-Formel in der Form

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}\, (x-a)^k \qquad (*)$$

ja auch gar nicht mehr die Koeffizienten der ursprünglichen Reihe, sondern nur noch die Tatsache, dass sich f überhaupt irgendwie als Potenzreihe schreiben lässt. Gilt die Formel $(*)$ also vielleicht sogar für jede unendlich oft differenzierbare Funktion f?

Leider ist (wie wir gleich sehen werden) die Antwort auf diese Frage nein. Für viele in der Praxis vorkommende Funktionen ist die Antwort allerdings auch ja – und daher lohnt es sich, die Sache doch noch weiter zu verfolgen. Wir geben der rechten Seite von $(*)$, bzw. den Partialsummen dieser Reihe, daher zunächst einen Namen.

Definition 11.11 (Taylor-Polynom und Taylor-Reihe)**.** Es seien $D \subset \mathbb{R}$ ein Intervall, $f\colon D \to \mathbb{R}$ eine Funktion, sowie $a \in D$ ein fest gewählter Punkt.

(a) Die Funktion f sei n-mal differenzierbar für ein $n \in \mathbb{N}$. Dann heißt die Polynomfunktion

$$T^n_{f,a}\colon D \to \mathbb{R},\ x \mapsto \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}\, (x-a)^k$$

das **n-te Taylor-Polynom** von f mit Entwicklungspunkt a; offensichtlich ist $\deg T^n_{f,a} \le n$.

(b) Ist f unendlich oft differenzierbar, so heißt die Potenzreihe in $x-a$

$$T_{f,a}(x) := \lim_{n\to\infty} T^n_{f,a}(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}\, (x-a)^k$$

die **Taylor-Reihe** von f mit Entwicklungspunkt a. Beachte, dass zunächst nicht klar ist, ob diese Potenzreihe einen Konvergenzradius größer als 0 hat, also ob sie überhaupt in irgendeinem Punkt x (außer a) konvergiert – und dass, selbst wenn sie konvergiert, nicht klar ist, ob sie als Funktion im Konvergenzgebiet mit f übereinstimmt.

Beispiel 11.12.

(a) Lässt sich eine Funktion f auf einem Intervall $(a-r, a+r)$ mit $a \in \mathbb{R}$ und $r > 0$ als Potenzreihe in $x-a$ schreiben, so besagt Satz 11.9 gerade, dass die Taylor-Reihe $T_{f,a}(x)$ genau diese Reihe ist, also $T_{f,a}(x) = f(x)$ für alle $x \in (a-r, a+r)$ gilt.

(b) Wir betrachten die Funktion

$$f\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto \log x$$

und bestimmen ihre Taylor-Reihe mit Entwicklungspunkt $a = 1$. Die Ableitungen von f sind einfach zu berechnen: Wegen $f'(x) = x^{-1}$ ist

$$f^{(k)}(x) = (-1)\cdot(-2)\cdots(-(k-1))\cdot x^{-k} = (-1)^{k-1}\cdot\frac{(k-1)!}{x^k}$$

und damit $f^{(k)}(1) = (-1)^{k-1}\cdot(k-1)!$ für alle $k > 0$. Die Taylor-Reihe von f mit Entwicklungspunkt 1 ist damit

$$\begin{aligned} T_{f,1}(x) = \log 1 + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}\cdot(k-1)!}{k!}(x-1)^k &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(x-1)^k \\ &= (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} \mp \cdots . \end{aligned}$$

Wie in Beispiel 7.29 (a) hat diese Potenzreihe in $x-1$ den Konvergenzradius 1, sie konvergiert also für $|x-1| < 1$, d. h. für $x \in (0,2)$, und divergiert für $|x-1| > 1$, also für $x < 0$ oder $x > 2$. Damit ist schon einmal klar, dass die Taylor-Reihe $T_{f,1}(x)$ für $x > 2$ sicher *nicht* die ursprüngliche Funktion $f(x) = \log x$ darstellt, da sie dort ja nicht einmal konvergiert. Aber auch für $x \in (0,2)$ ist noch nicht klar, dass wirklich $T_{f,1}(x) = f(x)$ gilt: Das folgende Beispiel zeigt, dass eine Taylor-Reihe auch im Fall der Konvergenz nicht mit der ursprünglichen Funktion übereinstimmen muss.

Aufgabe 11.13. Es sei

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} \exp\left(-\frac{1}{x^2}\right) & \text{für } x \neq 0, \\ 0 & \text{für } x = 0. \end{cases}$$

Zeige, dass f unendlich oft differenzierbar ist, und dass die Taylor-Reihe $T_{f,0}$ die Nullfunktion ist (also insbesondere zwar überall konvergiert, aber außer im Nullpunkt nirgends mit f übereinstimmt). Skizziere auch den Graphen von f.

(Hinweis: Man zeige mit vollständiger Induktion, dass alle Ableitungen von f in 0 gleich 0 und für $x \neq 0$ von der Form $\frac{p(x)}{q(x)}\exp(-\frac{1}{x^2})$ für gewisse Polynomfunktionen p und q sind.)

24

Wir benötigen also ein Kriterium, mit dem wir eine Funktion f mit ihren Taylor-Polynomen $T^n_{f,a}$ bzw. ihrer Taylor-Reihe $T_{f,a}$ vergleichen können, so dass wir letztlich nachprüfen können, ob eine (konvergierende) Taylor-Reihe auch wirklich gleich der ursprünglichen Funktion ist. Dies liefert der folgende Satz:

Satz 11.14 (Taylor-Formel). *Es seien $n \in \mathbb{N}$, $D \subset \mathbb{R}$ ein Intervall, und $f\colon D \to \mathbb{R}$ eine $(n+1)$-mal differenzierbare Funktion. Ferner seien $a, x \in D$. Dann gibt es ein c zwischen a und x mit*

$$f(x) - T^n_{f,a}(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

*Man bezeichnet $f(x) - T^n_{f,a}(x)$ auch als das **Restglied** des n-ten Taylor-Polynoms und schreibt es als $R^n_{f,a}(x)$.*

Beweis. Für den Fall $x = a$ ist die Aussage trivial, da dann beide Seiten der zu zeigenden Formel gleich 0 sind. Für $x \neq a$ behaupten wir, dass die Aussage unmittelbar aus dem Mittelwertsatz 10.24 (b) angewendet auf die beiden Funktionen

$$F\colon D \to \mathbb{R},\ t \mapsto T^n_{f,t}(x) = f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}(x-t)^2 + \dots + \frac{f^{(n)}(t)}{n!}(x-t)^n$$

$$\text{und} \quad G\colon D \to \mathbb{R},\ t \mapsto (x-t)^{n+1}$$

folgt. In der Tat sind F und G dann differenzierbar mit

$$\begin{aligned} F'(t) &= f'(t) + \Big(-\frac{f'(t)}{1!} + \frac{f''(t)}{1!}(x-t)\Big) + \Big(-\frac{f''(t)}{1!}(x-t) + \frac{f'''(t)}{2!}(x-t)^2\Big) + \dots \\ &\quad + \Big(-\frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} + \frac{f^{(n+1)}(t)}{n!}(x-t)^n\Big) \\ &= \frac{f^{(n+1)}(t)}{n!}(x-t)^n \\ \text{und} \quad G'(t) &= -(n+1)(x-t)^n \end{aligned}$$

sowie $F(x) - F(a) = f(x) - T^n_{f,a}(x) = R^n_{f,a}(x)$ und $G(x) - G(a) = -(x-a)^{n+1}$. Der Mittelwertsatz 10.24 (b) liefert also ein c zwischen a und x mit

$$-\frac{f^{(n+1)}(c)}{n!}(x-c)^n \cdot (x-a)^{n+1} = -(n+1)(x-c)^n \cdot R^n_{f,a}(x),$$

d. h. wie behauptet $R^n_{f,a}(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$. □

Bemerkung 11.15.

(a) Für $n = 0$ ist Satz 11.14 exakt der Mittelwertsatz 10.24 (a). Wir können die Taylor-Formel also auch als eine Verallgemeinerung des Mittelwertsatzes auffassen.

(b) Setzen wir in der Formel aus Satz 11.14 noch den Ausdruck aus Definition 11.11 (a) ein, so erhalten wir für jede $(n+1)$-mal differenzierbare Funktion f

$$f(x) = \underbrace{f(a) + \frac{f'(a)}{1!}(x-a) + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n}_{=T^n_{f,a}(x)} + \underbrace{\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}}_{=R^n_{f,a}(x)}$$

für ein c zwischen a und x. Das Restglied des n-ten Taylor-Polynoms hat also genau die Form des $(n+1)$-ten Gliedes der Taylor-Reihe – bis auf den Unterschied, dass man die Ableitung dort an einer Zwischenstelle c anstatt am Entwicklungspunkt a nehmen muss.

(c) Offensichtlich gilt genau dann $T_{f,a}(x) = f(x)$, wenn $\lim_{n\to\infty} R^n_{f,a}(x) = 0$. Wenn man die Taylor-Reihe oder die Taylor-Polynome mit der ursprünglichen Funktion vergleichen möchte, muss man also in irgendeiner Form das Restglied abschätzen. Hier sind zwei Beispiele dafür.

Beispiel 11.16 (Restgliedabschätzung)**.**

(a) Wenden wir Satz 11.14 auf die Taylor-Reihe der Funktion $f\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto \log x$ mit Entwicklungspunkt $a = 1$ aus Beispiel 11.12 (b) an, so erhalten wir mit den dort berechneten Ableitungen $f^{(n+1)}(x) = (-1)^n \frac{n!}{x^{n+1}}$ für jedes $x \in \mathbb{R}_{>0}$

$$R^n_{f,1}(x) = f(x) - T^n_{f,1}(x) = \frac{(-1)^n}{n+1} \cdot \frac{(x-1)^{n+1}}{c_n^{n+1}}$$

für ein c_n zwischen 1 und x (wir haben den Zwischenwert hier mit c_n statt c bezeichnet, da es natürlich für jedes n ein anderer sein wird). Ist nun $x \in [1,2]$, so ist aber stets $c_n \geq 1$ und $|x-1| \leq 1$, und wir erhalten die Abschätzung

$$|R^n_{f,1}(x)| \leq \frac{1}{n+1},$$

was mit $n \to \infty$ gegen 0 konvergiert. Also konvergieren die Taylor-Polynome für $n \to \infty$ zumindest auf $[1,2]$ wirklich gegen die Funktion f: Es gilt

$$\log x = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (x-1)^k \qquad \text{für alle } x \in [1,2]. \tag{$*$}$$

Insbesondere ergibt sich damit für $x = 2$ der Wert der alternierenden harmonischen Reihe zu

$$\log 2 = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \cdots .$$

Wir werden später in Beispiel 12.39 (a) übrigens sehen, dass die Gleichung $(*)$ sogar für $x \in (0,2]$ gilt – also für alle x, für die die Taylor-Reihe überhaupt konvergiert – aber mit unserer bisherigen Formel für das Restglied aus Satz 11.14 können wir das noch nicht beweisen.

(b) Wenn wir als Näherung einer Funktion nur an einem bestimmten Taylor-Polynom (und nicht an der kompletten Reihe) interessiert sind, kann uns Satz 11.14 sagen, wie groß der Fehler ist, den wir dabei machen. Betrachten wir z. B. die Sinusfunktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto \sin x$, so ist am Entwicklungspunkt 0

$$T^4_{f,0}(x) = x - \frac{x^3}{6}$$

(wie man aus Lemma 9.13 (b) sofort abliest, denn nach Satz 11.9 ist ja jede Potenzreihe ihre eigene Taylor-Reihe). Wegen $f^{(5)}(x) = \cos x$ besagt Satz 11.14 für $n = 4$ nun für alle $x \in \mathbb{R}$

$$R^4_{f,0}(x) = \sin x - \left(x - \frac{x^3}{6}\right) = \frac{\cos c}{5!} x^5$$

für ein c zwischen 0 und x. Wenn wir nun z. B. nur an Werten $x \in \mathbb{R}$ mit $|x| \le \frac{1}{2}$ interessiert sind, so können wir diesen Ausdruck wegen $|\cos c| \le 1$ abschätzen zu

$$|R^4_{f,0}(x)| \le \frac{(\frac{1}{2})^5}{5!} = \frac{1}{3840},$$

d. h. wenn wir für $|x| \le \frac{1}{2}$ den Sinus durch sein viertes Taylor-Polynom $x - \frac{x^3}{6}$ ersetzen, machen wir dabei einen Fehler von höchstens $\frac{1}{3840} \approx 0{,}0003$.

Aufgabe 11.17.

(a) Berechne das Taylor-Polynom $T^2_{f,1}$ für die Funktion $f(x) = \sqrt{x}$ und zeige die Restgliedabschätzung $|f(x) - T^2_{f,1}(x)| \le \frac{1}{20}$ für alle $x \in \left[\frac{1}{2}, \frac{3}{2}\right]$.

(b) Berechne $f^{(10)}(0)$ sowie das Taylor-Polynom $T^{10}_{f,0}$ für die Funktion $f(x) = \frac{\cos(x^5)}{1-2x^6}$.

Aufgabe 11.18. Es sei $f\colon \mathbb{R} \to \mathbb{R}$ eine zweimal differenzierbare Funktion mit $f'' = f$ sowie $f(0) = f'(0) = 1$. Berechne die Taylor-Reihe von f mit Entwicklungspunkt 0 und zeige durch eine Restgliedabschätzung, dass $f = \exp$ die Exponentialfunktion ist.

Als weitere Anwendung der Taylor-Formel wollen wir nun noch ein einfaches hinreichendes Kriterium für lokale Extrema geben. Wir hatten bisher ja nur in Lemma 10.21 gesehen, dass an einem lokalen Extremum, das nicht am Rand der Definitionsmenge liegt, ein kritischer Punkt vorliegen, also die erste Ableitung verschwinden muss – dass diese Bedingung aber nicht für ein lokales Extremum ausreicht. Mit Hilfe höherer Ableitungen und der Taylor-Formel können wir nun ein Kriterium angeben, das nahezu immer und ohne allzu großen Aufwand entscheiden kann, ob wirklich ein lokales Extremum vorliegt oder nicht:

Satz 11.19 (Extremwertkriterium). *Es seien $n \in \mathbb{N}_{>0}$ und $f\colon (a,b) \to \mathbb{R}$ eine n-mal stetig differenzierbare Funktion. Für ein $c \in (a,b)$ gelte $f'(c) = f''(c) = \cdots = f^{(n-1)}(c) = 0$ und $f^{(n)}(c) \neq 0$.*

(a) *Ist n gerade und $f^{(n)}(c) > 0$, so hat f in c ein isoliertes lokales Minimum.*

(b) *Ist n gerade und $f^{(n)}(c) < 0$, so hat f in c ein isoliertes lokales Maximum.*

(c) *Ist n ungerade, so hat f in c kein lokales Extremum.*

Beweis. Es sei ohne Beschränkung der Allgemeinheit $f^{(n)}(c) > 0$ (der Fall $f^{(n)}(c) < 0$ ist analog). Da $f^{(n)}$ nach Voraussetzung stetig ist, gilt nach Bemerkung 8.8 dann sogar $f^{(n)}(x) > 0$ in einer δ-Umgebung von c. Die Taylor-Formel aus Satz 11.14 besagt nun, dass es für alle $x \in (c-\delta, c+\delta)$ ein x' zwischen c und x gibt mit

$$f(x) - T_{f,c}^{n-1}(x) = \frac{f^{(n)}(x')}{n!}(x-c)^n.$$

Wegen $f'(c) = f''(c) = \cdots = f^{(n-1)}(c) = 0$ ist aber $T_{f,c}^{n-1}(x) = f(c)$, und damit also

$$f(x) - f(c) = \frac{f^{(n)}(x')}{n!}(x-c)^n. \qquad (*)$$

Da mit x auch x' in $(c-\delta, c+\delta)$ liegt, ist $f^{(n)}(x')$ in jedem Fall positiv. Also ist der Term $(*)$ für $x \in (c-\delta, c+\delta)\setminus\{c\}$

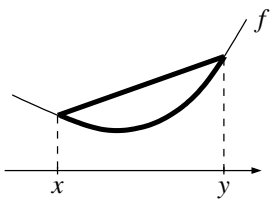
- immer größer als 0 falls n gerade ist; in diesem Fall hat f dann also ein isoliertes lokales Minimum in c;
- größer als 0 für $x > c$ und kleiner als Null für $x < c$ wenn n ungerade ist; in diesem Fall hat f also kein lokales Extremum in c. □

Bemerkung 11.20. Anschaulich kann man die Idee von Satz 11.19 kurz so zusammenfassen: Es sei f eine Funktion, von der wir an einer Stelle c wissen wollen, ob ein lokales Extremum vorliegt. Ist nun die n-te Ableitung von f die erste, die am Punkt c nicht verschwindet, so sagt uns die Idee der Taylor-Näherung, dass dann in der Nähe von c

$$f(x) \approx f(c) + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

sein sollte, denn dies ist ja gerade das n-te Taylor-Polynom. Da dieser Ausdruck auf der rechten Seite eine einfache Potenzfunktion ist, sieht man ihm aber natürlich sofort sein Verhalten um den Punkt c herum an: Für gerades n gibt es je nach Vorzeichen von $f^{(n)}(c)$ ein isoliertes lokales Minimum oder Maximum, und für ungerades n kein lokales Extremum. Mit dieser Idee lässt sich übrigens auch die Aussage des Satzes sehr leicht merken!

Aufgabe 11.21. In dieser Aufgabe wollen wir sehen, welche anschauliche Bedeutung das Vorzeichen der zweiten Ableitung für den Graphen einer Funktion hat.



Eine Funktion $f\colon [a,b] \to \mathbb{R}$ heißt *konvex*, wenn für alle $x, y \in [a,b]$ mit $x < y$ der Graph von $f|_{[x,y]}$ wie im Bild komplett auf oder unterhalb der Verbindungsgeraden von $(x, f(x))$ nach $(y, f(y))$ liegt.

Zeige, dass für eine zweimal stetig differenzierbare Funktion $f\colon [a,b] \to \mathbb{R}$ die folgenden Aussagen äquivalent sind:

(a) f ist konvex.

(b) Für alle $x, y \in [a,b]$ mit $x < y$ und $\lambda \in [0,1]$ gilt

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y).$$

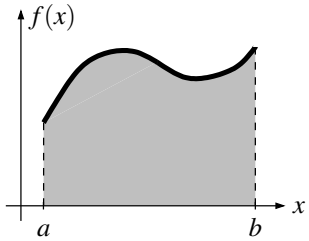
(c) $f''(x) \ge 0$ für alle $x \in (a,b)$.

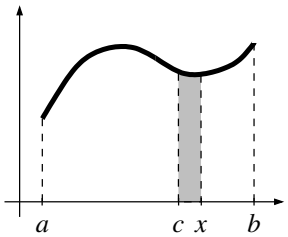
12. Integralrechnung

Als Abschluss der Analysis in einer Veränderlichen wollen wir nach der Differentiation nun noch die Integration betrachten. Da die Integralrechnung über $\mathbb{R}$ sehr verschieden von der über $\mathbb{C}$ ist, werden wir uns dabei nur mit reellen Integralen beschäftigen. In der Tat sind komplexe Integrale (oder allgemein die komplexe Analysis) der wesentliche Inhalt der Vorlesung „Einführung in die Funktionentheorie“, die ihr im zweiten Studienjahr hören könnt.

Die Integralrechnung kann man auf zweierlei Arten motivieren. Ist $f\colon [a,b] \to \mathbb{R}$ eine reelle Funktion, so können wir die folgenden beiden Fragestellungen betrachten:

- (Flächenberechnung) Wie groß ist die Fläche, die unter dem Graphen von f liegt (im Bild unten links grau eingezeichnet) – oder allgemeiner, wie kann man den Flächeninhalt gekrümmter Flächen berechnen?





- (Umkehrung der Differentiation) Gibt es eine differenzierbare Funktion $F\colon [a,b] \to \mathbb{R}$, deren Ableitung F' gleich f ist – und wenn ja, wie können wir ein solches F bestimmen? Diese Frage hat oft auch eine anschauliche Bedeutung: Beschreibt eine Funktion z. B. die Position eines Gegenstandes in Abhängigkeit von der Zeit, so ist die Ableitung dieser Funktion, also die lokale Positionsänderung pro Zeiteinheit, natürlich einfach die Geschwindigkeit des Gegenstandes. Wenn wir von der Ableitung auf die ursprüngliche Funktion zurück schließen wollen, möchten wir anschaulich also aus der Kenntnis der Geschwindigkeit zu jedem Zeitpunkt die von dem Gegenstand zurückgelegte Wegstrecke berechnen können.

Es ist leicht einzusehen, dass diese beiden Probleme sehr eng miteinander zusammenhängen: Bezeichnen wir für $c \in [a,b]$ mit $F(c)$ die Fläche, die über dem Intervall $[a,c]$ unter dem Graphen von f liegt, so ist $F(x) - F(c)$ für $x \in [a,b]$ natürlich gerade die Fläche unter f zwischen c und x (im Bild oben rechts grau eingezeichnet). Für x nahe bei c ist dies näherungsweise eine Rechteckfläche der Breite $x - c$ und Höhe $f(c)$, d. h. es ist

$$F(x) - F(c) \approx (x-c) \cdot f(c), \qquad \text{und damit} \qquad \frac{F(x) - F(c)}{x - c} \approx f(c).$$

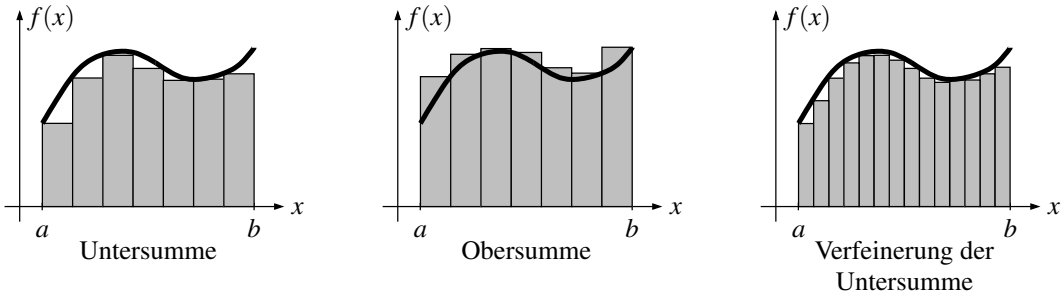
Im Grenzfall $x \to c$ sollte also $F' = f$ gelten, d. h. das Problem der Flächenberechnung unter dem Graphen einer Funktion sollte automatisch auch zur Umkehrung der Differentiation führen.

Wir werden uns im Folgenden zunächst in Abschnitt 12.A mit dem ersten Problem der Flächenberechnung beschäftigen, und daraufhin dann in Abschnitt 12.B den Zusammenhang zur Umkehrung der Differentiation herstellen.

12.A Das Riemann-Integral

Um den Flächeninhalt unter dem Graphen einer Funktion $f\colon [a,b] \to \mathbb{R}$ untersuchen zu können, müssen wir natürlich zunächst erst einmal mit einer exakten Definition dieses Konzepts beginnen. Die Idee hierfür ist einfach: Wir zerlegen das Intervall $[a,b]$ in viele kleine Teilintervalle, und approximieren die Fläche unter dem Graphen von f durch Rechteckflächen über diesen Teilintervallen,

indem wir wie im Bild unten als Höhe der Rechtecke einmal das Minimum und einmal das Maximum von f auf den betrachteten Teilintervallen wählen. Auf diese Art erhalten wir leicht zu berechnende Flächen, die im Fall des Minimums etwas kleiner und im Fall des Maximums etwas größer als die gesuchte Fläche sind. Wenn wir die Zerlegung in die Teilintervalle immer feiner machen (wie z. B. im Bild unten rechts), sollten diese Flächen dann von unten bzw. oben gegen den gesuchten Flächeninhalt unter dem Graphen von f konvergieren.



Wir wollen diese Idee nun mathematisch exakt definieren. Um die Theorie möglichst allgemein zu halten, wollen wir uns dabei nicht auf stetige Funktionen beschränken. Dies heißt natürlich, dass f auf den betrachteten Teilintervallen nicht mehr notwendig ein Minimum und Maximum hat (siehe Satz 8.25), sondern dass wir im Allgemeinen nur ein Infimum und Supremum erhalten – und das auch nur dann, wenn wir voraussetzen, dass f beschränkt ist.

Definition 12.1 (Zerlegungen, Unter- und Obersummen)**.** Es sei $f\colon [a,b] \to \mathbb{R}$ eine beschränkte Funktion.

(a) Eine endliche Teilmenge $Z = \{x_0, x_1, \dots, x_n\}$ von Punkten in $[a,b]$ mit $a,b \in Z$ bezeichnen wir als eine **Zerlegung** des Intervalls $[a,b]$. Wir vereinbaren, dass wir in dieser Schreibweise die $x_0, \dots, x_n$ immer so anordnen wollen, dass $a = x_0 < x_1 < \dots < x_n = b$ gilt. Sind Z, Z' zwei Zerlegungen von $[a,b]$ mit $Z \subset Z'$, so nennen wir Z' eine **Verfeinerung** von Z.

(b) Ist $Z = \{x_0, \dots, x_n\}$ eine Zerlegung von $[a,b]$, so heißt

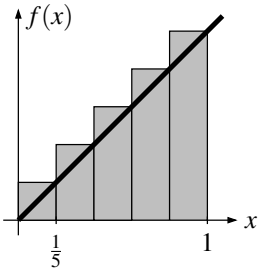
$$\mathrm{US}(f,Z) := \sum_{i=1}^{n} (x_i - x_{i-1}) \cdot \inf f([x_{i-1}, x_i]) \quad \text{die } \textbf{Untersumme}\text{, und analog}$$

$$\mathrm{OS}(f,Z) := \sum_{i=1}^{n} (x_i - x_{i-1}) \cdot \sup f([x_{i-1}, x_i]) \quad \text{die } \textbf{Obersumme}$$

von f bezüglich Z.

Beispiel 12.2. Wir betrachten die Funktion $f\colon [0,1] \to \mathbb{R}$, $x \mapsto x$, und für gegebenes $n \in \mathbb{N}_{>0}$ die Zerlegung $Z_n = \{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$. Natürlich ist das Supremum von f auf einem Teilintervall $\left[\frac{i-1}{n}, \frac{i}{n}\right]$ genau der Funktionswert $\frac{i}{n}$ an der rechten Intervallgrenze, und damit ist die Obersumme von f bezüglich Z_n (also die für den Fall $n = 5$ im Bild rechts eingezeichnete graue Fläche) gleich

$$\mathrm{OS}(f,Z_n) = \sum_{i=1}^{n} \left(\frac{i}{n} - \frac{i-1}{n}\right) \cdot \frac{i}{n} = \frac{1}{n^2} \sum_{i=1}^{n} i \overset{3.13}{=} \frac{1}{n^2} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2n}.$$



Analog müssen wir für die Untersumme jeweils den Funktionswert $\frac{i-1}{n}$ an der linken Intervallgrenze nehmen, und erhalten

$$\mathrm{US}(f,Z_n) = \sum_{i=1}^{n} \left(\frac{i}{n} - \frac{i-1}{n}\right) \cdot \frac{i-1}{n} = \frac{1}{n^2} \sum_{i=1}^{n} (i-1) = \frac{1}{n^2} \sum_{i=0}^{n-1} i = \frac{1}{n^2} \cdot \frac{(n-1)n}{2} = \frac{n-1}{2n}.$$

Lemma 12.3 (Eigenschaften von Unter- und Obersummen)**.** *Es seien $f\colon [a,b] \to \mathbb{R}$ eine beschränkte Funktion und Z, Z' zwei Zerlegungen von $[a,b]$. Dann gilt:*

(a) *Ist Z' eine Verfeinerung von Z, so ist* $\mathrm{US}(f,Z') \geq \mathrm{US}(f,Z)$ *und* $\mathrm{OS}(f,Z') \leq \mathrm{OS}(f,Z)$.

(b) $\mathrm{US}(f,Z) \leq \mathrm{OS}(f,Z')$.

25

Beweis.

(a) Da jede Verfeinerung von Z durch endliches Hinzufügen von weiteren Unterteilungspunkten entsteht, genügt es, den Fall zu betrachten, dass Z' durch Hinzufügen eines weiteren Punktes aus Z entsteht, also dass $Z = \{x_0, \ldots, x_n\}$ und $Z' = \{x_0, x_1, \ldots, x_{k-1}, x', x_k, \ldots, x_n\}$ ist. Nach Definition ist dann

$$\begin{aligned}\mathrm{US}(f,Z') = \sum_{i \neq k} (x_i - x_{i-1}) \cdot \inf f([x_{i-1}, x_i]) \\ + (x' - x_{k-1}) \cdot \inf f([x_{k-1}, x']) + (x_k - x') \cdot \inf f([x', x_k]).\end{aligned}$$

In dieser Summe sind nun die beiden Infima in der zweiten Zeile größer oder gleich dem Infimum der größeren Menge $f([x_{k-1}, x_k])$. Also erhalten wir wie gewünscht

$$\mathrm{US}(f,Z') \geq \sum_{i \neq k} (x_i - x_{i-1}) \cdot \inf f([x_{i-1}, x_i]) + (x_k - x_{k-1}) \cdot \inf f([x_{k-1}, x_k]) = \mathrm{US}(f,Z).$$

Die Aussage über die Obersumme beweist man natürlich analog.

(b) Da $Z \cup Z'$ eine gemeinsame Verfeinerung von Z und Z' ist, erhalten wir mit (a)

$$\mathrm{US}(f,Z) \leq \mathrm{US}(f,Z \cup Z') \leq \mathrm{OS}(f,Z \cup Z') \leq \mathrm{OS}(f,Z'),$$

wobei die mittlere Ungleichung gilt, weil das Infimum einer Menge immer kleiner oder gleich dem Supremum ist. □

Aufgabe 12.4. Es seien $f, g\colon [a,b] \to \mathbb{R}$ zwei beschränkte Funktionen, Z eine Zerlegung von $[a,b]$ und $c \in \mathbb{R}_{\geq 0}$. Man zeige:

(a) $\mathrm{OS}(f+g,Z) \leq \mathrm{OS}(f,Z) + \mathrm{OS}(g,Z)$;

(b) $\mathrm{OS}(cf,Z) = c \cdot \mathrm{OS}(f,Z)$;

(c) $\mathrm{OS}(|f|,Z) - \mathrm{US}(|f|,Z) \leq \mathrm{OS}(f,Z) - \mathrm{US}(f,Z)$.

Lemma 12.3 (b) besagt insbesondere, dass *jede* Obersumme eine obere Schranke für *alle* Untersummen ist. Die Menge aller Untersummen ist also nach oben beschränkt. Wir können damit das Supremum aller Untersummen (und genauso das Infimum aller Obersummen) bilden:

Definition 12.5 (Unter- und Oberintegral)**.** Es sei $f\colon [a,b] \to \mathbb{R}$ beschränkt. Dann heißt

$$\begin{aligned}\mathrm{UI}(f) &:= \sup\{\mathrm{US}(f,Z) : Z \text{ Zerlegung von } [a,b]\} && \text{das } \textbf{Unterintegral}\text{, und analog} \\ \mathrm{OI}(f) &:= \inf\{\mathrm{OS}(f,Z) : Z \text{ Zerlegung von } [a,b]\} && \text{das } \textbf{Oberintegral}\end{aligned}$$

von f.

Anschaulich bedeutet dies im Fall des Unterintegrals einfach, dass wir – wie in der Einleitung zu diesem Abschnitt erklärt – versuchen, die Untersummen (durch fortgesetztes Verfeinern der Zerlegungen) möglichst groß zu machen, so dass wir uns letztlich immer mehr dem eigentlich gesuchten Flächeninhalt unter dem Graphen von f nähern. Das Supremum dieser Untersummen, also das Unterintegral, sollte demnach bereits der gesuchte Flächeninhalt unter dem Graphen von f sein. Das gleiche gilt natürlich auch für das Oberintegral, so dass wir insgesamt erwarten würden, dass Unter- und Oberintegral übereinstimmen und gleich dem gesuchten Flächeninhalt sind.

Leider ist dies unter den schwachen Voraussetzungen, die wir bisher an f gestellt haben, im Allgemeinen nicht der Fall, wie wir gleich in Beispiel 12.9 (d) sehen werden. Für beliebiges f erhalten wir zunächst nur die folgende Ungleichung.

Lemma 12.6. *Für jede beschränkte Funktion $f\colon [a,b] \to \mathbb{R}$ gilt* $\mathrm{UI}(f) \leq \mathrm{OI}(f)$.

Beweis. Angenommen, es wäre $\mathrm{UI}(f) > \mathrm{OI}(f)$. Wir wählen dann eine beliebige Zahl $c \in \mathbb{R}$ mit $\mathrm{UI}(f) > c > \mathrm{OI}(f)$. Da $\mathrm{UI}(f)$ die kleinste obere Schranke für die Untersummen ist, ist c keine obere Schranke mehr, d. h. es gibt eine Zerlegung Z von $[a,b]$ mit $\mathrm{US}(f,Z) > c$. Analog finden wir auch eine Zerlegung Z' von $[a,b]$ mit $\mathrm{OS}(f,Z') < c$. Dann ist im Widerspruch zu Lemma 12.3 (b) aber $\mathrm{US}(f,Z) > c > \mathrm{OS}(f,Z')$. □

Definition 12.7 (Integrierbarkeit)**.** Es sei $f\colon [a,b] \to \mathbb{R}$ beschränkt. Gilt dann $\mathrm{UI}(f) = \mathrm{OI}(f)$, so nennen wir f **(Riemann-)integrierbar**, und definieren das **Integral** von f als diesen Wert

$$\int_a^b f(x)\,dx := \mathrm{UI}(f) = \mathrm{OI}(f).$$

Bemerkung 12.8.

(a) Die Schreibweise $\int_a^b f(x)\,dx$ ist an die Differentialschreibweise aus Notation 10.14 angelehnt und soll andeuten, dass man sich das Integral entsprechend unserer Konstruktion anschaulich als eine „unendliche Summe kleiner Rechteckflächen“ vorstellen kann. Dabei steht das Integralzeichen $\int$ als stilisiertes S weiterhin für eine Summe, und die aufsummierten Rechtecke haben die Höhe $f(x)$ und Breite dx (siehe Notation 10.14), also die Fläche $f(x)\,dx$. Die Integrationsvariable x ist damit analog zur Laufvariablen in einer Summe und kann daher auch durch einen anderen Buchstaben ersetzt werden, darf aber natürlich nicht gleichzeitig noch für etwas anderes (z. B. die Ober- oder Untergrenze) verwendet werden: Ein Ausdruck z. B. der Form $\int_a^x f(x)\,dx$ ergibt keinen Sinn, genauso wenig wie eine Summe $\sum_{n=1}^n a_n$.

(b) Es gibt mehrere Arten, den Flächeninhalt unter dem Graphen einer Funktion zu definieren. Neben der hier behandelten Riemannschen Integrationstheorie über Unter- und Obersummen, die wohl die einfachste Herangehensweise ist, ist die „zweitwichtigste“ Möglichkeit das sogenannte *Lebesgue-Integral*, das zwar komplizierter zu definieren ist, dafür aber allgemeiner ist in dem Sinne, dass eine größere Klasse von Funktionen integrierbar wird. Wir werden in dieser Vorlesung jedoch nur die Riemannsche Integrationstheorie behandeln und daher statt von Riemann-Integrierbarkeit einfach immer nur von Integrierbarkeit reden. Die Lebesguesche Integrationstheorie könnt ihr im zweiten Studienjahr in der Vorlesung „Maß- und Integrationstheorie“ kennenlernen.

Beispiel 12.9.

(a) Ist $f(x) = c$ (mit $c \in \mathbb{R}$) eine konstante Funktion, so sind die Infima und Suprema von f auf allen Teilintervallen gleich c. Damit ist dann $\mathrm{US}(f,Z) = \mathrm{OS}(f,Z) = c\,(b-a)$ für alle Unterteilungen Z und somit auch $\mathrm{UI}(f) = \mathrm{OI}(f) = c\,(b-a)$. Also ist f integrierbar mit $\int_a^b f(x)\,dx = c\,(b-a)$ (was natürlich auch genau der Flächeninhalt für $x \in [a,b]$ unter dem Graphen von f ist).

(b) Wie in Beispiel 12.2 betrachten wir noch einmal die Funktion $f\colon [0,1] \to \mathbb{R}$, $x \mapsto x$ mit den Zerlegungen $Z_n = \{0, \frac{1}{n}, \dots, 1\}$. Da das Unterintegral nach Definition eine obere Schranke für alle Untersummen (und analog das Oberintegral eine untere Schranke für alle Obersummen) ist, folgt aus der Rechnung von Beispiel 12.2 sowie Lemma 12.6

$$\frac{n-1}{2n} = \mathrm{US}(f,Z_n) \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le \mathrm{OS}(f,Z_n) = \frac{n+1}{2n},$$

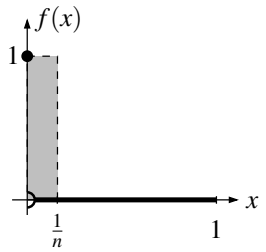
und damit durch Grenzwertbildung $n \to \infty$ nach Satz 5.23 (a)

$$\frac{1}{2} \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le \frac{1}{2},$$

d. h. $\mathrm{UI}(f) = \mathrm{OI}(f) = \frac{1}{2}$. Also ist f integrierbar mit $\int_0^1 f(x)\,dx = \frac{1}{2}$ – was anschaulich ja auch die Dreiecksfläche unter dem Graphen von f ist.

(c) Es sei f die (unstetige) Funktion

$$f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x = 0, \\ 0 & \text{für } x > 0. \end{cases}$$



Für die gleiche Zerlegung $Z_n = \{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$ wie in (b) ist diesmal $\mathrm{US}(f, Z_n) = 0$ und $\mathrm{OS}(f, Z_n) = \frac{1}{n}$ (im Bild rechts ist die Obersumme eingezeichnet). Also folgt wieder

$$0 = \mathrm{US}(f, Z_n) \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le \mathrm{OS}(f, Z_n) = \frac{1}{n},$$

und damit wie in (b) durch Grenzwertbildung für $n \to \infty$

$$0 \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le 0 \quad \Rightarrow \quad \mathrm{UI}(f) = \mathrm{OI}(f) = 0.$$

Damit ist f integrierbar mit $\int_0^1 f(x)\,dx = 0$ – was auch anschaulich einleuchtend ist, denn unter dem einen Punkt, an dem der Funktionswert gleich 1 ist, liegt ja kein Flächeninhalt größer als Null.

(d) Wir betrachten die Funktion

$$f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x \in \mathbb{Q}, \\ 0 & \text{für } x \notin \mathbb{Q}. \end{cases}.$$

Da in jedem Teilintervall von $[0,1]$ nach Aufgabe 5.35 sowohl rationale als auch irrationale Zahlen liegen, ist auf jedem solchen Teilintervall das Infimum von f gleich 0 und das Supremum gleich 1. Damit folgt $\mathrm{US}(f, Z) = 0$ und $\mathrm{OS}(f, Z) = 1$ für jede Zerlegung Z, d. h. es ist auch $\mathrm{UI}(f) = 0$ und $\mathrm{OI}(f) = 1$. Also ist f nicht integrierbar – mit unseren Definitionen können wir den Flächeninhalt unter dem Graphen von f nicht sinnvoll definieren.

Aufgabe 12.10. Zeige durch eine explizite Berechnung von Ober- und Untersummen, dass

$$\text{(a)} \quad \int_0^a \mathrm{e}^x\,dx = \mathrm{e}^a - 1 \qquad\qquad \text{(b)} \quad \int_0^a x^n\,dx = \frac{1}{n+1}\,a^{n+1}$$

für alle $a \in \mathbb{R}_{>0}$ und $n \in \mathbb{N}$. (Hinweis: Aufgabe 4.11 ist für (b) nützlich.)

Bevor wir die wichtigsten Eigenschaften integrierbarer Funktionen untersuchen, wollen wir zunächst noch ein einfaches Kriterium für die Integrierbarkeit beweisen, das implizit auch bereits in unseren Rechnungen von Beispiel 12.9 versteckt ist.

Lemma 12.11 (Riemannsches Integrabilitätskriterium). *Es sei $f\colon [a,b] \to \mathbb{R}$ eine beschränkte Funktion.*

(a) *f ist genau dann integrierbar, wenn es zu jedem $\varepsilon > 0$ eine Zerlegung Z von $[a,b]$ gibt mit $\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) < \varepsilon$.*

(b) *f ist genau dann integrierbar mit Integral $\int_a^b f(x)\,dx = c$, wenn es zu jedem $\varepsilon > 0$ Zerlegungen Z und Z' von $[a,b]$ gibt mit $\mathrm{OS}(f,Z) < c + \varepsilon$ und $\mathrm{US}(f,Z') > c - \varepsilon$.*

Beweis.

„$\Rightarrow$“ Es sei f integrierbar mit $\int_a^b f(x)\,dx = \mathrm{UI}(f) = \mathrm{OI}(f) = c$. Da $\mathrm{OI}(f)$ nach Definition die größte untere Schranke für die Obersummen von f ist, ist $c + \frac{\varepsilon}{2}$ keine untere Schranke mehr, d. h. es gibt eine Zerlegung Z von $[a,b]$ mit $\mathrm{OS}(f,Z) < c + \frac{\varepsilon}{2}$. Analog gibt es eine Zerlegung Z' von $[a,b]$ mit $\mathrm{US}(f,Z') > c - \frac{\varepsilon}{2}$, was bereits (b) zeigt. Außerdem erfüllt die Zerlegung $Z \cup Z'$ dann auch die Eigenschaft von (a), denn nach Lemma 12.3 (a) ist

$$\mathrm{OS}(f, Z \cup Z') - \mathrm{US}(f, Z \cup Z') \le \mathrm{OS}(f,Z) - \mathrm{US}(f,Z') < \left(c + \frac{\varepsilon}{2}\right) - \left(c - \frac{\varepsilon}{2}\right) = \varepsilon.$$

„$\Leftarrow$“ Für Teil (a) haben wir zu jedem $\varepsilon > 0$ eine Zerlegung Z wie in der Voraussetzung, und damit

$$\mathrm{OI}(f) - \mathrm{UI}(f) \le \mathrm{OS}(f,Z) - \mathrm{US}(f,Z) < \varepsilon,$$

da das Ober- bzw. Unterintegral eine untere bzw. obere Schranke für die Ober- bzw. Untersummen sind. Nimmt man hier den Grenzwert für $\varepsilon \to 0$, so ergibt sich $\mathrm{OI}(f) - \mathrm{UI}(f) \le 0$, mit Lemma 12.6 also $\mathrm{OI}(f) = \mathrm{UI}(f)$. Damit ist f dann integrierbar.

Für Teil (b) gilt stattdessen für jedes $\varepsilon > 0$

$$c - \varepsilon < \mathrm{US}(f,Z') \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le \mathrm{OS}(f,Z) < c + \varepsilon,$$

woraus im Grenzfall $\varepsilon \to 0$ die Ungleichungskette $c \le \mathrm{UI}(f) \le \mathrm{OI}(f) \le c$ folgt, d. h. f ist integrierbar mit Integral c. □

Als erste Anwendung dieses Kriteriums wollen wir nun untersuchen, wie die Integrierbarkeit mit der Stetigkeit einer Funktion zusammenhängt. Dazu haben wir in Beispiel 12.9 (c) schon gesehen, dass integrierbare Funktionen nicht notwendig stetig sein müssen. Die Umkehrung ist jedoch immer richtig:

Satz 12.12. *Ist $f\colon [a,b] \to \mathbb{R}$ stetig, so ist f auch integrierbar auf $[a,b]$.*

Beweis. Nach Satz 8.23 ist f beschränkt, so dass wir also die Begriffe dieses Kapitels anwenden können. Wir zeigen die Integrierbarkeit von f mit dem Kriterium aus Lemma 12.11 (a).

Es sei also $\varepsilon > 0$ gegeben. Da f auf dem abgeschlossenen Intervall $[a,b]$ stetig ist, ist f dort nach Satz 8.36 sogar gleichmäßig stetig. Es gibt also ein $\delta > 0$, so dass $|f(y) - f(z)| < \frac{\varepsilon}{b-a}$ für alle $y,z \in [a,b]$ mit $|y-z| < \delta$. Wir wählen nun eine Zerlegung $Z = \{x_0,\dots,x_n\}$ von $[a,b]$ mit $x_i - x_{i-1} < \delta$ für alle i, d. h. alle Teilintervalle sollen kürzer als δ sein. Dann gilt

$$\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) = \sum_{i=1}^{n} (x_i - x_{i-1}) \cdot \big(\sup f([x_{i-1},x_i]) - \inf f([x_{i-1},x_i])\big).$$

Als stetige Funktion nimmt f auf jedem Teilintervall $[x_{i-1},x_i]$ nach Satz 8.25 an einer Stelle y_i ein Maximum und an einer Stelle z_i ein Minimum an. Da y_i und z_i beide im Intervall $[x_{i-1},x_i]$ liegen, dessen Länge ja kleiner als δ ist, ist natürlich auch $|y_i - z_i| < \delta$ und damit $|f(y_i) - f(z_i)| < \frac{\varepsilon}{b-a}$ nach Wahl von δ. Wir können oben also weiterrechnen und erhalten

$$\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) = \sum_{i=1}^{n} (x_i - x_{i-1}) \cdot (f(y_i) - f(z_i)) < \sum_{i=1}^{n} (x_i - x_{i-1}) \cdot \frac{\varepsilon}{b-a} = \varepsilon,$$

woraus nun mit Lemma 12.11 (a) die Behauptung folgt. □

Als Nächstes wollen wir die wichtigsten elementaren Eigenschaften von integrierbaren Funktionen herleiten.

Satz 12.13 (Eigenschaften des Integrals)**.** *Es seien $f,g\colon [a,b] \to \mathbb{R}$ integrierbare Funktionen und $c \in \mathbb{R}$. Dann gilt:*

(a) *$f+g$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt*

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

(b) *$c\,f$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt*

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

(c) *Ist $f \le g$, d. h. $f(x) \le g(x)$ für alle $x \in [a,b]$, so ist $\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx$.*

(d) *$|f|$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt die **Dreiecksungleichung***

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx.$$

Beweis. Wir verwenden das Riemannsche Integrabilitätskriterium aus Lemma 12.11.

(a) Da f und g integrierbar sind, gibt es für alle $\varepsilon > 0$ nach Lemma 12.11 (b) Zerlegungen Z und Z' von $[a,b]$ mit

$$\mathrm{OS}(f,Z) < \int_a^b f(x)\,dx + \frac{\varepsilon}{2} \qquad \text{und} \qquad \mathrm{OS}(g,Z') < \int_a^b g(x)\,dx + \frac{\varepsilon}{2}.$$

Nach Aufgabe 12.4 (a) und Lemma 12.3 (a) folgt daraus

$$\begin{aligned}\mathrm{OS}(f+g,Z\cup Z') &\le \mathrm{OS}(f,Z\cup Z') + \mathrm{OS}(g,Z\cup Z') \le \mathrm{OS}(f,Z) + \mathrm{OS}(g,Z') \\ &< \int_a^b f(x)\,dx + \int_a^b g(x)\,dx + \varepsilon.\end{aligned}$$

Analog finden wir für die Untersummen Zerlegungen $\tilde{Z}$ und $\tilde{Z}'$ mit

$$\mathrm{US}(f+g,\tilde{Z}\cup\tilde{Z}') > \int_a^b f(x)\,dx + \int_a^b g(x)\,dx - \varepsilon.$$

Die Behauptung folgt nun aus Lemma 12.11 (b) angewendet auf $f+g$.

(b) Für $c = 0$ ist die Aussage trivial. Es sei nun $c > 0$. Zu gegebenem $\varepsilon > 0$ gibt es dann wieder eine Zerlegung Z von $[a,b]$ mit $\mathrm{OS}(f,Z) < \int_a^b f(x)\,dx + \frac{\varepsilon}{c}$. Damit folgt aus Aufgabe 12.4 (b) dann

$$\mathrm{OS}(cf,Z) = c\cdot\mathrm{OS}(f,Z) < c\cdot\int_a^b f(x)\,dx + \varepsilon.$$

Eine analoge Abschätzung bekommen wir natürlich auch wieder für die Untersummen. Damit folgt die Behauptung für $c > 0$ aus Lemma 12.11 (b).

Für $c < 0$ ergibt sich die Behauptung genauso aus der analog zu zeigenden Aussage $\mathrm{OS}(cf,Z) = c\cdot\mathrm{US}(f,Z)$.

(c) Aus $f \le g$ folgt sofort $\mathrm{OS}(f,Z) \le \mathrm{OS}(g,Z)$ für jede Zerlegung Z, und damit durch Übergang zum Infimum über alle Z auch $\int_a^b f(x)\,dx = \mathrm{OI}(f) \le \mathrm{OI}(g) = \int_a^b g(x)\,dx$.

(d) Wir zeigen zunächst die Integrierbarkeit von $|f|$. Dazu sei wieder $\varepsilon > 0$ gegeben; nach Lemma 12.11 (a) können wir eine Zerlegung Z von $[a,b]$ wählen mit $\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) < \varepsilon$. Mit Aufgabe 12.4 (c) folgt dann aber auch

$$\mathrm{OS}(|f|,Z) - \mathrm{US}(|f|,Z) \le \mathrm{OS}(f,Z) - \mathrm{US}(f,Z) < \varepsilon,$$

und damit ist $|f|$ nach Lemma 12.11 (a) integrierbar. Die Abschätzung des Integrals erhalten wir nun aus (c): Wegen $f \le |f|$ und $-f \le |f|$ ist sowohl

$$\int_a^b f(x)\,dx \le \int_a^b |f(x)|\,dx \qquad \text{als auch} \qquad -\int_a^b f(x)\,dx \overset{\text{(b)}}{=} \int_a^b -f(x)\,dx \le \int_a^b |f(x)|\,dx,$$

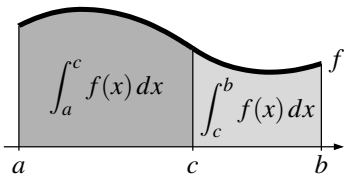
woraus sich die Behauptung ergibt, da $\left|\int_a^b f(x)\,dx\right|$ in jedem Fall eine dieser beiden linken Seiten ist. □

26

Eine weitere sehr anschauliche Eigenschaft von Integralen ist die sogenannte Additivität: für jede Zwischenstelle $c \in (a,b)$ ist die Fläche unter dem gesamten Graphen von $f\colon [a,b] \to \mathbb{R}$ gleich der Summe der Flächen von a bis c und von c bis b.

Satz 12.14 (Additivität des Integrals). *Es seien $f\colon [a,b] \to \mathbb{R}$ eine Funktion und $c \in (a,b)$. Ist f dann sowohl auf $[a,c]$ als auch auf $[c,b]$ integrierbar, so auch auf $[a,b]$, und es gilt*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$



Beweis. Der Beweis ist sehr ähnlich zu dem von Satz 12.13 (a). Es sei $\varepsilon > 0$ gegeben. Da f auf $[a,c]$ und $[c,b]$ integrierbar ist, gibt es Zerlegungen Z bzw. Z' dieser beiden Intervalle, so dass

$$\mathrm{OS}(f|_{[a,c]},Z) < \int_a^c f(x)\,dx + \frac{\varepsilon}{2} \qquad \text{und} \qquad \mathrm{OS}(f|_{[c,b]},Z') < \int_c^b f(x)\,dx + \frac{\varepsilon}{2}.$$

Die Obersumme von f bezüglich der Zerlegung $Z \cup Z'$ ist dann offensichtlich gerade die Summe dieser beiden Teilobersummen, d. h. wir haben

$$\mathrm{OS}(f,Z\cup Z') = \mathrm{OS}(f|_{[a,c]},Z) + \mathrm{OS}(f|_{[c,b]},Z') < \int_a^c f(x)\,dx + \int_c^b f(x)\,dx + \varepsilon,$$

und eine analoge Aussage auch genauso für die Untersummen. Damit folgt die Behauptung aus Lemma 12.11 (b). □

Notation 12.15 (Integrale mit vertauschten Grenzen)**.** Bisher haben wir Integrale $\int_a^b f(x)\,dx$ nur für $a \le b$ definiert. Ist hingegen $a > b$, so vereinbaren wir die Notation

$$\int_a^b f(x)\,dx := -\int_b^a f(x)\,dx, \tag{$*$}$$

wenn f auf $[b,a]$ integrierbar ist. Dies hat den Vorteil, dass die Formel aus Satz 12.14 (im Fall der Integrierbarkeit) dann nicht nur für $a \le c \le b$, sondern für beliebige a,b,c gilt: Ist z. B. $a < b < c$, so ist nach Satz 12.14

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx,$$

was (durch Subtraktion von $\int_b^c f(x)\,dx$ auf beiden Seiten) mit der Konvention $(*)$ wieder die gleiche Form

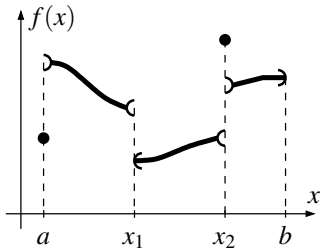
$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

wie in Satz 12.14 hat.

Beispiel 12.16 (Stückweise stetige Funktionen)**.** Es sei $f\colon [a,b] \to \mathbb{R}$ eine Funktion. Wir nennen einen Punkt $c \in [a,b]$ eine **Sprungstelle** von f, wenn die drei Zahlen

$$\lim_{\substack{x\to c\\ x<c}} f(x), \quad \lim_{\substack{x\to c\\ x>c}} f(x) \quad \text{und} \quad f(c)$$

existieren, aber nicht alle gleich sind (falls c einer der Randpunkte des Intervalls ist, gibt es den Grenzwert natürlich nur von einer der beiden Seiten). Man nennt f **stückweise stetig**, wenn f wie im Bild rechts stetig bis auf endlich viele Sprungstellen ist.



Eine solche stückweise stetige Funktion ist stets integrierbar:

(a) Es seien $a = x_0 < x_1 < \cdots < x_n = b$ die Sprungstellen und Randpunkte des Definitionsintervalls. Auf jedem Teilintervall $[x_{i-1},x_i]$ für $i = 1,\dots,n$ ist f dann eine stetige Funktion mit evtl. abgeänderten Funktionswerten an den Rändern, also die Summe aus einer stetigen Funktion und geeigneten Vielfachen der „Sprungfunktionen“

$$[x_{i-1},x_i] \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x = x_{i-1}, \\ 0 & \text{für } x > x_{i-1} \end{cases} \quad \text{und} \quad [x_{i-1},x_i] \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x = x_i, \\ 0 & \text{für } x < x_i. \end{cases}$$

Da eine stetige Funktion und diese Sprungfunktionen nach Satz 12.12 und Beispiel 12.9 (c) integrierbar sind, ist nach Satz 12.13 (a) und (b) auch $f|_{[x_{i-1},x_i]}$ integrierbar.

(b) Nach der Additivität aus Satz 12.14 ist f damit auch auf $[a,b]$ integrierbar.

Aufgabe 12.17. Zeige, dass die folgenden Funktionen integrierbar sind:

(a) eine beliebige monotone Funktion $f\colon [a,b] \to \mathbb{R}$;

(b) $f\colon [-1,1] \to \mathbb{R},\ x \mapsto \begin{cases} \sin\frac{1}{x} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0; \end{cases}$

(c) $f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} \frac{1}{q} & \text{für } x \in \mathbb{Q} \text{ mit gekürzter Darstellung } x = \frac{p}{q} \text{ für } p \in \mathbb{N} \text{ und } q \in \mathbb{N}_{>0}, \\ 0 & \text{für } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$

Aufgabe 12.18. Für eine Zerlegung $Z = \{x_0, \ldots, x_k\}$ eines Intervalls $[a,b]$ definieren wir die *Feinheit* $l(Z)$ als den größten Abstand $\max\{x_i - x_{i-1} : i = 1, \ldots, k\}$ zwischen zwei benachbarten Punkten von Z.

Es seien nun eine Folge (Z_n) von Zerlegungen $Z_n = \{x_{n,0}, \ldots, x_{n,k_n}\}$ von $[a,b]$ mit $\lim\limits_{n\to\infty} l(Z_n) = 0$ sowie Zwischenpunkte $\xi_{n,i} \in [x_{n,i-1}, x_{n,i}]$ für $n \in \mathbb{N}$ und $i = 1, \ldots, k_n$ gegeben. Zeige, dass dann für jede stetige Funktion $f\colon [a,b] \to \mathbb{R}$

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{k_n} (x_{n,i} - x_{n,i-1})\, f(\xi_{n,i})$$

gilt, also dass man das Integral statt mit Ober- und Untersummen auch mit einer beliebigen „Zwischensumme“ berechnen kann.

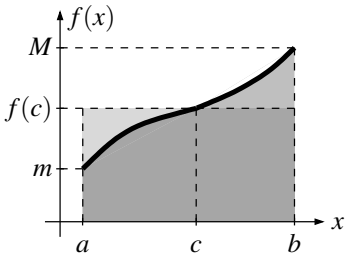
12.B Stammfunktionen

Unsere bisherigen Ergebnisse erlauben es uns zwar, von vielen Funktionen die Integrierbarkeit nachzuweisen, aber noch nicht, den Wert des Integrals dann auch explizit zu berechnen. Wie wir in der Einleitung dieses Kapitels schon motiviert haben, ist das zentrale Resultat für solche Berechnungen die Aussage, dass die Integration die Umkehrung der Differentiation ist. Um dies zu zeigen, benötigen wir zur Vorbereitung noch einen kleinen und sehr anschaulichen Hilfssatz.

Satz 12.19 (Mittelwertsatz der Integralrechnung). *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion. Dann gibt es einen Punkt $c \in [a,b]$ mit*

$$\int_a^b f(x)\,dx = f(c) \cdot (b-a)$$

(d. h. die Fläche unter dem Graphen ist wie im Bild rechts gleich der Fläche eines Rechtecks, dessen Höhe ein Funktionswert $f(c)$ auf dem betrachteten Intervall ist).



Beweis. Nach Satz 8.25 nimmt f als stetige Funktion auf einem abgeschlossenen Intervall ein Maximum M und Minimum m an. Also folgt aus Beispiel 12.9 (a) und Satz 12.13 (c)

$$m(b-a) = \int_a^b m\,dx \le \int_a^b f(x)\,dx \le \int_a^b M\,dx = M(b-a),$$

und damit

$$m \le \frac{1}{b-a} \cdot \int_a^b f(x)\,dx \le M.$$

Da f stetig ist, gibt es nun nach dem Zwischenwertsatz 8.21 ein $c \in [a,b]$ mit $f(c) = \frac{1}{b-a} \cdot \int_a^b f(x)\,dx$ – was genau die Behauptung war. □

Bemerkung 12.20. Mit Notation 12.15 gilt die Gleichung $\int_a^b f(x)\,dx = f(c)\,(b-a)$ für ein c zwischen a und b auch für den Fall $a > b$: Anwenden des Mittelwertsatzes 12.19 auf das Intervall $[b,a]$ liefert dann zunächst $\int_b^a f(x)\,dx = f(c)\,(a-b)$, woraus wir aber durch Multiplikation mit -1 wieder die Form $\int_a^b f(x)\,dx = f(c)\,(b-a)$ erhalten können.

Satz 12.21 (Hauptsatz der Differential- und Integralrechnung). *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion. Dann ist die Funktion*

$$F\colon [a,b] \to \mathbb{R},\ x \mapsto \int_a^x f(t)\,dt$$

(bei der wir also das Integral von f berechnen und dabei die Obergrenze als Variable nehmen) differenzierbar mit $F' = f$.

Beweis. Wir zeigen die Differenzierbarkeit von F in einem Punkt $c \in [a,b]$ mit dem Folgenkriterium. Es sei also $(x_n)_n$ eine beliebige Folge in $[a,b]\setminus\{c\}$ mit $x_n \to c$. Dann gilt für alle n

$$\begin{aligned} F(x_n) - F(c) &= \int_a^{x_n} f(t)\,dt - \int_a^c f(t)\,dt \\ &= \int_c^{x_n} f(t)\,dt && \text{(Satz 12.14 bzw. Notation 12.15)} \\ &= f(z_n)\,(x_n - c) && \text{(Satz 12.19 bzw. Bemerkung 12.20)} \end{aligned}$$

für ein z_n zwischen c und x_n. Beachte, dass wegen $x_n \to c$ auch $z_n \to c$ gilt, da z_n ja immer zwischen c und x_n liegt. Da f stetig ist, haben wir damit

$$\lim_{n\to\infty} \frac{F(x_n) - F(c)}{x_n - c} = \lim_{n\to\infty} f(z_n) = f(c)$$

nach dem Folgenkriterium für Stetigkeit aus Satz 8.11 (b). Wiederum nach dem Folgenkriterium gemäß Satz 8.11 (a) bedeutet dies nun aber gerade wie gewünscht

$$F'(c) = \lim_{x\to c} \frac{F(x) - F(c)}{x - c} = f(c). \qquad \square$$

Beispiel 12.22. Die Voraussetzung der Stetigkeit im Hauptsatz 12.21 ist wirklich notwendig: Betrachten wir z. B. noch einmal die unstetige Funktion

$$f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x = 0, \\ 0 & \text{für } x > 0 \end{cases}$$

aus Beispiel 12.9 (c), so ist hier

$$F(x) = \int_0^x f(t)\,dt = 0 \quad \text{für alle } x \in [0,1].$$

Diese Funktion ist zwar differenzierbar, hat als Ableitung jedoch die Nullfunktion und nicht f.

Für die Integralberechnung benötigen wir also Funktionen, deren Ableitung die ursprünglich gegebene Funktion ist. Wir geben solchen Funktionen daher einen besonderen Namen. Wegen Beispiel 12.22 beschränken wir uns dabei auf stetige Funktionen.

Definition 12.23 (Stammfunktionen)**.** Es seien $D \subset \mathbb{R}$ ein Intervall und $f\colon D \to \mathbb{R}$ eine stetige Funktion. Dann heißt eine differenzierbare Funktion $F\colon D \to \mathbb{R}$ mit $F' = f$ eine **Stammfunktion** von f.

Folgerung 12.24 (Integralberechnung mit Stammfunktionen)**.** *Es sei $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion. Dann gilt:*

(a) *f besitzt eine Stammfunktion.*

(b) *Sind F und G zwei Stammfunktionen von f, so unterscheiden sich diese nur um eine additive Konstante, d. h. es gibt ein $c \in \mathbb{R}$ mit $F - G = c$.*

(c) *Ist F eine Stammfunktion von f, so gilt*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Beweis.

(a) folgt unmittelbar aus Satz 12.21: Die dort angegebene Funktion $x \mapsto \int_a^x f(t)\,dt$ ist eine Stammfunktion von f.

(b) Nach Voraussetzung ist $(F-G)' = F' - G' = f - f = 0$. Damit ist $F - G$ nach Folgerung 10.25 (c) konstant.

(c) Nach dem Hauptsatz 12.21 sind sowohl F als auch $x \mapsto \int_a^x f(t)\,dt$ Stammfunktionen von f. Also gibt es nach (b) eine Konstante $c \in \mathbb{R}$ mit

$$F(x) - \int_a^x f(t)\,dt = c$$

für alle $x \in [a,b]$. Einsetzen von $x = a$ liefert $F(a) = c$, und damit

$$F(x) - F(a) = \int_a^x f(t)\,dt.$$

Für $x = b$ ergibt sich nun die Behauptung. □

Notation 12.25 (Unbestimmte Integrale)**.** Man schreibt die Differenz $F(b) - F(a)$ in Folgerung 12.24 (c) oft auch als $\big[F(x)\big]_{x=a}^b$ oder $F(x)\big|_{x=a}^b$ (wenn die Integrationsvariable aus dem Zusammenhang klar ist, auch als $\big[F(x)\big]_a^b$ oder $F(x)\big|_a^b$), und die dortige Gleichung damit als

$$\int_a^b f(x)\,dx = \big[F(x)\big]_a^b.$$

Da dies für beliebige Integrationsgrenzen a und b gilt, vereinfacht man diese Notation oft noch weiter und schreibt gemäß Folgerung 12.24 einfach

$$\int f(x)\,dx = F(x), \tag{$*$}$$

für die Aussage, dass F eine Stammfunktion von f ist. Man bezeichnet dies dann auch als ein **unbestimmtes Integral**. Beachte aber, dass $(*)$ nur eine symbolische Schreibweise ist, die erst nach dem Einsetzen von Grenzen zu einer echten Gleichheit in $\mathbb{R}$ wird! Dies kann man alleine schon daran sehen, dass x von der Notation her auf der linken Seite eine Integrationsvariable ist, auf der rechten Seite aber wie eine freie Variable aussieht (siehe Bemerkung 12.8 (a)). Außerdem ist mit F z. B. auch $F + 1$ eine Stammfunktion von f, d. h. wir können sowohl

$$\int f(x)\,dx = F(x) \qquad \text{als auch} \qquad \int f(x)\,dx = F(x) + 1$$

schreiben – was natürlich sofort zum Widerspruch $F(x) = F(x) + 1$ führen würde, wenn man dies als echte Gleichungen von Funktionen betrachten dürfte.

Beispiel 12.26. Wir können nun viele Integrale konkret berechnen, indem wir vom Integranden eine Stammfunktion suchen:

(a) Ist $f(x) = x^a$ für ein $a \in \mathbb{R}\backslash\{-1\}$, so ist $F(x) = \frac{1}{a+1}x^{a+1}$ nach Beispiel 10.30 (d) eine Stammfunktion von f, d. h. mit Notation 12.25 gilt

$$\int x^a\,dx = \frac{1}{a+1}x^{a+1} \qquad \text{für } a \neq -1.$$

Konkret können wir damit z. B. das Integral aus Beispiel 12.9 (b) auch ohne komplizierte Berechnung von Ober- und Untersummen bestimmen: Es ist einfach

$$\int_0^1 x\,dx = \left[\frac{1}{2}x^2\right]_{x=0}^1 = \frac{1}{2}\cdot 1^2 - \frac{1}{2}\cdot 0^2 = \frac{1}{2}.$$

(b) Für $a = -1$ ist die Formel aus (a) natürlich nicht anwendbar. Wir haben aber glücklicherweise mit dem Logarithmus schon eine Funktion kennengelernt, deren Ableitung nach Beispiel 10.30 (c) gleich $\frac{1}{x}$ ist: Es ist also

$$\int \frac{1}{x}\,dx = \log x$$

für Integrationsintervalle in $\mathbb{R}_{>0}$ (so dass der Logarithmus dort definiert ist). Falls das Integrationsintervall in $\mathbb{R}_{<0}$ liegt, können wir als Stammfunktion $\log(-x)$ nehmen, denn auch die Ableitung dieser Funktion ist ja gleich $\frac{1}{x}$. Insgesamt ist damit

$$\int \frac{1}{x}\,dx = \log|x|.$$

(c) Durch die Ableitungen der speziellen Funktionen, die wir in Beispiel 10.30 berechnet haben, sehen wir genauso z. B.

$$\int e^x\,dx = e^x, \qquad \int \cos x\,dx = \sin x \qquad \text{und} \qquad \int \sin x\,dx = -\cos x.$$

Nach der Einführung von Stammfunktionen wollen wir unseren Integralbegriff nun noch etwas erweitern. Bisher konnten wir Integrale nur auf abgeschlossenen Intervallen $[a,b]$ im Definitionsbereich des Integranden berechnen. Oft treten allerdings Fälle auf, in denen eine oder beide Integrationsgrenzen entweder nicht mehr im Definitionsbereich liegen oder aber gleich $\pm\infty$ sind. Auch in diesen Fällen kann man durch eine einfache Grenzwertbildung das Integral definieren.

Definition 12.27 (Uneigentliche Integrale)**.**

(a) Es sei $f\colon [a,b) \to \mathbb{R}$ eine reelle Funktion (wobei der Fall $b=\infty$ zugelassen ist). Wir nehmen weiterhin an, dass f auf jedem abgeschlossenen Intervall $[a,c]$ mit $a \le c < b$ integrierbar ist. Existiert dann der Grenzwert

$$\int_a^b f(x)\,dx := \lim_{\substack{c\to b\\ c<b}} \int_a^c f(x)\,dx \qquad \in \mathbb{R}\cup\{\pm\infty\},$$

so nennen wir ihn das **uneigentliche Integral** von f auf $[a,b)$. Liegt dieser Grenzwert zusätzlich in $\mathbb{R}$, so heißt das uneigentliche Integral $\int_a^b f(x)\,dx$ **konvergent**, andernfalls **divergent**. Analog definiert man uneigentliche Integrale im Fall $f\colon (a,b] \to \mathbb{R}$.

(b) Es sei nun $f\colon (a,b) \to \mathbb{R}$, wobei die Fälle $a=-\infty$ bzw. $b=\infty$ wieder zugelassen sind. Für ein $c \in (a,b)$ setzen wir dann

$$\int_a^b f(x)\,dx := \int_a^c f(x)\,dx + \int_c^b f(x)\,dx,$$

sofern die rechte Summe (von zwei uneigentlichen Integralen gemäß (a)) in $\mathbb{R}\cup\{\pm\infty\}$ existiert. Beachte, dass diese Summe dann wegen der Additivität des Integrals aus Satz 12.14 nicht von der Wahl des Zwischenpunktes c abhängt. Wie im Fall (a) spricht man auch hier von einem (beidseitig) uneigentlichen Integral bzw. von der Konvergenz oder Divergenz dieses Integrals.

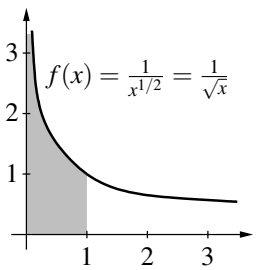
Beispiel 12.28.

(a) Für $a \in \mathbb{R}_{>0}\backslash\{1\}$ ist

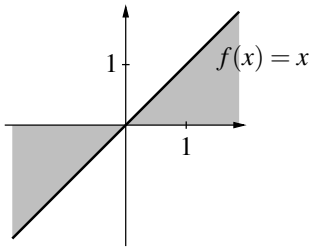
$$\int_0^1 \frac{1}{x^a}\,dx = \lim_{\substack{c\to 0\\ c>0}} \int_c^1 \frac{1}{x^a}\,dx = \lim_{\substack{c\to 0\\ c>0}} \left[\frac{1}{1-a}x^{1-a}\right]_c^1 = \frac{1}{1-a}\lim_{\substack{c\to 0\\ c>0}}(1-c^{1-a})$$

$$= \begin{cases} \frac{1}{1-a} & \text{für } a<1,\\ \infty & \text{für } a>1. \end{cases}$$

Das uneigentliche Integral konvergiert also genau für $a<1$. Anschaulich bedeutet dies, dass die Fläche von 0 bis 1 unter dem Graphen von $x \mapsto \frac{1}{x^a}$ in diesem Fall (wie im Bild unten links für $a=\frac{1}{2}$ dargestellt) endlich ist, obwohl sie nach oben eine unendliche Ausdehnung hat.



(a)



(b)

(b) Das uneigentliche Integral der Identität $f(x) = x$ auf $(-\infty, \infty)$ existiert nicht, denn bei der Wahl des Zwischenpunktes 0 erhalten wir den unbestimmten Ausdruck

$$\begin{aligned}\int_{-\infty}^{\infty} x\,dx &= \int_{-\infty}^{0} x\,dx + \int_{0}^{\infty} x\,dx = \lim_{c\to-\infty}\int_{c}^{0} x\,dx + \lim_{c\to\infty}\int_{0}^{c} x\,dx \\ &= \lim_{c\to-\infty}\left[\frac{1}{2}x^2\right]_c^0 + \lim_{c\to\infty}\left[\frac{1}{2}x^2\right]_0^c = \frac{1}{2}\lim_{c\to-\infty} -c^2 + \frac{1}{2}\lim_{c\to\infty} c^2 \\ &= -\infty + \infty.\end{aligned}$$

Auch anschaulich ist im Bild oben rechts ersichtlich, dass sich dieser Flächeninhalt aus einer unendlich großen negativen und positiven Fläche zusammensetzt. Beachte, dass wir nicht das gleiche Ergebnis erhalten hätten, wenn wir das uneigentliche Integral symmetrisch um die vertikale Achse als

$$\lim_{c\to\infty}\int_{-c}^{c} x\,dx = \lim_{c\to\infty}\left[\frac{1}{2}x^2\right]_{-c}^{c} = \frac{1}{2}\lim_{c\to\infty}(c^2 - c^2) = 0$$

definiert hätten!

27

12.C Integrationsregeln

In Abschnitt 12.B haben wir alle Stammfunktionen zur Berechnung von Integralen letztlich „durch Zufall“ gefunden – also weil wir uns einfach an eine Funktion erinnern konnten, deren Ableitung wir schon einmal berechnet haben und bei der für diese Ableitung dann die gegebene Funktion herauskam. Daher müssen wir uns jetzt natürlich fragen, wie man Stammfunktionen berechnen kann, wenn man nicht gerade zufällig eine solche sieht. Gibt es analog zur Berechnung von Ableitungen auch Regeln, mit denen man, wenn man die Stammfunktionen einiger spezieller Funktionen kennt, auch die Stammfunktionen z. B. ihrer Produkte, Quotienten oder Verkettungen berechnen kann?

Leider gibt es keine solchen universellen Regeln. Dies ist auch der Grund dafür, dass in mathematischen Formelsammlungen oft seitenweise Tabellen von Stammfunktionen stehen, während man für das Differenzieren aufgrund der Produkt-, Quotienten- und Kettenregel keine derartigen Tabellen benötigt. Es gibt jedoch auch für die Integration ein paar Regeln, mit denen man Integrale oft berechnen kann – nur ist es je nach der betrachteten Funktion mehr oder weniger schwierig (oder evtl. sogar unmöglich), einen Weg zu finden, um mit diesen Regeln ans Ziel zu kommen.

Wir wollen nun die wichtigsten derartigen Regeln behandeln. Die erste ist im wesentlichen nur die „umgekehrte Richtung“ der Produktregel der Differentiation:

Satz 12.29 (**Partielle Integration** bzw. **Produktintegration**). *Es seien $u, v\colon [a,b] \to \mathbb{R}$ stetig differenzierbare Funktionen. Dann gilt*

$$\int_a^b u'(x)v(x)\,dx = \big[u(x)v(x)\big]_a^b - \int_a^b u(x)v'(x)\,dx$$

(bzw. als unbestimmtes Integral $\int u'(x)v(x)\,dx = u(x)v(x) - \int u(x)v'(x)\,dx$).

Beweis. Nach der Produktregel aus Satz 10.8 (b) ist uv eine Stammfunktion von $(uv)' = u'v + uv'$. Also ist $\int_a^b (u'(x)v(x) + u(x)v'(x))\,dx = \big[u(x)v(x)\big]_a^b$. Die Behauptung folgt dann durch Subtraktion von $\int_a^b u(x)v'(x)\,dx$ nach Satz 12.13. □

Beispiel 12.30. Die Regel aus Satz 12.29 nennt sich „partielle Integration“, weil bei der Berechnung des Integrals auf der linken Seite mit Hilfe der rechten neben einem „ausintegrierten Anteil“ noch ein anderes Integral übrig bleibt – nämlich eines, bei dem wir von einem Faktor des ursprünglichen Integrals die Ableitung und vom anderen eine Stammfunktion gebildet haben. Die Anwendung dieser Regel ist also vor allem dann sinnvoll, wenn dieses neue Integral bereits bekannt oder zumindest einfacher als das ursprüngliche ist. Hier sind zwei Beispiele dafür.

(a) Zur Berechnung von $\int x\cos x\,dx$ setzen wir (mit den Notationen von Satz 12.29)

$$\begin{aligned} u(x) &= \sin x & u'(x) &= \cos x \\ v(x) &= x & v'(x) &= 1 \end{aligned}$$

und erhalten

$$\int x\cos x\,dx = x\sin x - \int \sin x\,dx = x\sin x + \cos x.$$

Wir haben bei der Anwendung der partiellen Integration also den Faktor x differenziert und den Faktor $\cos x$ integriert. Möchte man dies in der Rechnung deutlich machen (und die Funktionen u, u', v, v' nicht explizit hinschreiben), so notiert man dies auch oft als

$$\int \underset{\downarrow}{x}\,\underset{\uparrow}{\cos x}\,dx = x\sin x - \int \sin x\,dx = x\sin x + \cos x.$$

Beachte, dass die umgekehrte Wahl hier nicht zum Ziel geführt hätte: Die Rechnung

$$\int \underset{\uparrow}{x}\,\underset{\downarrow}{\cos x}\,dx = \frac{x^2}{2}\cos x + \int \frac{x^2}{2}\sin x\,dx$$

ist zwar korrekt, aber das neue Integral ist hier komplizierter als das ursprüngliche.

(b) Das Integral $\int \log x\,dx$ lässt sich mit einem Trick ebenfalls durch partielle Integration berechnen:

$$\int \log x\,dx = \int \underset{\uparrow}{1}\cdot\underset{\downarrow}{\log x}\,dx = x\log x - \int x\cdot\frac{1}{x}\,dx = x\log x - x.$$

Dieser Trick funktioniert hier, weil aus der (komplizierten) Logarithmusfunktion beim Ableiten die sehr viel einfachere Funktion $\frac{1}{x}$ entsteht. Auf die gleiche Art kann man übrigens auch die Integrale der Arkusfunktionen aus Definition 9.24 berechnen, da auch diese bei der Differentiation sehr viel einfacher werden (siehe Beispiel 10.30 (c)).

Die zweite wichtige Integrationsregel ergibt sich analog aus der Kettenregel der Differentiation.

Satz 12.31 (Substitutionsregel). *Es seien $f\colon [a,b]\to\mathbb{R}$ eine stetig differenzierbare Funktion und $g\colon D\to\mathbb{R}$ eine stetige Funktion mit $f([a,b])\subset D$. Dann gilt*

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

Beweis. Nach Folgerung 8.26 ist $f([a,b])$ ein abgeschlossenes Intervall, und damit hat g nach Folgerung 12.24 (a) dort eine Stammfunktion G. Die Kettenregel aus Satz 10.10 liefert dann $(G\circ f)'(x) = g(f(x))\,f'(x)$. Damit ist die linke Seite der zu beweisenden Gleichung

$$\int_a^b (G\circ f)'(x)\,dx = \big[G(f(x))\big]_{x=a}^b = G(f(b)) - G(f(a)),$$

und die rechte Seite ebenfalls

$$\int_{f(a)}^{f(b)} g(y)\,dy = \big[G(y)\big]_{y=f(a)}^{f(b)} = G(f(b)) - G(f(a)).$$ □

Bemerkung 12.32. Die Substitutionsregel nimmt in der Differentialschreibweise aus Notation 10.14 eine besonders einfache Form an: Setzen wir $y = f(x)$ und damit $\frac{dy}{dx} = f'(x)$, und bezeichnen wir die Integrationsgrenzen mit $x_1 = a$ und $x_2 = b$ bzw. $y_1 = f(a)$ und $y_2 = f(b)$, so schreibt sich die Substitutionsregel als

$$\int_{x_1}^{x_2} g(y)\,\frac{dy}{dx}\,dx = \int_{y_1}^{y_2} g(y)\,dy,$$

oder analog zu Notation 12.25 einfach als

$$\int g(y)\,\frac{dy}{dx}\,dx = \int g(y)\,dy,$$

wenn man darauf achtet, dass die Grenzen passend zur Integrationsvariablen gewählt werden. Die Regel sieht dann also einfach wie ein „formales Erweitern mit dx“ aus.

Beispiel 12.33. Die Substitutionsregel bietet sich natürlich immer dann an, wenn die zu integrierende Funktion eine Verkettung von zwei anderen Funktionen ist oder enthält – und insbesondere dann, wenn die Ableitung der inneren Funktion zusätzlich auch noch als Faktor im Integranden steht.

(a) Beim Integral $\int x\mathrm{e}^{x^2}\,dx$ stellen wir fest, dass sich im Integranden eine verkettete Funktion e^{x^2} befindet, und dass die Ableitung $2x$ der inneren Funktion x^2 auch (bis auf die Konstante 2) zusätzlich noch als Faktor im Integranden steht. Wir substituieren also $y = x^2$, so dass $\frac{dy}{dx} = 2x$ in der Notation von Bemerkung 12.32 gilt. Damit folgt also

$$\int x\mathrm{e}^{x^2}\,dx = \frac{1}{2}\int \frac{dy}{dx}\mathrm{e}^y\,dx \overset{12.31}{=} \frac{1}{2}\int \mathrm{e}^y\,dy = \frac{1}{2}\mathrm{e}^y = \frac{1}{2}\mathrm{e}^{x^2}.$$

Im Fall eines bestimmten Integrals hätten wir bei der Anwendung von Satz 12.31 die Grenzen mitsubstituieren müssen:

$$\int_a^b x\mathrm{e}^{x^2}\,dx = \frac{1}{2}\int_a^b \frac{dy}{dx}\mathrm{e}^y\,dx \overset{12.31}{=} \frac{1}{2}\int_{a^2}^{b^2} \mathrm{e}^y\,dy = \frac{1}{2}\left[\mathrm{e}^y\right]_{y=a^2}^{b^2} = \frac{1}{2}\left(\mathrm{e}^{b^2} - \mathrm{e}^{a^2}\right).$$

Beachte, dass der Faktor x im Integranden bei diesem Beispiel ganz wesentlich dafür war, dass die Substitutionsregel zum Ziel geführt hat: Ohne diesen Faktor hätten wir mit derselben Substitution

$$\int \mathrm{e}^{x^2}\,dx = \int \frac{1}{2x}\cdot\frac{dy}{dx}\mathrm{e}^y\,dx = \frac{1}{2}\int \frac{\mathrm{e}^y}{\sqrt{y}}\,dy$$

erhalten – was zwar auch richtig ist, aber nicht weiter hilft, weil das neue Integral auch nicht einfacher als das ursprüngliche zu berechnen ist. In der Tat kann man zeigen, dass sich die Stammfunktion von e^{x^2} nicht durch die uns bisher bekannten „speziellen Funktionen“ ausdrücken lässt.

(b) Besonders einfach wird die Substitutionsregel im Fall der sogenannten *linearen Substitution*: Ist f eine beliebige stetige Funktion, deren Stammfunktion F wir kennen, so können wir damit immer auch die Stammfunktion von $f(ax+b)$ mit $a,b \in \mathbb{R}$ und $a \neq 0$ bestimmen, da die innere Ableitung hier eine Konstante ist: Substituieren wir $y = ax+b$, so ergibt sich wegen $\frac{dy}{dx} = a$

$$\int f(ax+b)\,dx = \frac{1}{a}\int f(y)\frac{dy}{dx}\,dx \overset{12.31}{=} \frac{1}{a}\int f(y)\,dy = \frac{1}{a}F(y) = \frac{1}{a}F(ax+b).$$

Konkret ist also z. B.

$$\int \frac{1}{2x+3}\,dx = \frac{1}{2}\log|2x+3|,$$

da $x \mapsto \log|x|$ nach Beispiel 12.26 (b) eine Stammfunktion von $x \mapsto \frac{1}{x}$ ist.

Wir hatten in Beispiel 12.33 (a) bereits erwähnt, dass sich die Stammfunktionen von Funktionen, die aus unseren speziellen Funktionen zusammengesetzt sind, manchmal nicht wieder auf diese Art schreiben lassen. Für viele Klassen von Funktionen ist dies aber doch der Fall – z. B. für rationale Funktionen der Form $\frac{p}{q}$ für zwei Polynome p und q. Wir wollen dies hier nun zeigen, der Einfachheit halber allerdings nur für den Fall, dass das Nennerpolynom q in verschiedene Linearfaktoren zerfällt und größeren Grad als das Zählerpolynom p hat. Der Trick besteht in diesem Fall darin, den Ausdruck $\frac{p}{q}$ als Summe von Brüchen zu schreiben, die nur eine Konstante im Zähler und einen einzigen Linearfaktor im Nenner haben. Derartige Funktionen der Form $\frac{c}{x-a}$ lassen sich mit einer linearen Substitution wie in Beispiel 12.33 (b) dann einfach zu $c\log|x-a|$ integrieren.

Lemma 12.34 (Partialbruchzerlegung). *Es seien $n \in \mathbb{N}_{>0}$ und $a_1,\dots,a_n \in \mathbb{R}$ verschieden. Ferner sei p ein reelles Polynom mit $\deg p < n$. Dann gilt*

$$\frac{p(x)}{(x-a_1)\cdots(x-a_n)} = \sum_{i=1}^{n}\frac{c_i}{x-a_i} \quad \textit{für} \quad c_i := \frac{p(a_i)}{(a_i-a_1)\cdots(a_i-a_{i-1})(a_i-a_{i+1})\cdots(a_i-a_n)}$$

für alle $x \in \mathbb{R}\backslash\{a_1,\dots,a_n\}$.

Beweis. Offensichtlich ist

$$f(x) := p(x) - \sum_{i=1}^{n} p(a_i) \cdot \frac{(x-a_1)\cdots(x-a_{i-1})(x-a_{i+1})\cdots(x-a_n)}{(a_i-a_1)\cdots(a_i-a_{i-1})(a_i-a_{i+1})\cdots(a_i-a_n)} \qquad (*)$$

ein Polynom mit $\deg f < n$. Es hat aber jedes a_k für $k = 1,\dots,n$ als Nullstelle, denn es gilt

$$f(a_k) = p(a_k) - p(a_k) \cdot \frac{(a_k-a_1)\cdots(a_k-a_{k-1})(a_k-a_{k+1})\cdots(a_k-a_n)}{(a_k-a_1)\cdots(a_k-a_{k-1})(a_k-a_{k+1})\cdots(a_k-a_n)} = 0,$$

da nach Einsetzen von $x = a_k$ in der Summe über i in $(*)$ alle Terme mit $i \neq k$ einen Faktor 0 im Zähler haben und damit verschwinden. Nach Satz 3.18 (b) ist f also das Nullpolynom. Division von $(*)$ durch $(x-a_1)\cdots(x-a_n)$ liefert damit wie behauptet

$$0 = \frac{p(x)}{(x-a_1)\cdots(x-a_n)} - \sum_{i=1}^{n} \frac{c_i}{x-a_i}. \qquad \square$$

Bemerkung 12.35. Die Formel für die Koeffizienten c_i in Lemma 12.34 lässt sich leicht merken: Man erhält c_i, indem man $x = a_i$ im ursprünglichen Ausdruck $\frac{p(x)}{(x-a_1)\cdots(x-a_i)\cdots(x-a_n)}$ einsetzt – bis auf den Linearfaktor $x - a_i$ im Nenner, den man dabei weglassen muss, da er ansonsten ja auch zu einem Faktor 0 im Nenner führen würde.

Beispiel 12.36. Um das Integral $\int \frac{x}{x^2+3x+2}\,dx$ zu berechnen, führen wir eine Partialbruchzerlegung des Integranden durch: Mit $x^2+3x+2 = (x+1)(x+2)$ ist

$$\frac{x}{(x+1)(x+2)} = \frac{c_1}{x+1} + \frac{c_2}{x+2},$$

wobei sich $c_1 = \frac{-1}{1} = -1$ durch Einsetzen von $x = -1$ in $\frac{x}{x+2}$ und $c_2 = \frac{-2}{-1} = 2$ durch Einsetzen von $x = -2$ in $\frac{x}{x+1}$ ergibt. Also ist

$$\int \frac{x}{x^2+3x+2}\,dx = \int \Big(-\frac{1}{x+1} + 2\,\frac{1}{x+2}\Big)\,dx = -\log|x+1| + 2\log|x+2|$$

nach einer linearen Substitution wie in Beispiel 12.33 (b).

Als letzte Rechenregel zur Bestimmung von Integralen wollen wir nun noch untersuchen, wie man Integrale von Funktionen berechnen kann, die als Grenzwerte von Funktionenfolgen entstehen – also z. B. von Potenzreihen. Die Situation ist hier sehr viel einfacher als sie es bei der Differentiation in Satz 10.27 war.

Satz 12.37 (Vertauschbarkeit von Integration und Grenzwertbildung)**.** *Es seien* $f_n\colon [a,b] \to \mathbb{R}$ *stetige Funktionen, die gleichmäßig gegen eine (nach Satz 8.39 dann automatisch ebenfalls stetige) Grenzfunktion* $f\colon [a,b] \to \mathbb{R}$ *konvergieren. Dann gilt*

$$\int_a^b \Big(\lim_{n\to\infty} f_n(x)\Big)\,dx = \int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx,$$

d. h. „Grenzwertbildung und Integration können vertauscht werden“.

Beweis. Die erste behauptete Gleichheit ist natürlich nichts weiter als die Definition von f. Für die zweite sei $\varepsilon > 0$ gegeben. Wegen der gleichmäßigen Konvergenz von $(f_n)_n$ gibt es ein $n_0 \in \mathbb{N}$, so dass $|f_n(x) - f(x)| < \frac{\varepsilon}{2(b-a)}$ für alle $x \in [a,b]$ und $n \geq n_0$. Damit ergibt sich nach Satz 12.13

$$\begin{aligned}\left|\int_a^b f_n(x)\,dx - \int_a^b f(x)\,dx\right| &= \left|\int_a^b (f_n(x) - f(x))\,dx\right| \leq \int_a^b |f_n(x) - f(x)|\,dx \leq \int_a^b \frac{\varepsilon}{2(b-a)}\,dx \\ &= \frac{\varepsilon}{2} < \varepsilon.\end{aligned}$$

Nach Definition des Grenzwerts bedeutet dies genau $\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx$ für $n \to \infty$, was zu zeigen war. $\square$

Bemerkung 12.38 (Integration von Potenzreihen)**.** Insbesondere bedeutet Satz 12.37, dass Potenzreihen (die in jedem abgeschlossenen Intervall innerhalb des Konvergenzgebiets nach Satz 8.37 ja gleichmäßig konvergieren) gliedweise integriert werden können: Sind $f(x) = \sum_{k=0}^{\infty} c_k x^k$ eine Potenzreihe und $f_n(x) = \sum_{k=0}^{n} c_k x^k$ ihre Partialsummen, so folgt

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx = \lim_{n\to\infty} \left[\sum_{k=0}^{n} \frac{c_k}{k+1} x^{k+1}\right]_a^b = \left[\sum_{k=0}^{\infty} \frac{c_k}{k+1} x^{k+1}\right]_a^b$$

(sofern $[a,b]$ komplett im Konvergenzintervall der Potenzreihe liegt), d. h. als unbestimmtes Integral geschrieben ist

$$\int f(x)\,dx = \sum_{k=0}^{\infty} \frac{c_k}{k+1} x^{k+1}.$$

Dies zeigt noch einmal deutlich die besonders schönen Eigenschaften von Potenzreihen: Innerhalb ihres Konvergenzgebiets kann man mit ihnen praktisch „wie mit Polynomen rechnen“, d. h.

- sie können wie erwartet addiert und multipliziert werden (Lemma 7.4 und Bemerkung 7.36);
- sie sind beliebig oft differenzierbar und ihre Ableitungen können gliedweise berechnet werden (Folgerung 10.28 und Satz 11.9);
- Integrale können gliedweise berechnet werden;
- „viele Funktionen“ können als Potenzreihe (nämlich als ihre Taylor-Reihe, siehe Kapitel 11) geschrieben werden.

Beispiel 12.39.

(a) Wir betrachten die Funktion $f\colon \mathbb{R}_{>0} \to \mathbb{R}$, $x \mapsto \log x$. Die Ableitung dieser Funktion ist natürlich $f'(x) = \frac{1}{x}$. Nun können wir dies für $|x-1| < 1$ mit Hilfe der geometrischen Reihe (siehe Beispiel 7.3 (a)) als

$$f'(x) = \frac{1}{1+(x-1)} = 1 - (x-1) + (x-1)^2 - (x-1)^3 \pm \cdots$$

schreiben. Diese Potenzreihe kann jetzt aber nach Bemerkung 12.38 gliedweise integriert werden, und darum ist

$$(x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} \pm \cdots$$

für $|x-1| < 1$, also auf $(0,2)$, eine Stammfunktion von f'. Nach Folgerung 12.24 (b) kann sich diese von der ursprünglichen Funktion f nur um eine additive Konstante unterscheiden – Einsetzen von $x = 1$ liefert aber auch sofort, dass diese Konstante gleich 0 ist. Also erhalten wir die Darstellung

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} \pm \cdots$$

für alle $x \in (0,2)$ (die wir in Beispiel 11.16 (a) bereits für $x \in [1,2]$ bewiesen hatten).

(b) Eine analoge Rechnung können wir auch mit der Funktion $f\colon (-1,1) \to \mathbb{R}$, $x \mapsto \arctan x$ durchführen: Hier ist die Ableitung

$$f'(x) = \frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 \pm \cdots,$$

und damit ist

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} \pm \cdots$$

eine Stammfunktion von f', die sich von f wiederum nur um eine additive Konstante unterscheiden kann. Auch hier ist diese Konstante wegen $\arctan 0 = 0$ wieder gleich 0, und wir

erhalten auf $(-1,1)$ ohne irgendwelche komplizierte Rechnungen die Potenzreihendarstellung der Arkustangens-Funktion

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} \pm \cdots .$$

Aufgabe 12.40. Berechne die folgenden (z. T. unbestimmten bzw. uneigentlichen) Integrale:

(a) $\displaystyle\int_{-1}^{1} \frac{1}{\sqrt{2x+3}}\,dx$ (b) $\displaystyle\int_0^\infty x^2 \mathrm{e}^{-2x}\,dx$ (c) $\displaystyle\int \frac{x^3+x^2+1}{x^3-x}\,dx$

(d) $\displaystyle\int_{\mathrm{e}}^{\mathrm{e}^2} \frac{\log(\log x)}{x\log x}\,dx$ (e) $\displaystyle\int \frac{1}{1+\mathrm{e}^x}\,dx$ (f) $\displaystyle\int_0^{\frac{\pi}{2}} \sin^8 x \cos^3 x\,dx$

Aufgabe 12.41. Zeige mit Induktion über $n \in \mathbb{N}$, dass

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n x\,dx = \begin{cases} \pi \cdot \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \cdots \cdot \frac{n-1}{n} & \text{falls } n \text{ gerade,} \\ 2 \cdot \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{6}{7} \cdot \cdots \cdot \frac{n-1}{n} & \text{falls } n \text{ ungerade.} \end{cases}$$

Aufgabe 12.42 (Integralkriterium für Reihen). Es sei $f\colon \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 0}$ eine stetige und monoton fallende Funktion. Man zeige:

(a) Das uneigentliche Integral $\int_1^\infty f(x)\,dx$ hat das gleiche Konvergenzverhalten wie die Reihe $\sum_{n=1}^\infty f(n)$, d. h. es sind entweder beide konvergent oder beide divergent.

(b) Für $a \in \mathbb{R}$ konvergiert die Reihe $\sum_{n=1}^\infty \frac{1}{n^a}$ genau dann, wenn $a > 1$. (Dies ist eine Verallgemeinerung von Beispiel 7.3 (c) und 7.19 auf reelle Exponenten.)

Gilt die Aussage (a) auch ohne die Voraussetzung, dass f monoton fallend ist?

Aufgabe 12.43. Es sei $f\colon [0,1] \to \mathbb{R}$ eine stetige Funktion. Man zeige:

(a) $\lim_{n\to\infty} \int_0^1 f(x)\,x^n\,dx = 0$.

(b) Gilt $f(x) = f(1-x)$ für alle $x \in [0,1]$, so ist $\int_0^1 x f(x)\,dx = \frac{1}{2}\int_0^1 f(x)\,dx$.

Aufgabe 12.44. Es seien $a,b \in \mathbb{R}_{>0}$ und $f\colon [0,a] \to [0,b]$ eine bijektive, stetig differenzierbare Funktion. Man zeige:

(a) Ist f monoton wachsend mit $f(0) = 0$ und $f(a) = b$, dann gilt $\int_0^a f(x)\,dx + \int_0^b f^{-1}(x)\,dx = ab$.

(b) Ist f monoton fallend mit $f(0) = b$ und $f(a) = 0$, dann gilt $\int_0^a f(x)\,dx = \int_0^b f^{-1}(x)\,dx$.

Was bedeuten diese Aussagen geometrisch?

Aufgabe 12.45 (Binomische Reihe)**.** Für $a \in \mathbb{R}$ definieren wir die **verallgemeinerten Binomialkoeffizienten** durch

$$\binom{a}{n} := \frac{a\cdot(a-1)\cdot\cdots\cdot(a-n+1)}{n!}$$

für alle $n \in \mathbb{N}$. Wir betrachten nun auf $D = (-1,1)$ die Funktion $f\colon D \to \mathbb{R},\ f(x) = (1+x)^a$. Man zeige:

(a) Die Taylor-Reihe von f mit Entwicklungspunkt 0 ist gegeben durch $T_{f,0}(x) = \sum_{n=0}^\infty \binom{a}{n} x^n$ und konvergiert auf D.

(b) Es gilt sogar $(1+x)^a = \sum_{n=0}^\infty \binom{a}{n} x^n$ für alle $x \in D$, d. h. die Taylor-Reihe stellt wirklich die ursprüngliche Funktion dar. (Hinweis: Zeige zunächst, dass die Ableitung von $\frac{T_{f,0}}{f}$ gleich 0 ist.)

(c) Die Funktion arcsin lässt sich auf D als Potenzreihe schreiben. Berechne diese Potenzreihe explizit!

Was ergibt sich aus der binomischen Reihe in den Spezialfällen $a \in \mathbb{N}$ bzw. $a = -1$?

28

Grundlagen der Mathematik 1: Lineare Algebra

13. Vektorräume

Ausgehend von den elementaren Konzepten in den Kapiteln 1 bis 3 wollen wir in dieser Vorlesung zwei grundlegende Gebiete der Mathematik entwickeln: die Analysis und die lineare Algebra. Während sich die eindimensionale Analysis in den Kapiteln 4 bis 12 dabei hauptsächlich mit allgemeinen (in der Regel stetigen oder sogar differenzierbaren) Funktionen in einer reellen Variablen beschäftigt, wollen wir unabhängig davon im folgenden Teil des Skripts zunächst *lineare* Funktionen in *mehreren* Variablen studieren, wie sie in der Praxis z. B. in Form von linearen Gleichungssystemen auftreten. Dabei werden sich die ersten drei Kapitel 13 bis 15 hauptsächlich mit der allgemeinen Theorie beschäftigen, während in den folgenden drei Kapiteln 16 bis 18 auch die Frage konkreter Berechnungsmethoden eine zentrale Rolle spielt. Später werden wir die erarbeiteten Resultate dann mit der Analysis kombinieren, um Funktionen in mehreren Variablen mit Hilfe von Ableitungen linear approximieren zu können.

In der Analysis arbeitet man ja hauptsächlich über den reellen Zahlen, und das ist in der linearen Algebra auch nicht viel anders. Allerdings haben wir in Kapitel 3 ja auch schon gesehen, dass viele Dinge auch über einem beliebigen Körper funktionieren. Die lineare Algebra verhält sich hier sehr „gutartig“: Da wir letztlich nur lineare Funktionen bzw. Gleichungen betrachten werden, benötigen wir gar nicht mehr Struktur der reellen Zahlen als die Körperaxiome. Wir können daher nahezu die gesamte lineare Algebra über einem beliebigen Grundkörper studieren, also z. B. auch über $\mathbb{Q}$, dem Körper $\mathbb{Z}_2$ aus Beispiel 3.6 (b), oder anderen Körpern, die ihr vielleicht inzwischen kennt. Im Folgenden sei daher K immer ein fest gewählter Grundkörper (den ihr euch beim ersten Lesen aber durchaus gerne einfach als $\mathbb{R}$ vorstellen könnt).

13.A Der Vektorraumbegriff

Wie ihr ja sicher aus der Schule wisst, werden die Elemente von $\mathbb{R}^n$, die uns im Folgenden hauptsächlich interessieren werden, in der Regel *Vektoren* genannt. Aber was genau ist im Allgemeinen eigentlich ein Vektor? Genau wie bei Gruppen und Körpern in Kapitel 3.A werden Vektoren über die Operationen definiert, die man mit ihnen durchführen kann: In einer Gruppe gibt es *eine* Verknüpfung, die gewisse Eigenschaften erfüllt (siehe Definition 3.1), in einem Körper *zwei* Verknüpfungen „+“ und „ $\cdot$ “ mit den erwarteten Eigenschaften (siehe Definition 3.5). Was sind nun die analogen definierenden Verknüpfungen und Eigenschaften für Vektoren? Wir wissen alle aus der Schule, dass man Vektoren addieren und „strecken“, also mit einer reellen Zahl (bzw. mit einer Zahl des gewählten Grundkörpers K) multiplizieren kann. Genau diese beiden Strukturen definieren einen allgemeinen Vektorraum:

Definition 13.1 (Vektorräume)**.** Es sei K ein Körper. Ein **Vektorraum** über K (oder K-Vektorraum) ist eine Menge V zusammen mit zwei Verknüpfungen

$$\begin{aligned} &+ : V \times V \to V \quad (\textit{Vektoraddition}) \\ \text{und} \quad &\cdot : K \times V \to V \quad (\textit{Skalarmultiplikation}) \end{aligned}$$

so dass gilt:

(a) $(V,+)$ ist eine abelsche Gruppe (siehe Definition 3.1).

(b) (1. Distributivität) Für alle $\lambda, \mu \in K$ und $x \in V$ gilt $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$.

(c) (2. Distributivität) Für alle $\lambda \in K$ und $x, y \in V$ gilt $\lambda \cdot (x+y) = \lambda \cdot x + \lambda \cdot y$.

(d) (Assoziativität) Für alle $\lambda, \mu \in K$ und $x \in V$ gilt $(\lambda \cdot \mu) \cdot x = \lambda \cdot (\mu \cdot x)$.

(e) Für alle $x \in V$ gilt $1 \cdot x = x$.

Die Elemente von V heißen **Vektoren**, die Elemente von K **Skalare**.

Bemerkung 13.2.

(a) Beachte, dass man einen Vektorraum nur dann definieren kann, wenn man vorher einen Körper K gewählt hat. Wenn klar ist, welcher Körper gemeint ist, werden wir jedoch auch oft nur von einem Vektorraum (statt einem K-Vektorraum) sprechen.

(b) In Definition 13.1 haben wir mehrfach die gleichen Symbole für unterschiedliche Dinge verwendet: Es gibt z. B. zwei Additionen, die wir beide mit „+“ bezeichnet haben, nämlich die Addition $+\colon K \times K \to K$ zweier Körperelemente und die Addition $+\colon V \times V \to V$ der Vektoren. Da man aus der Art der verknüpften Elemente eindeutig ablesen kann, um welche Verknüpfung es sich handeln muss, können dadurch aber keine Mehrdeutigkeiten entstehen: So werden z. B. beim ersten Pluszeichen in Definition 13.1 (b) zwei Skalare, beim zweiten jedoch zwei Vektoren addiert. Nur wenn wir auch in der Notation explizit deutlich machen wollen, um welche der beiden Verknüpfungen es sich handelt, schreiben wir diese als $+_K$ bzw. $+_V$. Analog gibt es auch die Multiplikation zweimal, einmal als Multiplikation $\cdot_K$ in K und einmal als Skalarmultiplikation $\cdot_V$, und auch zweimal die Null, nämlich einmal als Null 0_K im Körper K und einmal als *Nullvektor* 0_V, d. h. als das neutrale Element von $(V,+)$. In dieser ausführlichen Notation könnte man z. B. die Bedingung aus Definition 13.1 (b) als

$$(\lambda +_K \mu) \cdot_V x = \lambda \cdot_V x +_V \mu \cdot_V x$$

schreiben. In der Regel werden wir diese Indizes K und V jedoch weglassen, genauso wie die Malzeichen sowohl für $\cdot_K$ als auch für $\cdot_V$.

Beispiel 13.3.

(a) Für jeden Körper K ist $V = \{0\}$ (mit den trivialen Verknüpfungen) ein K-Vektorraum, der sogenannte **Nullvektorraum**.

(b) Es seien K ein Körper und $n \in \mathbb{N}$. Wir betrachten die Menge

$$K^n = \underbrace{K \times \cdots \times K}_{n\text{-mal}} = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \dots, x_n \in K \right\}$$

aller „geordneten n-Tupel in K“, d. h. ein Element von K^n wird dadurch angegeben, dass man n Elemente $x_1, \dots, x_n$ von K angibt (die nicht notwendig verschieden sein müssen und auf deren Reihenfolge es ankommt). Dass wir die Elemente $x_1, \dots, x_n$ dabei untereinander und nicht nebeneinander schreiben, ist momentan eine reine Konvention, die sich später bei der Einführung von Matrizen in Abschnitt 16.A als nützlich erweisen wird. Definiert man nun auf K^n die komponentenweisen Verknüpfungen

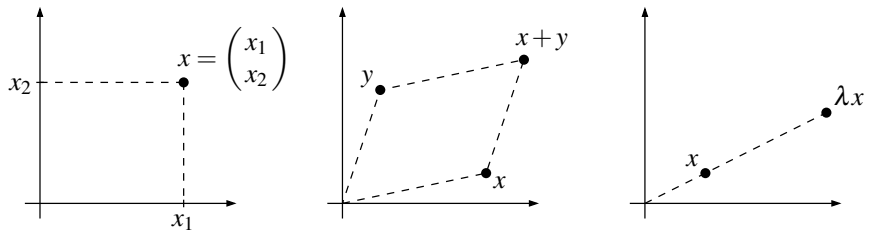
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \text{und} \quad \lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

für alle $\lambda \in K$, so ist K^n mit diesen Verknüpfungen ein K-Vektorraum. In der Tat folgen die Vektorraumeigenschaften alle aus den Körpereigenschaften von K; wir zeigen hier exemplarisch Teil (b) der Definition 13.1: Für alle $\lambda, \mu, x_1, \dots, x_n \in K$ gilt

$$(\lambda + \mu) \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} (\lambda + \mu) x_1 \\ \vdots \\ (\lambda + \mu) x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 + \mu x_1 \\ \vdots \\ \lambda x_n + \mu x_n \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \mu \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

wobei das mittlere Gleichheitszeichen genau die Distributivität in K ist und die beiden anderen aus der Definition der Vektoraddition und Skalarmultiplikation in K^n folgen.

Im Fall $K = \mathbb{R}$ und $n = 2$ ist $K^n = \mathbb{R}^2$ einfach die bekannte reelle Ebene, und die beiden Verknüpfungen entsprechen natürlich wie im folgenden Bild der aus der Schule bekannten Vektoraddition und Skalarmultiplikation.



Wenn wir im Folgenden vom Vektorraum K^n sprechen, werden wir diesen Raum immer als Vektorraum über dem Körper K betrachten (sofern wir nichts anderes angeben). Diese Vektorräume K^n für $n \in \mathbb{N}$ sind sicher die wichtigsten Beispiele für K-Vektorräume. Im Fall $n = 1$ erhält man $K^1 = K$, also K selbst als K-Vektorraum; der Fall $n = 0$ wird konventionsgemäß als der Nullvektorraum $K^0 = \{0\}$ aufgefasst.

(c) Sind V und W zwei K-Vektorräume, so ist (in Verallgemeinerung von (b)) auch ihr Produkt $V \times W$ mit komponentenweiser Addition und Skalarmultiplikation ein K-Vektorraum.

(d) Es seien K ein Körper und M eine Menge. Dann ist die Menge $\mathrm{Abb}(M,K) := \{f\colon M \to K\}$ aller Abbildungen von M nach K ein K-Vektorraum, indem wir Addition und Multiplikation punktweise definieren als

$$(f+g)(x) := f(x) + g(x) \quad \text{und} \quad (\lambda f)(x) := \lambda f(x)$$

für alle $\lambda \in K$, $x \in M$ und $f, g\colon M \to K$. In der Tat haben wir in Aufgabe 3.12 (a) bereits gesehen, dass $\mathrm{Abb}(M,K)$ eine abelsche Gruppe ist (der Nullvektor ist hierbei die Funktion, die jedes Element von M auf 0 abbildet, und das zu einer Funktion $f\colon M \to K$ additive Inverse die Funktion $-f\colon M \to K$, $x \mapsto -f(x)$). Die anderen Vektorraumeigenschaften zeigt man wieder analog.

Ein Spezialfall hiervon ist der Raum $\mathrm{Abb}(\mathbb{N},K)$, dessen Elemente $f \in \mathrm{Abb}(\mathbb{N},K)$ wir in Verallgemeinerung von (b) als „unendliche Folgen“ $(f(0), f(1), f(2), \dots)$ mit Elementen in K schreiben können. Im Fall $K = \mathbb{R}$ sind dies gerade die in der Analysis betrachteten reellen Zahlenfolgen.

Darüber hinaus lässt sich auf die gleiche Art auch die Menge $\mathrm{Abb}(M,W)$ aller Abbildungen von einer beliebigen Menge M in einen K-Vektorraum W zu einem Vektorraum machen.

(e) $\mathbb{R}$ ist ein $\mathbb{Q}$-Vektorraum. In der Tat kann man reelle Zahlen addieren und mit einer rationalen multiplizieren, und es ist klar, dass mit diesen Definitionen alle Vektorraumeigenschaften erfüllt sind.

Wir wollen nun zunächst ein paar elementare Eigenschaften von Vektorräumen zeigen. Sie haben einen ähnlichen Charakter wie die Axiome in Definition 13.1, folgen aber bereits aus diesen (so dass man sie nicht separat als Axiome fordern muss).

Lemma 13.4 (Eigenschaften von Vektorräumen)**.** *In jedem K-Vektorraum V gilt für alle $\lambda \in K$ und $x \in V$:*

(a) $0_K \cdot x = \lambda \cdot 0_V = 0_V$.

(b) *Ist $\lambda \cdot x = 0_V$, so ist $\lambda = 0_K$ oder $x = 0_V$.*

(c) $(-1) \cdot x = -x$.

Beweis.

(a) Der Beweis ist ganz analog zu dem von Lemma 3.8 (a): Wegen der Distributivität aus Definition 13.1 (b) gilt $0_K \cdot x = (0_K + 0_K) \cdot x = 0_K \cdot x + 0_K \cdot x$, nach Subtraktion von

$0_K \cdot x$ also wie behauptet $0_V = 0_K \cdot x$. Genauso erhalten wir mit der 2. Distributivität $\lambda \cdot 0_V = \lambda \cdot (0_V + 0_V) = \lambda \cdot 0_V + \lambda \cdot 0_V$, nach Subtraktion von $\lambda \cdot 0_V$ also $0_V = \lambda \cdot 0_V$.

(b) Ist $\lambda x = 0_V$ und $\lambda \neq 0_K$, so folgt

$$\begin{aligned} x &= 1 \cdot x && \text{(Definition 13.1 (e))} \\ &= (\lambda^{-1} \cdot \lambda) x \\ &= \lambda^{-1}(\lambda x) && \text{(Definition 13.1 (d))} \\ &= 0_V && \text{(Teil (a)).} \end{aligned}$$

(c) Es gilt

$$\begin{aligned} (-1) \cdot x + x &= (-1) \cdot x + 1 \cdot x && \text{(Definition 13.1 (e))} \\ &= (-1+1) \cdot x && \text{(Definition 13.1 (b))} \\ &= 0_K \cdot x = 0_V && \text{(Teil (a)),} \end{aligned}$$

also ist $(-1) \cdot x$ das additive Inverse zu x. □

Sehr viele weitere Beispiele von Vektorräumen können wir erhalten, indem wir in bereits bekannten Vektorräumen nach Teilmengen suchen, die mit der gegebenen Vektoraddition und Skalarmultiplikation selbst wieder die Axiome aus Definition 13.1 erfüllen. Solche Teilmengen werden als Untervektorräume bezeichnet.

Definition 13.5 (Untervektorräume)**.** Es sei V ein Vektorraum über einem Körper K. Eine Teilmenge $U \subset V$ heißt **Untervektorraum** oder **Unterraum** von V, in Zeichen $U \leq V$, wenn die folgenden Bedingungen gelten:

(a) $U \neq \emptyset$;

(b) für alle $x, y \in U$ ist $x + y \in U$;

(c) für alle $\lambda \in K$ und $x \in U$ ist $\lambda x \in U$.

Man sagt für (b) und (c) auch, dass U bezüglich Vektoraddition bzw. Skalarmultiplikation *abgeschlossen* ist.

Bemerkung 13.6.

(a) Durch fortgesetztes Anwenden der Eigenschaften (b) und (c) von Definition 13.5 sieht man sofort, dass in einem Unterraum U von V auch für alle $n \in \mathbb{N}_{>0}$ gilt:

$$\text{Für alle } \lambda_1, \dots, \lambda_n \in K \text{ und } x_1, \dots, x_n \in U \text{ ist } \lambda_1 x_1 + \dots + \lambda_n x_n \in U.$$

Beachte, dass $x_1, \dots, x_n$ hierbei im Gegensatz zu Beispiel 13.3 (b) verschiedene Vektoren, und nicht die Komponenten *eines* Vektors in K^n sind. In der Tat ist diese Indexnotation in der Praxis für beide Bedeutungen üblich. Aus dem Zusammenhang ist aber immer offensichtlich, was gemeint ist, da $x_1, \dots, x_n$ in Beispiel 13.3 (b) Elemente des Grundkörpers sind, hier jedoch Elemente des betrachteten Vektorraums.

(b) Jeder Unterraum U von V muss den Nullvektor enthalten: Nach Definition 13.5 (a) ist U nicht leer, enthält also ein Element $x \in U$. Damit liegt nach Definition 13.5 (c) auch $0 \cdot x = 0$ in U.

Die Abgeschlossenheit eines Untervektorraums bezüglich Vektoraddition und Skalarmultiplikation bedeutet gerade, dass sich diese beiden Verknüpfungen zu Verknüpfungen

$$+ : U \times U \to U \qquad \text{und} \qquad \cdot : K \times U \to U$$

auf U einschränken lassen. Wie oben bereits angekündigt wollen wir nun sehen, dass dies die Menge U selbst wieder zu einem Vektorraum macht.

Lemma 13.7. *Jeder Untervektorraum eines K-Vektorraums ist (mit den eingeschränkten Verknüpfungen) selbst wieder ein K-Vektorraum.*

Beweis. Wir müssen die Eigenschaften aus Definition 13.1 für U nachweisen. Dazu beginnen wir mit (a) und zeigen zunächst, dass $(U,+)$ eine Gruppe ist.

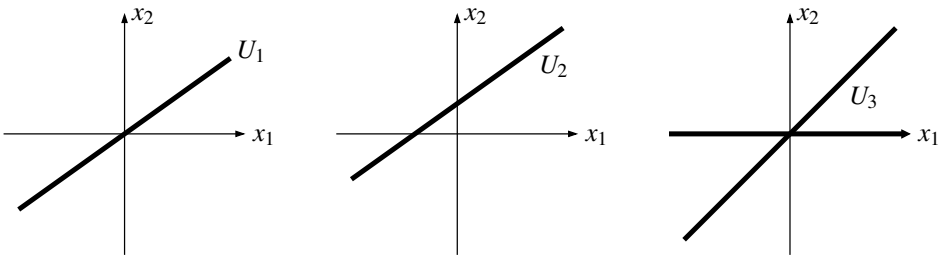
- (Assoziativität der Vektoraddition) Weil V ein Vektorraum ist, gilt $(x+y)+z = x+(y+z)$ für alle $x,y,z \in V$ und damit erst recht für alle $x,y,z \in U$. Die Assoziativität der Addition überträgt sich also direkt von V auf U.
- (Additives neutrales Element) Nach Bemerkung 13.6 (b) liegt der Nullvektor in U. Dieser erfüllt $x+0 = 0+x = x$ für alle $x \in V$ und damit auch für alle $x \in U$, und ist damit ein neutrales Element für die Addition in U.
- (Additive inverse Elemente) Für jedes $x \in U$ gilt nach Definition 13.5 (c) auch $(-1)\cdot x \in U$, und dies ist ja nach Lemma 13.4 (c) genau das additive inverse Element zu x.

Die übrigen Vektorraumeigenschaften sind alle von der Form, dass für alle Vektoren aus U eine bestimmte Gleichung gelten muss – und dies folgt nun genauso wie die Assoziativität oben sofort daraus, dass die betreffenden Gleichungen sogar für alle Vektoren aus V gelten. □

Beispiel 13.8.

(a) Für jeden Vektorraum V sind der Nullvektorraum $\{0\} \subset V$ und der gesamte Raum $V \subset V$ natürlich stets Unterräume von V. Sie werden die **trivialen Unterräume** genannt.

(b) Es seien $K = \mathbb{R}$ und $V = \mathbb{R}^2$. Für ein gegebenes $m \in \mathbb{R}$ betrachten wir die folgenden Teilmengen U_1, U_2, U_3 von V:



$$\left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = mx_1\right\} \qquad \left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = mx_1 + 1\right\} \qquad \left\{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = x_1 \text{ oder } x_2 = 0\right\}$$

Die Ursprungsgerade U_1 ist ein Untervektorraum von $\mathbb{R}^2$: Natürlich ist $U_1 \neq \emptyset$, und für alle $\lambda \in \mathbb{R}$ und $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$ gilt:

- Sind $x,y \in U_1$, also $x_2 = mx_1$ und $y_2 = my_1$, so ist $x_2 + y_2 = mx_1 + my_1 = m(x_1 + y_1)$ und damit auch $x+y \in U_1$.
- Ist $x \in U_1$, also $x_2 = mx_1$, so ist auch $\lambda x_2 = m\lambda x_1$, also $\lambda x \in U_1$.

U_2 ist nach Bemerkung 13.6 (b) kein Unterraum von $\mathbb{R}^2$, da $0 \notin U_2$. U_3 ist ebenfalls kein Unterraum, denn es liegen zwar $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in U_3, nicht aber deren Summe $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Beachte, dass man hier nicht nur an den Formeln, sondern auch an den Bildern oben schon sehen kann, ob die gegebenen Teilmengen abgeschlossen bezüglich Vektoraddition und Skalarmultiplikation, also ob sie Untervektorräume sind.

(c) Für eine gegebene Teilmenge $D \subset \mathbb{R}$ ist die Teilmenge $U \subset \mathrm{Abb}(D,\mathbb{R})$ aller Polynomfunktionen $f\colon D \to \mathbb{R}$ ein Unterraum des Vektorraums $\mathrm{Abb}(D,\mathbb{R})$ aus Beispiel 13.3 (d), denn mit f und g sind auch $f+g$ und λf für alle $\lambda \in \mathbb{R}$ Polynomfunktionen. Genauso ist für festes $k \in \mathbb{N}$ auch die Menge aller Polynomfunktionen vom Grad höchstens k ein Unterraum von $\mathrm{Abb}(D,\mathbb{R})$.

Bemerkung 13.9 (Durchschnitt und Vereinigung von Unterräumen)**.** Es seien U_1 und U_2 Unterräume eines K-Vektorraums V.

(a) Der Durchschnitt $U_1 \cap U_2$ ist ebenfalls wieder ein Unterraum von V:

- $U_1 \cap U_2$ ist nicht leer, denn nach Bemerkung 13.6 (b) liegt der Nullvektor in U_1 und U_2, und damit auch in $U_1 \cap U_2$.
- Sind $x, y \in U_1 \cap U_2$, so gilt insbesondere $x, y \in U_1$ und damit auch $x + y \in U_1$, da U_1 ein Unterraum ist. Genauso ergibt sich $x + y \in U_2$, insgesamt also $x + y \in U_1 \cap U_2$.
- Ist $x \in U_1 \cap U_2$ und $\lambda \in K$, so gilt insbesondere $x \in U_1$ und damit dann auch $\lambda x \in U_1$, da U_1 ein Unterraum ist. Genauso ergibt sich auch $\lambda x \in U_2$ und damit dann $\lambda x \in U_1 \cap U_2$.

Analog sieht man natürlich, dass auch der Durchschnitt $U_1 \cap \cdots \cap U_n$ von mehr als zwei Unterräumen wieder ein Unterraum ist.

(b) Die Vereinigung $U_1 \cup U_2$ ist im Allgemeinen kein Unterraum von V: Wir haben z. B. in Beispiel 13.8 (a) (mit $m = 1$ bzw. $m = 0$) gesehen, dass die beiden Geraden

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = x_1 \right\} \qquad \text{und} \qquad \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = 0 \right\}$$

Unterräume von $\mathbb{R}^2$ sind, ihre Vereinigung jedoch nicht.

Es gibt jedoch trotzdem eine Möglichkeit, aus zwei Unterräumen einen neuen zu erzeugen, der beide von ihnen enthält. Die korrekte Konstruktion hierfür ist nur nicht die Vereinigung, sondern ihre sogenannte Summe.

Lemma und Definition 13.10 (Summe von Unterräumen)**.** *Es seien U_1 und U_2 Unterräume eines K-Vektorraums V. Dann ist ihre **Summe***

$$U_1 + U_2 := \{x_1 + x_2 : x_1 \in U_1, x_2 \in U_2\}$$

ebenfalls ein Unterraum von V.

Beweis. Beachte zunächst, dass $U_1 + U_2$ nicht leer ist, weil beide Unterräume nach Bemerkung 13.6 (b) den Nullvektor enthalten und damit auch $0 = 0 + 0 \in U_1 + U_2$ gilt. Wir überprüfen nun die Kriterien aus Definition 13.5 für $U_1 + U_2$:

(a) Sind $x, y \in U_1 + U_2$, also $x = x_1 + x_2$ und $y = y_1 + y_2$ mit $x_1, y_1 \in U_1$ und $x_2, y_2 \in U_2$, so folgt auch

$$x + y = x_1 + x_2 + y_1 + y_2 = (x_1 + y_1) + (x_2 + y_2) \quad \in U_1 + U_2,$$

da wegen der Abgeschlossenheit von U_1 und U_2 auch $x_1 + y_1 \in U_1$ und $x_2 + y_2 \in U_2$ gilt.

(b) Sind analog $\lambda \in K$ und $x \in U_1 + U_2$, also wieder $x = x_1 + x_2$ mit $x_1 \in U_1$ und $x_2 \in U_2$, so ist auch

$$\lambda x = \lambda(x_1 + x_2) = \underbrace{\lambda x_1}_{\in U_1} + \underbrace{\lambda x_2}_{\in U_2} \quad \in U_1 + U_2. \qquad \square$$

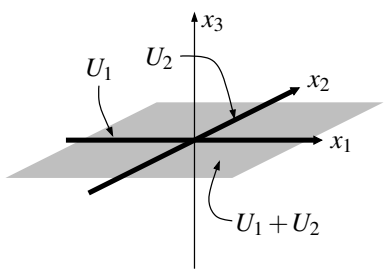
Bemerkung 13.11. Analog zu Definition 13.10 kann man natürlich auch die Summe $U_1 + \cdots + U_n$ von mehr als zwei Unterräumen $U_1, \ldots, U_n$ eines Vektorraums V definieren.

Beispiel 13.12. Für die Unterräume

$$U_1 = \left\{ \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} : x_1 \in \mathbb{R} \right\} \quad \text{und} \quad U_2 = \left\{ \begin{pmatrix} 0 \\ x_2 \\ 0 \end{pmatrix} : x_2 \in \mathbb{R} \right\}$$

von $\mathbb{R}^3$ ist ihre Summe wie im Bild rechts dargestellt die (x_1, x_2)-Ebene

$$U_1 + U_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\}.$$



Aufgabe 13.13. Welche der folgenden Teilmengen sind Unterräume von $\mathbb{R}^3$?

(a) $\left\{ \begin{pmatrix} -a \\ 0 \\ b-a \end{pmatrix} : a,b \in \mathbb{R} \right\}$;

(b) $\{x \in \mathbb{R}^3 : x_2 + a x_3 = 1 - a\}$ für ein festes, gegebenes $a \in \mathbb{R}$;

(c) $\{x \in \mathbb{R}^3 : x_1^3 = x_2^3 = x_3^3\}$;

(d) $\{x \in \mathbb{R}^3 : x_1, x_2, x_3 \in \mathbb{Z}\}$.

Hierbei sei jeweils $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, d. h. x_1, x_2, x_3 seien die Koordinaten des Vektors x.

Aufgabe 13.14. Es seien U_1 und U_2 Unterräume eines K-Vektorraums V. Zeige, dass $U_1 \cup U_2$ genau dann ein Unterraum von V ist, wenn $U_1 \subset U_2$ oder $U_2 \subset U_1$.

Aufgabe 13.15. Es seien

$$U = \{f \colon \mathbb{R} \to \mathbb{R} : f(-x) = f(x) \text{ für alle } x \in \mathbb{R}\}$$
$$\text{und} \qquad V = \{f \colon \mathbb{R} \to \mathbb{R} : f(-x) = -f(x) \text{ für alle } x \in \mathbb{R}\}$$

in $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ die Mengen aller sogenannten geraden bzw. ungeraden Funktionen. Man zeige:

(a) U und V sind Unterräume von $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$;

(b) $U \cap V = \{0\}$;

(c) $U + V = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$.

29

13.B Lineare Abbildungen

Im letzten Abschnitt haben wir Vektorräume eingeführt – also die grundlegende Struktur, mit der sich die lineare Algebra befasst. Um nun verschiedene Vektorräume miteinander in Verbindung setzen zu können, müssen wir als Nächstes Abbildungen zwischen Vektorräumen untersuchen, die mit den gegebenen Verknüpfungen (also der Vektoraddition und der Skalarmultiplikation) verträglich sind. Dies wollen wir in diesem Abschnitt tun.

In der Tat ist dies eine sehr allgemeine Vorgehensweise der Mathematik: Immer wenn man eine neue mathematische Struktur (wie z. B. Gruppen, Körper, oder jetzt hier die Vektorräume) einführt, wird man als Erstes zwei Dinge untersuchen:

- die sogenannten *Unterstrukturen*, d. h. Teilmengen, die selbst wieder die betrachtete Struktur haben (in unserem momentanen Fall also die Untervektorräume), und
- die sogenannten *Morphismen*, d. h. Abbildungen, die diese Struktur erhalten.

Wir haben dies in Abschnitt 3.A nur deswegen für Gruppen und Körper nicht getan, weil wir in dieser Vorlesung nur Vektorräume, aber nicht Gruppen und Körper ausführlich studieren wollen. Diejenigen von euch, die auch die Parallelvorlesung „Algebraische Strukturen“ hören, werden dort aber z. B. auch Untergruppen und Morphismen von Gruppen untersuchen (oder haben dies bereits getan) – und dann feststellen, dass sich Untervektorräume und Morphismen von Vektorräumen, die wir nun studieren werden, eigentlich „fast genauso“ verhalten.

Definition 13.16 (Lineare Abbildungen bzw. Morphismen)**.** Es seien V und W zwei Vektorräume über demselben Grundkörper K. Man nennt eine Abbildung $f \colon V \to W$ eine **lineare Abbildung** (oder **Morphismus** oder **(Vektorraum-)Homomorphismus**), wenn gilt:

(a) Für alle $x, y \in V$ ist $f(x+y) = f(x) + f(y)$ („f ist verträglich mit der Vektoraddition“).

(b) Für alle $\lambda \in K$ und $x \in V$ ist $f(\lambda x) = \lambda f(x)$ („f ist verträglich mit der Skalarmultiplikation“).

Die Menge aller solchen Morphismen mit Startraum V und Zielraum W wird mit $\mathrm{Hom}_K(V,W)$ bezeichnet (oder auch nur mit $\mathrm{Hom}(V,W)$, wenn der Grundkörper aus dem Zusammenhang klar ist).

Ist $V = K^n$, so schreiben wir statt $f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$ der Einfachheit halber oft nur $f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Beispiel 13.17.

(a) Für beliebige K-Vektorräume ist die Nullabbildung $f\colon V \to W$, $x \mapsto 0$ natürlich immer ein Morphismus, da sie die beiden Eigenschaften (a) und (b) aus Definition 13.16 erfüllt.

(b) Die Abbildung von $\mathbb{R}$-Vektorräumen

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$$

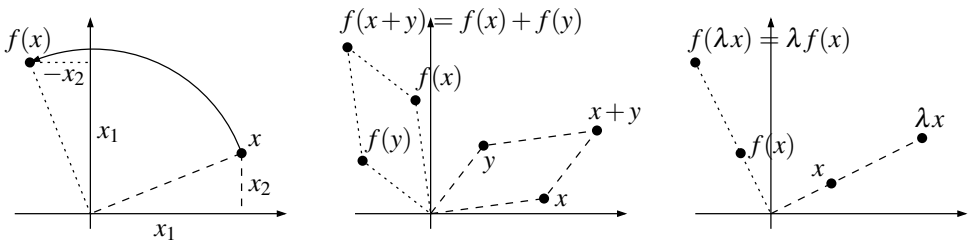
ist ein Morphismus, denn für alle $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2$ und $\lambda \in \mathbb{R}$ gilt

$$f(x+y) = f\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \end{pmatrix} = \begin{pmatrix} -x_2-y_2 \\ x_1+y_1 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} + \begin{pmatrix} -y_2 \\ y_1 \end{pmatrix} = f(x)+f(y)$$

und

$$f(\lambda x) = f\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix} = \begin{pmatrix} -\lambda x_2 \\ \lambda x_1 \end{pmatrix} = \lambda \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \lambda f(x).$$

Geometrisch beschreibt diese Abbildung wie im Bild unten links eine Vierteldrehung um den Ursprung. An den anderen beiden Bildern kann man die Morphismuseigenschaft auch gut anschaulich ablesen: Im mittleren Bild z. B. ist das gepunktete Parallelogramm aus der Drehung des gestrichelten entstanden, und der äußerste Punkt ergibt sich damit sowohl durch Addition der Punkte $f(x)$ und $f(y)$ als auch durch Drehung des Punktes $x+y$, d. h. es ist $f(x)+f(y) = f(x+y)$. Entsprechendes gilt für die Skalarmultiplikation im rechten Bild.



Anschaulich ist damit auch schon erkennbar, dass Drehungen um andere Winkel (um den Ursprung) ebenfalls Morphismen sind. Wir werden solche allgemeinen Drehungen später in Beispiel 19.13 und Abschnitt 22.A untersuchen.

(c) Die Abbildung

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1^2 + x_2$$

ist kein Morphismus, denn es ist z. B.

$$f\begin{pmatrix} -1 \\ 0 \end{pmatrix} + f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1+1 = 2 \neq 0 = f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = f\left(\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right).$$

(Beachte, dass es genügt, ein Gegenbeispiel anzugeben, um zu zeigen, dass die Verträglichkeiten aus Definition 13.16 nicht allgemein gelten.)

Wir werden in Abschnitt 16.A noch sehen, wie man *alle* Morphismen $f\colon K^n \to K^m$ beschreiben kann.

(d) Es sei $V \subset \mathrm{Abb}(\mathbb{R},\mathbb{R})$ der Vektorraum aller reellen Polynomfunktionen aus Beispiel 13.8 (c). Wir betrachten die Abbildung $f\colon V \to V,\ \varphi \mapsto \varphi'$, die jedem Polynom $\varphi\colon x \mapsto \sum_{k=0}^{n} a_k x^k$ seine *Ableitung*

$$\varphi'\colon x \mapsto \sum_{k=1}^{n} k\, a_k x^{k-1} \tag{$*$}$$

zuordnet.

Wenn ihr die Ableitung bereits aus der Analysis kennt, wisst ihr aus den Regeln in Beispiel 10.9 (c) schon, dass die Ableitung eines Polynoms durch $(*)$ gegeben ist, und dass f eine lineare Abbildung ist, da nach Satz 10.8 (a) und Beispiel 10.9 (b) für alle differenzierbaren Funktionen φ und ψ sowie $\lambda \in \mathbb{R}$ sowohl $(\varphi+\psi)' = \varphi' + \psi'$ als auch $(\lambda\varphi)' = \lambda\,\varphi'$ gilt.

Wenn ihr die Ableitung aus der Analysis noch nicht kennt, könnt ihr $(*)$ für die Zwecke der linearen Algebra einfach als *Definition* der Ableitung eines Polynoms ansehen. Man rechnet dann schnell nach, dass f wirklich eine lineare Abbildung ist: Für zwei Polynome $\varphi(x) = \sum_{k=0}^{n} a_k x^k$ und $\psi(x) = \sum_{k=0}^{n} b_k x^k$ (wobei n der größere der beiden Grade ist, so dass sowohl φ als auch ψ so geschrieben werden können) sowie $\lambda \in \mathbb{R}$ ist

$$(\varphi+\psi)'(x) = \sum_{k=1}^{n} k\,(a_k+b_k)x^{k-1} = \sum_{k=1}^{n} k\,a_k x^{k-1} + \sum_{k=1}^{n} k\,b_k x^{k-1} = \varphi'(x) + \psi'(x)$$

$$\text{und}\quad (\lambda\varphi)'(x) = \sum_{k=1}^{n} k\,(\lambda a_k)x^{k-1} = \lambda \sum_{k=1}^{n} k\,a_k x^{k-1} = \lambda\,\varphi'(x).$$

(e) Für den Vektorraum $V = \mathrm{Abb}(\mathbb{N},\mathbb{R})$ aller reellen Zahlenfolgen aus Beispiel 13.3 (d) sind die „Verschiebeabbildungen“

$$f\colon V \to V,\ (a_0,a_1,a_2,\dots) \mapsto (0,a_0,a_1,a_2,\dots)$$
$$\text{und}\qquad g\colon V \to V,\ (a_0,a_1,a_2,\dots) \mapsto (a_1,a_2,a_3,\dots)$$

Morphismen.

Bemerkung 13.18. Es sei $f\colon V \to W$ ein Morphismus von K-Vektorräumen.

(a) Durch fortgesetztes Anwenden der Eigenschaften aus Definition 13.16 erhält man sofort, dass auch für alle $\lambda_1,\dots,\lambda_n \in K$ und $x_1,\dots,x_n \in V$

$$f(\lambda_1 x_1 + \dots + \lambda_n x_n) = \lambda_1 f(x_1) + \dots + \lambda_n f(x_n)$$

gilt.

(b) Setzt man $\lambda = 0$ in der Eigenschaft 13.16 (b) ein, so erhält man mit Lemma 13.4 (a) sofort $f(0) = 0$.

Bemerkung 13.19 ($\mathrm{Hom}(V,W)$ als Vektorraum)**.** Die Menge $\mathrm{Hom}(V,W)$ der *linearen* Abbildungen zwischen zwei Vektorräumen V und W ist ein Unterraum des Vektorraums $\mathrm{Abb}(V,W)$ *aller* Abbildungen von V nach W aus Beispiel 13.3 (d):

- Nach Beispiel 13.17 (a) liegt die Nullabbildung in $\mathrm{Hom}(V,W)$.
- Sind $f,g\colon V \to W$ zwei lineare Abbildungen, so gilt für ihre Summe $f+g$ für alle $x+y \in V$ und $\mu \in K$

$$\begin{aligned}(f+g)(x+y) &= f(x+y)+g(x+y) && \text{(Definition von } f+g)\\ &= f(x)+f(y)+g(x)+g(y) && (f \text{ und } g \text{ sind linear})\\ &= (f+g)(x)+(f+g)(y) && \text{(Definition von } f+g)\end{aligned}$$

  sowie analog

$$(f+g)(\mu x) = f(\mu x)+g(\mu x) = \mu\, f(x) + \mu\, g(x) = \mu\,(f+g)(x).$$

  Also ist dann auch $f+g$ linear.
- Genauso zeigt man die Morphismuseigenschaften auch für die Abbildung λf mit $\lambda \in K$.

Gemäß Definition 13.5 ist $\mathrm{Hom}(V,W)$ damit ein Unterraum von $\mathrm{Abb}(V,W)$, also insbesondere nach Lemma 13.7 selbst wieder ein K-Vektorraum.

Aufgabe 13.20. Untersuche, ob die folgenden Abbildungen f zwischen $\mathbb{R}$-Vektorräumen linear sind:

(a) $f\colon \mathbb{R}^2 \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto ax_1 + bx_2 + c$ für fest gegebene $a,b,c \in \mathbb{R}$.

(b) V sei der Vektorraum der Polynomfunktionen auf $\mathbb{R}$, und

$$f\colon V \to V,\ \varphi \mapsto f(\varphi) \quad \text{mit} \quad f(\varphi)(x) = x^2\,\varphi(x) \quad \text{für alle } x \in \mathbb{R}.$$

Im Rest dieses Abschnitts wollen wir nun einige elementare Eigenschaften von Morphismen zeigen.

Lemma 13.21 (Bilder und Urbilder von Unterräumen)**.** *Es sei $f\colon V \to W$ eine lineare Abbildung. Dann gilt:*

(a) *Ist U ein Unterraum von V, so ist $f(U)$ ein Unterraum von W,*

(b) *Ist U ein Unterraum von W, so ist $f^{-1}(U)$ ein Unterraum von V,*

wobei $f(U)$ und $f^{-1}(U)$ das Bild bzw. Urbild von U unter f aus Definition 2.11 sind.

Beweis. Wir müssen die Eigenschaften aus Definition 13.5 überprüfen.

(a) Wegen $0 \in U$ ist nach Bemerkung 13.18 (b) zunächst $0 = f(0) \in f(U)$, also ist $f(U) \neq \emptyset$. Wir zeigen nun die Abgeschlossenheit von $f(U)$ bezüglich der Vektoraddition. Es seien dazu $x,y \in f(U)$, d. h. $x = f(u)$ und $y = f(v)$ für gewisse $u,v \in U$. Dann ist auch $u+v \in U$, und damit folgt $x+y = f(u)+f(v) = f(u+v) \in f(U)$. Genauso zeigt man die Abgeschlossenheit unter der Skalarmultiplikation.

(b) Wegen $f(0) = 0 \in U$ ist zunächst einmal $0 \in f^{-1}(U)$, d. h. es ist $f^{-1}(U) \neq \emptyset$. Wir zeigen jetzt die Abgeschlossenheit von $f^{-1}(U)$ unter der Vektoraddition. Dazu seien $x,y \in f^{-1}(U)$, d. h. $x,y \in V$ mit $f(x), f(y) \in U$. Dann ist auch $f(x+y) = f(x)+f(y) \in U$, also $x+y \in f^{-1}(U)$. Analog ergibt sich die Abgeschlossenheit unter der Skalarmultiplikation. □

Die wichtigsten Spezialfälle dieses Lemmas sind die folgenden:

Definition 13.22 (Bild und Kern eines Morphismus)**.** Es sei $f\colon V \to W$ ein lineare Abbildung von K-Vektorräumen.

(a) Die Menge $\mathrm{Im}\, f := f(V) = \{f(x) : x \in V\}$ heißt das **Bild** von f.

(b) Die Menge $\mathrm{Ker}\, f := f^{-1}(\{0\}) = \{x \in V : f(x) = 0\}$ heißt der **Kern** von f.

Die Bezeichnungen kommen von den englischen Begriffen „image“ und „kernel“. Nach Lemma 13.21 gilt offensichtlich $\mathrm{Im}\, f \leq W$ und $\mathrm{Ker}\, f \leq V$.

Beispiel 13.23. Für die lineare Abbildung

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1 + x_2$$

ist

$$\mathrm{Im}\, f = \left\{ x_1 + x_2 : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \right\} = \mathbb{R} \quad \text{und} \quad \mathrm{Ker}\, f = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_1 + x_2 = 0 \right\} \subset \mathbb{R}^2.$$

Offensichtlich ist eine lineare Abbildung $f\colon V \to W$ nach Definition genau dann surjektiv, wenn $\mathrm{Im}\, f = W$. Ein analoges Kriterium gibt es auch für die Injektivität:

Lemma 13.24. *Eine lineare Abbildung $f\colon V \to W$ ist genau dann injektiv, wenn* $\mathrm{Ker}\, f = \{0\}$.

Beweis.

„$\Rightarrow$“ Ist f injektiv, so hat der Nullvektor höchstens ein Urbild unter f. Wegen $f(0) = 0$ ist das Urbild des Nullvektors also genau der Nullvektor, d. h. es ist $\mathrm{Ker}\, f = \{0\}$.

„$\Leftarrow$“ Es sei $\operatorname{Ker} f = \{0\}$. Weiterhin seien $x, y \in V$ mit $f(x) = f(y)$. Wegen der Linearität von f gilt dann $f(x-y) = f(x) - f(y) = 0$, mit $\operatorname{Ker} f = \{0\}$ also $x - y = 0$. Damit folgt $x = y$, d. h. f ist injektiv. □

Lemma 13.25 (Umkehrabbildungen und Verkettungen)**.** *Es sei $f\colon V \to W$ ein Morphismus von K-Vektorräumen. Dann gilt:*

(a) *Ist f bijektiv, so ist auch die Umkehrabbildung f^{-1} ein Morphismus.*

(b) *Ist $g\colon W \to Z$ ein weiterer Morphismus von K-Vektorräumen, so ist auch $g \circ f\colon V \to Z$ ein Morphismus.*

Beweis.

(a) Es seien $x, y \in W$; wir setzen $u = f^{-1}(x)$ und $v = f^{-1}(y)$, also $x = f(u)$ und $y = f(v)$. Dann gilt

$$\begin{aligned} f^{-1}(x+y) &= f^{-1}(f(u)+f(v)) \\ &= f^{-1}(f(u+v)) && (f \text{ ist ein Morphismus}) \\ &= u+v && (f^{-1} \text{ ist Umkehrabbildung von } f) \\ &= f^{-1}(x)+f^{-1}(y). \end{aligned}$$

Analog zeigt man die Verträglichkeit mit der Skalarmultiplikation.

(b) Für $x, y \in V$ gilt

$$\begin{aligned} g(f(x+y)) &= g(f(x)+f(y)) && (f \text{ ist ein Morphismus}) \\ &= g(f(x))+g(f(y)) && (g \text{ ist ein Morphismus}). \end{aligned}$$

Genauso ergibt sich die Verträglichkeit mit der Skalarmultiplikation. □

Aufgabe 13.26. Man zeige: Sind $f, g\colon V \to W$ zwei lineare Abbildungen, so gilt

(a) $\operatorname{Ker} f \cap \operatorname{Ker} g \subset \operatorname{Ker}(f+g)$;

(b) $\operatorname{Im}(f+g) \subset \operatorname{Im} f + \operatorname{Im} g$.

Weiterhin gebe man in beiden Fällen ein Beispiel an, das zeigt, dass man im Allgemeinen nicht „$\subset$“ durch „$=$“ ersetzen kann.

Aufgabe 13.27. Es sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Ferner sei U ein Unterraum von V mit $U \cap \operatorname{Ker} f = \{0\}$ und $U + \operatorname{Ker} f = V$.

Zeige, dass die Abbildung $f|_U\colon U \to \operatorname{Im} f$ bijektiv ist.

Aufgabe 13.28. Für ein gegebenes $n \in \mathbb{N}_{>0}$ sei $V \leq \operatorname{Abb}(\mathbb{R}, \mathbb{R})$ der Unterraum aller Polynomfunktionen vom Grad kleiner als n. Zeige, dass die Abbildung

$$f\colon V \to \mathbb{R}^n,\ \varphi \mapsto (\varphi(1), \ldots, \varphi(n)).$$

linear ist, und bestimme Kern und Bild von f.

Aufgabe 13.29. Es sei V ein $\mathbb{R}$-Vektorraum. Eine lineare Abbildung $f\colon V \to V$ heißt *Projektion*, wenn $f \circ f = f$.

(a) Gib für den Fall $V = \mathbb{R}^2$ ein Beispiel für eine Projektion an, bei der sowohl $\operatorname{Ker} f$ als auch $\operatorname{Im} f$ nicht-triviale Unterräume von V sind.

(b) Man zeige: Ist $f\colon V \to V$ eine Projektion, so gilt $\operatorname{Ker} f \cap \operatorname{Im} f = \{0\}$ und $\operatorname{Ker} f + \operatorname{Im} f = V$.

(c) Es seien $f, g\colon V \to V$ zwei Projektionen. Beweise, dass $f+g$ genau dann ebenfalls eine Projektion ist, wenn $\operatorname{Im} f \subset \operatorname{Ker} g$ und $\operatorname{Im} g \subset \operatorname{Ker} f$.

Gilt diese Aussage auch für Vektorräume über einem beliebigen Körper?

Bijektive Morphismen wie in Lemma 13.25 (a) haben in der Praxis eine besondere Bedeutung, und daher auch einen besonderen Namen.

Definition 13.30 (Isomorphismen)**.** Es seien V und W zwei K-Vektorräume.

(a) Einen bijektiven Morphismus $f\colon V \to W$ (der nach Lemma 13.25 (a) also einen Umkehrmorphismus $f^{-1}\colon W \to V$ besitzt) bezeichnet man als **(Vektorraum-)Isomorphismus**.

(b) V und W heißen **isomorph** (in Zeichen: $V \cong W$), wenn es einen Isomorphismus $f\colon V \to W$ zwischen ihnen gibt.

Beispiel 13.31. Anschaulich bedeutet ein Isomorphismus f zwischen zwei Vektorräumen V und W, dass diese beiden Räume „als Vektorräume ununterscheidbar" sind: Die Objekte in V und W sind zwar unterschiedlich benannt, aber in allen Rechnungen können wir jederzeit mit der bijektiven Abbildung f bzw. der inversen Abbildung f^{-1} zwischen den beiden Darstellungen in V und W hin- und herwechseln, ohne das Endergebnis zu ändern. Die folgenden beiden Beispiele verdeutlichen dies.

(a) Der Unterraum

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\} \quad \text{ist mit} \quad f\colon V \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

isomorph zu $\mathbb{R}^2$. In der Tat ist in diesem Beispiel offensichtlich, dass f linear und bijektiv ist. Auch anschaulich ist in diesem Fall klar, dass V und $\mathbb{R}^2$ „im Prinzip ununterscheidbar" sind, denn beide Räume sind einfach die reelle Ebene – die im Fall von V lediglich als Koordinatenebene in den $\mathbb{R}^3$ eingebettet ist.

(b) Der Vektorraum V aller reellen Polynome vom Grad höchstens 2 ist mit der linearen Abbildung

$$f\colon V \to \mathbb{R}^3,\ a_0 + a_1 x + a_2 x^2 \mapsto \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

isomorph zu $\mathbb{R}^3$. Auch hier ist wieder klar, dass f linear ist, und dass der Vektor der Koeffizienten a_0, a_1, a_2 dieselben Informationen enthält wie das Polynom $a_0 + a_1 x + a_2 x^2$ (siehe Bemerkung 3.22).

14. Basen und Dimension

In diesem Kapitel wollen wir beginnen, die Struktur von Vektorräumen genauer zu untersuchen. Besonders zentral ist dabei der Begriff der Basis, den ihr ja wahrscheinlich schon aus der Schule kennt. Die Idee dabei ist, dass man aus jedem Vektorraum V eine (im Moment der Einfachheit halber als endlich angenommene) Familie von Vektoren $x_1, \dots, x_n$ auswählen kann, so dass sich jedes $x \in V$ auf eindeutige Art als eine sogenannte Linearkombination

$$x = \lambda_1 x_1 + \cdots + \lambda_n x_n$$

dieser ausgewählten Vektoren mit geeigneten Skalaren $\lambda_1, \dots, \lambda_n \in K$ schreiben lässt. Die $\lambda_1, \dots, \lambda_n$ lassen sich dann als „Koordinaten" von x bezüglich der Vektoren $x_1, \dots, x_n$ auffassen. Der Vorteil daran ist, dass man mit diesen Koordinaten in vielen Fällen deutlich einfacher rechnen kann als mit dem ursprünglichen Vektor x selbst, da sie ja einfach Elemente aus dem Grundkörper K sind und nicht aus einem abstrakten Vektorraum.

14.A Linearkombinationen und Basen

Als Erstes wollen wir den zentralen Begriff der Linearkombinationen einführen, also der Vektoren, die sich aus einer gegebenen Familie von Vektoren mit Hilfe beliebiger Skalarmultiplikationen und Vektoradditionen erzeugen lassen.

Definition 14.1 (Familien und Linearkombinationen)**.** Es sei V ein K-Vektorraum.

(a) Eine **Familie** von Vektoren in V ist eine Indexmenge I zusammen mit einem Vektor $x_i \in V$ für alle $i \in I$. Wir schreiben eine solche Familie als $B = (x_i)_{i \in I}$ und nennen die x_i ihre **Elemente**. Ist die Indexmenge I endlich, so sprechen wir auch von einer **endlichen Familie** B, wählen als Indexmenge in der Regel $I = \{1, \dots, n\}$, und schreiben die Familie dann einfach als $B = (x_1, \dots, x_n)$.

(b) Es sei $B = (x_i)_{i \in I}$ eine Familie von Vektoren in V. Eine **Linearkombination** der Vektoren aus B ist ein Vektor der Form

$$\sum_{i \in I} \lambda_i x_i \quad \in V, \tag{$*$}$$

wobei $\lambda_i \in K$ für alle $i \in I$ gilt und nur endlich viele dieser λ_i ungleich 0 sind (so dass der Ausdruck $(*)$ nach Weglassen aller Summanden, die 0 sind, also in jedem Fall nur eine endliche Summe ist). Im Fall der endlichen Indexmenge $I = \{1, \dots, n\}$ schreibt man die Linearkombination in der Regel als $\lambda_1 x_1 + \cdots + \lambda_n x_n$.

Wir bezeichnen die Menge aller Linearkombinationen von Vektoren aus B mit $\operatorname{Lin} B$.

Bemerkung 14.2.

(a) Beachte, dass eine *Familie* $B = (x_i)_{i \in I}$ nicht das gleiche ist wie eine *Menge* $\{x_i : i \in I\}$ von Vektoren: Im Gegensatz zu einer Menge sind die gegebenen Vektoren in einer Familie den Elementen einer Indexmenge zugeordnet, haben also z. B. im Fall einer endlichen Familie mit Indexmenge $\{1, \dots, n\}$ eine festgelegte Reihenfolge $x_1, \dots, x_n$. Für die obige Definition 14.1 (b) ist dies noch nicht wichtig (wir hätten $\operatorname{Lin} B$ genauso gut auf die gleiche Art für eine *Menge* B definieren können, und in der Tat wird das in der Literatur auch oft getan), aber später werden wir die gegebene Reihenfolge der Vektoren zwingend benötigen, um die Koordinaten eines Vektors den Elementen einer Basis zuordnen zu können.

(b) Im Folgenden werden wir uns in den meisten Fällen nur mit endlichen Familien beschäftigen. In diesem Fall ist die Bedingung in Definition 14.1 (b), dass nur endlich viele der Koeffizienten λ_i ungleich 0 sein dürfen, natürlich automatisch erfüllt und kann daher auch weggelassen

werden. Im Fall unendlicher Familien ist sie jedoch sehr wichtig, da der Ausdruck $\sum_{i\in I}\lambda_i x_i$ sonst eine unendliche Summe wäre und damit ohne Konvergenzbetrachtungen wie in Kapitel 7 (die nicht Gegenstand der linearen Algebra sind und in einem allgemeinen Körper auch gar nicht definiert werden können) keinen Sinn ergeben würde.

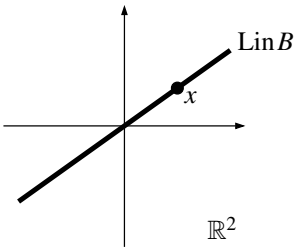
30

Beispiel 14.3.

(a) Für $m \in \mathbb{R}$ und die einelementige Familie $B = (x)$ in $\mathbb{R}^2$ mit $x = \begin{pmatrix} 1 \\ m \end{pmatrix}$ ist

$$\begin{aligned} \operatorname{Lin} B &= \left\{ \lambda \cdot \begin{pmatrix} 1 \\ m \end{pmatrix} : \lambda \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \lambda \\ \lambda m \end{pmatrix} : \lambda \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : x_2 = m x_1 \right\} \end{aligned}$$

genau der Unterraum U_1 aus Beispiel 13.8 (a). In der Tat werden wir gleich in Lemma 14.4 sehen, dass $\operatorname{Lin} B$ wie im Bild rechts eine sehr anschauliche Bedeutung hat: Es ist immer ein Untervektorraum, und zwar der kleinste, der alle Elemente von B enthält.



(b) Für die leere Familie $B = ()$ in einem beliebigen Vektorraum V ist $\operatorname{Lin} B = \{0\}$, da die leere Summe $\sum_{i\in\emptyset}\lambda_i x_i$ analog zu Notation 3.9 (c) konventionsgemäß gleich 0 ist.

(c) Im Vektorraum $V = \operatorname{Abb}(\mathbb{N},\mathbb{R})$ aller reellen Zahlenfolgen aus Beispiel 13.3 (d) betrachten wir die Familie $B = (e_i)_{i\in\mathbb{N}}$ aller „Einheitsfolgen“ $e_i = (0,\dots,0,1,0,0,\dots)$, wobei die 1 jeweils an der Stelle i steht. Wegen der Endlichkeitsbedingung in Definition 14.1 (b) ist $\operatorname{Lin} B$ dann *nicht* der gesamte Raum aller Folgen, sondern nur die Menge aller Folgen, bei denen nur endlich viele Folgenglieder ungleich 0 sind.

Lemma 14.4 (Erzeugte Unterräume)**.** *Für jede Familie B in einem Vektorraum V gilt:*

(a) $\operatorname{Lin} B$ *ist ein Unterraum von V, der alle Elemente von B enthält.*

(b) *Ist U ein beliebiger Unterraum von V, der alle Elemente von B enthält, so ist* $\operatorname{Lin} B \subset U$.

Anschaulich ist $\operatorname{Lin} B$ *also „der kleinste Unterraum von V, der B enthält“. Man nennt ihn daher auch den von B* ***erzeugten*** *oder* ***aufgespannten Unterraum****.*

Beweis. Es sei $B = (x_i)_{i\in I}$.

(a) Natürlich ist jedes x_i in $\operatorname{Lin} B$ enthalten, da wir es als Linearkombination mit Koeffizient 1 vor x_i und 0 vor allen anderen Elementen von B schreiben können. Wir müssen also nur noch die Unterraumkriterien aus Definition 13.5 überprüfen. Dies ist aber sehr einfach: $\operatorname{Lin} B$ enthält z. B. den Nullvektor $\sum_{i\in I} 0\cdot x_i$ und ist damit nicht leer, und für alle $x = \sum_{i\in I}\mu_i x_i$ und $y = \sum_{i\in I}\nu_i x_i$ in $\operatorname{Lin} B$ (mit also nur endlich vielen μ_i und ν_i ungleich 0) sowie $\lambda \in K$ gilt

$$x + y = \sum_{i\in I}(\mu_i+\nu_i)x_i \quad \in V \qquad \text{und} \qquad \lambda x = \sum_{i\in I}(\lambda\mu_i)x_i \quad \in V.$$

Da auch hier natürlich wieder nur endlich viele $\mu_i+\nu_i$ bzw. $\lambda\mu_i$ ungleich 0 sind, liegen diese Elemente ebenfalls in $\operatorname{Lin} B$.

(b) Ist U ein solcher Unterraum, so enthält er also alle x_i mit $i \in I$, und damit nach Bemerkung 13.6 (a) auch jede Linearkombination dieser Vektoren, also $\operatorname{Lin} B$. □

Bemerkung 14.5 (Summen als erzeugte Unterräume)**.** Es sei V ein K-Vektorraum. Ferner seien $U_1 = \operatorname{Lin}\left((x_i)_{i\in I_1}\right)$ und $U_2 = \operatorname{Lin}\left((x_i)_{i\in I_2}\right)$ zwei Untervektorräume von V, die von Vektoren x_i mit $i \in I_1$ bzw. $i \in I_2$ erzeugt werden. Dann sind die Elemente der Summe U_1+U_2 natürlich gerade die Linearkombinationen aus allen Vektoren x_i mit $i \in I_1$ oder $i \in I_2$ zusammen genommen. Es ist dann also $U_1+U_2 = \operatorname{Lin}\left((x_i)_{i\in I_1\cup I_2}\right)$.

Mit unseren Vorarbeiten können wir nun den Begriff der Basis eines Vektorraums definieren.

Definition 14.6 (Basen von Vektorräumen)**.** Es sei $B = (x_i)_{i \in I}$ eine Familie von Vektoren in einem K-Vektorraum V.

(a) Die Familie B heißt ein **Erzeugendensystem** von V, wenn $\operatorname{Lin} B = V$ gilt, d. h. wenn sich jeder Vektor $x \in V$ als Linearkombination der Vektoren in B schreiben lässt.

(b) Die Familie B heißt **linear abhängig**, wenn es eine Linearkombination $\sum_{i \in I} \lambda_i x_i = 0$ des Nullvektors gibt, in der mindestens ein λ_i ungleich 0 ist (man nennt dies auch eine *nicht-triviale Linearkombination* des Nullvektors).

Ist das Gegenteil der Fall, folgt aus der Linearkombination $\sum_{i \in I} \lambda_i x_i = 0$ des Nullvektors mit zunächst beliebigen $\lambda_i \in K$ also bereits, dass alle λ_i gleich 0 sein müssen, so heißt B **linear unabhängig**.

(c) Die Familie B heißt eine **Basis** von V, wenn B ein Erzeugendensystem von V und linear unabhängig ist.

Beispiel 14.7.

(a) Enthält eine Familie $B = (x_i)_{i \in I}$ den Nullvektor, d. h. ist $x_i = 0$ für ein $i \in I$, so ist B stets linear abhängig, denn dann ist ja $1 \cdot x_i = 0$ eine nicht-triviale Linearkombination des Nullvektors. Ebenso ist B immer linear abhängig, wenn die Familie einen Vektor mehrfach enthält, also wenn $x_i = x_j$ für gewisse $i, j \in I$ mit $i \neq j$ gilt, da dann $1 \cdot x_i - 1 \cdot x_j = 0$ eine nicht-triviale Linearkombination des Nullvektors ist.

(b) Ist $V = K^n$ und sind

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

die sogenannten **Einheitsvektoren** von K^n, so ist die Familie $B = (e_1, \ldots, e_n)$ aller dieser Einheitsvektoren eine Basis von V:

- B ist ein Erzeugendensystem von V, denn jeder Vektor in V hat die Form

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_1 e_1 + \cdots + \lambda_n e_n$$

  für gewisse $\lambda_1, \ldots, \lambda_n \in K$.

- B ist linear unabhängig, denn sind $\lambda_1, \ldots, \lambda_n \in K$ mit

$$0 = \lambda_1 e_1 + \cdots + \lambda_n e_n = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix},$$

  so ist natürlich notwendigerweise $\lambda_1 = \cdots = \lambda_n = 0$.

Man nennt B die **Standardbasis** von K^n.

Als Spezialfall davon für $n = 0$ ist die leere Familie eine Basis des Nullvektorraums (siehe Beispiel 14.3 (b)).

(c) In $V = \mathbb{R}^3$ ist die Familie

$$B = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right)$$

kein Erzeugendensystem, denn kein Vektor mit letztem Eintrag ungleich 0 kann eine Linearkombination der Elemente aus B sein. Die Familie B ist auch nicht linear unabhängig, denn

$$\begin{pmatrix}1\\0\\0\end{pmatrix}+\begin{pmatrix}0\\1\\0\end{pmatrix}-\begin{pmatrix}1\\1\\0\end{pmatrix}=0$$

ist eine nicht-triviale Linearkombination des Nullvektors aus den Elementen von B. Anschaulich bedeutet dies gerade, dass einer der drei Vektoren in der Ebene liegt, die von den beiden anderen erzeugt wird.

Beachte auch, dass je zwei der drei Vektoren dieser Familie B jedoch stets linear unabhängig sind. Die lineare Unabhängigkeit einer Familie kann also *nicht* überprüft werden, indem man immer nur zwei ihrer Vektoren miteinander vergleicht.

(d) Im Vektorraum $V \leq \mathrm{Abb}(\mathbb{R},\mathbb{R})$ aller Polynomfunktionen aus Beispiel 13.8 (c) ist die Familie $B=(x^i)_{i\in\mathbb{N}}$ aller Potenzfunktionen mit natürlichen Exponenten eine Basis von V:

- B ist ein Erzeugendensystem, da jede Polynomfunktion nach Definition eine (endliche!) Linearkombination der Potenzfunktionen x^i ist;
- B ist linear unabhängig, da eine nicht-triviale Linearkombination der x^i nach dem Koeffizientenvergleich aus Lemma 3.21 nie die Nullfunktion sein kann.

Genauso ist natürlich $(x^0,x^1,\dots,x^n)$ eine Basis des Vektorraums aller Polynomfunktionen vom Grad höchstens n.

(e) Umformuliert bedeutet Beispiel 14.3 (c), dass die Einheitsfolgen $e_i=(0,\dots,0,1,0,0,\dots)$, wobei die 1 an Position i steht, kein Erzeugendensystem des Folgenraums $\mathrm{Abb}(\mathbb{N},\mathbb{R})$ bilden. Sie sind jedoch linear unabhängig, denn ist $\sum_{i\in\mathbb{N}}\lambda_i e_i=(0,0,0,\dots)$ (für eine gewisse Summe mit nur endlich vielen $\lambda_i\neq 0$), so folgt durch Vergleich des i-ten Folgengliedes natürlich sofort $\lambda_i=0$ für alle i.

Aufgabe 14.8. Untersuche die Familie B in den folgenden Fällen auf lineare Unabhängigkeit im Vektorraum V:

(a) $V=\mathbb{R}^3$; $B=\left(\begin{pmatrix}1\\2\\3\end{pmatrix},\begin{pmatrix}4\\5\\6\end{pmatrix},\begin{pmatrix}7\\8\\9\end{pmatrix}\right)$.

(b) V ein beliebiger Vektorraum; $B=(x+y,x+z,y+z)$ für drei linear unabhängige Vektoren x,y,z.

(c) $V=\mathrm{Abb}(\mathbb{R},\mathbb{R})$; $B=(f,f')$ für ein gegebenes Polynom f (wobei f' die Ableitung von f wie in Beispiel 13.17 (d) ist).

(d) $V=\mathrm{Abb}(\mathbb{R},\mathbb{R})$; $B=(f_a)_{a\in\mathbb{R}}$ mit

$$f_a\colon \mathbb{R}\to\mathbb{R},\ x\mapsto\begin{cases}0 & \text{für } x\leq a,\\ x-a & \text{für } x\geq a.\end{cases}$$

Wie wir in der Einleitung zu diesem Kapitel schon erwähnt haben, ist das Schöne an einer Basis eines Vektorraums V nun, dass sich jeder Vektor aus V auf eindeutige Art als Linearkombination dieser Vektoren schreiben lässt. Dies wird insbesondere bei der Untersuchung linearer Abbildungen in Abschnitt 16.B noch eine große Rolle spielen.

Lemma und Definition 14.9 (Koordinatendarstellung)**.** *Es sei $B=(x_i)_{i\in I}$ eine Basis eines K-Vektorraums V. Dann lässt sich jeder Vektor $x\in V$ auf eindeutige Art als Linearkombination*

$$x=\sum_{i\in I}\lambda_i x_i$$

*mit $\lambda_1,\dots,\lambda_n\in K$ (von denen nur endlich viele ungleich 0 sind) schreiben. Man nennt die λ_i die **Koordinaten** von x bezüglich der Basis B.*

Beweis. Es sei $x \in V$. Da B ein Erzeugendensystem von V ist, ist $x \in \operatorname{Lin} B$ also auch eine Linearkombination der Vektoren aus B, d. h. es gibt eine Darstellung der gewünschten Form $x = \sum_{i \in I} \lambda_i x_i$.

Um auch noch die Eindeutigkeit dieser Darstellung zu zeigen, sei $x = \sum_{i \in I} \mu_i x_i$ eine weitere Darstellung von x als Linearkombination der gegebenen Basisvektoren. Dann erhalten wir durch Subtraktion der beiden Darstellungen

$$0 = x - x = \sum_{i \in I} \lambda_i x_i - \sum_{i \in I} \mu_i x_i = \sum_{i \in I} (\lambda_i - \mu_i) x_i.$$

Da B linear unabhängig ist, folgt daraus für alle i aber sofort $\lambda_i - \mu_i = 0$ und damit $\lambda_i = \mu_i$, d. h. die beiden Darstellungen müssen gleich sein. □

14.B Existenz von Basen

Um Vektorräume mit Hilfe von Basen untersuchen zu können, müssen wir uns jetzt natürlich noch fragen, ob denn überhaupt jeder Vektorraum eine Basis besitzt – aus den Vektorraumaxiomen ist das ja nicht offensichtlich. In der Tat ist das aber so: Man kann beweisen, dass jeder beliebige Vektorraum eine Basis hat. Wir wollen uns beim Beweis dieser Tatsache allerdings auf Vektorräume beschränken, die von *endlich vielen* Elementen erzeugt werden können und dann auch eine *endliche* Basis besitzen (den allgemeinen Beweis könnt ihr z. B. in [GK] Proposition II.2.22 finden). Dies liegt zum einen daran, dass der Beweis für Vektorräume mit unendlichen Basen deutlich komplizierter und abstrakter ist. Zum anderen – und das ist fast der wichtigere Grund – ist der Beweis im allgemeinen Fall *nicht konstruktiv* und daher eigentlich nur von theoretischem Interesse. Betrachten wir z. B. noch einmal den $\mathbb{R}$-Vektorraum $\operatorname{Abb}(\mathbb{N}, \mathbb{R})$ aller reellen Zahlenfolgen aus Beispiel 13.3 (d) und Beispiel 14.7 (e). Wir werden in Beispiel 14.20 (c) sehen, dass dieser Vektorraum keine *endliche* Basis besitzt. Man weiß nun zwar aufgrund des oben angegebenen Satzes, dass dieser Folgenraum dennoch eine (unendliche) Basis hat, aber niemand kann eine solche Basis konkret angeben! Versucht doch einmal, eine Basis zu finden – ihr werdet sehr schnell merken, dass das aussichtslos ist. Zur Erinnerung: Ihr müsstet dazu eine (unendliche) Familie von Folgen hinschreiben, so dass *jede beliebige* Folge auf *eindeutige* Art eine *endliche* Linearkombination der Folgen ist, die ihr ausgewählt habt.

Formal bedeutet dies, dass wir uns in Zukunft in der Regel auf Vektorräume mit der folgenden Zusatzbedingung beschränken wollen.

Definition 14.10 (Endlich erzeugte Vektorräume)**.** Ein K-Vektorraum V heißt **endlich erzeugt**, wenn er ein Erzeugendensystem aus endlich vielen Vektoren besitzt.

Für derartige Vektorräume wollen wir nun die Existenz von Basen zeigen. In der Tat werden wir die stärkeren Aussagen beweisen, dass man aus jedem Erzeugendensystem (das evtl. noch nicht linear unabhängig ist) eine Basis auswählen und jede linear unabhängig Menge (die evtl. noch kein Erzeugendensystem ist) zu einer Basis ergänzen kann.

Satz 14.11 (**Basisauswahl**)**.** *Es sei B ein endliches Erzeugendensystem eines Vektorraums V. Dann kann man aus B eine Basis von V auswählen.*

Insbesondere besitzt also jeder endlich erzeugte Vektorraum eine (endliche) Basis.

Beweis. Es sei $B = (x_1, \ldots, x_n)$ ein Erzeugendensystem von V. Ist B bereits linear unabhängig, so sind wir fertig. Andernfalls gibt es eine Linearkombination $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$, bei der nicht alle $\lambda_1, \ldots, \lambda_n$ gleich Null sind. Wir können also ein $i \in \{1, \ldots, n\}$ wählen mit $\lambda_i \neq 0$. Mit diesem i setzen wir nun $B' = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ und behaupten, dass B' immer noch ein Erzeugendensystem von V ist.

In der Tat ist dies leicht einzusehen: Wegen der vorausgesetzten Linearkombination des Nullvektors und $\lambda_i \neq 0$ ist ja

$$x_i = \frac{1}{\lambda_i} \cdot \big(-\lambda_1 x_1 - \cdots - \lambda_{i-1} x_{i-1} - \lambda_{i+1} x_{i+1} - \cdots - \lambda_n x_n \big) \quad \in \operatorname{Lin}(B'),$$

da sich dieser Vektor als Linearkombination von $x_1,\dots,x_{i-1},x_{i+1},\dots,x_n$ darstellen lässt. Damit enthält der Unterraum $\operatorname{Lin}(B')$ alle Vektoren $x_1,\dots,x_n$, nach Lemma 14.4 (b) also auch den von diesen Vektoren erzeugten Unterraum $\operatorname{Lin}(x_1,\dots,x_n)=\operatorname{Lin}(B)=V$. Es ist daher $\operatorname{Lin}(B')=V$, d. h. B' ist immer noch ein Erzeugendensystem von V.

Ist B' nun linear unabhängig, so sind wir fertig. Andernfalls wiederholen wir das obige Verfahren so lange, bis die resultierende Familie linear unabhängig und damit eine Basis von V ist (dies muss spätestens nach n Schritten passieren, da dann keine Vektoren mehr übrig sind und die leere Familie natürlich linear unabhängig ist). □

Beispiel 14.12. Beachte, dass der Beweis von Satz 14.11 auch konstruktiv ist, d. h. ein Verfahren angibt, wie man aus einem Erzeugendensystem B eine Basis bekommt: Solange es eine nicht-triviale Linearkombinationen der Null in B gibt, lässt man einen beliebigen Vektor aus B weg, der in dieser Linearkombination mit einem Vorfaktor ungleich Null auftaucht. So gibt es z. B. in dem Erzeugendensystem

$$B=\left(\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix},\begin{pmatrix}0\\2\end{pmatrix}\right)$$

von $\mathbb{R}^2$ die nicht-triviale Linearkombination

$$2\cdot\begin{pmatrix}0\\1\end{pmatrix}-1\cdot\begin{pmatrix}0\\2\end{pmatrix}=0,$$

also können wir aus B einen der Vektoren $\begin{pmatrix}0\\1\end{pmatrix}$ und $\begin{pmatrix}0\\2\end{pmatrix}$ weglassen. Lassen wir in diesem Fall z. B. den Vektor $\begin{pmatrix}0\\2\end{pmatrix}$ weg, so haben wir damit (nach Beispiel 14.7 (b)) bereits eine Basis von $\mathbb{R}^2$ erhalten, nämlich die Standardbasis.

Als Nächstes wollen wir unterschiedliche Basen eines Vektorraums miteinander vergleichen können. Hierfür beweisen wir zunächst ein Hilfsresultat.

Lemma 14.13 (Austauschlemma). *Es sei $B=(x_1,\dots,x_n)$ eine (endliche) Basis eines Vektorraums V. Weiterhin sei $y\in V$ ein beliebiger Vektor, den wir natürlich (nach Lemma 14.9) als Linearkombination $y=\lambda_1x_1+\dots+\lambda_nx_n$ schreiben können. Ist dann $i\in\{1,\dots,n\}$ mit $\lambda_i\neq 0$, so ist auch $B'=(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n)$ eine Basis von V (wir können in der Basis also x_i durch y ersetzen).*

Beweis. Nach evtl. Umbenennung der Vektoren können wir ohne Beschränkung der Allgemeinheit $i=1$, also $\lambda_1\neq 0$ und $B'=(y,x_2,\dots,x_n)$ annehmen. Dann gilt:

- B' erzeugt V: Wegen
$$x_1=\frac{1}{\lambda_1}\cdot\big(y-\lambda_2x_2-\dots-\lambda_nx_n\big)\quad\in\operatorname{Lin}(B')$$
enthält $\operatorname{Lin}(B')=\operatorname{Lin}(y,x_2,\dots,x_n)$ alle Vektoren $x_1,\dots,x_n$ und nach Lemma 14.4 (b) damit auch den davon erzeugen Unterraum $\operatorname{Lin}(x_1,\dots,x_n)=\operatorname{Lin}(B)=V$. Also ist B' ein Erzeugendensystem von V.
- B' ist linear unabhängig: Ist
$$\mu_1y+\mu_2x_2+\dots+\mu_nx_n=0$$
eine Linearkombination der Null mit Vektoren aus B', so folgt daraus
$$\begin{aligned}0&=\mu_1(\lambda_1x_1+\lambda_2x_2+\dots+\lambda_nx_n)+\mu_2x_2+\dots+\mu_nx_n\\&=(\mu_1\lambda_1)x_1+(\mu_1\lambda_2+\mu_2)x_2+\dots+(\mu_1\lambda_n+\mu_n)x_n.\end{aligned}$$
Dies ist nun eine Linearkombination der Null mit Vektoren aus B. Da B linear unabhängig ist, müssen darin alle Vorfaktoren verschwinden, d. h. es gilt $\mu_1\lambda_1=0$ (und damit $\mu_1=0$ wegen $\lambda_1\neq 0$) und dann auch $\mu_1\lambda_i+\mu_i=\mu_i=0$ für alle $i=2,\dots,n$. Also ist B' linear unabhängig. □

31

Beispiel 14.14. Wir betrachten den Vektorraum $V = \mathbb{R}^3$ mit der Standardbasis $B = (e_1, e_2, e_3)$ aus Beispiel 14.7 (b). Weiterhin sei $y = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = e_1 + 2e_2$. Da in dieser Linearkombination die Vektoren e_1 und e_2 mit Vorfaktoren ungleich Null vorkommen, folgt aus dem Austauschlemma 14.13 also, dass wir einen dieser Vektoren durch y ersetzen können, d. h. dass auch (y, e_2, e_3) und (e_1, y, e_3) Basen von $\mathbb{R}^3$ sind. Im Gegensatz dazu ist (e_1, e_2, y) jedoch keine Basis von V (wie aus der Linearkombination $e_1 + 2e_2 - y = 0$ natürlich auch sofort folgt).

Der folgende Satz ergibt sich nun einfach, indem man Lemma 14.13 mehrmals nacheinander anwendet.

Satz 14.15 (Steinitzscher Austauschsatz). *Es seien $B = (x_1, \dots, x_n)$ eine Basis und $(y_1, \dots, y_r)$ eine linear unabhängige Familie in einem Vektorraum V.*

Dann ist $r \leq n$, und die $x_1, \dots, x_n$ lassen sich so umnummerieren, dass $B' = (y_1, \dots, y_r, x_{r+1}, \dots, x_n)$ ebenfalls eine Basis von V ist (man kann in der Basis also r der Vektoren $x_1, \dots, x_n$ durch die gegebenen Vektoren $y_1, \dots, y_r$ ersetzen).

Beweis. Wir beweisen den Satz mit Induktion über r; für $r = 0$ ist nichts zu zeigen.

Für den Induktionsschritt $r \to r+1$ sei nun $(y_1, \dots, y_{r+1})$ linear unabhängig. Da dann natürlich auch $(y_1, \dots, y_r)$ linear unabhängig ist, gilt nach Induktionsvoraussetzung $r \leq n$, und nach geeigneter Umnummerierung der Vektoren in B ist $(y_1, \dots, y_r, x_{r+1}, \dots, x_n)$ eine Basis von V. Wir können den Vektor y_{r+1} also als Linearkombination

$$y_{r+1} = \lambda_1 y_1 + \dots + \lambda_r y_r + \lambda_{r+1} x_{r+1} + \dots + \lambda_n x_n$$

schreiben. Dabei muss mindestens einer der Vorfaktoren $\lambda_{r+1}, \dots, \lambda_n$ ungleich Null sein, denn andernfalls wäre

$$\lambda_1 y_1 + \dots + \lambda_r y_r - y_{r+1} = 0$$

im Widerspruch dazu, dass die Familie $(y_1, \dots, y_{r+1})$ linear unabhängig ist. Insbesondere muss also $r + 1 \leq n$ gelten, und nach evtl. Umbenennung der $x_{r+1}, \dots, x_n$ können wir annehmen, dass $\lambda_{r+1} \neq 0$ ist. Dann ist nach dem Austauschlemma 14.13 aber auch $(y_1, \dots, y_{r+1}, x_{r+2}, \dots, x_n)$ eine Basis von V. Damit ist der Satz mit Induktion bewiesen. □

Der Steinitzsche Austauschsatz hat zwei wichtige Konsequenzen: zum einen das bereits angekündigte Gegenstück zur Basisauswahl in Satz 14.11, und zum anderen die Tatsache, dass alle Basen eines endlich erzeugten Vektorraums gleich viele Elemente haben.

Folgerung 14.16 (Basisergänzung). *Es sei B eine linear unabhängige Familie eines endlich erzeugten Vektorraums V. Dann ist B endlich, und man kann B zu einer endlichen Basis von V ergänzen.*

Beweis. Nach Satz 14.11 besitzt V eine endliche Basis $(x_1, \dots, x_n)$. Wegen Satz 14.15 kann es dann keine linear unabhängige Teilmenge von V mit mehr als n Elementen geben; insbesondere ist B also endlich und von der Form $B = (y_1, \dots, y_r)$ für ein $r \leq n$. Satz 14.15 besagt nun, dass wir diese Vektoren $y_1, \dots, y_r$ in die Basis $(x_1, \dots, x_n)$ hineintauschen können, so dass wir also eine Basis von V bekommen, die die Vektoren aus B enthält. □

Aufgabe 14.17. Es sei $U = \mathrm{Lin}(x_1, x_2, x_3) \leq \mathbb{R}^3$ mit

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Finde eine Basis von U und ergänze sie zu einer Basis von $\mathbb{R}^3$.

Folgerung 14.18. *Alle Basen eines endlich erzeugten Vektorraums sind endlich und haben gleich viele Elemente.*

Beweis. Es sei V ein endlich erzeugter Vektorraum.Nach Folgerung 14.16 ist zunächst einmal jede Basis von V endlich. Sind nun $B = (x_1, \dots, x_n)$ und $B' = (y_1, \dots, y_r)$ zwei Basen von V, so folgt $r \leq n$ aus Satz 14.15, und aus Symmetriegründen durch Vertauschen von B und B' dann natürlich auch $n \leq r$. Also ist $n = r$, d. h. B und B' haben gleich viele Elemente. □

Da jeder endlich erzeugte Vektorraum nach Satz 14.11 auch wirklich eine Basis besitzt, können wir also definieren:

Definition 14.19 (Dimension von Vektorräumen)**.** Für einen endlich erzeugten Vektorraum V ist die **Dimension**, geschrieben $\dim_K V$ oder einfach $\dim V$, definiert als die Anzahl der Elemente in einer (beliebigen) Basis von V.

Ist V nicht endlich erzeugt (und hat damit natürlich auch keine endliche Basis), so schreiben wir formal $\dim V = \infty$. Ein endlich erzeugter Vektorraum wird daher oft auch als **endlich-dimensionaler Vektorraum** bezeichnet.

Beispiel 14.20.

(a) Nach Beispiel 14.7 (b) ist $\dim K^n = n$ für alle $n \in \mathbb{N}$.

(b) Nach Beispiel 14.7 (d) hat der Raum aller reellen Polynomfunktionen vom Grad höchstens n eine Basis $(x^0, x^1, \dots, x^n)$, und damit Dimension $n+1$.

(c) Es sei $V = \mathrm{Abb}(\mathbb{N}, \mathbb{R})$ der Raum aller reellen Zahlenfolgen aus Beispiel 13.3 (d). Nach Beispiel 14.7 (d) besitzt V eine unendliche linear unabhängige Familie von Vektoren und kann somit nach Folgerung 14.16 nicht endlich erzeugt sein. Es ist also $\dim V = \infty$.

Unsere erste Anwendung des Dimensionsbegriffs ist ein vereinfachtes Kriterium dafür, ob eine gegebene Familie B eines endlich-dimensionalen Vektorraums V eine Basis ist: Wenn wir bereits wissen, dass B die richtige Anzahl von Vektoren enthält, dann brauchen wir nicht mehr zu überprüfen, dass B ein Erzeugendensystem *und* linear unabhängig ist, sondern es reicht bereits *eine* dieser beiden Eigenschaften.

Satz 14.21 (Basiskriterium)**.** *Es sei $B = (x_1, \dots, x_n)$ eine Familie von n Vektoren in einem endlich erzeugten Vektorraum V.*

(a) *Ist B ein Erzeugendensystem von V, so gilt $n \geq \dim V$.*

(b) *Ist B linear unabhängig, so gilt $n \leq \dim V$.*

Ist eine dieser beiden Voraussetzungen erfüllt und sogar $n = \dim V$, so ist B bereits eine Basis von V.

Beweis. Dies folgt unmittelbar mit Hilfe der Basisauswahl und -ergänzung:

(a) Nach Satz 14.11 können wir aus B eine Basis B' von V auswählen. Da diese Basis B' dann aus $\dim V$ Vektoren besteht, muss natürlich $\dim V \leq n$ gelten.

(b) Nach Folgerung 14.16 können wir B zu einer Basis B' von V ergänzen. Da diese Basis wieder aus $\dim V$ Vektoren besteht, folgt $\dim V \geq n$.

In beiden Fällen folgt im Fall $n = \dim V$ natürlich $B' = B$, d. h. B ist bereits eine Basis von V. □

Wir wollen nun sehen, wie man endlich erzeugte Vektorräume mit Hilfe des Dimensionsbegriffs klassifizieren kann. Immer wenn man eine neue mathematische Struktur eingeführt und ein paar Beispiele untersucht hat, fragt man sich in der Regel, ob man vielleicht sogar eine *vollständige* Liste aller Beispiele angeben kann – in unserem Fall also, ob wir eine vollständige Liste aller (endlich erzeugten) Vektorräume hinschreiben können. Dabei soll „vollständig“ immer „vollständig bis auf Isomorphie“ bedeuten, da wir ja in Beispiel 13.31 schon gesehen haben, dass isomorphe Vektorräume von ihrer Struktur her ohnehin ununterscheidbar sind, so dass es uns bei isomorphen Vektorräumen natürlich reichen sollte, wenn einer von ihnen in unserer Liste steht.

Bei vielen mathematischen Strukturen ist eine derartige Klassifikation schlichtweg aussichtslos, weil es viel zu viele Beispiele gibt, die auch keinem ersichtlichen Schema folgen. Dies ist z. B. bei Gruppen (siehe Definition 3.1) der Fall – niemand kann eine vollständige Liste aller Gruppen (bis auf Isomorphie) angeben. Es ist eine Besonderheit der linearen Algebra, dass dies bei endlich erzeugten Vektorräumen anders ist: Diese Vektorräume sind genau durch ihre Dimension klassifiziert, d. h. zu jedem Körper K und jeder natürlichen Zahl $n \in \mathbb{N}$ gibt es bis auf Isomorphie *genau einen* K-Vektorraum dieser Dimension, nämlich K^n.

Dieses Ergebnis, das wir jetzt zeigen wollen, wird uns die Arbeit später ganz wesentlich vereinfachen: Wenn wir Aussagen über beliebige endlich erzeugte Vektorräume beweisen wollen, genügt es deswegen nämlich, nur die Vektorräume K^n zu betrachten – und mit denen lässt es sich natürlich deutlich leichter arbeiten als mit abstrakten Vektorräumen.

Um dieses Klassifikationsresultat zu zeigen, beginnen wir mit der einfachen Aussage, dass isomorphe Vektorräume dieselbe Dimension haben. Dies sollte, da man sich isomorphe Vektorräume ja als „ununterscheidbar“ vorstellen kann, nicht weiter erstaunlich sein.

Satz 14.22. *Es seien $f\colon V \to W$ ein Morphismus zwischen K-Vektorräumen, $B = (x_1, \dots, x_n)$ eine endliche Familie in V und $f(B) := (f(x_1), \dots, f(x_n))$ die zugehörige Familie der Bildvektoren in W. Dann gilt:*

(a) *Ist f surjektiv und B ein Erzeugendensystem von V, so ist $f(B)$ ein Erzeugendensystem von W.*

(b) *Ist f injektiv und B linear unabhängig, so ist auch $f(B)$ linear unabhängig.*

(c) *Ist f ein Isomorphismus und B eine Basis von V, so ist $f(B)$ eine Basis von W. Insbesondere haben isomorphe endlich erzeugte Vektorräume also die gleiche Dimension.*

Beweis.

(a) Es sei $y \in W$ beliebig. Da f surjektiv ist, gibt es ein $x \in V$ mit $f(x) = y$, und weil B ein Erzeugendensystem von V ist, können wir $x = \lambda_1 x_1 + \dots + \lambda_n x_n$ für gewisse $\lambda_1, \dots, \lambda_n \in K$ schreiben. Damit ist aber auch
$$y = f(x) = f(\lambda_1 x_1 + \dots + \lambda_n x_n) = \lambda_1 f(x_1) + \dots + \lambda_n f(x_n)$$
eine Linearkombination von Vektoren aus $f(B)$. Also ist $f(B)$ ein Erzeugendensystem von W.

(b) Es seien $\lambda_1, \dots, \lambda_n \in K$ mit $\lambda_1 f(x_1) + \dots + \lambda_n f(x_n) = 0$, also (weil f ein Morphismus ist) mit
$$f(\lambda_1 x_1 + \dots + \lambda_n x_n) = 0.$$
Nach Voraussetzung ist f injektiv, d. h. es ist $\operatorname{Ker} f = \{0\}$ aufgrund von Lemma 13.24. Also folgt bereits $\lambda_1 x_1 + \dots + \lambda_n x_n = 0$ und damit auch $\lambda_1 = \dots = \lambda_n = 0$, da B linear unabhängig ist. Die Familie $f(B)$ ist somit linear unabhängig.

(c) ist einfach nur (a) kombiniert mit (b). □

Nach Teil (c) dieses Satzes ist also $K^n \ncong K^m$ für $n \neq m$. Um unser angekündigtes Klassifikationsresultat zu zeigen, müssen wir also nur noch beweisen, dass jeder K-Vektorraum der Dimension n bereits isomorph zu K^n ist.

Satz 14.23 (Klassifikation endlich-dimensionaler Vektorräume)**.** *Es sei V ein endlich-dimensionaler K-Vektorraum mit* $\dim V = n$. *Dann ist V isomorph zu K^n.*

Beweis. Wir wählen eine Basis $(x_1, \dots, x_n)$ von V. Dann ist die Abbildung
$$f\colon K^n \to V, \quad \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \mapsto \lambda_1 x_1 + \dots + \lambda_n x_n$$
ein Isomorphismus (es ist klar, dass f linear ist, und f ist nach Lemma 14.9 bijektiv). □

Folgerung 14.24 (Dimension von Produkten)**.** *Sind V und W zwei endlich-dimensionale Vektorräume, so gilt*

$$\dim(V \times W) = \dim V + \dim W.$$

Eine analoge Aussage gilt natürlich auch für Produkte von mehr als zwei Vektorräumen.

Beweis. Nach Satz 14.23 gibt es Isomorphismen $f\colon K^n \to V$ und $g\colon K^m \to W$ mit $n = \dim V$ und $m = \dim W$. Dann ist aber auch

$$K^{n+m} = K^n \times K^m \to V \times W,\ (x,x') \mapsto (f(x), g(x'))$$

ein Isomorphismus (mit Umkehrabbildung $(y,y') \mapsto (f^{-1}(y), g^{-1}(y'))$), und damit folgt

$$\dim(V \times W) = \dim K^{n+m} = n+m = \dim V + \dim W$$

nach Satz 14.22 (c) und Beispiel 14.20 (a). □

Aufgabe 14.25. Es sei (a_n) die sogenannte *Fibonacci-Folge*, die durch $a_0 = a_1 = 1$ sowie die Rekursionsgleichung

$$a_{n+2} = a_{n+1} + a_n \qquad \text{für alle } n \in \mathbb{N} \tag{$*$}$$

gegeben ist, also die Folge $(1,1,2,3,5,8,13,21,\dots)$. Wir wollen in dieser Aufgabe eine explizite Formel für das n-te Folgenglied a_n herleiten.

Es sei dazu $V \leq \mathrm{Abb}(\mathbb{N},\mathbb{R})$ der Unterraum aller reellen Zahlenfolgen, die die Rekursionsgleichung $(*)$ erfüllen (ihr braucht nicht nachzuweisen, dass dies wirklich ein Unterraum ist, solltet euch aber trotzdem kurz überlegen, warum das so ist).

(a) Für welche $q \in \mathbb{R}$ liegt die Folge $(q^n)_{n\in\mathbb{N}} = (1,q,q^2,q^3,\dots)$ in V? (Die Existenz von Quadratwurzeln aus nicht-negativen reellen Zahlen wie in Folgerung 5.29 könnt ihr dabei voraussetzen.)

(b) Zeige, dass $\dim V = 2$, und bestimme eine Basis von V.

(c) Berechne eine explizite nicht-rekursive Formel für die Glieder a_n der Fibonacci-Folge.

Zum Abschluss dieses Kapitels wollen wir noch zeigen, dass Unterräume endlich erzeugter Vektorräume stets wieder endlich erzeugt sind, also dass wir die schöne Welt der endlich-dimensionalen Vektorräume nicht verlassen, wenn wir Teilmengen gegebener Vektorräume betrachten. Dieses Resultat ist vermutlich nicht wirklich überraschend, aber dennoch auch nicht völlig offensichtlich, da es keine einfache Möglichkeit gibt, aus einem Erzeugendensystem eines Vektorraums ein Erzeugendensystem eines gegebenen Unterraums zu konstruieren.

Lemma 14.26. *Es sei U ein Unterraum eines endlich-dimensionalen K-Vektorraums V. Dann ist auch U endlich-dimensional, und es ist* $\dim U \leq \dim V$.

Gilt sogar $\dim U = \dim V$*, so ist* $U = V$.

Beweis. Wäre U nicht endlich erzeugt, so könnten wir rekursiv eine (unendliche) Folge $x_1,x_2,x_3,\dots$ in U wählen mit $x_n \notin \mathrm{Lin}(x_1,\dots,x_{n-1})$ für alle $n \geq 1$, denn es ist ja stets $\mathrm{Lin}(x_1,\dots,x_{n-1}) \subsetneq U$. Die Familie $(x_n)_{n\in\mathbb{N}_{>0}}$ ist dann linear unabhängig: Gäbe es eine Relation $\lambda_1 x_1 + \dots + \lambda_n x_n = 0$ mit $\lambda_1,\dots,\lambda_n \in K$, die nicht alle 0 sind, so könnten wir darin n so wählen, dass $\lambda_n \neq 0$ ist, und damit den Widerspruch $x_n = -\frac{1}{\lambda_n}(\lambda_1 x_1 + \dots + \lambda_{n-1}x_{n-1}) \in \mathrm{Lin}(x_1,\dots,x_{n-1})$ erhalten. Also ist $(x_n)_{n\in\mathbb{N}_{>0}}$ eine unendliche linear unabhängige Familie in U, und damit auch in V. Dies ist aber ein Widerspruch zu Folgerung 14.16, da V als endlich erzeugt vorausgesetzt wurde. Also ist U endlich erzeugt.

Nach Satz 14.11 gibt es damit eine (endliche) Basis $(x_1,\dots,x_n)$ von U, wobei $n = \dim U$. Da diese Vektoren insbesondere linear unabhängig (in V) sind, können wir sie nach Folgerung 14.16 zu einer Basis $(x_1,\dots,x_n,y_1,\dots,y_m)$ von V ergänzen, wobei $n+m = \dim V$. Dann gilt aber

$$\dim U = n \leq n+m = \dim V.$$

Ist darüberhinaus $\dim U = \dim V$, also $m = 0$, so wird V bereits von den Vektoren $x_1,\dots,x_n$ erzeugt, d. h. es gilt $V = \mathrm{Lin}(x_1,\dots,x_n) = U$. □

32

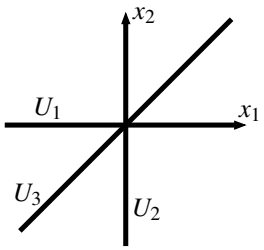
15. Quotientenräume und Dimensionsformeln

Im letzten Kapitel haben wir den zentralen Begriff der Dimension eines Vektorraums eingeführt. Für einen irgendwie konstruierten Vektorraum, z. B. als Durchschnitt oder Summe zweier Unterräume, oder als Bild oder Kern einer gegebenen linearen Abbildung, haben wir bisher jedoch kaum effiziente Möglichkeiten, seine Dimension (oder gar eine Basis) explizit zu bestimmen. Wir wollen nun einige weitere nützliche Konstruktionen und Resultate der Vektorraumtheorie behandeln, mit denen sich unter anderem Dimensionen in vielen Fällen einfach ablesen lassen. Ab dem nächsten Kapitel werden wir dann auch algorithmische Verfahren untersuchen, die selbst in komplizierten Fällen die explizite Berechnung nahezu aller unserer Konstruktionen ermöglichen.

15.A Komplemente und Quotientenräume

Als Erstes wollen wir uns wie in Definition 13.10 und Bemerkung 13.11 noch einmal die Konstruktion der Summe $U_1 + \cdots + U_n$ von Unterräumen $U_1, \ldots, U_n$ eines Vektorraums V anschauen.

Bemerkung 15.1 (Eindeutigkeit der Summendarstellung)**.** Jeder Vektor in einer Summe $U_1 + \cdots + U_n$ von Unterräumen eines Vektorraums V lässt sich nach Definition als $x_1 + \cdots + x_n$ mit $x_i \in U_i$ für alle $i = 1, \ldots, n$ schreiben. Allerdings ist diese Darstellung im Allgemeinen nicht eindeutig: Betrachten wir wie im Bild rechts die drei Ursprungsgeraden



$$U_1 = \operatorname{Lin}\begin{pmatrix}1\\0\end{pmatrix}, \quad U_2 = \operatorname{Lin}\begin{pmatrix}0\\1\end{pmatrix}, \quad U_3 = \operatorname{Lin}\begin{pmatrix}1\\1\end{pmatrix},$$

in $\mathbb{R}^2$, so hat z. B. der Vektor

$$\begin{pmatrix}1\\1\end{pmatrix} = \underbrace{\begin{pmatrix}1\\0\end{pmatrix}}_{\in U_1} + \underbrace{\begin{pmatrix}0\\1\end{pmatrix}}_{\in U_2} + \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_3} = \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_1} + \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_2} + \underbrace{\begin{pmatrix}1\\1\end{pmatrix}}_{\in U_3} \quad \in U_1 + U_2 + U_3 = \mathbb{R}^2$$

zwei verschiedene Darstellungen dieser Art. Ist die Darstellung jedoch immer eindeutig, so geben wir dieser Situation einen besonderen Namen:

Definition 15.2 (Direkte Summe von Unterräumen)**.** Es seien $U_1, \ldots, U_n$ Untervektorräume eines K-Vektorraums V und $U = U_1 + \cdots + U_n$. Hat jedes $x \in U$ eine *eindeutige* Darstellung der Form $x = x_1 + \cdots + x_n$ mit $x_i \in U_i$ für alle $i = 1, \ldots, n$, so nennt man die Summe **direkt**. Möchte man dies auch in der Notation andeuten, so schreibt man dafür $U = U_1 \oplus \cdots \oplus U_n$.

Die Summe in Bemerkung 15.1 oben ist also nicht direkt – was natürlich einfach daran liegt, dass die drei aufspannenden Vektoren von U_1, U_2 und U_3 linear abhängig sind. In der Tat kann man sich direkte Summen als eine Verallgemeinerung des Konzepts der linearen Abhängigkeit auf Unterräume vorstellen.

Lemma 15.3 (Dimension direkter Summen)**.** *Die Summe $U_1 + \cdots + U_n$ von Unterräumen $U_1, \ldots, U_n$ eines K-Vektorraums V ist genau dann direkt, wenn die Abbildung*

$$U_1 \times \cdots \times U_n \to U_1 + \cdots + U_n, \ (x_1, \ldots, x_n) \mapsto x_1 + \cdots + x_n$$

ein Isomorphismus ist. Ist V endlich-dimensional, so gilt in diesem Fall insbesondere

$$\dim(U_1 \oplus \cdots \oplus U_n) = \dim U_1 + \cdots + \dim U_n.$$

Beweis. Es ist klar, dass die gegebene Abbildung in jedem Fall linear ist; außerdem ist sie nach Definition der Summe $U_1+\dots+U_n$ stets surjektiv. Injektiv ist sie genau dann, wenn für alle $x_i,y_i \in U_i$ aus $x_1+\dots+x_n = y_1+\dots+y_n$ bereits $(x_1,\dots,x_n)=(y_1,\dots,y_n)$, also $x_i=y_i$ für alle i folgt. Dies bedeutet nach Definition 15.2 aber genau, dass die Summe direkt ist.

Ist V darüber hinaus endlich-dimensional, so gilt dies nach Lemma 14.26 auch für die Unterräume $U_1,\dots,U_n$. Da endlich erzeugte isomorphe Vektorräume nach Satz 14.22 (c) die gleiche Dimension haben, ergibt sich aus Folgerung 14.24 also

$$\dim(U_1\oplus\dots\oplus U_n) = \dim(U_1\times\dots\times U_n) = \dim U_1+\dots+\dim U_n. \qquad \square$$

Im Fall von nur zwei Unterräumen kann man besonders einfach feststellen, ob ihre Summe direkt ist:

Lemma 15.4. *Die Summe U_1+U_2 von zwei Unterräumen U_1 und U_2 eines K-Vektorraums V ist genau dann direkt, wenn $U_1\cap U_2=\{0\}$.*

Beweis. Nach Lemma 15.3 ist die Summe U_1+U_2 genau dann direkt, wenn die Abbildung

$$f\colon U_1\times U_2 \to U_1+U_2,\ (x_1,x_2)\mapsto x_1+x_2$$

ein Isomorphismus ist. Wir hatten im Beweis dieses Lemmas aber auch schon gesehen, dass f stets linear und surjektiv ist. Also ist die Summe U_1+U_2 genau dann direkt, wenn f injektiv ist, d. h. nach Lemma 13.24 genau dann, wenn $\operatorname{Ker} f=\{(0,0)\}$. Nun ist aber

$$\begin{aligned}\operatorname{Ker} f &= \{(x_1,x_2): x_1\in U_1, x_2\in U_2, x_1+x_2=0\}\\ &= \{(x_1,-x_1): x_1\in U_1, -x_1\in U_2\}\\ &= \{(x_1,-x_1): x_1\in U_1\cap U_2\},\end{aligned}$$

und damit ist wie behauptet genau dann $\operatorname{Ker} f=\{(0,0)\}$, wenn $U_1\cap U_2=\{0\}$. $\square$

Beispiel 15.5.

(a) Die Summe U_1+U_2 der x_1-Achse und der x_2-Achse in $\mathbb{R}^3$ in Beispiel 13.12 ist direkt, denn in diesem Fall ist natürlich $U_1\cap U_2=\{0\}$. In der Tat sieht man in diesem Beispiel auch sofort, dass sich jeder Vektor in der x_1-x_2-Ebene U_1+U_2 eindeutig als Summe von einem Vektor in U_1 und einem in U_2 schreiben lässt.

(b) Die Summe $U_1+U_2+U_3$ in Bemerkung 15.1 ist hingegen nicht direkt, wie wir dort bereits gesehen hatten. Allerdings ist in diesem Fall trotzdem $U_1\cap U_2\cap U_3=\{0\}$ – was zeigt, dass sich die Aussage von Lemma 15.4 nicht genauso auf mehr als zwei Summanden übertragen lässt. Die zu Lemma 15.4 analoge Aussage für allgemeine Summen ist stattdessen die folgende:

Aufgabe 15.6. Zeige, dass die folgenden Aussagen für Unterräume $U_1,\dots,U_n$ eines K-Vektorraums V äquivalent sind:

(a) $U=U_1\oplus\dots\oplus U_n$;

(b) $U=U_1+\dots+U_n$ und $U_i\cap(U_1+\dots+U_{i-1}+U_{i+1}+\dots+U_n)=\{0\}$ für alle $i=1,\dots,n$.

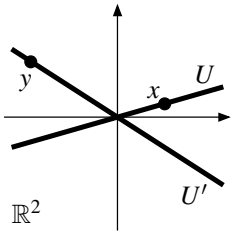
Einen besonderen Namen hat die Situation, wenn die Summe zweier Unterräume direkt und gleich dem gesamten Vektorraum ist.

Definition 15.7 (Komplemente)**.** Es sei U ein Unterraum eines K-Vektorraums V. Ein Unterraum $U'\le V$ heißt **Komplement** von U in V, wenn $U\oplus U'=V$ (nach Lemma 15.4 also $U+U'=V$ und $U\cap U'=\{0\}$) gilt.

Bemerkung 15.8 (Dimension von Komplementen)**.** Nach Lemma 15.3 gilt für jedes Komplement U' eines Untervektorraums U in einem endlich-dimensionalen Vektorraum V die Dimensionsformel $\dim U+\dim U'=\dim V$, also $\dim U'=\dim V-\dim U$.

Beispiel 15.9 (Nichteindeutigkeit von Komplementen)**.** Wie im Bild unten rechts seien $U = \mathrm{Lin}(x)$ und $U' = \mathrm{Lin}(y)$ zwei verschiedene Ursprungsgeraden in $\mathbb{R}^2$. Da x und y dann keine Vielfachen voneinander sind, sind diese beiden Vektoren also linear unabhängig; mit Satz 14.21 bilden sie daher eine Basis des zweidimensionalen Vektorraums $\mathbb{R}^2$.

Nach Bemerkung 14.5 ist damit $U + U' = \mathrm{Lin}(x, y) = \mathbb{R}^2$, außerdem gilt offensichtlich $U \cap U' = \{0\}$. Also ist U' ein Komplement von U.



Da es zu einer gegebenen Ursprungsgeraden U in $\mathbb{R}^2$ natürlich unendlich viele Geraden $U' \neq U$ gibt, folgt daraus insbesondere, dass Komplemente von Unterräumen in der Regel nicht eindeutig sind. Wir wollen nun aber sehen, dass Komplemente zumindest im endlich-dimensionalen Fall stets existieren:

Satz 15.10 (Existenz von Komplementen)**.** *Jeder Unterraum U eines endlich-dimensionalen K-Vektorraums V besitzt ein Komplement.*

Beweis. Nach Lemma 14.26 ist U endlich erzeugt, hat also nach Satz 14.11 eine Basis $(x_1, \ldots, x_n)$. Wir ergänzen sie gemäß Folgerung 14.16 zu einer Basis $(x_1, \ldots, x_n, y_1, \ldots, y_m)$ von V und behaupten, dass $U' := \mathrm{Lin}(y_1, \ldots, y_m)$ dann ein Komplement von U ist.

- Es ist offensichtlich $U + U' = V$, denn nach Bemerkung 14.5 ist
$$U + U' = \mathrm{Lin}(x_1, \ldots, x_n) + \mathrm{Lin}(y_1, \ldots, y_m) = \mathrm{Lin}(x_1, \ldots, x_n, y_1, \ldots, y_m) = V.$$
- Es ist $U \cap U' = \{0\}$: Für $x \in U \cap U'$, also $x = \lambda_1 x_1 + \cdots + \lambda_n x_n = \mu_1 y_1 + \cdots + \mu_m y_m$ für gewisse $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_m \in K$, erhalten wir durch Subtraktion
$$0 = x - x = \lambda_1 x_1 + \cdots + \lambda_n x_n - \mu_1 y_1 - \cdots - \mu_m y_m.$$
Da die Familie $(x_1, \ldots, x_n, y_1, \ldots, y_m)$ linear unabhängig ist, ist dies aber nur möglich für $\lambda_1 = \cdots = \lambda_n = \mu_1 = \cdots = \mu_m = 0$, also für $x = 0$.

Also ist U' ein Komplement von U in V. □

Bemerkung 15.11 (Berechnung von Komplementen)**.** Beachte, dass der Beweis von Satz 15.10 konstruktiv ist, d. h. auch die konkrete Berechnung eines Komplements ermöglicht: Möchte man ein Komplement U' zu einem Unterraum U eines endlich-dimensionalen Vektorraums V berechnen, muss man (z. B. mit Hilfe des Verfahrens aus Satz 14.15) eine Basis von U zu einer Basis von V ergänzen; die dafür hinzu genommenen Vektoren bilden dann eine Basis eines Komplements U'.

Ist z. B. $y = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \in \mathbb{R}^3$ und $U = \mathrm{Lin}(y)$, so haben wir in Beispiel 14.14 gesehen, dass man y z. B. für e_1 in die Standardbasis (e_1, e_2, e_3) hineintauschen kann, dass also (y, e_2, e_3) ebenfalls eine Basis von $\mathbb{R}^3$ ist. Damit ist $U' = \mathrm{Lin}(e_2, e_3)$ ein Komplement von U.

Komplemente von Unterräumen sind in der Praxis sehr nützlich und haben auch eine einfache anschauliche Deutung: Ist U' ein Komplement eines Unterraums U in einem Vektorraum V, so lässt sich ja jeder Vektor aus V nach Definition 15.7 eindeutig als Summe aus einem Vektor in U und einem „Restvektor“ in U' schreiben. Wir können uns U' also in gewissem Sinne als einen Unterraum vorstellen, der diese „Restteile“ von Vektoren in V misst, wenn man ihren Anteil in U heraus nimmt. Unschön ist an dieser Konstruktion allerdings, dass ein Komplement nach Beispiel 15.9 nicht eindeutig bestimmt und damit ein recht unnatürliches Objekt ist. Was im obigen Sinne der Restteil eines Vektors in V nach Herausnehmen des Anteils in U ist, lässt sich also nicht beantworten, solange man nicht eine (letztlich willkürliche) Wahl eines Komplements von U in V getroffen hat.

Wir wollen nun eine deutlich schönere Konstruktion einführen, die solche Restteile auch ohne willkürliche Wahlen messen kann. Der Preis dafür ist, dass der Vektorraum, der diese Restteile auf ganz natürliche Art beschreibt, kein *Unterraum* von V mehr ist, sondern ein sogenannter *Quotientenraum*: ein Raum von Äquivalenzklassen von Vektoren in V wie in Abschnitt 2.B, wobei wir zwei Vektoren

in V miteinander identifizieren wollen, wenn sie sich um ein Element von U voneinander unterscheiden. Diejenigen von euch, die auch die Vorlesung „Algebraische Strukturen“ besuchen, kennen diese Idee vielleicht bereits von den Faktorgruppen [G, Kapitel 6].

Lemma und Definition 15.12. *Es seien V ein K-Vektorraum und $U \leq V$ ein fest gewählter Unterraum. Dann ist durch*

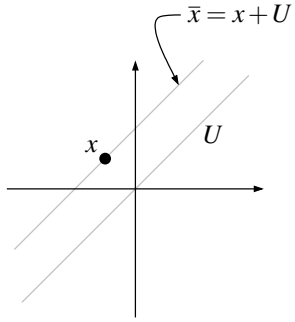
$$x \sim y \quad :\Leftrightarrow \quad x - y \in U \qquad \text{für alle } x, y \in V$$

eine Äquivalenzrelation auf V definiert. Für die Äquivalenzklasse eines Vektors $x \in V$ bezüglich dieser Relation gilt

$$\overline{x} = x + U := \{x + u : u \in U\}.$$

*Man nennt diese Menge (wie im Bild rechts) einen **affinen** bzw. **verschobenen Unterraum** mit Aufpunkt x.*

Die Menge $V/\sim$ aller Äquivalenzklassen bezüglich dieser Relation bezeichnet man mit V/U.



Beweis. Wir zeigen zunächst, dass die gegebene Relation eine Äquivalenzrelation wie in Definition 2.27 ist.

Reflexivität: Für alle $x \in V$ gilt $x - x = 0 \in U$ nach Bemerkung 13.6 (b), und damit $x \sim x$.

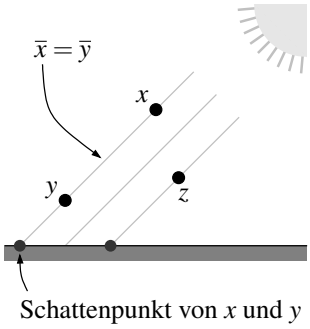
Symmetrie: Es seien $x, y \in V$ mit $x \sim y$, also $x - y \in U$. Dann ist nach Definition 13.5 (c) auch $(-1)(x - y) = y - x \in U$ und damit $y \sim x$.

Transitivität: Sind $x, y, z \in V$ mit $x \sim y$ und $y \sim z$, also $x - y \in U$ und $y - z \in U$, so ist nach Definition 13.5 (b) auch $(x - y) + (y - z) = x - z \in U$ und damit $x \sim z$.

Also ist $\sim$ eine Äquivalenzrelation. Für die Klasse $\overline{x}$ eines Vektors x gilt nun nach Definition 2.27 (c)

$$\overline{x} = \{y \in V : y - x \in U\} = \{y \in V : y - x = u \text{ für ein } u \in U\} = \{x + u : u \in U\}. \qquad \square$$

Bemerkung 15.13 (Anschauliche Deutung von V/U)**.** Die geometrische Bedeutung des Raumes V/U lässt sich am besten wie im Bild rechts erläutern, in dem $V = \mathbb{R}^2$ und $U = \operatorname{Lin}\begin{pmatrix}1\\1\end{pmatrix}$ ist. Dort scheint die Sonne mit parallelen (hell eingezeichneten) Strahlen in Richtung von U und wirft dabei von jedem Punkt in V einen Schatten auf den Boden. In diesem Bild ist die Klasse $\overline{x} \in V/U$ eines Punktes $x \in V$ gerade der Sonnenstrahl durch x. Zwei Punkte in V bestimmen also genau dann den gleichen Punkt in V/U, wenn sie auf dem gleichen Sonnenstrahl liegen, d. h. denselben Schattenpunkt auf dem Boden werfen. Im Bild rechts ist also $\overline{x} = \overline{y} \neq \overline{z}$.



In diesem Sinne kann man sich V/U damit als eine „Schattenwelt“ von V vorstellen, die zwar jeden Punkt von V sieht, aber nur mit einem Teil seiner Informationen: Der „Abstand zur Sonne“ eines Punktes in V ist anhand des Schattenbildes nicht mehr zu rekonstruieren. Für einen Vektor $x \in V$ nimmt die Klasse $\overline{x}$ also wie beabsichtigt „den Anteil in U heraus“.

Bemerkung 15.14. Für zwei Vektoren $x, y \in V$ gilt nach Satz 2.29 (a) genau dann $\overline{x} = \overline{y}$ in V/U, wenn $x \sim y$ ist. Wir sehen mit Definition 15.12 also für alle $x, y \in V$ in V/U:

$$\boxed{\begin{array}{lcll} & \overline{x} = \overline{y} & \Leftrightarrow & x - y \in U, \\ \text{insbesondere also} & \overline{x} = \overline{0} & \Leftrightarrow & x \in U. \end{array}}$$

Mit diesen Rechenregeln kann man Gleichungen zwischen Äquivalenzklassen in V/U immer auf Aussagen über die Repräsentanten in V zurückführen. So ist in der Situation von Bemerkung 15.13

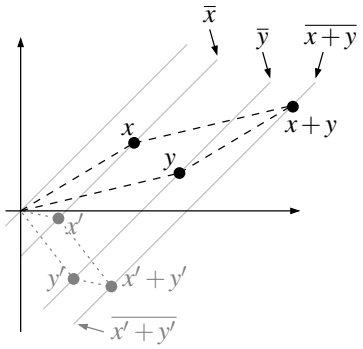
beispielsweise

$$\overline{\begin{pmatrix}3\\5\end{pmatrix}} = \overline{\begin{pmatrix}6\\8\end{pmatrix}}, \quad \text{weil} \quad \begin{pmatrix}3\\5\end{pmatrix} - \begin{pmatrix}6\\8\end{pmatrix} = \begin{pmatrix}-3\\-3\end{pmatrix} \in U.$$

Allerdings fehlt uns noch ein letzter Schritt: Bisher ist der Raum V/U nur eine Menge ohne weitere Struktur. Um ihn im Rahmen der linearen Algebra untersuchen zu können, müssen wir ihn selbst wieder zu einem Vektorraum machen, also auf ihm eine Vektoraddition und Skalarmultiplikation definieren und zeigen, dass damit dann die Vektorraumeigenschaften für V/U gelten.

Die Idee hierfür ist sehr einfach und im Bild rechts dargestellt: Wollen wir die verschobenen Unterräume $\overline{x}$ und $\overline{y}$ in V/U addieren, so addieren wir einfach wie im oberen Teil des Bildes die Aufpunkte x und y und verwenden den so erhaltenen Punkt $x+y$ als Aufpunkt für die Summe, d. h. wir setzen

$$\overline{x}+\overline{y} := \overline{x+y}. \qquad (*)$$



Allerdings müssen wir dabei etwas aufpassen: Wir hätten für dieselben verschobenen Unterräume statt x und y ja auch wie im unteren Teil des Bildes genauso gut andere Aufpunkte x' bzw. y' wählen können und hätten dann als Ergebnis den verschobenen Unterraum $\overline{x'+y'}$ erhalten!

Damit die Vorschrift $(*)$ wirklich widerspruchsfrei eine Verknüpfung auf V/U definiert, müssen wir also überprüfen, dass der verschobene Unterraum $\overline{x'+y'}$ derselbe ist wie $\overline{x+y}$, d. h. dass *das Endergebnis nicht von der Wahl der Aufpunkte abhängt*. Man sagt dazu auch, dass wir die *Wohldefiniertheit* von $(*)$ überprüfen müssen. Eine solche Überprüfung ist immer dann nötig, wenn wir eine Funktion auf einer Menge von Äquivalenzklassen (hier: V/U) definieren wollen und bei der Konstruktion die Wahl eines Repräsentanten einer Äquivalenzklasse (hier: eines Aufpunkts eines verschobenen Unterraums) verwenden. Die allgemeine Situation ist die folgende:

Notation 15.15 (Wohldefiniertheit)**.** Es sei $\sim$ eine Äquivalenzrelation auf einer Menge M. Möchte man auf der Menge $M/\sim$ der Äquivalenzklassen eine Abbildung in eine andere Menge N definieren, so ist die Idee hierfür in der Regel, dass man eine Abbildung $g\colon M \to N$ wählt und dann

$$f\colon M/\sim \to N,\ f(\overline{x}) := g(x) \qquad (*)$$

setzt. Man möchte das Bild einer Äquivalenzklasse unter f also dadurch definieren, dass man einen Repräsentanten dieser Klasse wählt und diesen dann mit g abbildet. Damit dies nun f widerspruchsfrei definiert, brauchen wir offensichtlich, dass das Ergebnis dieser Vorschrift nicht von der Wahl des Repräsentanten abhängt: Sind $x,y \in M$ äquivalent zueinander, sind sie also Repräsentanten derselben Äquivalenzklasse, so muss $g(x) = g(y)$ gelten. Mit anderen Worten benötigen wir

$$g(x) = g(y) \quad \text{für alle } x,y \in M \text{ mit } \overline{x} = \overline{y},$$

damit die Definition $(*)$ widerspruchsfrei ist. Statt „widerspruchsfrei“ sagt man in diesem Fall wie oben schon erwähnt in der Regel, dass f durch die Vorschrift $(*)$ **wohldefiniert** ist. Die Wohldefiniertheit einer Funktion muss man also immer dann nachprüfen, wenn der Startraum der Funktion eine Menge von Äquivalenzklassen ist und die Funktionsvorschrift Repräsentanten dieser Klassen benutzt. Oder noch etwas allgemeiner: Wenn eine Funktionsvorschrift an irgendeiner Stelle eine Wahl beinhaltet, muss man sich vergewissern, dass der letztliche Funktionswert von dieser Wahl unabhängig ist.

33

Nach diesen Vorbemerkungen können wir die Menge V/U nun wie angekündigt zu einem Vektorraum machen:

Satz und Definition 15.16 (Quotientenräume)**.** *Es sei U ein Unterraum eines K-Vektorraums V. Dann sind die Verknüpfungen*

$$\overline{x}+\overline{y} := \overline{x+y} \qquad \textit{und} \qquad \lambda\cdot\overline{x} := \overline{\lambda x} \qquad \textit{für } x,y \in V \textit{ und } \lambda \in K$$

auf V/U wohldefiniert und machen V/U zu einem K-Vektorraum. Man nennt ihn den ***Quotientenraum*** *bzw.* ***Faktorraum*** *von V nach U.*

Man spricht V/U auch als „V ***modulo*** *U" und nennt $\overline{x} \in V/U$ für ein $x \in V$ die* ***Restklasse*** *von V modulo U.*

Beweis. Wir zeigen zunächst die Wohldefiniertheit der Addition: Sind $x, x', y, y' \in V$ mit $\overline{x} = \overline{x'}$ und $\overline{y} = \overline{y'}$, so bedeutet dies nach Bemerkung 15.14 genau $x - x' \in U$ und $y - y' \in U$. Nach Definition 13.5 (b) ist dann aber auch $(x+y) - (x'+y') = (x-x') + (y-y') \in U$ – was wiederum nach Bemerkung 15.14 genau $\overline{x+y} = \overline{x'+y'}$ bedeutet. Also ist die Addition auf V/U wohldefiniert.

Genauso zeigt man die Wohldefiniertheit der Skalarmultiplikation: Sind $\lambda \in K$ und $x, x' \in V$ mit $\overline{x} = \overline{x'}$, also $x - x' \in U$, so ist nach Definition 13.5 (c) auch $\lambda x - \lambda x' = \lambda(x - x') \in U$ und damit $\overline{\lambda x} = \overline{\lambda x'}$.

Die Vektorraumaxiome für V/U ergeben sich nun unmittelbar aus denen von V. So erhält man z. B. die Assoziativität der Vektoraddition durch die einfache Rechnung

$$\left(\overline{x} + \overline{y}\right) + \overline{z} = \overline{x+y} + \overline{z} = \overline{(x+y)+z} = \overline{x+(y+z)} = \overline{x} + \overline{y+z} = \overline{x} + \left(\overline{y} + \overline{z}\right)$$

für alle $x, y, z \in V$, wobei die mittlere Gleichheit die Assoziativität in V ist und sich die anderen Gleichungen aus der Definition der Addition in V/U ergeben. Die übrigen Eigenschaften überprüft man genauso; der Nullvektor in V/U ist die Klasse $\overline{0}$ des Nullvektors in V bzw. der unverschobene Unterraum U, das additive Inverse eines Elements $\overline{x} \in V/U$ ist $\overline{-x}$. □

Aufgrund der anschaulichen Deutung von Komplementen und Quotientenräumen sollte es nicht verwundern, dass wir nun zeigen können, dass diese beiden Konzepte letztlich das gleiche beschreiben, also als Vektorräume isomorph zueinander sind. Wie oben schon erwähnt ist der Vorteil des Komplements lediglich, dass es als Unterraum des ursprünglichen Vektorraums anschaulich leichter zu verstehen ist; der Vorteil des Quotientenraums ist dagegen, dass er ohne willkürliche Wahlen konstruiert werden kann und damit aus mathematischer Sicht das natürlichere Objekt ist.

Satz 15.17 (Dimension von Quotientenräumen)**.** *Es seien U ein Untervektorraum eines K-Vektorraums V und U' ein Komplement von U. Dann ist die Abbildung*

$$f\colon U' \to V/U,\ x \mapsto \overline{x}$$

ein Isomorphismus.

Ist V zusätzlich endlich-dimensional, so gilt also insbesondere $\dim V/U = \dim V - \dim U$.

Beweis. Um zu zeigen, dass f ein Isomorphismus ist, müssen wir die folgenden Dinge überprüfen:

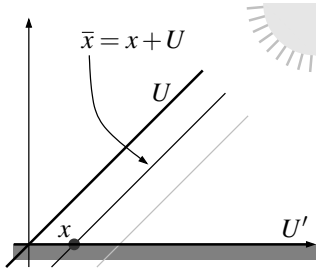
- f ist eine lineare Abbildung, denn für alle $x, y \in U'$ ist

  $$f(x+y) = \overline{x+y} = \overline{x} + \overline{y} = f(x) + f(y),$$

  und eine analoge Aussage gilt natürlich für die Skalarmultiplikation.
- f ist injektiv: Es sei $x \in U'$ mit $f(x) = \overline{x} = \overline{0}$, also $x \in U$ nach Bemerkung 15.14. Dann ist aber $x \in U \cap U' = \{0\}$. Damit ist f nach Lemma 13.24 injektiv.
- f ist surjektiv: Es sei $\overline{x} \in V/U$ beliebig, also $x \in V$. Wegen $V = U + U'$ können wir $x = x_1 + x_2$ mit $x_1 \in U$ und $x_2 \in U'$ schreiben. Dann liegt x_2 in der Definitionsmenge U' von f, und es gilt $f(x_2) = \overline{x_2} = \overline{x}$ nach Bemerkung 15.14, da $x - x_2 = x_1 \in U$. Also ist f surjektiv.

Die Zusatzaussage folgt nun sofort daraus, dass das Komplement U' nach Bemerkung 15.8 die Dimension $\dim V - \dim U$ hat. □

Bemerkung 15.18. Das Bild rechts illustriert in der Situation von Bemerkung 15.13 noch einmal, dass der Morphismus f aus Satz 15.17 bijektiv ist: Die Bodenlinie U' ist nach Beispiel 15.9 ein Komplement der Richtung U der Sonnenstrahlen. Die Abbildung f ordnet nun jedem Punkt $x \in U'$ auf dem Boden den Sonnenstrahl $\overline{x} \in V/U$ durch diesen Punkt zu, und liefert offensichtlich eine Bijektion zwischen den Bodenpunkten und der Menge der Sonnenstrahlen. Wenn wir in Bemerkung 15.13 gesagt haben, dass V/U die „Schattenwelt" auf dem Boden ist, haben wir dabei also schon den Isomorphismus zwischen dem eigentlichen Quotientenraum V/U und dem Boden U' verwendet.



Bemerkung 15.19 (Basen von Quotientenräumen)**.** Nach Satz 15.17 (und Satz 14.22 (c)) erhalten wir eine Basis des Quotientenraums V/U, indem wir eine Basis eines Komplements von U wählen und die Restklassen dieser Vektoren in V/U nehmen. Kombinieren wir dies mit dem Verfahren aus Bemerkung 15.11, so bedeutet dies: Wir können eine Basis von V/U konstruieren, indem wir eine Basis von U zu einer Basis von V ergänzen und dann die Restklassen der hinzu genommenen Vektoren wählen. Im Beispiel von Bemerkung 15.11 ist also z. B. $(\overline{e_2}, \overline{e_3})$ eine Basis von V/U.

Aufgabe 15.20. Es sei $U = \operatorname{Lin}\begin{pmatrix}1\\2\\0\end{pmatrix} \leq \mathbb{R}^3$. Man zeige:

(a) Die Familie $B = \left(\overline{\begin{pmatrix}1\\0\\0\end{pmatrix}}, \overline{\begin{pmatrix}0\\1\\0\end{pmatrix}}\right)$ in $\mathbb{R}^3/U$ ist weder linear unabhängig noch ein Erzeugendensystem.

(b) $f\colon \mathbb{R}^3/U \to \mathbb{R}^3$, $\overline{\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}} \mapsto \begin{pmatrix}2x_1 - x_2 - 3x_3\\0\\2x_3\end{pmatrix}$ ist eine wohldefinierte lineare Abbildung.

Aufgabe 15.21 (Invariante Unterräume)**.** Es sei $f\colon V \to V$ eine lineare Abbildung. Ein Unterraum $U \subset V$ heißt *f-invariant*, wenn $f(U) \subset U$.

Zeige, dass U genau dann f-invariant ist, wenn $g\colon V/U \to V/U$, $\overline{x} \mapsto \overline{f(x)}$ eine lineare Abbildung definiert.

Aufgabe 15.22. Es sei U ein Unterraum eines endlich erzeugten K-Vektorraums V.

Man zeige: Ist $(x_1, \ldots, x_n)$ eine Basis von U und sind $y_1, \ldots, y_m \in V$ so dass $(\overline{y_1}, \ldots, \overline{y_m})$ eine Basis von V/U ist, so ist $(x_1, \ldots, x_n, y_1, \ldots, y_m)$ eine Basis von V.

15.B Isomorphiesätze und Dimensionsformeln

Als Anwendung der Quotientenräume werden wir nun einige nützliche Isomorphismen zwischen unseren bisherigen Konstruktionen herleiten, aus denen dann weitere wichtige Dimensionsformeln folgen. Als Erstes wollen wir dazu den sogenannten Homomorphiesatz beweisen, der aus jeder linearen Abbildung $f\colon V \to W$ „einen Isomorphismus machen kann". Die Idee hierfür ist sehr einfach: Natürlich kann man f zunächst einmal surjektiv machen, indem man den Zielraum W durch den Unterraum $\operatorname{Im} f$ ersetzt. Um f auch noch injektiv zu machen, also gemäß Lemma 13.24 den Kern zu $\{0\}$ zu machen, können wir einfach den Startraum V durch den Quotientenraum $V/\operatorname{Ker} f$ ersetzen: Auf diese Art werden alle Elemente des Kerns von f miteinander identifiziert, so dass der Kern der neuen Abbildung auf dem Quotientenraum nur noch aus dem einen Element $\overline{0} = \operatorname{Ker} f$ besteht.

Satz 15.23 (**Homomorphiesatz**)**.** *Für jede lineare Abbildung $f\colon V \to W$ ist die Abbildung*

$$g\colon V/\operatorname{Ker} f \to \operatorname{Im} f,\ \overline{x} \mapsto f(x)$$

(wohldefiniert und) ein Isomorphismus.

Beweis. Wir müssen einige Dinge überprüfen:

- Die Abbildung g ist wohldefiniert: Sind $x, y \in V$ mit $\overline{x} = \overline{y}$, also $x - y \in \operatorname{Ker} f$ nach Bemerkung 15.14, so ist $f(x-y) = f(x) - f(y) = 0$ und damit $f(x) = f(y)$.
- Die Abbildung g ist linear: Für $x, y \in V$ gilt
$$g(\overline{x}+\overline{y}) = g(\overline{x+y}) = f(x+y) = f(x) + f(y) = g(\overline{x}) + g(\overline{y});$$
analog folgt auch die Verträglichkeit mit der Skalarmultiplikation.
- Die Abbildung g ist surjektiv: Dies ist klar nach Definition von $\operatorname{Im} f$, denn jedes Element in $\operatorname{Im} f$ ist ja von der Form $f(x) = g(\overline{x})$ für ein $x \in V$.
- Die Abbildung g ist injektiv: Nach Lemma 13.24 genügt es dafür zu zeigen, dass $\operatorname{Ker} g = \{\overline{0}\}$. Es sei also $x \in V$ mit $g(\overline{x}) = 0$. Dann ist $f(x) = 0$, also $x \in \operatorname{Ker} f$ und damit $\overline{x} = \overline{0} \in V/\operatorname{Ker} f$ nach Bemerkung 15.14. □

Beispiel 15.24 (Anschauliche Deutung des Homomorphiesatzes)**.** Als anschauliches Beispiel für den Homomorphiesatz können wir noch einmal die „Schattenwelt“ aus Bemerkung 15.13 und Bemerkung 15.18 betrachten. Ist $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die lineare Abbildung, die einen Punkt auf seinen Schattenpunkt auf den Boden abbildet, so ist $\operatorname{Ker} f = U$ der Sonnenstrahl durch 0 und $\operatorname{Im} f = U'$ der Boden. Satz 15.23 gibt uns dann den Isomorphismus $g\colon \mathbb{R}^2/U \to U'$, der jeden Sonnenstrahl auf seinen Bodenpunkt abbildet und genau die Umkehrung des Isomorphismus aus Satz 15.17 ist.

Aufgabe 15.25. Die lineare Abbildung, die der Situation in Beispiel 15.24 entspricht, ist
$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix}.$$
Überprüfe den Homomorphiesatz in diesem Fall explizit, d. h. zeige durch eine direkte Rechnung, dass die Abbildung
$$g\colon \mathbb{R}^2/\operatorname{Lin}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \to \operatorname{Lin}\begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \overline{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \mapsto \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix}$$
wohldefiniert, linear, surjektiv und injektiv ist.

Folgerung 15.26 (Dimensionsformel für Morphismen)**.** *Für jeden Morphismus $f\colon V \to W$ zwischen endlich-dimensionalen K-Vektorräumen gilt* $\dim \operatorname{Im} f + \dim \operatorname{Ker} f = \dim V$.

Beweis. Nach dem Homomorphiesatz 15.23 ist $V/\operatorname{Ker} f$ isomorph zu $\operatorname{Im} f$. Also gilt
$$\dim \operatorname{Im} f = \dim(V/\operatorname{Ker} f) = \dim V - \dim \operatorname{Ker} f$$
nach Satz 15.17. □

Folgerung 15.27. *Es seien V ein endlich-dimensionaler K-Vektorraum und $f\colon V \to V$ ein Morphismus. Dann gilt*
$$f \text{ ist injektiv} \quad\Leftrightarrow\quad f \text{ ist surjektiv} \quad\Leftrightarrow\quad f \text{ ist bijektiv}.$$

Beweis. Es genügt zu zeigen, dass f genau dann injektiv ist, wenn f surjektiv ist:
$$\begin{aligned} f \text{ ist injektiv} &\Leftrightarrow \operatorname{Ker} f = \{0\} && \text{(nach Lemma 13.24)} \\ &\Leftrightarrow \dim \operatorname{Ker} f = 0 \\ &\Leftrightarrow \dim V - \dim \operatorname{Im} f = 0 && \text{(nach Folgerung 15.26)} \\ &\Leftrightarrow \dim \operatorname{Im} f = \dim V \\ &\Leftrightarrow \operatorname{Im} f = V && \text{(nach Lemma 14.26)} \\ &\Leftrightarrow f \text{ ist surjektiv.} \end{aligned}$$
□

Beispiel 15.28. Folgerung 15.27 ist für unendlich-dimensionale Vektorräume falsch: Betrachten wir z. B. auf dem Vektorraum $V = \mathrm{Abb}(\mathbb{N},\mathbb{R})$ aller reellen Zahlenfolgen die Morphismen

$$f\colon V \to V,\ (a_0,a_1,a_2,\dots) \mapsto (0,a_0,a_1,a_2,\dots)$$
$$\text{und} \qquad g\colon V \to V,\ (a_0,a_1,a_2,\dots) \mapsto (a_1,a_2,a_3,\dots)$$

aus Beispiel 13.17 (e), so ist f injektiv aber nicht surjektiv, und g surjektiv aber nicht injektiv.

Nach Kernen und Bildern linearer Abbildungen betrachten wir als Nächstes Summen und Durchschnitte von Unterräumen. Auch dafür gibt es zunächst wieder eine nützliche Isomorphieaussage.

Satz 15.29 (**Isomorphiesatz** für Summen und Durchschnitte)**.** *Es seien U_1 und U_2 zwei Unterräume eines K-Vektorraums V. Dann ist die Abbildung*

$$g\colon U_1/(U_1 \cap U_2) \to (U_1+U_2)/U_2,\ \overline{x} \mapsto \overline{x}$$

(wohldefiniert und) ein Isomorphismus.

Beweis. Wir betrachten zunächst die Abbildung

$$f\colon U_1 \to (U_1+U_2)/U_2,\ x \mapsto \overline{x}.$$

Sie ist offensichtlich linear (und Wohldefiniertheit muss hier nicht überprüft werden, da der Startraum kein Quotientenraum ist). Wir bestimmen das Bild und den Kern von f:

- $\operatorname{Im} f = (U_1+U_2)/U_2$, d. h. f ist surjektiv: Für ein beliebiges Element $\overline{x_1+x_2} \in (U_1+U_2)/U_2$ (mit $x_1 \in U_1$ und $x_2 \in U_2$) ist $x_1 \in U_1$ ein Urbild unter f, denn $f(x_1) = \overline{x_1} = \overline{x_1+x_2}$.
- $\operatorname{Ker} f = U_1 \cap U_2$: Ein Element $x \in U_1$ liegt genau dann im Kern von f, wenn $f(x) = \overline{x} = \overline{0}$ ist, was nach Bemerkung 15.14 äquivalent ist zu $x \in U_2$. Der Kern von f besteht also aus allen Elementen von U_1, die auch in U_2 liegen, und ist damit gleich $U_1 \cap U_2$.

Der Homomorphiesatz 15.23 für f liefert nun den Isomorphismus $g\colon U_1/\operatorname{Ker} f \to \operatorname{Im} f$, $\overline{x} \mapsto f(x)$, also wie behauptet

$$g\colon U_1/(U_1 \cap U_2) \to (U_1+U_2)/U_2,\ \overline{x} \mapsto \overline{x}. \qquad \square$$

Bemerkung 15.30. In der Literatur sind für die Sätze 15.23 und 15.29 z. T. recht unterschiedliche Bezeichnungen üblich. Manche Bücher bezeichnen sie als den 1. und 2. Isomorphiesatz, während andere aber auch Satz 15.29 den 1. Isomorphiesatz nennen.

Folgerung 15.31 (Dimensionsformel für Summen und Durchschnitte)**.** *Sind U_1 und U_2 endlich erzeugte Unterräume eines K-Vektorraums V, so gilt*

$$\dim(U_1+U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2.$$

Beweis. Nach dem Isomorphiesatz 15.29 haben $U_1/(U_1 \cap U_2)$ und $(U_1+U_2)/U_2$ dieselbe Dimension. Aus der Dimensionsformel für Quotienten in Satz 15.17 ergibt sich also

$$\dim U_1 - \dim(U_1 \cap U_2) = \dim(U_1+U_2) - \dim U_2$$

und somit die Behauptung. $\square$

Beispiel 15.32. Beachte, dass nur die Summe der Dimensionen von $U_1 \cap U_2$ und U_1+U_2 durch $\dim U_1$ und $\dim U_2$ bestimmt sind, nicht aber die Dimensionen von $U_1 \cap U_2$ und U_1+U_2 selbst. Als einfaches Beispiel hierfür seien U_1 und U_2 zwei Geraden (durch den Ursprung) in $\mathbb{R}^2$. Dann gibt es zwei Möglichkeiten:

(a) Ist $U_1 = U_2$, so ist $U_1+U_2 = U_1 \cap U_2 = U_1 = U_2$, und die Dimensionsformel ergibt

$$\dim(U_1+U_2) + \dim(U_1 \cap U_2) = 1+1 = 1+1 = \dim U_1 + \dim U_2.$$

(b) Ist hingegen $U_1 \neq U_2$, so ist $U_1+U_2 = \mathbb{R}^2$ und $U_1 \cap U_2 = \{0\}$, und damit

$$\dim(U_1+U_2) + \dim(U_1 \cap U_2) = 2+0 = 1+1 = \dim U_1 + \dim U_2.$$

Aufgabe 15.33. Es seien V ein K-Vektorraum und $U_1, U_2 \leq V$ mit $\dim V = 6$, $\dim U_1 = 5$ und $\dim U_2 = 3$.

Welche Dimension kann $U_1 \cap U_2$ haben? Gib für jede solche Möglichkeit ein konkretes Beispiel für U_1, U_2 und V an.

Aufgabe 15.34. Wir betrachten die Unterräume

$$U_1 = \operatorname{Lin}\left(\begin{pmatrix}0\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\\0\end{pmatrix}\right) \qquad \text{und} \qquad U_2 = \operatorname{Lin}\left(\begin{pmatrix}0\\1\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\1\\0\end{pmatrix}\right).$$

von $\mathbb{R}^4$. Bestimme jeweils die Dimension und eine Basis von $U_1 + U_2$, von $U_1 \cap U_2$, von einem Komplement von U_1 und vom Quotientenraum V/U_1.

Aufgabe 15.35. Es seien $U_1, \dots, U_n$ Unterräume eines endlich erzeugten Vektorraums V. Zeige in Ergänzung zu Lemma 15.3, dass die Summe $U_1 + \dots + U_n$ *genau dann* direkt ist, wenn

$$\dim(U_1 + \dots + U_n) = \dim U_1 + \dots + \dim U_n,$$

und dass man in diesem Fall eine Basis von $U_1 + \dots + U_n$ erhält, indem man Basen von $U_1, \dots, U_n$ vereinigt.

Aufgabe 15.36 (Alternative Beweise für Dimensionsformeln)**.** Gib einen alternativen Beweis der Dimensionsformel für Morphismen (siehe Folgerung 15.26) sowie für Summen und Durchschnitte (siehe Folgerung 15.31) an, der keine Quotientenräume benötigt, indem du die folgenden Aussagen zeigst und benutzt:

(a) Es sei $f\colon V \to W$ ein Morphismus endlich-dimensionaler Vektorräume. Wir wählen eine Basis $(x_1, \dots, x_n)$ von $\operatorname{Ker} f$ und ergänzen sie zu einer Basis $(x_1, \dots, x_n, y_1, \dots, y_m)$ von V. Dann ist $(f(y_1), \dots, f(y_m))$ eine Basis von $\operatorname{Im} f$.

(b) Es seien U_1 und U_2 zwei endlich-dimensionale Unterräume eines Vektorraums V. Wir wählen eine Basis $(x_1, \dots, x_n)$ von $U_1 \cap U_2$, und ergänzen diese sowohl zu einer Basis $(x_1, \dots, x_n, y_1, \dots, y_m)$ von U_1 als auch zu einer Basis $(x_1, \dots, x_n, z_1, \dots, z_p)$ von U_2. Dann ist $(x_1, \dots, x_n, y_1, \dots, y_m, z_1, \dots, z_p)$ eine Basis von $U_1 + U_2$.

16. Lineare Abbildungen als Matrizen

Wir wollen uns jetzt genauer mit linearen Abbildungen beschäftigen. Sind V und W zwei endlich-dimensionale K-Vektorräume, so wissen wir aus Bemerkung 13.19 bereits, dass die Menge

$$\mathrm{Hom}(V,W) = \{f\colon V \to W \text{ ist Morphismus}\}$$

der linearen Abbildungen von V nach W selbst wieder ein K-Vektorraum ist. Wir wollen diesen Vektorraum nun genau beschreiben, d. h. angeben, wie *alle* Morphismen von V nach W konkret aussehen.

Die entscheidende Idee hierfür ist, dass V und W nach Satz 14.23 isomorph zu K^n bzw. K^m (mit $n = \dim V$ und $m = \dim W$) sind. Gemäß der Idee, dass isomorphe Vektorräume praktisch ununterscheidbar sind, werden wir gleich in Abschnitt 16.A also zunächst einmal den Fall von linearen Abbildungen von K^n nach K^m genau untersuchen. Danach können wir dann in Abschnitt 16.B studieren, wie wir diese Ergebnisse mit Hilfe von Isomorphismen auf lineare Abbildungen zwischen beliebigen endlich-dimensionalen Vektorräumen V und W übertragen können.

16.A Matrizen

Lineare Abbildungen von K^n nach K^m können am besten in der Sprache der Matrizen angegeben werden, die wir jetzt einführen wollen.

Definition 16.1 (Matrizen)**.** Es seien K ein Körper und $m,n \in \mathbb{N}$. Eine $m \times n$-**Matrix** mit Einträgen in K ist ein rechteckiges Zahlenschema

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \quad \text{mit } a_{i,j} \in K \text{ für } 1 \le i \le m \text{ und } 1 \le j \le n$$

mit m Zeilen und n Spalten. Analog zur Schreibweise für Zahlenfolgen bezeichnen wir eine solche Matrix auch kurz mit

$$\big(a_{i,j}\big)_{\substack{1\le i\le m\\ 1\le j\le n}} \quad \text{oder einfach} \quad (a_{i,j})_{i,j}.$$

Es ist eine Konvention, dass wir in diesem Fall hinter der Klammer immer zuerst den Zeilenindex und dann den Spaltenindex schreiben (Merkregel: „Zeile zuerst, Spalte später“). Im Fall $m = n$ nennen wir die Matrix A **quadratisch**.

Die Menge aller $m \times n$-Matrizen mit Einträgen in K wird mit $\mathrm{Mat}(m \times n, K)$ bezeichnet – auch hier steht in der Bezeichnung stets zuerst die Anzahl der Zeilen und dann die Anzahl der Spalten.

Definition 16.2 (Matrixoperationen)**.** Für $\lambda \in K$ sowie zwei Matrizen $A = (a_{i,j})_{i,j}$ und $B = (b_{i,j})_{i,j}$ in $\mathrm{Mat}(m \times n, K)$ definieren wir

(a) die Addition $A + B := (a_{i,j} + b_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$;

(b) die Skalarmultiplikation $\lambda A := (\lambda a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$;

(c) die **transponierte Matrix** $A^\mathsf{T} := (a_{j,i})_{i,j} \in \mathrm{Mat}(n \times m, K)$.

Beispiel 16.3. Die Addition und Skalarmultiplikation für Matrizen ist einfach komponentenweise definiert, es ist also z. B.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 \\ 6 & 7 & 8 \end{pmatrix} \qquad \text{und} \qquad 2 \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}$$

in $\mathrm{Mat}(2\times 3,\mathbb{R})$. Die Transposition hingegen vertauscht die Rolle von Zeilen und Spalten in der Matrix, wie z. B. in

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

Bemerkung 16.4.

(a) Offensichtlich ist $\mathrm{Mat}(m\times n,K)$ mit der Addition und Skalarmultiplikation aus Definition 16.2 ein K-Vektorraum. In der Tat unterscheidet sich dieser Raum von $K^{m\cdot n}$ ja nur dadurch, dass wir die $m\cdot n$ Einträge der Matrix nicht untereinander, sondern in einem rechteckigen Schema anordnen. Mathematisch exakt bedeutet dies, dass die Abbildung

$$\mathrm{Mat}(m\times n,K)\to K^{m\cdot n}, \qquad \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \mapsto \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{1,n} \\ \vdots \\ a_{m,1} \\ \vdots \\ a_{m,n} \end{pmatrix}$$

ein Isomorphismus ist. So ist also $\dim \mathrm{Mat}(m\times n,K) = \dim K^{m\cdot n} = m\cdot n$, und die vier Matrizen

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

bilden z. B. eine Basis von $\mathrm{Mat}(2\times 2,K)$ (nämlich die Basis, die im K^4 der Standardbasis entspricht): Jede 2×2-Matrix lässt sich auf eindeutige Art als Linearkombination

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = a_{1,1}A_1 + a_{1,2}A_2 + a_{2,1}A_3 + a_{2,2}A_4$$

dieser vier Matrizen schreiben.

(b) Der Nullvektor im Vektorraum $\mathrm{Mat}(m\times n,K)$ ist offensichtlich die Matrix, in der alle Einträge gleich 0 sind. Diese Matrix wird dementsprechend auch die **Nullmatrix** genannt und einfach als 0 geschrieben.

(c) Matrizen in $\mathrm{Mat}(m\times 1,K)$ mit nur einer Spalte haben in ihrer Schreibweise die gleiche Form wie Vektoren in K^m. In der Tat werden wir $m\times 1$-Matrizen im Folgenden in der Regel mit Vektoren in K^m identifizieren.

(d) Offensichtlich ist die Transpositionsabbildung

$$\mathrm{Mat}(m\times n,K)\to \mathrm{Mat}(n\times m,K),\ \ A\mapsto A^{\mathsf{T}}$$

ein Isomorphismus: Es gilt $(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}}$ und $(\lambda A)^{\mathsf{T}} = \lambda A^{\mathsf{T}}$ für alle $\lambda\in K$ und $A,B\in \mathrm{Mat}(m\times n,K)$, und die Umkehrabbildung ist durch erneute Transposition gegeben.

34

Bisher gibt es bis auf die Art der Anordnung der Zahlen keinen nennenswerten Unterschied zwischen den Matrizen in $\mathrm{Mat}(m\times n,K)$ und den Vektoren in K^{mn}. Es gibt jedoch eine sehr wichtige weitere Operation, die auf Matrizen, jedoch nicht auf Vektoren in K^{mn} definiert ist, nämlich die sogenannte *Matrixmultiplikation*:

Definition 16.5 (Matrixmultiplikation). Für $m,n,p\in\mathbb{N}$ seien $A=(a_{i,j})_{i,j}\in \mathrm{Mat}(m\times n,K)$ und $B=(b_{j,k})_{j,k}\in \mathrm{Mat}(n\times p,K)$, d. h. die Matrix B habe so viele Zeilen wie A Spalten. Dann definieren wir das Matrixprodukt AB als

$$AB := A\cdot B := \Big(\sum_{j=1}^{n} a_{i,j}b_{j,k}\Big)_{i,k} \in \mathrm{Mat}(m\times p,K)$$

(merke: Es wird über die „mittleren Indizes“ summiert, also über den Spaltenindex der ersten und den Zeilenindex der zweiten Matrix). Das Produkt AB hat also so viele Zeilen wie die erste Matrix und so viele Spalten wie die zweite.

Beispiel 16.6.

(a) Es ist z. B.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1\cdot 5+2\cdot 7 & 1\cdot 6+2\cdot 8 \\ 3\cdot 5+4\cdot 7 & 3\cdot 6+4\cdot 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

(hier ist $m = n = p = 2$). Im Gegensatz dazu ist das Matrixprodukt

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

nicht definiert, weil die erste Matrix nur eine Spalte, die zweite aber zwei Zeilen hat. Der Einfachheit halber werden wir in Zukunft bei einem Matrixprodukt AB stets voraussetzen, dass die zweite Matrix so viele Zeilen hat wie die erste Spalten, und dies nicht jedesmal wieder erwähnen.

(b) Ist $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$ und $x \in K^n$ mit Komponenten $x_1, \ldots, x_n$, so können wir x gemäß Bemerkung 16.4 (c) als Matrix mit nur einer Spalte auffassen und erhalten das Matrixprodukt

$$Ax = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n \end{pmatrix} \quad \in \mathrm{Mat}(m \times 1, K) = K^m.$$

(c) Ein einfacher, aber oft vorkommender und daher wichtiger Spezialfall von (b) ist, wenn $x = e_j$ für $j = 1, \ldots, n$ der j-te Einheitsvektor ist: In diesem Fall ist Ae_j gerade die j-te Spalte von A.

Wie üblich nach dem Einführen einer neuen Struktur wollen wir auch hier zunächst einmal die grundlegenden Eigenschaften der Matrixmultiplikation angeben bzw. beweisen.

Lemma 16.7 (Eigenschaften der Matrixmultiplikation)**.** *Für alle Matrizen A, B, C passender Größe (d. h. so dass die betrachteten Summen und Produkte definiert sind) sowie $\lambda \in K$ gilt:*

(a) *(Distributivität)* $(A+B)C = AC + BC$ *und* $A(B+C) = AB + AC$.

(b) *(Verträglichkeit mit der Skalarmultiplikation)* $(\lambda A)B = A(\lambda B) = \lambda(AB)$.

(c) *(Assoziativität)* $(AB)C = A(BC)$.

(d) *(Verträglichkeit mit der Transposition)* $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$.

Das Matrixprodukt ist jedoch im Allgemeinen nicht kommutativ (aufgrund der Größenbedingung ist das Produkt AB ja noch nicht einmal genau dann definiert, wenn BA es ist).

Beweis. Der Beweis ergibt sich durch einfaches Nachrechnen. Wir zeigen exemplarisch den zweiten Teil von (a): Für $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$ und $B = (b_{j,k})_{j,k}, C = (c_{j,k})_{j,k} \in \mathrm{Mat}(n \times p, K)$ gilt

$$\begin{aligned} A(B+C) &= \Big(\sum_{j=1}^n a_{i,j}(b_{j,k} + c_{j,k}) \Big)_{i,k} \\ &= \Big(\sum_{j=1}^n a_{i,j} b_{j,k} + \sum_{j=1}^n a_{i,j} c_{j,k} \Big)_{i,k} \\ &= \Big(\sum_{j=1}^n a_{i,j} b_{j,k} \Big)_{i,k} + \Big(\sum_{j=1}^n a_{i,j} c_{j,k} \Big)_{i,k} \\ &= AB + AC. \end{aligned}$$

□

Aufgabe 16.8. Für ein gegebenes $a \in \mathbb{R}\backslash\{0\}$ sei

$$A = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} \quad \in \operatorname{Mat}(2 \times 2, \mathbb{R}).$$

(a) Berechne A^n für alle $n \in \mathbb{N}_{>0}$.

(b) Bestimme eine Basis des Quotientenraums $\operatorname{Mat}(2 \times 2, \mathbb{R}) / \operatorname{Lin}((A^n)_{n \in \mathbb{N}_{>0}})$.

Aufgabe 16.9 (Blockmatrixmultiplikation). Es seien $A \in \operatorname{Mat}(m \times n, K)$ und $B \in \operatorname{Mat}(n \times p, K)$ zwei Matrizen, die in „Blockform“

$$A = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right) \quad \text{bzw.} \quad B = \left(\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \end{array}\right)$$

mit $A_1 \in \operatorname{Mat}(m_1 \times n_1, K)$ und $B_1 \in \operatorname{Mat}(n_1 \times p_1, K)$ gegeben sind. Zeige, dass das Matrixprodukt AB dann ebenfalls in Blockform

$$AB = \left(\begin{array}{c|c} A_1B_1 + A_2B_3 & A_1B_2 + A_2B_4 \\ \hline A_3B_1 + A_4B_3 & A_3B_2 + A_4B_4 \end{array}\right)$$

berechnet werden kann (wobei die Blöcke formal genauso multipliziert und addiert werden, als würde man das Produkt zweier Matrizen der Größe 2×2 berechnen). Analog funktioniert diese Rechenregel auch für eine Aufteilung in noch mehr Blöcke.

Mit Hilfe der Matrixmultiplikation können wir jetzt sehr einfach lineare Abbildungen von K^n nach K^m konstruieren.

Konstruktion 16.10. Es sei $A \in \operatorname{Mat}(m \times n, K)$. Dann ist

$$f\colon K^n \to K^m,\ x \mapsto Ax$$

eine lineare Abbildung, denn für alle $x, y \in K^n$ und $\lambda \in K$ gilt nach Lemma 16.7

$$f(x+y) = A(x+y) = Ax + Ay = f(x) + f(y) \quad \text{sowie} \quad f(\lambda x) = A(\lambda x) = \lambda Ax = \lambda f(x).$$

In Koordinaten ist diese Funktionsvorschrift also durch den Ausdruck in Beispiel 16.6 (b) gegeben.

In der Tat wollen wir jetzt sehen, dass bereits *alle* linearen Abbildungen von K^n nach K^m diese Form haben, und Konstruktion 16.10 damit eine Bijektion zwischen dem Raum aller Matrizen und dem Raum aller linearen Abbildungen liefert.

Satz und Definition 16.11 (Lineare Abbildungen $K^n \to K^m$ und Matrizen)**.** *Es seien $n, m \in \mathbb{N}$.*

(a) *Zu jeder Matrix $A \in \operatorname{Mat}(m \times n, K)$ gibt es genau einen Morphismus $f \in \operatorname{Hom}(K^n, K^m)$ mit*

$$f(x) = Ax \quad \textit{für alle } x \in K^n.$$

Wir bezeichnen ihn mit f_A.

(b) *Zu jedem Morphismus $f \in \operatorname{Hom}(K^n, K^m)$ gibt es genau eine Matrix $A \in \operatorname{Mat}(m \times n, K)$ mit*

$$f(x) = Ax \quad \textit{für alle } x \in K^n,$$

*nämlich $A = (\, f(e_1) \mid \cdots \mid f(e_n)\,)$. Wir nennen sie die **Abbildungsmatrix** von f und bezeichnen sie mit A_f.*

Dies liefert einen Isomorphismus $\operatorname{Mat}(m \times n, K) \to \operatorname{Hom}(K^n, K^m)$, $A \mapsto f_A$ mit Umkehrabbildung $f \mapsto A_f$. Insbesondere ist also $\dim \operatorname{Hom}(K^n, K^m) = mn$.

Beweis.

(a) Die geforderte Bedingung definiert offensichtlich eine eindeutige Abbildung $f\colon K^n \to K^m$, die nach Konstruktion 16.10 auch linear ist.

(b) Die geforderte Bedingung legt nach Beispiel 16.6 (c) für alle $j = 1, \dots, n$ die j-te Spalte von A fest zu $Ae_j = f(e_j)$. Damit ist A eindeutig bestimmt als $(\, f(e_1) \mid \cdots \mid f(e_n)\,)$. Da f linear ist, folgt aus dieser Beziehung für die Einheitsvektoren weiterhin auch für alle $x \in K^n$ mit Koordinaten $x_1, \dots, x_n$

$$f(x) = f(x_1 e_1 + \cdots + x_n e_n) = x_1 f(e_1) + \cdots + x_n f(e_n) = x_1 A e_1 + \cdots + x_n A e_n = Ax.$$

Da die geforderte Bedingung in (a) und (b) dieselbe ist, ist außerdem klar, dass die Abbildungen $A \mapsto f_A$ und $f \mapsto A_f$ invers zueinander sind. Weiterhin ist die Abbildung $A \mapsto f_A$ linear, denn für alle $A, B \in \operatorname{Mat}(m \times n, K)$ und $\lambda \in K$ ist

$$f_{A+B}(x) = (A+B)x = Ax + Bx = f_A(x) + f_B(x) \quad \text{und} \quad f_{\lambda A}(x) = (\lambda A)x = \lambda Ax = \lambda f_A(x),$$

d. h. es ist $f_{A+B} = f_A + f_B$ und $f_{\lambda A} = \lambda f_A$. Damit ist $\operatorname{Mat}(m \times n, K) \to \operatorname{Hom}(K^n, K^m)$, $A \mapsto f_A$ ein Isomorphismus. □

Lemma 16.12. *Unter dem Isomorphismus aus Satz 16.11 entspricht die Matrixmultiplikation der Verkettung von Morphismen, d. h. für alle $A \in \operatorname{Mat}(m \times n, K)$ und $B \in \operatorname{Mat}(n \times p, K)$ gilt*

$$f_{AB} = f_A \circ f_B \colon K^p \to K^m.$$

Beweis. Für alle $x \in K^p$ gilt

$$f_{AB}(x) = (AB)x = A(Bx) = Af_B(x) = f_A(f_B(x)) = (f_A \circ f_B)(x),$$

und damit ist $f_{AB} = f_A \circ f_B$. □

Bemerkung 16.13. Genau wie in Beispiel 13.31 bedeutet auch dieser Isomorphismus anschaulich formuliert wieder, dass Matrizen in $\operatorname{Mat}(m \times n, K)$ und lineare Abbildungen von K^n nach K^m „im Prinzip dasselbe“ sind: Man kann jederzeit zwischen Matrizen und linearen Abbildungen hin- und herwechseln ohne etwas am Ergebnis zu ändern; dabei entspricht eine Matrix A der linearen Abbildung $f_A \colon x \mapsto Ax$, und eine lineare Abbildung $f \colon K^n \to K^m$ der Matrix A_f, die man erhält, wenn man in die Spalten die Bilder der Einheitsvektoren unter f schreibt.

Beispiel 16.14. Zur Matrix

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \in \operatorname{Mat}(2 \times 2, \mathbb{R})$$

ist die zugehörige lineare Abbildung nach Satz 16.11

$$f_A \colon \mathbb{R}^2 \to \mathbb{R}^2, \ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_2 \end{pmatrix}.$$

Ist umgekehrt $f = f_A$ diese Abbildung, so erhält man daraus die zugehörige Abbildungsmatrix wieder durch

$$A_f = (\, f(e_1) \mid f(e_2)\,) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$$

zurück.

Da Matrizen linearen Abbildungen entsprechen, können wir nun auch ursprünglich für lineare Abbildungen definierte Konzepte wie Bilder, Kerne und Isomorphismen in die Sprechweise der Matrizen übertragen. Dies wollen wir im restlichen Teil dieses Abschnitts tun.

Definition 16.15 (Bild und Kern einer Matrix)**.** Es sei $A \in \operatorname{Mat}(m \times n, K)$ eine Matrix mit zugehöriger linearer Abbildung $f_A \colon K^n \to K^m$, $x \mapsto Ax$.

(a) Das **Bild** von A ist definiert als das Bild

$$\operatorname{Im} A := \operatorname{Im} f_A = \{Ax : x \in K^n\};$$

der zugehörigen linearen Abbildung; nach Lemma 13.21 (a) ist dies ein Unterraum von K^m.

(b) Der **Kern** von A ist definiert als der Kern

$$\operatorname{Ker} A := \operatorname{Ker} f_A = \{x \in K^n : Ax = 0\};$$

der zugehörigen linearen Abbildung; nach Lemma 13.21 (b) ist dies ein Unterraum von K^n.

Bemerkung 16.16.

(a) (Dimensionsformel für Matrizen) Die Dimensionsformel für Morphismen aus Folgerung 15.26 überträgt sich natürlich sofort auf Matrizen: Für alle $A \in \mathrm{Mat}(m \times n, K)$ (die ja einer Abbildung von K^n nach K^m entsprechen) gilt

$$\dim \mathrm{Im} A + \dim \mathrm{Ker} A = n.$$

(b) Ist $f\colon K^n \to K^m$ ein Morphismus mit zugehöriger Matrix $A = A_f \in \mathrm{Mat}(m \times n, K)$, so ist der Morphismus $f\colon K^n \to \mathrm{Im} f$, der daraus durch Einschränkung des Zielraums entsteht, natürlich surjektiv und bildet damit nach Satz 14.22 (a) die Basis $(e_1, \dots, e_n)$ von K^n auf ein Erzeugendensystem $(f(e_1), \dots, f(e_n))$ von $\mathrm{Im} f = \mathrm{Im} A$ ab. Da $f(e_1), \dots, f(e_n)$ nach Satz 16.11 aber genau die Spalten von A sind, erhalten wir die einfache Aussage:

$$\boxed{\text{Das Bild einer Matrix ist der von ihren Spalten erzeugte Unterraum.}}$$

So ist z. B. das Bild der Matrix A aus Beispiel 16.14

$$\mathrm{Im} A = \mathrm{Lin}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}\right) = \mathbb{R}^2.$$

Der Kern von A hingegen ist nach Definition

$$\mathrm{Ker} A = \{x \in \mathbb{R}^2 : Ax = 0\} = \left\{\begin{pmatrix}x_1\\x_2\end{pmatrix} : \begin{pmatrix}x_1 + 2x_2\\x_2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}\right\} = \left\{\begin{pmatrix}0\\0\end{pmatrix}\right\}$$

(insbesondere ist hier die Dimensionsformel aus (a) also als $2 + 0 = 2$ erfüllt).

Um den Begriff der Isomorphismen zu übertragen, müssen wir uns offensichtlich auf quadratische Matrizen beschränken, da isomorphe Vektorräume nach Satz 14.22 (c) die gleiche Dimension haben, ein Isomorphismus $K^n \to K^m$ also nur für $m = n$ existieren kann. Zunächst müssen wir dazu die identische Abbildung in die Sprache der Matrizen übertragen.

Konstruktion 16.17 (Einheitsmatrix)**.** Für $n \in \mathbb{N}$ definieren wir die **Einheitsmatrix** der Größe $n \times n$ als

$$E_n := (e_1 \mid \cdots \mid e_n) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathrm{Mat}(n \times n, K).$$

Wir schreiben sie oft auch einfach als E, wenn ihre Größe aus dem Zusammenhang klar ist. In der Literatur ist auch die Bezeichnung I_n oder I üblich.

Da in den Spalten von E_n gerade die Einheitsvektoren stehen, entspricht sie gemäß Satz 16.11 der Identität auf K^n. Die Gleichungen $\mathrm{id}_{K^m} \circ f = f \circ \mathrm{id}_{K^n} = f$ für eine beliebige (lineare) Abbildung $f\colon K^n \to K^m$ liefern also sofort

$$E_m A = A E_n = A \quad \text{für alle } A \in \mathrm{Mat}(m \times n, K),$$

weil die Verkettung linearer Abbildungen nach Lemma 16.12 der Matrixmultiplikation entspricht (natürlich könnte man dies auch direkt mit Definition 16.5 nachrechnen). Die Einheitsmatrizen der passenden Größe sind in diesem Sinne also neutrale Elemente für die Matrixmultiplikation.

Damit können wir nun das Konzept von Isomorphismen, also invertierbaren Morphismen, ganz analog zur Definition inverser Elemente in Definition 3.1 (c) in die Sprache der Matrizen übersetzen:

Definition 16.18 (Invertierbare Matrizen)**.** Eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ heißt **invertierbar**, wenn es eine Matrix $A' \in \mathrm{Mat}(n \times n, K)$ gibt mit $A'A = AA' = E_n$.

Die Menge aller invertierbaren Matrizen in $\mathrm{Mat}(n \times n, K)$ wird mit $\mathrm{GL}(n, K)$ bezeichnet (der Name kommt von der englischen Bezeichnung „general linear group“ und rührt daher, dass $\mathrm{GL}(n, K)$ in der Tat eine Gruppe ist, wie wir gleich in Lemma 16.20 zeigen werden).

Lemma und Definition 16.19 (Inverse Matrizen)**.** *Es sei* $A \in \mathrm{Mat}(n \times n, K)$.

(a) *Die Matrix A ist genau dann invertierbar, wenn ihre zugehörige lineare Abbildung f_A ein Isomorphismus ist. In diesem Fall ist die Matrix A' in Definition 16.18 eindeutig bestimmt als die Matrix zur Umkehrabbildung f_A^{-1}. Wir nennen sie die zu A* ***inverse Matrix*** *und schreiben sie als A^{-1}.*

(b) *Für die Invertierbarkeit von A genügt es bereits, in Definition 16.18 eine der beiden Gleichungen $A'A = E_n$ und $AA' = E_n$ vorauszusetzen.*

Beweis.

(a) Die Abbildung f_A ist genau dann ein Isomorphismus, wenn es einen Umkehrmorphismus g mit $g \circ f_A = f_A \circ g = \mathrm{id}_{K^n}$ gibt (der dann natürlich nach Lemma 2.20 (c) eindeutig bestimmt ist). In Matrixsprechweise übersetzt bedeutet dies nach Lemma 16.12 und Konstruktion 16.17 aber gerade $A'A = AA' = E_n$ (wobei A' die Abbildungsmatrix von g ist), also genau die Definition 16.18 der Invertierbarkeit.

(b) Es sei $A'A = E_n$ oder $AA' = E_n$. Nach Lemma 16.12 und Konstruktion 16.17 gilt dann für die zugehörigen linearen Abbildungen $f_{A'} \circ f_A = \mathrm{id}_{K^n}$ oder $f_A \circ f_{A'} = \mathrm{id}_{K^n}$. Dies bedeutet nach Lemma 2.20 aber genau, dass f_A injektiv oder surjektiv, und damit nach Folgerung 15.27 bereits ein Isomorphismus ist. Die Behauptung folgt damit aus Teil (a). □

Wir werden später in Algorithmus 17.18 noch sehen, wie man von einer gegebenen quadratischen Matrix A bestimmen kann, ob sie invertierbar ist, und wie man in diesem Fall die inverse Matrix A^{-1} konkret berechnen kann. Für den Moment begnügen wir uns damit, die folgenden einfachen Eigenschaften invertierbarer Matrizen zu zeigen:

Lemma 16.20 (Eigenschaften invertierbarer Matrizen)**.** *Es seien $A, B \in \mathrm{GL}(n, K)$ invertierbare Matrizen. Dann gilt:*

(a) *AB ist ebenfalls invertierbar, und es ist $(AB)^{-1} = B^{-1}A^{-1}$.*

(b) *A^{-1} ist ebenfalls invertierbar, und es ist $(A^{-1})^{-1} = A$.*

(c) *A^T ist ebenfalls invertierbar, und es ist $(A^\mathsf{T})^{-1} = (A^{-1})^\mathsf{T}$.*

Insbesondere ist $\mathrm{GL}(n, K)$ eine Gruppe bezüglich der Matrixmultiplikation (mit neutralem Element E und zu A inversem Element A^{-1}).

Beweis. Alle Aussagen ergeben sich durch einfaches Nachrechnen mit Hilfe von Lemma 16.19 (b):

(a) Es gilt $(B^{-1}A^{-1})(AB) = B^{-1}A^{-1}AB = B^{-1}EB = B^{-1}B = E$. Also ist AB invertierbar mit inverser Matrix $B^{-1}A^{-1}$.

(b) Die Gleichung $A^{-1}A = E$ besagt natürlich auch gerade, dass A die inverse Matrix zu A^{-1} ist.

(c) Nach Lemma 16.7 (d) ist $(A^{-1})^\mathsf{T}A^\mathsf{T} = (AA^{-1})^\mathsf{T} = E^\mathsf{T} = E$. Also ist $(A^{-1})^\mathsf{T}$ die inverse Matrix zu A^T. □ 35

Beispiel 16.21. Wir betrachten noch einmal die Matrix A mit zugehöriger linearer Abbildung f_A aus Beispiel 16.14, also

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad f_A \colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_2 \end{pmatrix}.$$

Offensichtlich ist diese Abbildung bijektiv: Die Umkehrung ist gegeben durch $x_1 = y_1 - 2y_2$ und $x_2 = y_2$, also

$$f_A^{-1} \colon \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \begin{pmatrix} y_1 - 2y_2 \\ y_2 \end{pmatrix} \quad \text{mit zugehöriger Matrix} \quad A^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

Zur Kontrolle kann man natürlich auch das Matrixprodukt

$$A^{-1}A = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E$$

(und genauso $AA^{-1} = E$) sofort mit Definition 16.5 nachrechnen.

16.B Lineare Abbildungen zwischen endlich erzeugten Vektorräumen

Nachdem wir mit Satz 16.11 jetzt lineare Abbildungen zwischen K^n und K^m durch Matrizen beschreiben können, wollen wir nun den Fall von Morphismen zwischen beliebigen (endlich-dimensionalen) Vektorräumen darauf zurückführen. Dazu erinnern wir uns daran, dass jeder n-dimensionale Vektorraum V isomorph zu K^n ist, indem wir einen Vektor auf seine Koordinaten bezüglich einer gewählten Basis abbilden:

Definition 16.22 (Koordinatenabbildungen und Koordinatenvektoren)**.** Es sei $B = (x_1, \dots, x_n)$ eine Basis eines n-dimensionalen Vektorraums V, so dass sich also jeder Vektor $x \in V$ eindeutig als Linearkombination $x = \lambda_1 x_1 + \dots + \lambda_n x_n$ für gewisse $\lambda_1, \dots, \lambda_n \in K$ schreiben lässt. Dann heißt die Abbildung

$$\Phi_B \colon V \to K^n,\ \lambda_1 x_1 + \dots + \lambda_n x_n \mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

(die nach Satz 14.23 ein Isomorphismus ist), die **Koordinatenabbildung** bezüglich B; dementsprechend heißt $\Phi_B(x) \in K^n$ für ein $x \in V$ der **Koordinatenvektor** von x bezüglich B.

Der Koordinatenvektor enthält also exakt die gleichen Informationen wie der ursprüngliche Vektor selbst. Wollen wir nun eine lineare Abbildung $f \colon V \to W$, $x \mapsto f(x)$ zwischen zwei endlich-dimensionalen Vektorräumen beschreiben, so können wir daher genauso gut Basen von V und W wählen und die zugehörige lineare Abbildung betrachten, die den Koordinatenvektor von x auf den von $f(x)$ abbildet. Dies ist dann ein Morphismus von K^n nach K^m mit $n = \dim V$ und $m = \dim W$, den wir also genau wie bisher durch Multiplikation mit einer Abbildungsmatrix erhalten können.

Formal bedeutet dies einfach, dass wir Satz 16.11 mehr oder weniger abschreiben können, wenn wir in den Matrixprodukten nur alle Vektoren (sowohl im Start- als auch im Zielraum) durch ihre Koordinatenvektoren ersetzen.

Satz und Definition 16.23 (Lineare Abbildungen $V \to W$ und Matrizen)**.** *Es seien V und W zwei endlich-dimensionale Vektorräume mit $n := \dim V$ und $m := \dim W$. Ferner seien $B = (x_1, \dots, x_n)$ und $C = (y_1, \dots, y_m)$ fest gewählte Basen von V bzw. W.*

(a) *Zu jeder Matrix $A \in \mathrm{Mat}(m \times n, K)$ gibt es genau einen Morphismus $f \in \mathrm{Hom}(V, W)$ mit*

$$\Phi_C(f(x)) = A \cdot \Phi_B(x) \quad \textit{für alle } x \in V.$$

Wir bezeichnen ihn mit $f_A^{B,C}$.

(b) *Zu jedem Morphismus $f \in \mathrm{Hom}(V, W)$ gibt es genau eine Matrix $A \in \mathrm{Mat}(m \times n, K)$ mit*

$$\Phi_C(f(x)) = A \cdot \Phi_B(x) \quad \textit{für alle } x \in V,$$

*nämlich $A = (\,\Phi_C(f(x_1)) \mid \cdots \mid \Phi_C(f(x_n))\,)$. Wir nennen sie die **Abbildungsmatrix** von f bezüglich B und C und bezeichnen sie mit $A_f^{B,C}$.*

Dies liefert einen Isomorphismus $\mathrm{Mat}(m \times n, K) \to \mathrm{Hom}(V, W)$, $A \mapsto f_A^{B,C}$ mit Umkehrabbildung $f \mapsto A_f^{B,C}$. Insbesondere ist also $\dim \mathrm{Hom}(V, W) = \dim V \cdot \dim W$.

Beweis. Man könnte diesen Satz zeigen, indem man ihn mit Hilfe der Koordinatenabbildungen auf Satz 16.11 zurückführt. Da der Beweis so kurz ist, ist es aber vermutlich übersichtlicher, ihn einfach nochmal neu aufzuschreiben:

(a) Die geforderte Bedingung definiert f eindeutig zu $f(x) = \Phi_C^{-1}(A \cdot \Phi_B(x))$ für alle $x \in V$, was (als Verkettung linearer Abbildungen) offensichtlich ein Morphismus ist.

(b) Da der Koordinatenvektor von x_j bezüglich B gerade der j-te Einheitsvektor ist, bestimmt die geforderte Bedingung nach Beispiel 16.6 (c) für alle $j = 1, \dots, n$ die j-te Spalte von A zu $Ae_j = A\Phi_B(x_j) = \Phi_C(f(x_j))$. Damit ist A eindeutig festgelegt. Mit der Linearität von f und den Koordinatenabbildungen folgt aus dieser Beziehung weiterhin auch für alle $x \in V$ mit Koordinaten $\lambda_1, \dots, \lambda_n$ bezüglich B

$$\begin{aligned}\Phi_C(f(x)) &= \Phi_C(f(\lambda_1 x_1 + \dots + \lambda_n x_n)) = \lambda_1 \Phi_C(f(x_1)) + \dots + \lambda_n \Phi_C(f(x_n)) \\ &= \lambda_1 A e_1 + \dots + \lambda_n A e_n = A \cdot \Phi_B(x).\end{aligned}$$

Da die Bedingung in (a) und (b) dieselbe ist, ist natürlich klar, dass die Abbildungen $A \mapsto f_A^{B,C}$ und $f \mapsto A_f^{B,C}$ invers zueinander sind. Die Linearität von $A \mapsto f_A^{B,C}$ überprüft man unmittelbar genau wie im Beweis von Satz 16.11. □

Bemerkung 16.24.

(a) Ist bereits $V = K^n$ und $W = K^m$, und sind B und C die Standardbasen dieser Vektorräume, so sind die Koordinatenabbildungen Φ_B und Φ_C die Identität auf V bzw. W. Satz 16.23 stimmt dann also genau mit Satz 16.11 überein.

(b) Man kann sich die Satz 16.23 zugrunde liegende Idee auch an dem rechts dargestellten Diagramm verdeutlichen: In der oberen Zeile ist eine lineare Abbildung $f\colon V \to W$ dargestellt. Mit Hilfe der vertikal dargestellten Koordinatenabbildungen (die ja Isomorphismen sind, was durch die Schlangen am Beginn der Pfeile angedeutet werden soll) können wir daraus eine Abbildung $g = \Phi_C \circ f \circ \Phi_B^{-1}$ von K^n nach K^m konstruieren, indem wir im Diagramm „den Umweg über f nehmen“.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\Phi_B", "\wr"'] & W \arrow[d, "\Phi_C"] \\ K^n \arrow[r, "g"'] & K^m \end{tikzcd}$$

Dieser neue Morphismus $g\colon K^n \to K^m$ ist nun genau die am Anfang dieses Abschnitts beschriebene Abbildung, die für alle $x \in V$ den Koordinatenvektor von x auf den von $f(x)$ abbildet. Wir können sie mit einer gewöhnlichen Abbildungsmatrix A im Sinne von Satz 16.11 beschreiben, es ist also

$$g(v) = Av, \quad \text{d. h.} \quad \Phi_C(f(\Phi_B^{-1}(v))) = Av \quad \text{für alle } v \in K^n.$$

Mit der Substitution $v = \Phi_B(x)$ ist dies nun aber äquivalent zu

$$\Phi_C(f(x)) = A \cdot \Phi_B(x) \quad \text{für alle } x \in V,$$

also genau zu unserer Bedingung aus Satz 16.23.

Beispiel 16.25. Es seien $V = \mathbb{R}^2$ mit der Standardbasis $B = (e_1, e_2)$ und

$$W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\} \leq \mathbb{R}^3 \quad \text{mit} \quad C = \left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right).$$

Man überprüft sofort, dass C eine Basis von W ist und die Abbildung

$$f\colon V \to W, \ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 - x_2 \\ x_1 + x_2 \\ -2x_1 \end{pmatrix}$$

linear ist sowie wirklich V nach W abbildet (denn die Summe der Koordinaten der Bildvektoren ist ja immer gleich 0). Wir wollen nun die Abbildungsmatrix $A_f^{B,C}$ bestimmen. Dazu müssen wir nach Definition 16.11 einfach nur die Basisvektoren e_1 und e_2 von B mit f abbilden und die Ergebnisse

als Linearkombinationen der Basisvektoren von C ausdrücken:

$$f\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\1\\-2\end{pmatrix} = -1\cdot\begin{pmatrix}1\\-1\\0\end{pmatrix} + 2\cdot\begin{pmatrix}1\\0\\-1\end{pmatrix}$$

$$\text{und}\quad f\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-1\\1\\0\end{pmatrix} = -1\cdot\begin{pmatrix}1\\-1\\0\end{pmatrix} + 0\cdot\begin{pmatrix}1\\0\\-1\end{pmatrix}.$$

Die dabei benötigten Vorfaktoren sind nun also die Koordinatenvektoren von $f(e_1)$ und $f(e_2)$ bezüglich C; wir schreiben sie als Spaltenvektoren in die Abbildungsmatrix und erhalten

$$A_f^{B,C} = \begin{pmatrix}-1 & -1\\ 2 & 0\end{pmatrix}.$$

Beachte, dass diese Matrix wegen $\dim V = \dim W = 2$ die Größe 2×2 hat (obwohl W ein Unterraum von $\mathbb{R}^3$ ist). Natürlich hätten wir f hier auch als Morphismus von $\mathbb{R}^2$ nach $\mathbb{R}^3$ auffassen und ihm damit auch ohne Basiswahl eine Abbildungsmatrix (der Größe 3×2) wie in Definition 16.11 zuordnen können. Deutlicher wird die Notwendigkeit der Verwendung von Basen im allgemeinen Fall im folgenden Beispiel.

Beispiel 16.26. Es sei V der Vektorraum aller reellen Polynome vom Grad höchstens 2 mit der Basis $B = (1,x,x^2)$, und W der Vektorraum aller Polynome vom Grad höchstens 1 mit der Basis $C = (1,x)$ (siehe Beispiel 14.7 (d)). Wir betrachten wie in Beispiel 13.17 (d) die lineare Abbildung $f\colon V \to W,\ \varphi \mapsto \varphi'$, die einem Polynom seine Ableitung zuordnet.

(a) Um die Abbildungsmatrix $A_f^{B,C}$ zu bestimmen, müssen wir nach Definition 16.23 wieder die Basisvektoren von B abbilden und als Linearkombinationen der Basiselemente von C schreiben:

$$\begin{aligned} f(1) &= 1' &&= 0 &&= 0\cdot 1 + 0\cdot x,\\ f(x) &= x' &&= 1 &&= 1\cdot 1 + 0\cdot x,\\ f(x^2) &= (x^2)' &&= 2x &&= 0\cdot 1 + 2\cdot x. \end{aligned}$$

Die Koeffizienten dieser Linearkombinationen bilden also die Koordinatenvektoren von $f(1)$, $f(x)$ und $f(x^2)$; wir schreiben sie als Spalten in eine Matrix und erhalten

$$A_f^{B,C} = \begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 2\end{pmatrix}.$$

(b) Auch die umgekehrte Richtung, aus der Abbildungsmatrix die Abbildung zu rekonstruieren, ist nicht weiter schwierig, wenn man sich daran erinnert, dass die Matrix immer nur Koordinatenvektoren sieht. Angenommen, wir wollen $f(\varphi)$ für $\varphi = 2x^2+3x+4$, also letztlich die Ableitung φ', nur aus der Kenntnis der Abbildungsmatrix $A_f^{B,C}$ bestimmen. Dann brauchen wir zunächst den Koordinatenvektor $\Phi_B(\varphi)$ und können diesen dann an die Abbildungsmatrix multiplizieren: Wegen $\varphi = 4\cdot 1 + 3\cdot x + 2\cdot x^2$ ist

$$\Phi_B(\varphi) = \begin{pmatrix}4\\3\\2\end{pmatrix},\quad \text{und damit}\quad A_f^{B,C}\cdot\Phi_B(\varphi) = \begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 2\end{pmatrix}\cdot\begin{pmatrix}4\\3\\2\end{pmatrix} = \begin{pmatrix}3\\4\end{pmatrix}.$$

Nach Satz 16.23 ist dies nun der Koordinatenvektor $\Phi_C(f(\varphi))$ des Bildes $f(\varphi)$. Damit ist $f(\varphi) = 3\cdot 1 + 4\cdot x = 4x+3$ (was in der Tat die Ableitung von φ ist).

Eine einfache Folgerung aus Satz 16.23 ist, dass sich lineare Abbildungen stets auf einer Basis des Startraums beliebig vorgeben lassen und dann eindeutig bestimmt sind:

Folgerung 16.27. *Es seien V und W endlich-dimensionale K-Vektorräume, $B = (x_1,\dots,x_n)$ eine Basis von V und $(v_1,\dots,v_n)$ eine beliebige Familie (mit gleich vielen Elementen) in W.*

Dann gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $f(x_i) = v_i$ für alle $i = 1,\dots,n$.

Beweis. Wähle eine Basis C von W, und betrachte die Abbildungsmatrix $A_f^{B,C}$ der gesuchten Morphismen. Ihre i-te Spalte für $i = 1, \ldots, n$ ist nach Beispiel 16.6 (c) und Satz 16.23 gerade

$$A_f^{B,C} e_i = A_f^{B,C} \cdot \Phi_B(x_i) = \Phi_C(f(x_i)).$$

Die Bedingungen $f(x_i) = v_i$, also $\Phi_C(f(x_i)) = \Phi_C(v_i)$, sind daher äquivalent dazu, dass die Abbildungsmatrix gleich $A_f^{B,C} = (\,\Phi_C(v_1) \mid \cdots \mid \Phi_C(v_n)\,)$ ist, und liefern damit genau einen solchen Morphismus. □

Nach Konstruktion hängen die gerade eingeführten Abbildungsmatrizen $A_f^{B,C}$ natürlich von der (letztlich willkürlichen) Wahl der Basen B und C im Start- bzw. Zielraum der Abbildung f ab. Wir wollen nun untersuchen, wie sich diese Abbildungsmatrizen ändern, wenn man zu anderen Basen übergeht. Dazu benötigen wir die sogenannten Basiswechselmatrizen, die letztlich ein Spezialfall von Abbildungsmatrizen sind.

Definition 16.28 (Basiswechselmatrizen)**.** Es seien $B = (x_1, \ldots, x_n)$ und $B' = (x'_1, \ldots, x'_n)$ zwei Basen eines endlich-dimensionalen K-Vektorraums V. Dann heißt die Abbildungsmatrix der Identität id_V bezüglich der Startbasis B' und Zielbasis B, nach Definition 16.23 also

$$A^{B',B} := A_{\mathrm{id}}^{B',B} = (\,\Phi_B(x'_1) \mid \cdots \mid \Phi_B(x'_n)\,) \quad \in \mathrm{Mat}(n \times n, K),$$

die **Basiswechselmatrix** von B' nach B.

Bemerkung 16.29. Es seien B und B' zwei Basen eines endlich-dimensionalen Vektorraums V.

(a) Offensichtlich ist stets $A^{B,B} = E$.

(b) Nach Satz 16.23 ist $A^{B',B}$ die eindeutig bestimmte Matrix mit $\Phi_B(x) = A^{B',B} \cdot \Phi_{B'}(x)$ für alle $x \in V$. Die Basiswechselmatrix wandelt also einfach nur einen Koordinatenvektor bezüglich B' in einen bezüglich B um – was auch den Namen erklärt.

Beispiel 16.30. Wollen wir für die beiden Basen

$$B = \left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \qquad \text{und} \qquad B' = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$$

von $\mathbb{R}^2$ die Basiswechselmatrix $A^{B',B}$ bestimmen, so müssen wir die beiden Basisvektoren von B' nach Definition 16.28 als Linearkombination der Vektoren aus B schreiben: Es ist

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= 1 \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \text{und} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= (-1) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

Die Koeffizienten dieser Linearkombinationen bilden wieder die Koordinatenvektoren bezüglich B, wir schreiben sie also in die Spalten der gesuchten Matrix

$$A^{B',B} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 16.31. *Es sei $B = (x_1, \ldots, x_n)$ eine Basis eines endlich erzeugten Vektorraums V. Dann gilt:*

(a) *Für jede weitere Basis B' von V ist $A^{B',B}$ invertierbar mit $(A^{B',B})^{-1} = A^{B,B'}$.*

(b) *Für jede invertierbare Matrix $T \in \mathrm{GL}(n,K)$ gibt es eine Basis B' von V mit $A^{B',B} = T$.*

(c) *Für jede invertierbare Matrix $T \in \mathrm{GL}(n,K)$ gibt es eine Basis B' von V mit $A^{B,B'} = T$.*

Beweis.

(a) Nach Bemerkung 16.29 (b) ist $\Phi_B(x) = A^{B',B} \cdot \Phi_{B'}(x)$ und $\Phi_{B'}(x) = A^{B,B'} \cdot \Phi_B(x)$ für alle $x \in V$. Setzen wir dies ineinander ein, erhalten wir für alle $x \in V$

$$\Phi_B(x) = A^{B',B} \cdot A^{B,B'} \cdot \Phi_B(x).$$

Wiederum nach Bemerkung 16.29 (b) ist $A^{B',B} \cdot A^{B,B'}$ also die Basiswechselmatrix von B nach B, d. h. nach Bemerkung 16.29 (a) gilt $A^{B',B} \cdot A^{B,B'} = E$ und damit $(A^{B',B})^{-1} = A^{B,B'}$.

(b) Wir setzen $x_i' = \Phi_B^{-1}(Te_i)$ für $i = 1,\dots,n$. Da T invertierbar und Φ_B^{-1} ein Isomorphismus ist, ist auch $K^n \to V$, $x \mapsto \Phi_B^{-1}(Tx)$ ein Isomorphismus, und bildet nach Satz 14.22 (c) damit die Standardbasis $(e_1,\dots,e_n)$ auf eine Basis $B' := (x_1',\dots,x_n')$ von V ab.

Für $i = 1,\dots,n$ ist die i-te Spalte der Basiswechselmatrix $A^{B',B}$ nun nach Definition

$$\Phi_B(x_i') = \Phi_B(\Phi_B^{-1}(Te_i)) = Te_i,$$

also die i-te Spalte von T. Damit ist wie gewünscht $A^{B',B} = T$.

(c) Nach (b) gibt es eine Basis B' mit $A^{B',B} = T^{-1}$, nach (a) also mit $A^{B,B'} = T$. □

Mit diesen Basiswechselmatrizen können wir nun konkret angeben, wie sich Abbildungsmatrizen bei einem Basiswechsel transformieren.

Satz 16.32 (Verhalten von Abbildungsmatrizen unter Basiswechsel)**.** *Es sei $f\colon V \to W$ ein Morphismus zwischen endlich erzeugten Vektorräumen mit gegebenen Basen B bzw. C.*

(a) *Sind B' und C' zwei weitere Basen von V bzw. W, so gilt*

$$A_f^{B',C'} = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B}.$$

(b) *Sind umgekehrt $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$ zwei invertierbare Matrizen, so gibt es Basen B' und C' von V bzw. W, so dass $A^{C,C'} = S$ und $A^{B',B} = T$, und damit*

$$A_f^{B',C'} = S \cdot A_f^{B,C} \cdot T.$$

Beweis.

(a) Nach Definition 16.23 bzw. Bemerkung 16.29 (b) gilt

$$\begin{aligned} \Phi_{C'}(y) &= A^{C,C'} \cdot \Phi_C(y) && \text{für alle } y \in W, \\ \Phi_C(f(x)) &= A_f^{B,C} \cdot \Phi_B(x) && \text{für alle } x \in V, \\ \text{und} \qquad \Phi_B(x) &= A^{B',B} \cdot \Phi_{B'}(x) && \text{für alle } x \in V. \end{aligned}$$

Setzen wir dies für $y = f(x)$ ineinander ein, so erhalten wir

$$\Phi_{C'}(f(x)) = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B} \cdot \Phi_{B'}(x) \quad \text{für alle } x \in V,$$

und damit nach Satz 16.11 wie gewünscht $A_f^{B',C'} = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B}$.

(b) Nach Lemma 16.31 (b) und (c) existieren Basen B' und C' mit $A^{C,C'} = S$ und $A^{B',B} = T$; die behauptete Formel für die Abbildungsmatrix ergibt sich dann aus (a). □

Aufgabe 16.33.

(a) Wir betrachten die Basen

$$B = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right) \quad \text{und} \quad C = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$$

von $\mathbb{R}^3$ bzw. $\mathbb{R}^2$ sowie die lineare Abbildung $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ mit

$$A_f^{B,C} = \begin{pmatrix} 2 & -1 & 0 \\ 4 & 0 & 2 \end{pmatrix}.$$

Berechne die Abbildungsmatrix A_f von f bezüglich der Standardbasen von $\mathbb{R}^3$ und $\mathbb{R}^2$.

(b) Es sei $f\colon \mathbb{R}^3 \to \mathbb{R}^3/U$, $x \mapsto \overline{x}$ mit $U = \mathrm{Lin}(e_1 - 2e_2 + e_3)$. Bestimme eine Basis B von $\mathbb{R}^3/U$ sowie die zugehörige Abbildungsmatrix $A_f^{E,B}$ für die Standardbasis E von $\mathbb{R}^3$.

Aufgabe 16.34. Es seien V der Vektorraum aller reellen Polynome vom Grad höchstens 2 mit der Basis $B = (1, x, x^2)$, und $f\colon V \to V$ die Abbildung mit $f(\varphi)(x) = \varphi(x+1)$ für alle $\varphi \in V$.

(a) Überprüfe, dass f linear ist, und berechne die Abbildungsmatrix $A_f^{B,B}$.

(b) Zeige, dass $A_f^{B,B}$ invertierbar ist und berechne die inverse Matrix $\left(A_f^{B,B}\right)^{-1}$.

(c) Zeige, dass es zu jeder Basis C' von V eine Basis C von V gibt mit

$$A_f^{C,C'} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Aufgabe 16.35. Man zeige:

(a) Sind $f\colon V \to V'$ und $g\colon V' \to V''$ zwei lineare Abbildungen zwischen endlich-dimensionalen Vektorräumen sowie B, B' und B'' Basen von V, V' bzw. V'', so gilt

$$A_{g\circ f}^{B,B''} = A_g^{B',B''} \cdot A_f^{B,B'}.$$

(b) Ist $A \in \mathrm{Mat}(m \times n, K)$ eine Matrix mit $\dim(\mathrm{Im}A) = r$, so gibt es Matrizen $A' \in \mathrm{Mat}(m \times r, K)$ und $A'' \in \mathrm{Mat}(r \times n, K)$ mit $A = A' \cdot A''$.

16.C Äquivalente Matrizen und Normalformen

Im letzten Abschnitt haben wir gesehen, dass man einen Morphismus $f\colon V \to W$ zwischen endlich erzeugten Vektorräumen erst dann durch eine Abbildungsmatrix beschreiben kann, wenn man Basen B und C von V bzw. W gewählt hat. Wir wollen nun untersuchen, in wie weit man aber zumindest durch eine geschickte Wahl dieser Basen erreichen kann, dass die resultierende Abbildungsmatrix $A_f^{B,C}$ eine möglichst einfache Form hat (also z. B. viele Nullen enthält).

Dazu erinnern wir uns daran, dass ein Basiswechsel in B oder C nach Satz 16.32 dazu führt, dass die Abbildungsmatrix von rechts bzw. links mit einer invertierbaren Matrix multipliziert wird, und dass auch umgekehrt jede Multiplikation von $A_f^{B,C}$ mit einer invertierbaren Matrix von rechts oder links einem Basiswechsel entspricht. Wir definieren daher:

Definition 16.36 (Äquivalente Matrizen)**.** Zwei Matrizen $A, A' \in \mathrm{Mat}(m \times n, K)$ heißen **äquivalent**, wenn es invertierbare Matrizen $S \in \mathrm{GL}(m, K)$ und $T \in \mathrm{GL}(n, K)$ gibt mit $A' = SAT$.

Bemerkung 16.37.

(a) Wie man leicht nachprüfen kann, ist die Äquivalenz von Matrizen in der Tat eine Äquivalenzrelation (siehe Definition 2.27).

(b) Nach Satz 16.32 sind zwei Matrizen genau dann äquivalent, wenn sie dieselbe lineare Abbildung, nur bezüglich evtl. verschiedener Basen im Start- und Zielraum beschreiben.

36

Unser Ziel ist es also, zu einer gegebenen Matrix eine äquivalente Matrix möglichst einfacher Gestalt zu finden. Da wir durch die beidseitige Multiplikation mit beliebigen invertierbaren Matrizen sehr viele Möglichkeiten haben, eine gegebene Matrix zu einer äquivalenten umzuformen, könnte man sich vielleicht sogar fragen, ob zwei Matrizen (derselben Größe) womöglich immer zueinander äquivalent sind. Dies ist jedoch nicht der Fall – es gibt eine sogenannte „Invariante“, d. h. eine Zahl, die man einer Matrix zuordnen kann, und die beim Übergang zu einer beliebigen äquivalenten Matrix immer gleich bleibt. Hat man also zwei Matrizen, bei denen diese Invariante verschieden ist, so können diese damit nicht äquivalent zueinander sein. Diese Invariante ist der sogenannte Rang einer Matrix, den wir jetzt einführen wollen.

Definition 16.38 (Rang eines Morphismus bzw. einer Matrix)**.**

(a) Ist $f\colon V \to W$ ein Morphismus zwischen endlich erzeugten Vektorräumen, so heißt die Zahl $\mathrm{rk}\, f := \dim \mathrm{Im}\, f$ der **Rang** von f (die Bezeichnung kommt vom englischen Wort „rank“).

(b) Für eine Matrix $A \in \mathrm{Mat}(m \times n, K)$ heißt die Zahl $\mathrm{rk}A := \dim \mathrm{Im}A = \mathrm{rk}\, f_A$ der **Rang** von A.

Bemerkung 16.39.

(a) Ist $f\colon V \to W$ ein Morphismus, so ist natürlich $\operatorname{rk} f = \dim \operatorname{Im} f \le \dim W$. Nach der Dimensionsformel für Morphismen aus Folgerung 15.26 gilt aber auch

$$\operatorname{rk} f = \dim \operatorname{Im} f = \dim V - \dim \operatorname{Ker} f \le \dim V.$$

Eine analoge Aussage gilt natürlich auch für Matrizen: Für $A \in \operatorname{Mat}(m \times n, K)$ ist stets sowohl $\operatorname{rk} A \le m$ als auch $\operatorname{rk} A \le n$ (denn $\operatorname{rk} A = \operatorname{rk} f_A$ mit $f_A\colon K^n \to K^m$).

(b) Für eine quadratische Matrix $A \in \operatorname{Mat}(n \times n, K)$ gilt:

$$\begin{aligned} \operatorname{rk} A = n &\Leftrightarrow \dim \operatorname{Im} f_A = n \\ &\Leftrightarrow f_A \text{ ist surjektiv} \\ &\Leftrightarrow f_A \text{ ist ein Isomorphismus} && \text{(Folgerung 15.27)} \\ &\Leftrightarrow A \in \operatorname{GL}(n, K) && \text{(Lemma 16.19 (a)).} \end{aligned}$$

(c) Es sei $A \in \operatorname{Mat}(m \times n, K)$ eine Matrix mit Spalten $x_1, \dots, x_n \in K^m$. Nach Bemerkung 16.16 (b) ist dann $\operatorname{Im} A = \operatorname{Lin}(x_1, \dots, x_n)$. Wir können aus diesen n Vektoren also eine Basis von $\operatorname{Im} f$ auswählen, die dann natürlich so viele Elemente hat, wie es linear unabhängige Vektoren unter den $x_1, \dots, x_n$ gibt. Also ist $\operatorname{rk} A = \dim \operatorname{Lin}(x_1, \dots, x_n)$ die Maximalzahl linear unabhängiger Vektoren in $(x_1, \dots, x_n)$, d. h. die Maximalzahl linear unabhängiger Spalten in A.

Beispiel 16.40. Nach Bemerkung 16.39 (c) ist z. B.

$$\operatorname{rk}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 0, \quad \operatorname{rk}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix} = 1 \quad \text{und} \quad \operatorname{rk}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} = 2.$$

Wie oben schon erwähnt hat der Rang einer Matrix die wichtige Eigenschaft, sich beim Übergang zu einer äquivalenten Matrix nicht zu ändern. Dies wollen wir jetzt beweisen.

Lemma 16.41. *Es sei $f\colon V \to W$ ein Morphismus zwischen endlich erzeugten Vektorräumen. Dann gilt*

$$\operatorname{rk} A_f^{B,C} = \operatorname{rk} f$$

für alle Basen B und C von V bzw. W, d. h. der Rang der Abbildungsmatrix von f ist unabhängig von der Wahl der Basen.

Insbesondere haben äquivalente Matrizen also denselben Rang.

Beweis. Nach Satz 16.23 ist $A_f^{B,C} = (\,\Phi_C(f(x_1)) \mid \cdots \mid \Phi_C(f(x_n))\,)$ mit $B = (x_1, \dots, x_n)$. Also gilt

$$\begin{aligned} \operatorname{rk} A_f^{B,C} &= \dim \operatorname{Lin}(\Phi_C(f(x_1)), \dots, \Phi_C(f(x_n))) && \text{(Bemerkung 16.16 (b))} \\ &= \dim \Phi_C(f(V)) && \text{(Satz 14.22 (a))} \\ &= \dim f(V) && (\Phi_C \text{ ist ein Isomorphismus}) \\ &= \operatorname{rk} f. \end{aligned}$$

□

Wir können nun zeigen, dass wir zu jedem Morphismus $f\colon V \to W$ Basen vom Start- und Zielraum so finden können, dass die zugehörige Abbildungsmatrix eine sehr einfache Form hat:

Satz und Definition 16.42 (Normalform von Abbildungsmatrizen)**.** *Es sei $f\colon V \to W$ ein Morphismus vom Rang r zwischen endlich erzeugten K-Vektorräumen mit $n = \dim V$ und $m = \dim W$. Dann gibt es Basen B und C von V bzw. W, so dass die Abbildungsmatrix von f bezüglich dieser Basen die Form*

$$A_f^{B,C} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1 & \\ \hline & 0 & & 0 \end{array}\right)$$

hat, d. h. so dass Einsen genau auf den ersten r Diagonalpositionen stehen, und sonst überall Nullen. Man sagt, dass eine solche Abbildungsmatrix in ***Normalform*** *ist. (Beachte dabei, dass unter der Einheitsmatrix E_r genau $m-r$ Nullzeilen, rechts von der Einheitsmatrix hingegen $n-r$ Nullspalten stehen – die Matrix $A_f^{B,C}$ ist also nicht notwendig quadratisch.)*

Beweis. Nach der Dimensionsformel aus Folgerung 15.26 ist $\dim \operatorname{Ker} f = \dim V - \dim \operatorname{Im} f = n - r$. Wähle nun eine Basis $(x_{r+1}, \ldots, x_n)$ von $\operatorname{Ker} f$ und ergänze diese zu einer Basis $B = (x_1, \ldots, x_n)$ von V. Dann ist $(\overline{x_1}, \ldots, \overline{x_r})$ nach Bemerkung 15.19 eine Basis des Quotientenraums $V/\operatorname{Ker} f$.

Nach dem Homomorphiesatz 15.23 ist nun aber die Abbildung $V/\operatorname{Ker} f \to \operatorname{Im} f$, $\overline{x} \mapsto f(x)$ ein Isomorphismus, und damit ist $(y_1, \ldots, y_r)$ mit $y_i := f(x_i)$ für $i = 1, \ldots, r$ nach Satz 14.22 (c) eine Basis von $\operatorname{Im} f$. Wir ergänzen diese schließlich noch zu einer Basis $C = (y_1, \ldots, y_m)$ von W. Wegen

$$f(x_i) = \begin{cases} y_i & \text{für } i \leq r, \\ 0 & \text{für } i > r \quad \text{(denn dann ist } x_i \in \operatorname{Ker} f) \end{cases}$$

hat die Matrix $A_f^{B,C}$ nach Definition 16.23 dann die gewünschte Form (z. B. ist die erste Spalte dieser Matrix der Koordinatenvektor von $f(x_1) = y_1$ bezüglich C, also der erste Einheitsvektor). □

Bemerkung 16.43.

(a) Nach Definition 16.36 können wir Satz 16.42 auch so formulieren: Ist $A \in \operatorname{Mat}(m \times n, K)$ eine beliebige Matrix vom Rang r, so ist A äquivalent zu der Matrix

$$\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right).$$

Analog zu Satz 16.42 nennt man dies auch die **Normalform** von A (bezüglich der Äquivalenz von Matrizen); sie ist die „einfachste“ Matrix in der Äquivalenzklasse von A.

(b) Beachte, dass der Beweis von Satz 16.42 konstruktiv ist, also auch die konkrete Berechnung der Basen B und C ermöglicht.

Eine erste wichtige und überraschende Folgerung aus Satz 16.42 ist die Aussage, dass der Rang einer Matrix A mit dem Rang der transponierten Matrix A^T übereinstimmt, dass $\operatorname{rk} A$ also (gemäß Bemerkung 16.39 (c)) nicht nur die Maximalzahl linear unabhängiger Spalten, sondern auch die Maximalzahl linear unabhängiger Zeilen in A ist.

Folgerung 16.44. *Für jede Matrix $A \in \operatorname{Mat}(m \times n, K)$ gilt $\operatorname{rk}(A^\mathsf{T}) = \operatorname{rk} A$.*

Beweis. Nach Bemerkung 16.43 (a) gibt es $S \in \operatorname{GL}(m, K)$ und $T \in \operatorname{GL}(n, K)$, so dass

$$SAT = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

mit $r = \operatorname{rk} A$. Nach Lemma 16.7 (d) gilt dann auch

$$T^\mathsf{T} A^\mathsf{T} S^\mathsf{T} = (SAT)^\mathsf{T} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)^\mathsf{T} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

(wobei sich beim letzten Gleichheitszeichen die Größe der Matrix von $m \times n$ auf $n \times m$ ändert). Nach Lemma 16.20 (c) sind mit S und T ferner auch S^T und T^T invertierbar, d. h. $T^\mathsf{T} A^\mathsf{T} S^\mathsf{T}$ ist äquivalent zu A^T. Damit ergibt sich aus Lemma 16.41

$$\operatorname{rk} A^\mathsf{T} = \operatorname{rk}(T^\mathsf{T} A^\mathsf{T} S^\mathsf{T}) = \operatorname{rk}\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) = r = \operatorname{rk} A. \qquad \square$$

Aufgabe 16.45. Wir betrachten den Morphismus $f\colon \operatorname{Mat}(2 \times 2, \mathbb{R}) \to \operatorname{Mat}(2 \times 2, \mathbb{R})$, $M \mapsto M + M^\mathsf{T}$.

(a) Berechne die Abbildungsmatrix von f bezüglich der „Standardbasis“ von $\operatorname{Mat}(2 \times 2, \mathbb{R})$ aus Bemerkung 16.4 (a) als Start- und Zielbasis.

(b) Finde Basen B und C von $\operatorname{Mat}(2 \times 2, \mathbb{R})$, so dass $A_f^{B,C}$ in Normalform ist.

Aufgabe 16.46. Es sei $A \in \mathrm{Mat}(m \times n, K)$ eine Matrix vom Rang r. Für $k \leq \min(m,n)$ verstehen wir unter einer $k \times k$-*Teilmatrix* von A einer Matrix, die man erhält, indem man aus A eine beliebige Auswahl von Zeilen und Spalten herausstreicht, so dass eine quadratische Matrix der Größe $k \times k$ übrig bleibt.

Zeige, dass es genau dann eine invertierbare $k \times k$-Teilmatrix von A gibt, wenn $k \leq r$ gilt.

17. Das Gauß-Verfahren

Nachdem wir jetzt viele Probleme der linearen Algebra (z. B. Basen von Vektorräumen zu konstruieren, Morphismen durch lineare Abbildungen darzustellen oder den Rang einer Matrix zu bestimmen) theoretisch lösen können, wollen wir in diesem Kapitel sehen, wie man alle diese Rechnungen mit Hilfe des sogenannten Gauß-Verfahrens auch dann praktisch durchführen kann, wenn sie numerisch so kompliziert werden, dass man das Ergebnis nicht mehr einfach sehen kann.

17.A Lineare Gleichungssysteme

Wie wir in diesem Kapitel sehen werden, führen alle bisher betrachteten konkreten Rechnungen der linearen Algebra letztlich auf *lineare Gleichungssysteme*, d. h. auf die folgende Situation: Für eine gegebene Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$ und $b_1, \dots, b_m \in K$ suchen wir alle $x_1, \dots, x_n \in K$, für die die Gleichungen

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

gelten. Mit Hilfe des Matrixprodukts ausgedrückt wollen wir also die Gleichung $Ax = b$ nach x auflösen, wobei $b_1, \dots, b_m$ und $x_1, \dots, x_n$ die Komponenten der Vektoren $b \in K^m$ und $x \in K^n$ bezeichnen. Bevor wir aber konkret angeben, wie man ein solches Gleichungssystem numerisch lösen kann, wollen wir zunächst etwas über die Struktur der Lösungsmenge dieses Systems aussagen.

Satz 17.1 (Existenz und Eindeutigkeit von Lösungen linearer Gleichungssysteme)**.** *Für ein lineares Gleichungssystem $Ax = b$ mit $A \in \mathrm{Mat}(m \times n, K)$ und $b \in K^m$ gilt:*

(a) *(Existenz) Das Gleichungssystem hat genau dann eine Lösung $x \in K^n$, wenn $\mathrm{rk}(A\,|\,b) = \mathrm{rk}A$ gilt, wobei $(A\,|\,b)$ die sogenannte* erweiterte Koeffizientenmatrix

$$(A\,|\,b) = \left(\begin{array}{ccc|c} a_{1,1} & \dots & a_{1,n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m,1} & \dots & a_{m,n} & b_m \end{array}\right) \in \mathrm{Mat}(m \times (n+1), K)$$

bezeichnet.

(b) *(Eindeutigkeit) Ist das Gleichungssystem lösbar und $v \in K^n$ eine Lösung (d. h. gilt $Av = b$), so ist die gesamte Lösungsmenge des Gleichungssystems gegeben durch*

$$\{x \in K^n : Ax = b\} = v + \mathrm{Ker}A.$$

Die Lösungsmenge ist dann also ein verschobener Unterraum der Dimension

$$\dim \mathrm{Ker}A = n - \mathrm{rk}A.$$

Insbesondere ist die Lösung v des Gleichungssystems also genau dann eindeutig, wenn $\mathrm{rk}A = n$.

Beweis.

(a) Bezeichnen wir mit $x_1, \dots, x_n$ die Komponenten von x und mit $a_i \in K^m$ die i-te Spalte von A (für $i = 1, \dots, n$), so können wir das Gleichungssystem $Ax = b$ auch schreiben als

$$x_1a_1 + \cdots + x_na_n = b \quad \in K^m.$$

Es gilt dann

$$\begin{aligned}
Ax = b \text{ lösbar} &\Leftrightarrow \text{es gibt } x_1,\dots,x_n \text{ mit } x_1a_1+\dots+x_na_n = b \\
&\Leftrightarrow b \in \operatorname{Lin}(a_1,\dots,a_n) \\
&\Leftrightarrow \operatorname{Lin}(a_1,\dots,a_n,b) = \operatorname{Lin}(a_1,\dots,a_n) \\
&\Leftrightarrow \dim\operatorname{Lin}(a_1,\dots,a_n,b) = \dim\operatorname{Lin}(a_1,\dots,a_n) \quad \text{(Lemma 14.26)} \\
&\Leftrightarrow \operatorname{rk}(A\,|\,b) = \operatorname{rk}A \quad \text{(Bemerkung 16.39 (c)).}
\end{aligned}$$

(b) Ist $Av = b$, so ist für ein $x \in K^n$ genau dann $Ax = b$, wenn $A(x-v) = 0$ gilt, also $x - v \in \operatorname{Ker}A$ und damit $x \in v + \operatorname{Ker}A$ ist. □

Bemerkung 17.2 (Eindeutige Lösbarkeit linearer Gleichungssysteme)**.**

(a) Fassen wir die beiden Teile von Satz 17.1 zusammen, so sehen wir, dass das lineare Gleichungssystem $Ax = b$ mit $A \in \operatorname{Mat}(m \times n, K)$ und $b \in K^m$ genau dann eine eindeutige Lösung $x \in K^n$ besitzt, wenn $\operatorname{rk}(A\,|\,b) = \operatorname{rk}A = n$ ist.

(b) Im Fall $m = n$ eines Gleichungssystems mit genau so vielen Variablen wie Gleichungen vereinfacht sich das Kriterium aus (a) noch etwas: Da der Rang der erweiterten Koeffizientenmatrix $(A\,|\,b)$ nicht kleiner als $\operatorname{rk}A$ und (nach Bemerkung 16.39 (a)) nicht größer als $m = n$ sein kann, ist die eindeutige Lösbarkeit des Gleichungssystems $Ax = b$ in diesem Fall bereits äquivalent zur Bedingung $\operatorname{rk}A = n$ – was nach Bemerkung 16.39 (b) genau bedeutet, dass A invertierbar ist.

In der Tat können wir die eindeutige Lösung des Gleichungssystems $Ax = b$ dann natürlich durch Multiplikation mit A^{-1} von links auch sofort als $x = A^{-1}b$ schreiben. Momentan hilft uns dies allerdings nicht besonders viel, weil wir noch nicht wissen, wie man die inverse Matrix A^{-1} im Allgemeinen konkret berechnen kann.

Das Hauptziel dieses Abschnitts ist es nun zu sehen, wie man lineare Gleichungssysteme explizit numerisch lösen kann. Wie ihr natürlich wisst, besteht die Strategie hierbei darin, die gegebenen Gleichungen auf geschickte Art z. B. durch Addition, Subtraktion oder Multiplikation mit Skalaren so umzuformen und zu kombinieren, dass hinterher ein äquivalentes Gleichungssystem entsteht, von dem man die Lösung leicht ablesen kann. Schreiben wir das Gleichungssystem in Matrixform als $Ax = b$, so entspricht nun jede Gleichung einer Zeile der Matrix A, und demzufolge wollen wir also die Zeilen der Matrix umformen und miteinander kombinieren können. Diese Zeilenumformungen entsprechen in unserer Schreibweise einfachen Matrixmultiplikationen, die wir jetzt untersuchen wollen.

Konstruktion 17.3 (Elementarmatrizen)**.** Es sei $A \in \operatorname{Mat}(m \times n, K)$.

(a) Für $k \in \{1,\dots,m\}$ und $\lambda \in K \backslash \{0\}$ setzen wir

$$F_k(\lambda) := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \lambda & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in \operatorname{Mat}(m \times m, K),$$

wobei der Eintrag λ in Zeile und Spalte k steht. Die Matrix $F_k(\lambda)$ ist also nichts weiter als die Einheitsmatrix, bei der der Eintrag 1 in der k-ten Zeile und Spalte durch ein $\lambda \neq 0$ ersetzt wurde.

Mit dieser Matrix ist das Produkt $F_k(\lambda)\cdot A$ nun gegeben durch

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \lambda & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ \lambda a_{k,1} & \cdots & \lambda a_{k,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix},$$

d. h. die Multiplikation einer Matrix $A \in \mathrm{Mat}(m\times n, K)$ mit $F_k(\lambda)$ von links entspricht genau der Multiplikation der k-ten Zeile von A mit λ.

(b) Für $k,l \in \{1,\dots,m\}$ mit $k \neq l$ und $\lambda \in K$ setzen wir

$$F_{k,l}(\lambda) := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \lambda & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in \mathrm{Mat}(m\times m, K),$$

wobei der Eintrag λ in Zeile k und Spalte l steht, d. h. diesmal haben wir in der Einheitsmatrix den Eintrag 0 in Zeile k und Spalte l durch λ ersetzt. (Beachte, dass der Eintrag λ für $k < l$ oberhalb und für $k > l$ unterhalb der Diagonalen steht; wir haben in der Matrix oben der Einfachheit halber nur den Fall $k < l$ dargestellt.) In diesem Fall ist das Matrixprodukt $F_{k,l}(\lambda)\cdot A$ gleich

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \lambda & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,n} \\ \vdots & & \vdots \\ a_{l,1} & \cdots & a_{l,n} \\ \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & & \vdots \\ a_{k,1}+\lambda a_{l,1} & \cdots & a_{k,n}+\lambda a_{l,n} \\ \vdots & & \vdots \\ a_{l,1} & \cdots & a_{l,n} \\ \vdots & & \vdots \end{pmatrix},$$

d. h. die Multiplikation einer Matrix $A \in \mathrm{Mat}(m\times n, K)$ mit $F_{k,l}(\lambda)$ von links entspricht genau der Addition des λ-fachen von Zeile l zu Zeile k.

Die Matrizen $F_k(\lambda)$ für $\lambda \in K\backslash\{0\}$ sowie $F_{k,l}(\lambda)$ für $k \neq l$ und $\lambda \in K$ heißen **Elementarmatrizen**. Es gibt sie in jeder (quadratischen) Größe $m\times m$; zur Vereinfachung der Schreibweise deuten wir diese Größe in der Notation $F_k(\lambda)$ bzw. $F_{k,l}(\lambda)$ aber nicht an.

Man sagt, dass $F_k(\lambda)\cdot A$ und $F_{k,l}(\lambda)\cdot A$ aus A durch eine **elementare Zeilenumformung** entstehen.

Bemerkung 17.4.

(a) Die Elementarmatrizen sind invertierbar mit

$$(F_k(\lambda))^{-1} = F_k\left(\frac{1}{\lambda}\right) \quad \text{und} \quad (F_{k,l}(\lambda))^{-1} = F_{k,l}(-\lambda).$$

Dies folgt auch ohne weitere Rechnung direkt aus Konstruktion 17.3: Wenn wir z. B. das Matrixprodukt $F_k(\frac{1}{\lambda})\cdot F_k(\lambda) = F_k(\frac{1}{\lambda})\cdot F_k(\lambda)\cdot E$ bilden, multiplizieren wir die k-te Zeile in der Einheitsmatrix zuerst mit λ und dann mit $\frac{1}{\lambda}$, d. h. es kommt insgesamt wieder die Einheitsmatrix heraus – also ist $F_k(\frac{1}{\lambda})\cdot F_k(\lambda) = E$. Genauso ergibt sich auch $F_k(\lambda)\cdot F_k(\frac{1}{\lambda})$, und damit ist $F_k(\frac{1}{\lambda})$ das Inverse von $F_k(\lambda)$. Analog erhält man auch das Produkt $F_{k,l}(-\lambda)\cdot F_{k,l}(\lambda)$, indem man in der Einheitsmatrix das λ-fache der l-ten Zeile zur k-ten addiert und dann gleich wieder davon subtrahiert, so dass auch hier wieder die Einheitsmatrix herauskommt.

(b) Führen wir mit einer Matrix A mehrere elementare Zeilenumformungen aus, so ist das Ergebnis nach Konstruktion 17.3 eine Matrix $F_1\cdot\cdots\cdot F_r\cdot A$ mit Elementarmatrizen $F_1,\dots,F_r$. Da diese Elementarmatrizen nach (a) invertierbar sind, ist das Produkt $F := F_1\cdot\cdots\cdot F_r$ nach Lemma 16.20 (a) ebenfalls wieder invertierbar. Das Ergebnis der Zeilenumformungen ist also die Matrix FA, d. h. wir haben A von links mit einer invertierbaren Matrix multipliziert.

(c) Das Vertauschen zweier Zeilen $k,l \in \{1,\dots,m\}$ mit $k \neq l$ in einer Matrix $A \in \mathrm{Mat}(m \times n, K)$ lässt sich als Folge von elementaren Zeilenumformungen realisieren: Sind $a_1,\dots,a_m$ die Zeilen von A, so erhalten wir

$$A = \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} \xrightarrow{F_{k,l}(1)\cdot} \begin{pmatrix} \vdots \\ a_k + a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix} \xrightarrow{F_{l,k}(-1)\cdot} \begin{pmatrix} \vdots \\ a_k + a_l \\ \vdots \\ -a_k \\ \vdots \end{pmatrix} \xrightarrow{F_{k,l}(1)\cdot} \begin{pmatrix} \vdots \\ a_l \\ \vdots \\ -a_k \\ \vdots \end{pmatrix} \xrightarrow{F_l(-1)\cdot} \begin{pmatrix} \vdots \\ a_l \\ \vdots \\ a_k \\ \vdots \end{pmatrix}.$$

(d) Man kann die Elementarmatrizen auch verwenden, um analog *elementare Spaltenumformungen* durchzuführen: Multipliziert man eine Matrix $A \in \mathrm{Mat}(m \times n, K)$ mit einer Elementarmatrix $F_k(\lambda)$ (der Größe n) von *rechts*, so entspricht dies der Multiplikation der k-ten *Spalte* von A mit λ. Analog addiert man das λ-fache von Spalte k zu Spalte l von A, wenn man das Produkt $A \cdot F_{k,l}(\lambda)$ bildet. Wie in (b) lassen sich elementare Spaltenumformungen also durch Multiplikation mit einer invertierbaren Matrix von rechts beschreiben.

Mit derartigen Zeilenumformungen wollen wir jetzt eine Matrix bzw. ein lineares Gleichungssystem auf eine möglichst einfache Form bringen. Die folgende Form wird dabei herauskommen.

Definition 17.5 (Zeilenstufenform)**.** Wir sagen, dass eine Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, K)$ **Zeilenstufenform** hat, wenn sie von der Form

$$\begin{array}{c} \begin{array}{ccccc} k_1 & k_2 & & k_r & \\ \downarrow & \downarrow & & \downarrow & \end{array} \\ \begin{pmatrix} 1 & & & & \\ & 1 & & * & \\ & & \ddots & & \\ 0 & & & 1 & \\ & & & & \end{pmatrix} \end{array}$$

ist, wobei das Symbol „$*$“ für beliebige Einträge steht. Mit anderen Worten gibt es ein $r \in \{0,\dots,m\}$ und Spalten $1 \leq k_1 < \dots < k_r \leq n$ (die sogenannten **Stufenspalten**), so dass gilt:

- $a_{i,k_i} = 1$ für alle $i = 1,\dots,r$;
- rechts hiervon stehen beliebige Einträge (d. h. $a_{i,j}$ ist beliebig falls $i \leq r$ und $j > k_i$);
- alle anderen Einträge sind 0.

Sind zusätzlich noch außer den geforderten Einsen alle anderen Einträge in den Stufenspalten gleich Null, ist A also sogar von der Form

$$\begin{pmatrix} 1 & * \cdots * & 0 & * \cdots * & & 0 & * \cdots * \\ & & 1 & * \cdots * & & 0 & * \cdots * \\ & & & & \ddots & \vdots & \vdots \\ 0 & & & & & 1 & * \cdots * \\ & & & & & & \end{pmatrix}$$

(d. h. $a_{l,k_i} = 0$ für alle $l < i \leq r$) so sagt man, dass A **reduzierte Zeilenstufenform** hat.

37

Der wesentliche Satz in diesem Kapitel besagt nun, dass man jede Matrix mit elementaren Zeilenumformungen auf diese Formen bringen kann. Dabei sind beide Formen in der Praxis wichtig: Die reduzierte Zeilenstufenform ist natürlich „schöner“, weil sie mehr Nullen enthält und damit einfacher ist – andererseits werden wir aber sehen, dass die normale, nicht-reduzierte Zeilenstufenform für viele Zwecke ausreichend und mit weniger Rechenaufwand zu erzielen ist.

Satz 17.6 (Gauß-Verfahren). *Jede Matrix $A \in \mathrm{Mat}(m \times n, K)$ lässt sich durch elementare Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen. Mit anderen Worten gibt es also ein Produkt $F \in \mathrm{GL}(m, K)$ von Elementarmatrizen, so dass FA (reduzierte) Zeilenstufenform hat (siehe Bemerkung 17.4 (b)).*

Bei diesem Satz ist der Beweis – also das Verfahren, wie man die (reduzierte) Zeilenstufenform erreicht – mindestens genauso wichtig wie die eigentliche Aussage:

Beweis. Wir beweisen den Satz mit Induktion über die Zahl n der Spalten von A. Da für $n = 0$ nichts zu zeigen ist, müssen wir nur den Induktionsschritt durchführen. Es sei also $A \in \mathrm{Mat}(m \times n, K)$ beliebig vorgegeben. Wir wollen A mit elementaren Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen und nehmen nach Induktionsvoraussetzung an, dass wir dies für alle Matrizen mit weniger als n Spalten bereits können. Wir unterscheiden dabei zwei Fälle:

Fall 1: Alle Einträge in der ersten Spalte von A sind 0, d. h. A hat die Form

$$A = \left(\begin{array}{c|c} 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$$

für eine Matrix $A' \in \mathrm{Mat}(m \times (n-1), K)$. Nach Induktionsvoraussetzung können wir A' dann durch Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen. Da diese Zeilenumformungen an der ersten Nullspalte aber nichts ändern, haben wir damit auch A auf (reduzierte) Zeilenstufenform gebracht.

Fall 2: Es sind nicht alle Einträge in der ersten Spalte von A gleich 0.

(a) Falls der Eintrag $a_{1,1}$ links oben in A gleich Null ist, vertauschen wir zwei Zeilen von A so, dass dieser Eintrag nicht mehr gleich 0 ist (nach Bemerkung 17.4 (c) ist dies durch elementare Zeilenumformungen machbar).

(b) Wir dividieren die erste Zeile durch $a_{1,1}$ und erhalten eine Matrix der Form

$$\left(\begin{array}{c|c} 1 & \\ a_{2,1} & \\ \vdots & * \\ a_{m,1} & \end{array}\right).$$

(c) Wir subtrahieren von jeder Zeile $k \in \{2, \ldots, m\}$ das $a_{k,1}$-fache der ersten Zeile und bekommen dadurch

$$\left(\begin{array}{c|c} 1 & * \cdots * \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$$

mit einer Matrix $A' \in \mathrm{Mat}((m-1) \times (n-1), K)$.

(d) Nach Induktionsvoraussetzung können wir jetzt A' durch elementare Zeilenumformungen auf Zeilenstufenform bringen. Damit hat aber auch die gesamte Matrix bereits Zeilenstufenform – wollen wir also nur diese normale, nicht-reduzierte Zeilenstufenform erreichen, so sind wir damit fertig. Andernfalls bringen wir A' gemäß der Induktionsvoraussetzung sogar auf reduzierte Zeilenstufenform und bekommen eine Matrix der Form

$$\left(\begin{array}{c|cccccc} 1 & * \cdots * & * & * \cdots * & & * & * \cdots * \\ \hline 0 & & 1 & * \cdots * & & 0 & * \cdots * \\ \vdots & & & & \ddots & & \\ \vdots & & 0 & & & 1 & * \cdots * \\ 0 & & & & & & \end{array}\right).$$

Subtrahieren wir nun geeignete Vielfache der Zeilen $2,\dots,m$ von der ersten Zeile, so können wir damit dann noch die Einträge in den Stufenspalten der ersten Zeile zu Null machen und erhalten so auch die reduzierte Zeilenstufenform. $\square$

Beispiel 17.7. Wir wollen die Matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \in \operatorname{Mat}(3\times 3,\mathbb{R})$$

mit elementaren Zeilenumformungen auf Zeilenstufenform bringen. Dazu können wir nach dem Verfahren im Beweis von Satz 17.6 wie folgt vorgehen (die Notation $Z3 \to Z3 - 3Z2$ bedeutet z. B., dass wir das Dreifache der zweiten Zeile von der dritten subtrahieren):

$$\begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \xrightarrow{Z1\to -Z1} \begin{pmatrix} 1 & -2 & -4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \xrightarrow[Z3\to Z3-Z1]{Z2\to Z2+Z1} \begin{pmatrix} 1 & -2 & -4 \\ 0 & 1 & 3 \\ 0 & 3 & 9 \end{pmatrix}$$

$$\xrightarrow{Z3\to Z3-3Z2} \begin{pmatrix} 1 & -2 & -4 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Möchten wir eine reduzierte Zeilenstufenform, so addieren wir schließlich noch das Doppelte der zweiten Zeile zur ersten und erhalten

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Ein erster Nutzen einer Zeilenstufenform besteht darin, dass wir an ihr sofort den Rang einer Matrix ablesen können:

Algorithmus 17.8 (Rang einer Matrix)**.** Haben wir eine Matrix $A \in \operatorname{Mat}(m\times n,K)$ wie in Satz 17.6 mit einem Produkt $F \in \operatorname{GL}(m,K)$ von Elementarmatrizen auf Zeilenstufenform FA gebracht, so ist FA äquivalent zu A und hat damit nach Lemma 16.41 insbesondere auch den gleichen Rang wie A. Der Rang (also nach Bemerkung 16.39 (c) und Folgerung 16.44 die maximale Anzahl linear unabhängiger Zeilen) von FA ist aber natürlich einfach die Zahl r der Stufen wie in Definition 17.5:

$$\boxed{\text{Der Rang einer Matrix } A \text{ ist die Anzahl der Stufen in einer Zeilenstufenform von } A.}$$

Für die Matrix A aus Beispiel 17.7 gilt demnach z. B. $\operatorname{rk} A = 2$.

Die wichtigste Anwendung des Gauß-Verfahrens ist aber sicher, dass wir damit wie folgt sehr einfach lineare Gleichungssysteme lösen können.

Algorithmus 17.9 (Lineare Gleichungssysteme)**.** Es seien $A \in \operatorname{Mat}(m\times n,K)$ und $b \in K^m$.

Um das lineare Gleichungssystem $Ax = b$ zu lösen, bringen wir die Matrix A mit elementaren Zeilenumformungen auf reduzierte Zeilenstufenform und führen dabei alle Zeilenumformungen mit der erweiterten Koeffizientenmatrix $(A\,|\,b) \in \operatorname{Mat}(m\times(n+1),K)$ durch. Wir erhalten am Ende also eine Matrix $(FA\,|\,Fb)$, wobei $F \in \operatorname{GL}(m,K)$ ein Produkt von Elementarmatrizen ist und FA reduzierte Zeilenstufenform hat.

Da F invertierbar ist, ist das ursprüngliche lineare Gleichungssystem $Ax = b$ äquivalent zu $FAx = Fb$, d. h. wir haben die Lösungsmenge des Gleichungssystems durch den Übergang von $(A\,|\,b)$ zu $(FA\,|\,Fb)$ nicht geändert. Das neue Gleichungssystem lässt sich nun aber leicht lösen: Die resultierende Matrix ist von der Form

$$\left(\begin{array}{cccccccccc|c}
 & k_1 & & & k_2 & & & k_r & & & \\
 & \downarrow & & & \downarrow & & & \downarrow & & & \\
 & 1 & * & \cdots & * & 0 & * \cdots * & 0 & * & \cdots * & b'_1 \\
 & & & & & 1 & * \cdots * & 0 & * & \cdots * & b'_2 \\
 & & & & & & & \vdots & & \vdots & \vdots \\
 & 0 & & & & & & 1 & * & \cdots * & b'_r \\
 & & & & & & & & & & \vdots \\
 & & & & & & & & & & b'_m
\end{array}\right),$$

wobei $b'_1, \ldots, b'_m \in K$ die Komponenten von Fb sind. Wir unterscheiden nun zwei Fälle:

(a) Ist $b'_i \neq 0$ für ein $i \in \{r+1, \ldots, m\}$, so liefert die i-te Zeile des umgeformten Gleichungssystems den Widerspruch $0 = b'_i$. Das Gleichungssystem ist dann also nicht lösbar.

(b) Ist $b'_{r+1} = \cdots = b'_m = 0$, so sind die letzten $m - r$ Gleichungen des umgeformten Systems einfach $0 = 0$ und damit immer erfüllt. Wir brauchen also nur die ersten r Zeilen zu betrachten. Von diesen drückt die i-te Zeile offensichtlich die Stufenvariable x_{k_i} eindeutig durch die Variablen x_j mit $j \notin \{k_1, \ldots, k_r\}$ aus. Wir können also alle Variablen, die nicht an den Stufenspalten stehen, frei wählen, und erhalten dann für jede solche Wahl genau eine Lösung, bei der die restlichen Variablen $x_{k_1}, \ldots, x_{k_r}$ durch Verwendung der ersten r Gleichungen eindeutig bestimmt sind.

Alternativ kann man A auch nur mit elementaren Zeilenumformungen auf nicht notwendig reduzierte Zeilenstufenform bringen (was natürlich weniger Arbeit ist). Dann drückt die i-te Zeile der Zeilenstufenform die Variable x_{k_i} durch alle freien Variablen x_j mit $j \notin \{k_1, \ldots, k_r\}$ *und die weiter rechts stehenden Stufenvariablen* $x_{k_{i+1}}, \ldots, x_{k_r}$ aus. Wir müssen diese Zeilen dann also von unten nach oben durchgehen und so der Reihe nach $x_{k_r}, \ldots, x_{k_1}$ aus den freien Variablen *und den bereits bestimmten Stufenvariablen* berechnen (was dann wieder etwas mehr Arbeit ist).

Beispiel 17.10. Wir wollen das lineare Gleichungssystem

$$\begin{aligned} x_1 + 4x_2 \phantom{{}+ x_3} + 5x_4 &= 1 \\ 2x_1 + 8x_2 + x_3 + 7x_4 &= 3 \\ 2x_3 - 6x_4 &= 2 \end{aligned}$$

über $\mathbb{R}$ lösen, also das System $Ax = b$ mit

$$A = \begin{pmatrix} 1 & 4 & 0 & 5 \\ 2 & 8 & 1 & 7 \\ 0 & 0 & 2 & -6 \end{pmatrix} \in \mathrm{Mat}(3 \times 4, \mathbb{R}) \quad \text{und} \quad b = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \in \mathbb{R}^3.$$

Dazu bringen wir die Matrix A auf reduzierte Zeilenstufenform, führen aber alle Umformungen mit der erweiterten Koeffizientenmatrix $(A\,|\,b)$ durch:

$$\left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 2 & 8 & 1 & 7 & 3 \\ 0 & 0 & 2 & -6 & 2 \end{array}\right) \xrightarrow{Z2 \to Z2 - 2Z1} \left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 2 & -6 & 2 \end{array}\right) \xrightarrow{Z3 \to Z3 - 2Z2} \left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right). \tag{1}$$

Wir haben also das neue äquivalente Gleichungssystem

$$\begin{aligned} x_1 + 4x_2 \phantom{{}- 3x_3} + 5x_4 &= 1 \\ x_3 - 3x_4 &= 1 \\ 0 &= 0 \end{aligned}$$

erhalten, das wir sofort lösen können: Die Variablen x_2 und x_4 (die nicht an den Stufenspalten stehen) können wir als freie reelle Parameter wählen, und die beiden nicht-trivialen Gleichungen des umgeformten Systems bestimmen dann die restlichen beiden Variablen eindeutig zu $x_1 = 1 - 4x_2 - 5x_4$ bzw. $x_3 = 1 + 3x_4$. Wir erhalten also die allgemeine Lösung

$$\{x \in \mathbb{R}^4 : Ax = b\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -5 \\ 0 \\ 3 \\ 1 \end{pmatrix} : x_2, x_4 \in \mathbb{R} \right\}. \tag{2}$$

Algorithmus 17.11 (Basis vom Kern einer Matrix)**.** Ein unmittelbarer Spezialfall von Algorithmus 17.9 zur Lösung eines linearen Gleichungssystems ist die Berechnung des Kerns einer Matrix $A \in \mathrm{Mat}(m \times n, K)$, denn $\mathrm{Ker}A$ ist ja genau die Lösungsmenge des Gleichungssystems $Ax = 0$. Im Gauß-Verfahren ist also die rechte Seite immer 0 (weswegen wir sie natürlich gar nicht erst mit hinschreiben müssen).

Die Dimension des Kerns ist offensichtlich gleich der Anzahl der freien Parameter in der Lösung, also die Anzahl der Nichtstufenspalten $j \notin \{k_1, \dots, k_r\}$ in der Notation von Algorithmus 17.9 (beachte, dass diese Anzahl in Übereinstimmung mit Bemerkung 16.16 (a) und Algorithmus 17.8 genau $n - r = n - \mathrm{rk}A = \dim \mathrm{Ker}A$ ist). Um eine Basis von $\mathrm{Ker}A$ zu bestimmen, können wir für jedes solche j den Vektor der Lösungsmenge bilden, bei dem wir $x_j = 1$ und alle anderen freien Variablen gleich 0 setzen. Im Beispiel 17.10 ist für die dort angegebene Matrix A z. B.

$$\left(\begin{pmatrix} -4 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ 0 \\ 3 \\ 1 \end{pmatrix} \right)$$

(also die Vektoren, die in der Lösungsmenge (2) bei den freien Variablen x_2 und x_4 stehen) eine Basis von $\mathrm{Ker}A$.

Dieses Verfahren lässt sich durch einen kleinen Rechentrick noch etwas vereinfachen. Dazu seien $a_{i,j}$ die Einträge der reduzierten Zeilenstufenform von A. Wenn wir jetzt für den Basisvektor zur Nichtstufenspalte j die freie Variable x_j gleich -1 statt 1 setzen (was den Basisvektor nur mit -1 multipliziert), bestimmt die i-te Zeile des umgeformten Systems für $i = 1, \dots, r$ die zugehörige i-te Stufenvariable zu

$$1 \cdot x_{k_i} + a_{i,j} x_j = 0 \quad \Rightarrow \quad x_{k_i} = -a_{i,j} x_j = -a_{i,j} \cdot (-1) = a_{i,j},$$

also genau zu einem der Einträge der umgeformten Matrix. Schreibt man die i-te Zeile dieser reduzierten Zeilenstufenform nun noch in Zeile k_i einer $n \times n$-Matrix A' (so dass also die 1 der Stufe auf der Diagonalen dieser Matrix steht), so steht dieser Eintrag $x_{k_i} = a_{i,j}$ jetzt statt in Zeile i und Spalte j in Zeile k_i und Spalte j, und damit bereits an der richtigen Stelle dafür, dass sich der konstruierte Basisvektor in Spalte j dieser neuen Matrix ergibt. Wir erhalten damit das folgende nützliche „Kochrezept“:

> Zur Berechnung einer Basis von $\mathrm{Ker}A$ bringe man A auf reduzierte Zeilenstufenform. Man schreibe die Stufenzeilen dieser Matrix dann so in eine $n \times n$-Matrix A', dass die Einsen der Stufen auf der Diagonalen der Matrix stehen; die restlichen Zeilen fülle man mit -1 auf der Diagonalen und 0 sonst auf. Die Spalten von A' mit den Einträgen -1 auf der Diagonalen bilden dann eine Basis von $\mathrm{Ker}A$.

In unserem obigen Zahlenbeispiel bilden wir also eine neue 4×4-Matrix, in der wir die beiden Zeilen der reduzierten Zeilenstufenform (1) in Zeile 1 und 3 schreiben (damit die Einsen der Stufen auf der Diagonalen der Matrix sind, unten im Bild grau hinterlegt) und die restlichen Einträge mit -1 auf der Diagonalen und 0 sonst aufgefüllt sind. Die zweite und vierte Spalte (unten im Bild eingekreist) bilden dann eine Basis B des Kerns:

$$A' = \begin{pmatrix} 1 & 4 & 0 & 5 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad \Rightarrow \quad B = \left(\begin{pmatrix} 4 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \\ -3 \\ -1 \end{pmatrix} \right)$$

Aufgabe 17.12 (Eindeutigkeit der reduzierten Zeilenstufenform)**.** Zeige die folgende Verallgemeinerung von Lemma 17.17 (b): Führt man mit einer Matrix $A \in \mathrm{Mat}(m \times n, K)$ auf verschiedene Arten elementare Zeilenumformungen durch, so dass das Ergebnis jeweils eine Matrix in reduzierter Zeilenstufenform ist, so ergibt sich dabei immer dieselbe Matrix.

17.B Weitere Algorithmen mit dem Gauß-Verfahren

Wie bereits erwähnt lassen sich alle konkreten Rechnungen der linearen Algebra letztlich auf das Gauß-Verfahren zurückführen. Im Rest dieses Kapitels wollen wir dies für noch ein paar weitere häufig auftretende Aufgabenstellungen explizit tun. Das wichtigste Grundprinzip hierbei ist:

Lemma 17.13. *Für jede Matrix $A \in \mathrm{Mat}(m \times n, K)$ gilt:*

(a) *Elementare Zeilenumformungen in A ändern den Kern von A nicht.*

(b) *Elementare Spaltenumformungen in A ändern das Bild von A nicht.*

Beweis.

(a) Nach Bemerkung 17.4 (b) erhalten wir durch die Zeilenumformungen eine Matrix FA mit $F \in \mathrm{GL}(m, K)$. Dann ist aber

$$\mathrm{Ker}(FA) = \{x \in K^n : FAx = 0\} = \{x \in K^n : Ax = 0\} = \mathrm{Ker}A.$$

(b) In diesem Fall erhalten wir nach Bemerkung 17.4 (d) eine Matrix AF mit $F \in \mathrm{GL}(n, K)$. Mit der Substitution $y = Fx$, also $x = F^{-1}y$, ist dann

$$\mathrm{Im}(AF) = \{AFx : x \in K^n\} = \{Ay : y \in K^n\} = \mathrm{Im}A. \qquad \square$$

Algorithmus 17.14 (Basis vom Bild einer Matrix)**.** Die Bestimmung einer Basis von $\mathrm{Im}A$ für eine Matrix $A \in \mathrm{Mat}(m \times n, K)$ ist einfacher als die einer Basis von $\mathrm{Ker}A$. Es gibt dafür zwei Möglichkeiten:

(a) Nach Lemma 17.13 (b) können wir A mit elementaren Spaltenumformungen auf eine „Spaltenstufenform“ mit $r = \mathrm{rk}A$ Stufen bringen, ohne das Bild der Matrix zu ändern. Die ersten r Spalten (die nicht 0 sind) sind dann offensichtlich linear unabhängig und bilden damit eine Basis von $\mathrm{Im}A$. Alternativ können wir natürlich auch Zeilenumformungen mit der transponierten Matrix A^T durchführen.

Als Beispiel dafür betrachten wir noch einmal die Matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \in \mathrm{Mat}(3 \times 3, \mathbb{R})$$

aus Beispiel 17.7 und bringen A^T mit Zeilenumformungen auf Zeilenstufenform:

$$\begin{pmatrix} -1 & -1 & 1 \\ 2 & 3 & 1 \\ 4 & 7 & 5 \end{pmatrix} \xrightarrow{Z1 \to -Z1} \begin{pmatrix} 1 & 1 & -1 \\ 2 & 3 & 1 \\ 4 & 7 & 5 \end{pmatrix} \xrightarrow[Z3 \to Z3 - 4Z1]{Z2 \to Z2 - 2Z1} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 3 \\ 0 & 3 & 9 \end{pmatrix}$$

$$\xrightarrow{Z3 \to Z3 - 3Z2} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Damit ist also $\left(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \right)$ eine Basis von $\mathrm{Im}A$.

(b) (Basisauswahl) Nach Bemerkung 16.16 (b) wird $\operatorname{Im} A$ von den Spalten von A erzeugt. Wollen wir nicht nur eine beliebige Basis von $\operatorname{Im} A$ bestimmen, sondern wie in Satz 14.11 eine Basis aus den Spalten von A auswählen, so können wir alternativ folgendes Verfahren anwenden: Es seien $x_1, \dots, x_n$ die Spalten von A. Wir bringen nun A mit einem Produkt F von Elementarmatrizen auf Zeilenstufenform FA mit Stufenspalten $k_1 < \dots < k_r$ wie in Definition 17.5; natürlich ist dabei $r = \operatorname{rk} A$ nach Algorithmus 17.8.

Hätten wir diese Zeilenumformungen nur mit den Stufenspalten $k_1, \dots, k_r$ durchgeführt, so hätten wir stattdessen auch von der Zeilenstufenform nur diese Stufenspalten erhalten, also eine Matrix der Form

$$F \cdot (x_{k_1} \mid \dots \mid x_{k_r}) = \begin{pmatrix} 1 & & & * \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \in \operatorname{Mat}(m \times r, K)$$

bekommen, die offensichtlich denselben Rang r wie A besitzt. Also ist auch

$$\dim \operatorname{Lin}(x_{k_1}, \dots, x_{k_r}) = r = \operatorname{rk} A = \dim \operatorname{Im} A.$$

Da der Teilraum von $\operatorname{Im} A$, der von den ausgewählten Spalten $x_{k_1}, \dots, x_{k_r}$ erzeugt wird, dieselbe Dimension wie $\operatorname{Im} A$ hat, muss er nach Lemma 14.26 gleich $\operatorname{Im} A$ sein, d. h. $(x_{k_1}, \dots, x_{k_r})$ ist ein Erzeugendensystem und nach Satz 14.21 (a) wegen $r = \dim \operatorname{Im} A$ sogar eine Basis von $\operatorname{Im} A$.

So haben wir z. B. für die Matrix A aus (a) bereits in 17.7 eine (reduzierte) Zeilenstufenform berechnet und gesehen, dass die ersten beiden Spalten dabei die Stufenspalten waren. Also sind die ersten beiden Spalten

$$\left(\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \right)$$

von A eine Basis von $\operatorname{Im} A$.

Algorithmus 17.15 (Basisergänzung)**.** Genau wie die Basisauswahl in Algorithmus 17.14 (b) lässt sich auch die Basisergänzung mit dem Gauß-Verfahren explizit durchführen. Hat man nämlich mit Algorithmus 17.11 oder 17.14 (a) eine Basis vom Kern oder Bild einer Matrix $A \in \operatorname{Mat}(m \times n, K)$ bestimmt, so ist es in beiden Fällen sehr einfach, diese zu einer Basis vom gesamten Raum K^n bzw. K^m zu ergänzen:

(a) In Algorithmus 17.11 sind alle Spalten der quadratischen Matrix A' linear unabhängig (weil A' stets Einträge ± 1 auf der Diagonalen und nur Einträge 0 darunter hat). Die Einheitsvektoren e_i für alle Stufenspalten i (also die im Bild nicht eingekreisten Spalten in A') ergänzen damit die gefundene Basis von $\operatorname{Ker} A$ zu einer Basis von K^n – im konkreten Beispiel in Algorithmus 17.11 also e_1 und e_3.

(b) In Algorithmus 17.14 (a) ist aus der Form der gefundenen Basis B klar, dass B durch die Einheitsvektoren, die nicht zu den Stufenzeilen von A^T gehören, zu einer Basis von K^m ergänzt wird – im konkreten Beispiel oben also durch e_3.

Algorithmus 17.16 (Operationen mit Unterräumen)**.** Wir betrachten zwei durch erzeugende Vektoren gegebene Unterräume $U_1 = \operatorname{Lin}(x_1, \dots, x_k)$ und $U_2 = \operatorname{Lin}(y_1, \dots, y_l)$ von K^n. Mit Hilfe des Gauß-Verfahrens können wir dann die folgenden Konstruktionen explizit durchführen:

(a) (Basis eines Unterraums) Um eine Basis von U_1 zu bestimmen, berechnen wir einfach mit Algorithmus 17.14 eine Basis von $U_1 = \operatorname{Im}(x_1 \mid \dots \mid x_k)$.

(b) (Komplement von U_1, Quotientenraum K^n/U_1) Aufbauend auf (a) ergänzen wir die gefundene Basis von U_1 wie in Algorithmus 17.15 durch geeignete Einheitsvektoren zu einer Basis von K^n. Diese hinzugenommenen Einheitsvektoren bilden dann nach Bemerkung 15.11 eine Basis eines Komplements von U_1, ihre Klassen in K^n/U_1 nach Bemerkung 15.19 eine Basis von K^n/U_1.

(c) (Summe U_1+U_2) Nach Bemerkung 14.5 ist

$$U_1+U_2 = \mathrm{Lin}(x_1,\dots,x_k,y_1,\dots,y_l) = \mathrm{Im}\,(x_1 \mid \cdots \mid x_k \mid y_1 \mid \cdots \mid y_l\,).$$

Um eine Basis der Summe zu finden, können wir wie in (a) also wieder Algorithmus 17.14 anwenden.

(d) (Durchschnitt $U_1 \cap U_2$) Die Vektoren im Schnitt $U_1 \cap U_2$ erhält man offensichtlich durch Gleichsetzen

$$\lambda_1 x_1 + \cdots + \lambda_k x_k = -\mu_1 y_1 - \cdots - \mu_l y_l \tag{1}$$

der Elemente von U_1 und U_2, und damit durch Auflösen des linearen Gleichungssystems

$$\lambda_1 x_1 + \cdots + \lambda_k x_k + \mu_1 y_1 + \cdots + \mu_l y_l = 0 \tag{2}$$

(die Vorzeichen spielen hierbei keine Rolle, da mit $y \in U_2$ auch $-y \in U_2$ liegt, und sind daher so gewählt, dass sie sich im resultierenden Gleichungssystem wegheben). Die sich als Lösung ergebenden Werte für $\lambda_1,\dots,\lambda_k$ können dann in die linke Seite von (1) eingesetzt werden und liefern die gesuchten Vektoren im Schnitt.

Als konkretes Beispiel betrachten wir die Unterräume

$$U_1 = \mathrm{Lin}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right) \quad \text{und} \quad U_2 = \mathrm{Lin}\left(\begin{pmatrix}0\\1\\2\end{pmatrix}, \begin{pmatrix}1\\1\\3\end{pmatrix}\right)$$

von $\mathbb{R}^3$. Der Ansatz (2) führt zum folgenden Gleichungssystem in $\lambda_1,\lambda_2,\mu_1,\mu_2$:

$$\begin{pmatrix}1&0&0&1\\0&1&1&1\\0&1&2&3\end{pmatrix} \xrightarrow{Z3\to Z3-Z2} \begin{pmatrix}1&0&0&1\\0&1&1&1\\0&0&1&2\end{pmatrix} \xrightarrow{Z2\to Z2-Z3} \begin{pmatrix}1&0&0&1\\0&1&0&-1\\0&0&1&2\end{pmatrix}.$$

Der Lösungsraum ist also eindimensional und wird (z. B. nach Algorithmus 17.11) von $(\lambda_1,\lambda_2,\mu_1,\mu_2) = (1,-1,2,-1)$ erzeugt. Einsetzen von λ_1 und λ_2 (oder alternativ μ_1 und μ_2) in (1) zeigt also, dass der eindimensionale Unterraum $U_1 \cap U_2$ erzeugt wird von

$$\lambda_1 x_1 + \lambda_2 x_2 = 1 \cdot \begin{pmatrix}1\\0\\0\end{pmatrix} - 1 \cdot \begin{pmatrix}0\\1\\1\end{pmatrix} = \begin{pmatrix}1\\-1\\-1\end{pmatrix}.$$

38

Zum Abschluss dieses Kapitels wollen wir nun noch untersuchen, wie inverse Matrizen algorithmisch bestimmt werden können. Dazu benötigen wir zunächst das folgende oft nützliche Kriterium für Invertierbarkeit.

Lemma 17.17. *Für eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ sind äquivalent:*

(a) $A \in \mathrm{GL}(n,K)$.

(b) *Bringt man A durch elementare Zeilenumformungen auf reduzierte Zeilenstufenform, so ist das Ergebnis stets die Einheitsmatrix E_n.*

(c) *A ist ein Produkt von Elementarmatrizen.*

Beweis.

(a) $\Rightarrow$ (b): Ist $A \in \mathrm{GL}(n,K)$, so gilt $\mathrm{rk}A = n$ nach Bemerkung 16.39 (b). Bringt man A auf reduzierte Zeilenstufenform, so hat das Resultat nach Algorithmus 17.8 also n Stufen. Die einzige $n \times n$-Matrix in reduzierter Zeilenstufenform, die n Stufen hat, ist aber die Einheitsmatrix E_n.

(b) $\Rightarrow$ (c): Nach Voraussetzung können wir A mit elementaren Zeilenumformungen auf die Einheitsmatrix bringen, d. h. es gibt Elementarmatrizen $F_1,\dots,F_m$ mit $F_1\cdots F_m A = E_n$. Nach Bemerkung 17.4 (a) sind Elementarmatrizen aber invertierbar und ihre Inversen wieder Elementarmatrizen, und damit ist auch $A = F_m^{-1}\cdots F_1^{-1}$ ein Produkt von Elementarmatrizen.

(c) $\Rightarrow$ (a): Dies haben wir bereits in Bemerkung 17.4 (b) gesehen. $\square$

Algorithmus 17.18 (Inverse Matrix)**.** Gegeben sei eine quadratische Matrix $A \in \mathrm{Mat}(n\times n,K)$. Wir wollen überprüfen, ob A invertierbar ist, und in diesem Fall die inverse Matrix A^{-1} berechnen. Dazu starten wir mit der $n\times(2n)$-Matrix $(A\,|\,E_n)$, in der wir A und die Einheitsmatrix der gleichen Größe nebeneinander schreiben. Wir bringen dann die Matrix A mit dem Gauß-Verfahren auf reduzierte Zeilenstufenform, machen die dafür benötigten Zeilenumformungen aber in allen $2n$ Spalten der Matrix. Ist $F \in \mathrm{GL}(n,K)$ das Produkt der Elementarmatrizen, das den durchgeführten Umformungen entspricht, so erhalten wir also die Matrix $(FA\,|\,FE_n) = (FA\,|\,F)$, wobei in der linken Hälfte die reduzierte Zeilenstufenform FA von A steht.

Nach Lemma 17.17 ist A genau dann invertierbar, wenn für diese reduzierte Zeilenstufenform die Einheitsmatrix herausgekommen ist, wenn also $FA = E_n$ ist. In diesem Fall ist aber offensichtlich $F = A^{-1}$ die gesuchte inverse Matrix – und diese steht nach den Umformungen genau in der rechten Hälfte unseres Diagramms.

Als konkretes Beispiel wollen wir die Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} \quad \in \mathrm{Mat}(2\times 2,\mathbb{R})$$

wählen. Wir wenden also wie gewohnt das Gauß-Verfahren auf die Matrix A an, führen aber alle Umformungen mit der 2×4-Matrix $(A\,|\,E_2)$ durch:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \xrightarrow{Z2\to Z2-3Z1} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -3 & 1 \end{array}\right) \xrightarrow{Z1\to Z1-Z2} \left(\begin{array}{cc|cc} 1 & 0 & 4 & -1 \\ 0 & 1 & -3 & 1 \end{array}\right).$$

Da die herausgekommene reduzierte Zeilenstufenform von A (die linke Hälfte dieser Matrix) die Einheitsmatrix ist, ist A invertierbar. Die inverse Matrix steht nun in der rechten Hälfte des Diagramms: Es ist

$$A^{-1} = \begin{pmatrix} 4 & -1 \\ -3 & 1 \end{pmatrix}.$$

Aufgabe 17.19.

(a) Berechne Basen von $\mathrm{Im}A$ und $\mathrm{Ker}A$ für die reelle Matrix $A = \begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & 2 & -1 \end{pmatrix}$.

(b) Berechne den Rang der reellen Matrix $A = \begin{pmatrix} 0 & b & b & b \\ a & 0 & b & b \\ a & a & 0 & b \end{pmatrix}$ in Abhängigkeit von $a,b\in\mathbb{R}$.

Aufgabe 17.20. In $\mathbb{R}^4$ betrachten wir die Unterräume $U_1 = \mathrm{Lin}\left(\begin{pmatrix} 1 \\ 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 4 \\ -5 \\ 6 \end{pmatrix}\right)$ und $U_2 = \{x\in\mathbb{R}^4 : x_1 - x_2 - x_3 = x_1 + x_2 - 3x_4 = 0\}$, wobei $x_1,\dots,x_4$ wie üblich die Koordinaten von $x\in\mathbb{R}^4$ sind. Berechne Basen von U_1, U_2, U_1+U_2, $U_1\cap U_2$, und einem Komplement von U_1+U_2.

Aufgabe 17.21. Untersuche, ob die reellen Matrizen $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ und $B = \begin{pmatrix} 3 & 1 & 3 \\ -1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$ zueinander äquivalent sind, und finde gegebenenfalls invertierbare Matrizen S und T mit $B = SAT$.

Aufgabe 17.22 (Elementare Codierungstheorie)**.** In dieser Aufgabe sei $K = \mathbb{Z}_2$ der Körper mit 2 Elementen aus Beispiel 3.6 (b).

(a) Es sei $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \in \mathrm{Mat}(7 \times 4, K)$. Berechne eine Matrix $B \in \mathrm{Mat}(3 \times 7, K)$ mit $\mathrm{Im}A = \mathrm{Ker}B$.

(b) Zeige: Für alle $y \in K^7$ gibt es genau ein $x \in K^4$, für das sich y von Ax in höchstens einer Koordinate unterscheidet. (Hinweis: der Vektor $By \in K^3$ sollte erkennen lassen, wo dieser Unterschied liegt.)

Diese Aufgabe beinhaltet die Grundidee der sogenannten Codierungstheorie. Angenommen, wir wollen einen Vektor $x \in K^4$ (also 4 Bits) über eine gestörte Leitung an einen Empfänger übertragen. Wenn wir dann statt der 4 Bits x die 7 Bits Ax übertragen und beim Empfänger statt Ax fälschlicherweise y ankommt, kann dieser nach (b) die ursprüngliche Information x auch dann noch aus y rekonstruieren, wenn bei ihm ein Bit falsch angekommen ist. Auf jeder CD oder DVD sind die Daten in einer solchen Art codiert, damit z. B. durch Kratzer verursachte Lesefehler korrigiert werden können.

Aufgabe 17.23. Es seien $n \in \mathbb{N}_{>0}$ und $a_0, \ldots, a_{n-1} \in K$. Zeige, dass die Matrix

$$(a_j^i)_{0 \le i,j \le n-1} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_0 & a_1 & \cdots & a_{n-1} \\ \vdots & \vdots & & \vdots \\ a_0^{n-1} & a_1^{n-1} & \cdots & a_{n-1}^{n-1} \end{pmatrix}$$

genau dann invertierbar ist, wenn alle $a_0, \ldots, a_{n-1}$ verschieden sind.

Aufgabe 17.24 (Basiswechselmatrix für K^n)**.** Es seien $B = (x_1, \ldots, x_n)$ und $C = (y_1, \ldots, y_n)$ zwei Basen von K^n. Wir bilden daraus die Matrix $A = (\, y_1 \mid \cdots \mid y_n \mid x_1 \mid \cdots \mid x_n \,) \in \mathrm{Mat}(n \times 2n, K)$. Nun bringen wir die linke Hälfte dieser Matrix mit elementaren Zeilenumformungen auf reduzierte Zeilenstufenform, führen die Umformungen dabei aber wieder mit allen Spalten der Matrix durch.

Zeige, dass dann in der rechten Hälfte der Matrix genau die Basiswechselmatrix $A^{B,C}$ aus Definition 16.28 steht.

18. Determinanten

Nachdem wir nun schon recht ausführlich Matrizen und lineare Gleichungssysteme studiert haben, wollen wir jetzt die sogenannten Determinanten einführen, die beim Rechnen mit Matrizen ein unverzichtbares Hilfsmittel darstellen. Determinanten haben sehr viele schöne Eigenschaften und können demzufolge auch auf viele verschiedene Arten motiviert werden. Eine mögliche Herangehensweise ist, dass man nach einem einfachen Kriterium für die Invertierbarkeit quadratischer Matrizen sucht, so wie in dem folgenden einfachen Lemma für 2×2-Matrizen:

Lemma 18.1. *Eine 2×2-Matrix*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

über einem Körper K ist genau dann invertierbar, wenn $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$.

Beweis. Wir unterscheiden zwei Fälle:

1. Fall: Ist $a_{1,1} = 0$, so ist A genau dann invertierbar, wenn $a_{1,2} \neq 0$ und $a_{2,1} \neq 0$ gilt – denn wenn diese beiden Einträge ungleich Null sind, sind die beiden Spalten von A offensichtlich linear unabhängig (so dass dann $\operatorname{rk} A = 2$ gilt), während A andernfalls eine Nullzeile oder Nullspalte enthält und damit höchstens Rang 1 haben kann. Nach Bemerkung 16.39 (b) ist A also genau dann invertierbar, wenn $a_{1,2} \neq 0$ und $a_{2,1} \neq 0$, also wenn $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$.

2. Fall: Ist $a_{1,1} \neq 0$, so wenden wir Algorithmus 17.8 an, um $\operatorname{rk} A$ zu berechnen:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \xrightarrow{Z1 \to \frac{1}{a_{1,1}} Z1} \begin{pmatrix} 1 & \frac{a_{1,2}}{a_{1,1}} \\ a_{2,1} & a_{2,2} \end{pmatrix} \xrightarrow{Z2 \to Z2 - a_{2,1} Z1} \begin{pmatrix} 1 & \frac{a_{1,2}}{a_{1,1}} \\ 0 & a_{2,2} - \frac{a_{1,2}a_{2,1}}{a_{1,1}} \end{pmatrix}.$$

Diese Matrix hat genau dann Rang 2, ist also genau dann invertierbar, wenn $a_{2,2} - \frac{a_{1,2}a_{2,1}}{a_{1,1}} \neq 0$, d. h. wenn $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$ gilt. □

Die Zahl $a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$ werden wir später die *Determinante* $\det A$ von A nennen (weil sie „determiniert“, ob A invertierbar ist oder nicht).

Unser Ziel in diesem Kapitel ist es, Lemma 18.1 auf größere quadratische Matrizen zu verallgemeinern, also zu jeder Matrix $A \in \operatorname{Mat}(n \times n, K)$ eine Zahl $\det A \in K$ zu definieren, die ein Polynom in den Einträgen von A ist und (neben vielen anderen schönen Eigenschaften) genau dann ungleich Null ist, wenn A invertierbar ist.

18.A Konstruktion der Determinante

Leider ist eine direkte Angabe der Determinante einer quadratischen Matrix $A \in \operatorname{Mat}(n \times n, K)$ als polynomialer Ausdruck in den Einträgen von A so wie in Lemma 18.1 für allgemeines n zwar möglich (siehe Bemerkung 18.14), aber auch recht kompliziert. Wir wollen daher hier den für euch wahrscheinlich etwas ungewohnten Zugang wählen, die Determinante über ihre Eigenschaften zu definieren, d. h. als eine Funktion $A \mapsto \det A$ auf den $n \times n$-Matrizen, die eine gewisse „Wunschliste“ elementarer Eigenschaften erfüllt. Im Anschluss werden wir dann zeigen, dass unsere Wunschliste wirklich erfüllbar ist und die Determinante in der Tat auch eindeutig bestimmt.

Hier ist nun unsere Wunschliste:

Definition 18.2 (Determinante)**.** Es seien K ein Körper und $n \in \mathbb{N}_{>0}$ gegeben. Eine Abbildung $\det\colon \operatorname{Mat}(n \times n, K) \to K$ heißt **Determinante** (von $n \times n$-Matrizen), wenn gilt:

(a) („det ist multilinear“) Die Funktion det ist *linear in jeder Zeile*, d. h. für alle $k \in \{1,\dots,n\}$ und $\lambda \in K$ gilt

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_k + a_k' \\ \vdots \\ a_n \end{pmatrix} = \det\begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix} + \det\begin{pmatrix} a_1 \\ \vdots \\ a_k' \\ \vdots \\ a_n \end{pmatrix} \quad \text{und} \quad \det\begin{pmatrix} a_1 \\ \vdots \\ \lambda\, a_k \\ \vdots \\ a_n \end{pmatrix} = \lambda \cdot \det\begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix},$$

wobei $a_1,\dots,a_k,a_k',\dots,a_n \in \mathrm{Mat}(1\times n,K)$ die Zeilen der jeweiligen (quadratischen) Matrizen bezeichnen. (Halten wir also alle Zeilen bis auf die k-te fest, so haben wir genau eine lineare Abbildung in der k-ten Zeile im Sinne von Definition 13.16.)

(b) („det ist alternierend“) Stimmen zwei Zeilen von $A \in \mathrm{Mat}(n\times n,K)$ überein, so ist $\det A = 0$.

(c) („det ist normiert“) Es gilt $\det(E_n) = 1$.

Beispiel 18.3. Die Funktion

$$\det\colon\ \mathrm{Mat}(2\times 2,K) \to K,\ \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \mapsto a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$$

aus Lemma 18.1 ist eine Determinante:

(a) det ist multilinear: Die Additivität in der ersten Zeile ergibt sich z. B. aus der Rechnung

$$\begin{aligned}\det\begin{pmatrix} a_{1,1}+a_{1,1}' & a_{1,2}+a_{1,2}' \\ a_{2,1} & a_{2,2} \end{pmatrix} &= (a_{1,1}+a_{1,1}')\,a_{2,2} - (a_{1,2}+a_{1,2}')\,a_{2,1} \\ &= a_{1,1}a_{2,2} - a_{1,2}a_{2,1} + a_{1,1}'a_{2,2} - a_{1,2}'a_{2,1} \\ &= \det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} + \det\begin{pmatrix} a_{1,1}' & a_{1,2}' \\ a_{2,1} & a_{2,2} \end{pmatrix};\end{aligned}$$

die anderen Linearitätseigenschaften folgen natürlich genauso.

(b) det ist alternierend: Sind die beiden Zeilen der Matrix gleich, so ist

$$\det\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{1,1} & a_{1,2} \end{pmatrix} = a_{1,1}a_{1,2} - a_{1,2}a_{1,1} = 0.$$

(c) det ist normiert, denn natürlich ist $\det(E_2) = 1$.

Bemerkung 18.4.

(a) Wir werden in Folgerung 18.8 und Satz 18.12 sehen, dass es zu jedem Körper K und jedem $n \in \mathbb{N}_{>0}$ in der Tat genau eine Determinante $\det\colon\ \mathrm{Mat}(n\times n,K) \to K$ gibt, dass Definition 18.2 die Determinante also widerspruchsfrei und eindeutig festlegt. Solange wir dies noch nicht gezeigt haben, sollten wir aber korrekterweise immer von *einer* Determinante (und nicht von *der* Determinante) sprechen.

(b) Enthält A eine Nullzeile, so können wir aus dieser Zeile den Faktor 0 herausziehen und erhalten aus der Linearitätseigenschaft in dieser Zeile sofort, dass dann $\det A = 0$ sein muss.

(c) Aus Eigenschaft (b) der Definition 18.2 einer Determinante folgt, dass sich $\det A$ beim Vertauschen zweier Zeilen mit -1 multipliziert, also genau das Vorzeichen ändert (daher kommt auch der Name „alternierend“ für diese Eigenschaft): Für alle $k,l \in \{1,\dots,n\}$ mit $k \neq l$ ergibt

sich zusammen mit der Multilinearität nämlich

$$\underbrace{\det\begin{pmatrix}\vdots\\ a_k+a_l\\ \vdots\\ a_k+a_l\\ \vdots\end{pmatrix}}_{=0} = \det\begin{pmatrix}\vdots\\ a_k\\ \vdots\\ a_k+a_l\\ \vdots\end{pmatrix} + \det\begin{pmatrix}\vdots\\ a_l\\ \vdots\\ a_k+a_l\\ \vdots\end{pmatrix}$$

$$= \underbrace{\det\begin{pmatrix}\vdots\\ a_k\\ \vdots\\ a_k\\ \vdots\end{pmatrix}}_{=0} + \det\begin{pmatrix}\vdots\\ a_k\\ \vdots\\ a_l\\ \vdots\end{pmatrix} + \det\begin{pmatrix}\vdots\\ a_l\\ \vdots\\ a_k\\ \vdots\end{pmatrix} + \underbrace{\det\begin{pmatrix}\vdots\\ a_l\\ \vdots\\ a_l\\ \vdots\end{pmatrix}}_{=0},$$

und damit

$$\det\begin{pmatrix}\vdots\\ a_k\\ \vdots\\ a_l\\ \vdots\end{pmatrix} = -\det\begin{pmatrix}\vdots\\ a_l\\ \vdots\\ a_k\\ \vdots\end{pmatrix}.$$

(d) Analog zu (c) wollen wir jetzt untersuchen, was mit einer Determinante passiert, wenn wir in einer Matrix A für gegebenes $k \in \{1,\dots,n\}$ die k-te Zeile unter Beibehaltung der Reihenfolge der anderen Zeilen ganz nach oben schieben. Wir können dies wie folgt durch $k-1$ Vertauschungen zweier benachbarter Zeilen erreichen:

$$A = \begin{pmatrix}\vdots\\ a_{k-2}\\ a_{k-1}\\ a_k\\ a_{k+1}\\ \vdots\end{pmatrix} \longrightarrow \begin{pmatrix}\vdots\\ a_{k-2}\\ a_k\\ a_{k-1}\\ a_{k+1}\\ \vdots\end{pmatrix} \longrightarrow \cdots \longrightarrow \begin{pmatrix}a_k\\ a_1\\ \vdots\\ a_{k-1}\\ a_{k+1}\\ \vdots\end{pmatrix}.$$

Da sich bei jeder dieser Vertauschungen nach (c) das Vorzeichen der Determinante ändert, ändert das gesamte Verschieben der k-ten Zeile ganz nach oben die Determinante von A also um einen Faktor $(-1)^{k-1}$.

Um die weiteren Eigenschaften von Determinanten zu untersuchen, beginnen wir zunächst mit den Elementarmatrizen.

Lemma 18.5 (Determinanten von Elementarmatrizen)**.** *Es sei* $\det\colon \mathrm{Mat}(n\times n,K) \to K$ *eine Determinante. Dann gilt für alle* $A \in \mathrm{Mat}(n\times n,K)$ *sowie für alle* $n\times n$*-Elementarmatrizen* $F_k(\lambda)$ *und* $F_{k,l}(\lambda)$ *aus Konstruktion 17.3:*

(a) $\det(F_k(\lambda)\cdot A) = \lambda \det A$.

(b) $\det(F_{k,l}(\lambda)\cdot A) = \det A$.

Insbesondere gilt also $\det F_k(\lambda) = \lambda$ *und* $\det F_{k,l}(\lambda) = 1$*, und damit* $\det(FA) = \det F \cdot \det A$ *für jede Elementarmatrix* F *und jede beliebige quadratische Matrix* A*.*

Beweis. Es seien $a_1,\dots,a_n \in \mathrm{Mat}(1\times n,K)$ die Zeilen von A. Nach Konstruktion 17.3 entspricht eine Multiplikation von A mit einer Elementarmatrix von links genau einer elementaren Zeilenumformung. Damit erhalten wir mit den Eigenschaften (a) und (b) aus Definition 18.2

$$\det(F_k(\lambda)\cdot A) = \det\begin{pmatrix} \vdots \\ \lambda\, a_k \\ \vdots \end{pmatrix} = \lambda \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \end{pmatrix} = \lambda \det A$$

und

$$\det(F_{k,l}(\lambda)\cdot A) = \det\begin{pmatrix} \vdots \\ a_k+\lambda\, a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix} = \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} + \lambda \underbrace{\det\begin{pmatrix} \vdots \\ a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix}}_{=0} = \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} = \det A,$$

was die beiden Teile des Lemmas zeigt. Die Determinanten der Elementarmatrizen erhält man daraus für $A = E_n$. □

Aus diesem einfachen Lemma folgt nun bereits die wahrscheinlich wichtigste Eigenschaft von Determinanten:

Satz 18.6 (**Produktsatz** für Determinanten)**.** *Es seien* $\det\colon \mathrm{Mat}(n\times n,K)\to K$ *eine Determinante und* $A,B\in\mathrm{Mat}(n\times n,K)$. *Dann gilt:*

(a) $\det(AB) = \det A\cdot\det B$.

(b) *A ist genau dann invertierbar, wenn* $\det A\neq 0$. *In diesem Fall ist* $\det(A^{-1}) = \frac{1}{\det A}$.

Beweis. Wir unterscheiden zwei Fälle:

1. Fall: A ist invertierbar. Dann ist $A = F_1\cdot\dots\cdot F_k$ nach Lemma 17.17 ein Produkt von Elementarmatrizen. Durch k-fache Anwendung von Lemma 18.5 erhält man dann

$$\begin{aligned} \det(AB) &= \det(F_1\cdot\dots\cdot F_k\cdot B) &&= \det F_1\cdot\dots\cdot\det F_k\cdot\det B \\ \text{sowie}\qquad \det A &= \det(F_1\cdot\dots\cdot F_k) &&= \det F_1\cdot\dots\cdot\det F_k, \end{aligned}$$

und damit wie behauptet $\det(AB) = \det A\cdot\det B$. Setzt man hier $B = A^{-1}$ ein, so ergibt sich insbesondere $\det A\cdot\det A^{-1} = \det(AA^{-1}) = \det E_n = 1$, d. h. es ist $\det A\neq 0$ und $\det(A^{-1}) = \frac{1}{\det A}$.

2. Fall: A ist nicht invertierbar. Dann bringen wir A mit einem Produkt F von Elementarmatrizen auf Zeilenstufenform FA. Nach Bemerkung 16.39 (b) und Algorithmus 17.8 hat diese Zeilenstufenform weniger als n Stufen und damit am Ende (mindestens) eine Nullzeile. Also ist $\det(FA) = 0$ nach Bemerkung 18.4 (b). Da F als Produkt von Elementarmatrizen nach Lemma 17.17 invertierbar ist, bedeutet dies nach dem bereits gezeigten 1. Fall auch $\det F\cdot\det A = 0$, wegen $\det F\neq 0$ also $\det A = 0$.

Mit FA hat aber auch FAB eine Nullzeile. Damit folgt genauso wie oben auch $\det(AB) = 0$, also insbesondere $\det(AB) = \det A\cdot\det B$. □

Bemerkung 18.7. Im Gegensatz zu Produkten gibt es *keine* Formel für die Determinante $\det(A+B)$ einer Summe von zwei Matrizen – insbesondere ist im Allgemeinen $\det(A+B)\neq\det A+\det B$!

Als Folgerung aus dem Produktsatz können wir nun bereits beweisen, dass die Eigenschaften aus Definition 18.2 eine Determinante eindeutig festlegen.

Folgerung 18.8 (Eindeutigkeit der Determinante)**.** *Zu jedem Körper K und $n\in\mathbb{N}_{>0}$ gibt es höchstens eine Determinante* $\det\colon \mathrm{Mat}(n\times n,K)\to K$.

Beweis. Es sei $A \in \operatorname{Mat}(n \times n, K)$. Ist A nicht invertierbar, so ist nach Satz 18.6 notwendigerweise $\det A = 0$. Andernfalls ist $A = F_1 \cdot \dots \cdot F_k$ nach Lemma 17.17 ein Produkt von Elementarmatrizen, und damit ist nach Satz 18.6 (a)

$$\det A = \det F_1 \cdot \dots \cdot \det F_k.$$

Da die Determinante der Elementarmatrizen nach Lemma 18.5 aber durch Definition 18.2 eindeutig bestimmt ist, ist damit auch $\det A$ durch diese Definition eindeutig festgelegt. □

Auf ganz ähnliche Art wollen wir nun zeigen, dass sich eine Determinante beim Transponieren der Matrizen nicht ändert.

Folgerung 18.9. *Ist $A \in \operatorname{Mat}(n \times n, K)$ und* $\det\colon \operatorname{Mat}(n \times n, K) \to K$ *eine Determinante, so gilt* $\det(A^\mathsf{T}) = \det A$.

Beweis. Ist A nicht invertierbar, also $\operatorname{rk} A < n$, so ist nach Folgerung 16.44 auch A^T nicht invertierbar, und damit ist $\det(A^\mathsf{T}) = 0 = \det A$ nach Satz 18.6 (b).

Andernfalls ist $A = F_1 \cdot \dots \cdot F_k$ nach Lemma 17.17 wieder ein Produkt von Elementarmatrizen. Da die zu zeigende Aussage für Elementarmatrizen aus Lemma 18.5 offensichtlich ist (es ist nämlich $(F_k(\lambda))^\mathsf{T} = F_k(\lambda)$ und $(F_{k,l}(\lambda))^\mathsf{T} = F_{l,k}(\lambda)$), folgt somit nach Lemma 16.7 (d) und Satz 18.6 (a)

$$\det(A^\mathsf{T}) = \det((F_1 \cdots F_k)^\mathsf{T}) = \det(F_k^\mathsf{T} \cdots F_1^\mathsf{T}) = \det(F_k^\mathsf{T}) \cdots \det(F_1^\mathsf{T}) = \det F_1 \cdots \det F_k = \det A. \qquad \square$$

Bemerkung 18.10. Folgerung 18.9 besagt anschaulich, dass alle Eigenschaften, die für die Zeilen einer Determinante gelten, analog auch für die Spalten gelten. So ist eine Determinante z. B. auch linear in jeder Spalte (vgl. Definition 18.2 (a)) und ändert ihr Vorzeichen beim Vertauschen zweier Spalten (vgl. Bemerkung 18.4 (c)).

Um sicherzustellen, dass wir mit Definition 18.2 keine in sich widersprüchliche Wunschliste aufgeschrieben haben, kommen wir nun aber endlich zum bereits angekündigten Resultat, dass eine Determinante mit den geforderten Eigenschaften auch wirklich existiert. Wir werden die Funktionen $\det\colon \operatorname{Mat}(n \times n, K) \to K$ rekursiv über n definieren und verwenden dazu die folgende Konstruktion, um Matrizen der Größe n auf solche der Größe $n-1$ zurückzuführen.

Definition 18.11 (Streichungsmatrix)**.** Zu $A = (a_{i,j})_{i,j} \in \operatorname{Mat}(n \times n, K)$ sowie $k, l \in \{1, \dots, n\}$ sei

$$A'_{k,l} := \begin{pmatrix} a_{1,1} & \cdots & a_{1,l-1} & a_{1,l+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k-1,1} & \cdots & a_{k-1,l-1} & a_{k-1,l+1} & \cdots & a_{k-1,n} \\ a_{k+1,1} & \cdots & a_{k+1,l-1} & a_{k+1,l+1} & \cdots & a_{k+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,l-1} & a_{n,l+1} & \cdots & a_{n,n} \end{pmatrix} \in \operatorname{Mat}((n-1) \times (n-1), K)$$

die Matrix, die man erhält, wenn man aus A die k-te Zeile und l-te Spalte herausstreicht. Wir bezeichnen diese Matrizen als **Streichungsmatrizen** zu A.

Satz 18.12 (Existenz der Determinante)**.** *Für alle $n \in \mathbb{N}_{>0}$ definieren wir* $\det\colon \operatorname{Mat}(n \times n, K) \to K$ *rekursiv über n durch die folgende Vorschrift:*

- *Für $n = 1$ setzen wir* $\det(a_{1,1}) := a_{1,1}$.
- *Für $n > 1$ setzen wir*

$$\det A := \sum_{k=1}^{n} (-1)^{k+1} a_{k,1} \det A'_{k,1},$$

wobei wie üblich $a_{k,1}$ die Einträge der ersten Spalte von A und $A'_{k,1}$ die zu diesen Einträgen gehörigen Streichungsmatrizen sind.

Dann ist det *eine (und damit nach Folgerung 18.8 „die“) Determinante für alle n.*

39

Bevor wir diesen Satz beweisen, wollen wir uns ein paar Beispiele anschauen, um die angegebene rekursive Formel besser zu verstehen.

Beispiel 18.13 (Determinante von 2×2- und 3×3-Matrizen)**.**

(a) Für $n=2$ besagt die Formel aus Satz 18.12

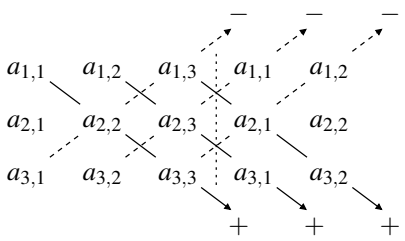
$$\begin{aligned}\det\begin{pmatrix} a_{1,1} & a_{1,2}\\ a_{2,1} & a_{2,2}\end{pmatrix} &= (-1)^{1+1}a_{1,1}\det(a_{2,2}) + (-1)^{2+1}a_{2,1}\det(a_{1,2})\\ &= a_{1,1}a_{2,2} - a_{2,1}a_{1,2}\end{aligned}$$

und reproduziert damit die Formel aus Lemma 18.1.

(b) Für $n=3$ ergibt sich unter Benutzung des Ergebnisses aus (a)

$$\begin{aligned}\det\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3}\\ a_{2,1} & a_{2,2} & a_{2,3}\\ a_{3,1} & a_{3,2} & a_{3,3}\end{pmatrix} &= (-1)^{1+1}a_{1,1}\det\begin{pmatrix} a_{2,2} & a_{2,3}\\ a_{3,2} & a_{3,3}\end{pmatrix} + (-1)^{2+1}a_{2,1}\det\begin{pmatrix} a_{1,2} & a_{1,3}\\ a_{3,2} & a_{3,3}\end{pmatrix}\\ &\quad + (-1)^{3+1}a_{3,1}\det\begin{pmatrix} a_{1,2} & a_{1,3}\\ a_{2,2} & a_{2,3}\end{pmatrix}\\ &= a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{2,3}a_{3,2} - a_{2,1}a_{1,2}a_{3,3} + a_{2,1}a_{1,3}a_{3,2}\\ &\quad + a_{3,1}a_{1,2}a_{2,3} - a_{3,1}a_{1,3}a_{2,2}.\end{aligned}$$

Am einfachsten kann man sich diese Formel nach der sogenannten **Regel von Sarrus** merken: Bilden wir die 3×5-Matrix, in der wir neben der Matrix A die beiden ersten Spalten noch einmal wiederholen, so ergeben sich die 6 Terme der Determinante mit ihren Vorzeichen aus dem folgenden Schema:



Beachte aber, dass diese einfache Merkregel *nur für* $n=3$ gilt – für größere n ist der komplett ausmultiplizierte Ausdruck für $\det A$ deutlich komplizierter (und für konkrete numerische Berechnungen in der Tat auch nicht mehr geeignet).

Bemerkung 18.14. Diejenigen von euch, die aus der Parallelvorlesung „Algebraische Strukturen“ die symmetrische Gruppe S_n aller Permutationen von $\{1,\dots,n\}$ kennen [G, Kapitel 2], können die Formel für die Determinante auch nicht-rekursiv als

$$\det A = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1,\sigma(1)}\cdot\dots\cdot a_{n,\sigma(n)} \qquad (*)$$

hinschreiben. Man sieht an dieser Darstellung also, dass die Determinante aus einer Summe von $n!$ Termen besteht. Dabei ist jeder Term ein Produkt von genau n Einträgen von A, und zwar aus jeder Zeile und jeder Spalte genau einem. Aufsummiert wird über alle Möglichkeiten, n Einträge von A eben gerade so auszuwählen, dass man aus jeder Zeile und Spalte einen Eintrag genommen hat. Die Vorzeichen der einzelnen Terme sind immer genau das Vorzeichen der entsprechenden Permutation.

Wir werden die Formel $(*)$ in dieser Vorlesung aber nicht benötigen und daher auch nicht beweisen, dass sie wirklich mit der rekursiven Definition aus Satz 18.12 übereinstimmt bzw. die Eigenschaften von Definition 18.2 erfüllt.

Wir kommen nun aber endlich zum Beweis des Existenzsatzes 18.12.

Beweis von Satz 18.12. Wir überprüfen die drei Eigenschaften aus Definition 18.2 mit Induktion über n. Für $n = 1$ sind alle Aussagen klar. Wir können also annehmen, dass $n > 1$ ist und wir die Eigenschaften von Definition 18.2 für Matrizen der Größe $n-1$ bereits gezeigt haben; wir müssen sie nun für Matrizen der Größe n zeigen.

det ist multilinear: Der Ausdruck $a_{1,1}\det A'_{1,1}$ ist linear in der ersten Zeile, da $a_{1,1}$ natürlich linear in der ersten Zeile ist und $A'_{1,1}$ nicht von der ersten Zeile abhängt. Die Ausdrücke $a_{k,1}\det A'_{k,1}$ für $k > 1$ sind ebenfalls linear in der ersten Zeile, da $a_{k,1}$ nicht von der ersten Zeile abhängt und $\det A'_{k,1}$ nach Induktionsvoraussetzung linear in der ersten Zeile ist. Damit ist auch $\det A$ als Linearkombination dieser Ausdrücke linear in der ersten Zeile. Die Linearität in den anderen Zeilen folgt natürlich analog.

det ist alternierend: Wir bezeichnen die Zeilen von A mit $a_1,\dots,a_n \in \mathrm{Mat}(1\times n,K)$. Weiterhin seien $a'_1,\dots,a'_n \in \mathrm{Mat}(1\times(n-1),K)$ die Zeilen von A, bei denen man jeweils den ersten Eintrag herausgestrichen hat. Wir nehmen nun an, dass zwei Zeilen a_i und a_j von A übereinstimmen, und müssen zeigen, dass $\det A = 0$ folgt. Ohne Beschränkung der Allgemeinheit sei dazu $i > j$.

Beachte, dass dann auch in den Streichungsmatrizen $A'_{k,1}$ mit $k \neq i$ und $k \neq j$, bei denen wir also weder die i-te noch die j-te Zeile herausgestrichen haben, jeweils zwei Zeilen übereinstimmen. Nach Induktionsvoraussetzung ist die Determinante aller dieser Streichungsmatrizen gleich 0, und damit bleibt in der rekursiven Formel für $\det A$ nur der Ausdruck

$$\det A = (-1)^{i+1} a_{i,1}\det A'_{i,1} + (-1)^{j+1} a_{j,1}\det A'_{j,1} \qquad (*)$$

übrig. Nun können wir

$$\text{sowohl}\quad A'_{i,1} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_{j-1} \\ a'_j \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_{i+1} \\ \vdots \end{pmatrix} \quad \text{als auch}\quad A'_{j,1} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_{j-1} \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_i \\ a'_{i+1} \\ \vdots \end{pmatrix} \quad \text{auf die Form}\quad A' := \begin{pmatrix} a'_i \\ a'_1 \\ \vdots \\ a'_{j-1} \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_{i+1} \\ \vdots \end{pmatrix}$$

bringen, indem wir die Zeile a'_j bzw. a'_i unter Beibehaltung der Reihenfolge der anderen Zeilen ganz nach oben schieben. Da det für Matrizen der Größe $n-1$ nach Induktionsvoraussetzung eine Determinante ist, ändern sich dadurch die Vorzeichen von $\det A'_{i,1}$ und $\det A'_{j,1}$ wie in Bemerkung 18.4 (d): Da wir in $A'_{i,1}$ die Zeile mit der Nummer j, in $A'_{j,1}$ jedoch die Zeile mit der Nummer $i-1$ nach oben schieben (im letzteren Fall fehlt ja die Zeile a'_j oberhalb von a'_i), ist also

$$\det A'_{i,1} = (-1)^{j-1}\det A' \quad \text{und} \quad \det A'_{j,1} = (-1)^{i-2}\det A'$$

und damit nach $(*)$

$$\det A = (-1)^{i+j} a_{i,1}\det A' + (-1)^{i+j-1} a_{j,1}\det A' = 0$$

wegen $a_{i,1} = a_{j,1}$.

det ist normiert: In der ersten Spalte der Einheitsmatrix sind natürlich der erste Eintrag gleich 1 und alle anderen gleich 0. Weiterhin ist die Streichungsmatrix des Eintrags links oben gerade E_{n-1}. Also folgt sofort

$$\det E_n = (-1)^{1+1}\cdot 1\cdot \det E_{n-1} = 1.$$

Damit ist alles gezeigt. □

Insgesamt haben wir jetzt also gesehen, dass es für alle Körper K und $n \in \mathbb{N}$ genau eine Determinante $\det\colon \mathrm{Mat}(n\times n,K) \to K$ gibt. In Zukunft werden wir daher immer von *der* Determinante quadratischer Matrizen sprechen.

18.B Eigenschaften der Determinante

Im letzten Abschnitt haben wir die Determinante quadratischer Matrizen definiert und auch bereits ihre ersten wichtigen Eigenschaften gesehen. Wir wollen diese Untersuchung der Determinante jetzt fortsetzen und uns dabei als Erstes um ihre praktische Berechnung kümmern. In der Tat ist hierfür die rekursive Formel aus Satz 18.12 bereits sehr nützlich. Wir können sie allerdings noch etwas erweitern, denn dort ist ja momentan die erste Spalte der Matrix ausgezeichnet – obwohl aufgrund von Definition 18.2 natürlich klar sein sollte, dass die erste Spalte der Matrix keine besondere Rolle spielt. Wir sollten eine ähnliche Rekursionsformel also auch für die anderen Spalten (und aufgrund von Folgerung 18.9 in der Tat auch für die Zeilen) erwarten können. Dies besagt der folgende Satz.

Satz 18.15 (**Laplacescher Entwicklungssatz**). *Es sei* $A \in \mathrm{Mat}(n \times n, K)$.

(a) *Für alle* $l \in \{1, \dots, n\}$ *gilt* $\det A = \sum_{k=1}^{n} (-1)^{k+l} \cdot a_{k,l} \cdot \det A'_{k,l}$.

(b) *Für alle* $k \in \{1, \dots, n\}$ *gilt* $\det A = \sum_{l=1}^{n} (-1)^{k+l} \cdot a_{k,l} \cdot \det A'_{k,l}$.

Benutzt man diese Formeln, so sagt man auch, dass man die Determinante von A nach der l-ten Spalte bzw. k-ten Zeile entwickelt.

Beweis.

(a) Es sei $B = (b_{i,j})_{i,j}$ die Matrix, die man aus A erhält, indem man die Spalte l unter Beibehaltung der Reihenfolge der anderen Spalten ganz nach links schiebt. Nach den Bemerkungen 18.4 (d) und 18.10 ist dann $\det A = (-1)^{l-1} \det B$. Andererseits ist natürlich $b_{k,1} = a_{k,l}$ und $B'_{k,1} = A'_{k,l}$ für alle $k \in \{1, \dots, n\}$. Damit folgt wie behauptet nach Satz 18.12 angewendet auf B

$$\det A = (-1)^{l-1} \det B = (-1)^{l-1} \sum_{k=1}^{n} (-1)^{k+1} b_{k,1} \det B'_{k,1} = \sum_{k=1}^{n} (-1)^{k+l} a_{k,l} \det A'_{k,l}.$$

(b) Dies ergibt sich mit Bemerkung 18.10 sofort aus (a). □

Beispiel 18.16 (Berechnung von Determinanten). Die Entwicklung nach Laplace ist oft die geschickteste Art, die Determinante einer Matrix A konkret zu berechnen – insbesondere wenn man nach einer Spalte oder Zeile entwickeln kann, in der bereits viele Einträge gleich 0 sind, so dass die entsprechenden Terme in der Summe wegfallen. In der Praxis empfiehlt es sich daher, zunächst mit elementaren Spalten- oder Zeilenumformungen eine Spalte oder Zeile zu erzeugen, in der nur ein Eintrag ungleich Null ist, und dann nach dieser Spalte bzw. Zeile zu entwickeln. Beachte, dass die Determinante dabei nach Lemma 18.5 ...

- mit λ multipliziert wird, wenn wir eine Spalte oder Zeile mit λ multiplizieren; und
- unverändert bleibt, wenn wir ein Vielfaches einer Spalte bzw. Zeile zu einer anderen addieren.

Hier ist ein Beispiel, bei dem wir der Reihe nach die erste von der dritten Spalte subtrahieren, nach der dritten Spalte entwickeln, und noch einmal nach der zweiten Zeile entwickeln: Es ist

$$\det \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 3 \end{pmatrix} = (-1)^{3+3} \cdot 2 \cdot \det \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 3 \end{pmatrix} = 2 \cdot (-1)^{2+1} \cdot 1 \cdot \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} = -4.$$

Ein besonders einfacher Fall – der aber dennoch häufig vorkommt – sind die sogenannten Dreiecksmatrizen, bei denen oberhalb oder unterhalb der Diagonalen nur Nullen stehen.

Definition 18.17 (Dreiecksmatrizen)**.** Eine quadratische Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$ heißt **obere Dreiecksmatrix**, falls $a_{i,j} = 0$ für alle $i > j$ gilt, und **untere Dreiecksmatrix**, falls $a_{i,j} = 0$ für alle $i < j$ gilt. Obere bzw. untere Dreiecksmatrizen haben also die Form

$$\begin{pmatrix} a_{1,1} & & * \\ & \ddots & \\ 0 & & a_{n,n} \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} a_{1,1} & & 0 \\ & \ddots & \\ * & & a_{n,n} \end{pmatrix}.$$

Sind zusätzlich noch alle Einträge $a_{i,i}$ auf der Diagonalen gleich Null, so heißt A **echte (obere bzw. untere) Dreiecksmatrix**.

Folgerung 18.18 (Determinante von Dreiecksmatrizen)**.** *Ist $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$ eine (obere oder untere) Dreiecksmatrix, so ist ihre Determinante gleich dem Produkt ihrer Einträge auf der Diagonalen*

$$\det A = a_{1,1} \cdot \dots \cdot a_{n,n}.$$

Beweis. Da untere Dreiecksmatrizen beim Transponieren in obere übergehen, reicht es nach Folgerung 18.9, die Aussage für obere Dreiecksmatrizen zu zeigen. Wir beweisen die Aussage in diesem Fall mit Induktion über n; der Fall $n = 1$ ist dabei trivial. Für $n > 1$ entwickeln wir $\det A$ gemäß Satz 18.15 nach der 1. Spalte: Da hier nur der erste Eintrag ungleich Null ist, ergibt sich sofort nach Induktionsvoraussetzung

$$\det A = (-1)^{1+1} a_{1,1} \det A'_{1,1} = a_{1,1} \cdot (a_{2,2} \cdot \dots \cdot a_{n,n}),$$

da auch $A'_{1,1} \in \mathrm{Mat}((n-1) \times (n-1), K)$ eine obere Dreiecksmatrix (mit Diagonaleinträgen $a_{2,2}, \dots, a_{n,n}$) ist. □

Aufgabe 18.19.

(a) Berechne die Determinante sowie die inverse Matrix von

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \quad \in \mathrm{Mat}(3 \times 3, \mathbb{R}).$$

(b) Für $a_1, \dots, a_n \in K \backslash \{0\}$ zeige man

$$\det \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & a_1 & 0 & \cdots & 0 \\ 1 & 0 & a_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & a_n \end{pmatrix} = -\left(\prod_{i=1}^n a_i\right) \cdot \left(\sum_{i=1}^n \frac{1}{a_i}\right).$$

Wir wollen nun noch zwei Ergebnisse zu Determinanten beweisen, die mehr aus theoretischer als aus rechentechnischer Sicht interessant sind. Das erste betrifft inverse Matrizen: Ist A eine invertierbare Matrix, so haben wir in Algorithmus 17.18 ja bereits gesehen, wie man A^{-1} konkret numerisch berechnen kann. Mit Hilfe von Determinanten können wir nun auch eine explizite Formel für A^{-1} angeben – die allerdings den Nachteil hat, dass sie bei konkreten Berechnungen relativ aufwändig ist, weil für jeden Eintrag von A^{-1} eine eigene Determinante berechnet werden muss.

Satz 18.20 (Explizite Formel für die inverse Matrix)**.** *Es sei $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$.*

(a) *Ist $C = (c_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$ die Matrix mit Einträgen*

$$c_{i,j} = (-1)^{i+j} \det A'_{j,i}$$

(beachte die Vertauschung von Spalten- und Zeilenindizes bei der Streichungsmatrix!), so ist $CA = AC = (\det A) \cdot E_n$.

(b) *Ist A invertierbar, so ist die inverse Matrix von A gegeben durch*

$$A^{-1} = \frac{1}{\det A} \cdot C = \left((-1)^{i+j} \frac{\det A'_{j,i}}{\det A} \right)_{i,j}$$

mit C wie in (a).

Beweis. Für alle $i,k = 1,\dots,n$ überprüfen wir den (i,k)-Eintrag des Matrixprodukts CA: Nach Definition 16.5 ist dies

$$\sum_{j=1}^{n} c_{i,j} a_{j,k} = \sum_{j=1}^{n} (-1)^{i+j} a_{j,k} \det A'_{j,i} \overset{18.15}{=} \det \begin{array}{c} \text{Spalte } i \\ \downarrow \\ \begin{pmatrix} a_{1,1} & \cdots & a_{1,k} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,k} & \cdots & a_{n,n} \end{pmatrix} \end{array},$$

wobei die zweite Gleichung genau die Entwicklung nach Spalte i ist, und die Matrix auf der rechten Seite aus A entsteht, indem die Einträge aus Spalte k auch in Spalte i geschrieben werden. Die Determinante dieser Matrix ist aber 0 für $i \neq k$ (da dann zwei gleiche Spalten existieren) und $\det A$ für $i = k$ (denn dann ist diese Matrix gleich A). Damit ist $CA = (\det A) E_n$. Analog zeigt man auch $AC = (\det A) E_n$ und damit Teil (a). Die Formel in (b) folgt daraus natürlich sofort mit Division durch $\det A$. □

Beispiel 18.21. Für eine 2×2-Matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

hat die Matrix C aus Satz 18.20 die Einträge

$$c_{1,1} = (-1)^{1+1} \det(a_{2,2}) = a_{2,2}, \qquad c_{1,2} = (-1)^{1+2} \det(a_{1,2}) = -a_{1,2},$$

und genauso $c_{2,1} = -a_{2,1}$ und $c_{2,2} = a_{1,1}$. Damit ist nach Satz 18.20 (b) im Fall einer invertierbaren Matrix also

$$A^{-1} = \frac{1}{\det A} \cdot \begin{pmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{pmatrix}.$$

Eine konkrete Anwendung von Satz 18.20 ergibt sich bei der Lösung linearer Gleichungssysteme: Sind $A \in \mathrm{GL}(n,K)$ eine invertierbare Matrix und $b \in K^n$, so wissen wir bereits, dass das Gleichungssystem $Ax = b$ für x die eindeutige Lösung $x = A^{-1}b$ hat. Da wir gerade mit Hilfe von Determinanten eine explizite Formel für die inverse Matrix A^{-1} gefunden haben, überrascht es nicht, dass wir auch für die Koordinaten dieses Lösungsvektors $x = A^{-1}b$ eine ähnliche explizite Formel herleiten können:

Satz 18.22 (Cramersche Regel). *Es seien $A \in \mathrm{GL}(n,K)$ und $b \in K^n$. Wir bezeichnen die Spalten von A mit $a_1,\dots,a_n \in K^n$. Dann ist die (nach Bemerkung 17.2 (b) eindeutige) Lösung des Gleichungssystems $Ax = b$ der Vektor $x \in K^n$ mit den Komponenten*

$$x_i = \frac{\det(a_1 \mid \cdots \mid a_{i-1} \mid b \mid a_{i+1} \mid \cdots \mid a_n)}{\det A}$$

für $i = 1,\dots,n$.

Beweis. Nach Satz 18.20 (b) und Definition 16.5 der Matrixmultiplikation ist x_i, also die i-te Komponente des Matrixprodukts $A^{-1}b$, gleich

$$x_i = \sum_{j=1}^{n} (-1)^{i+j} \frac{\det A'_{j,i}}{\det A} \cdot b_j \overset{18.15}{=} \frac{1}{\det A} \det(a_1 \mid \cdots \mid a_{i-1} \mid b \mid a_{i+1} \mid \cdots \mid a_n),$$

wobei die zweite Gleichheit die Entwicklung nach der i-ten Spalte ist. □

Beispiel 18.23. Wir wollen das lineare Gleichungssystem

$$\begin{aligned} x_1 + \ \ x_2 &= \ \ 2 \\ x_1 - 2x_2 &= -1 \end{aligned} \quad \text{, also} \quad \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

mit der Cramerschen Regel lösen. Dies ist sehr einfach: Es ist

$$x_1 = \frac{\det \begin{pmatrix} 2 & 1 \\ -1 & -2 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}} = \frac{-3}{-3} = 1 \quad \text{und} \quad x_2 = \frac{\det \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}} = \frac{-3}{-3} = 1.$$

Beachte jedoch, dass es für konkrete Gleichungssysteme mit mehr als zwei Variablen meistens sehr rechenaufwändig ist, die Cramersche Regel zu verwenden. Man wird diese Regel daher meistens nur für theoretische Überlegungen verwenden, in denen man eine konkrete Formel für die Lösung (und nicht nur ein Lösungsverfahren) braucht. Für numerische Berechnungen ist das Gauß-Verfahren in Algorithmus 17.9 wesentlich effizienter.

Aufgabe 18.24. Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, die eine Blockgestalt der Form

$$A = \left(\begin{array}{c|c} B & * \\ \hline 0 & C \end{array} \right)$$

hat, wobei $B \in \mathrm{Mat}(m \times m, K)$ und $C \in \mathrm{Mat}((n-m) \times (n-m), K)$ selbst quadratische Matrizen sind. Zeige, dass dann $\det A = \det B \cdot \det C$ gilt.

Aufgabe 18.25. Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine invertierbare Matrix mit ganzzahligen Einträgen.

40 Zeige, dass A^{-1} genau dann ebenfalls nur ganzzahlige Einträge hat, wenn $\det A = \pm 1$ gilt.

Grundlagen der Mathematik 2: Lineare Algebra

19. Endomorphismen

In Kapitel 16 haben wir ausführlich untersucht, wie man lineare Abbildungen $f\colon V \to W$ zwischen zwei (endlich-dimensionalen) Vektorräumen V und W beschreiben kann. Wir haben gesehen, dass sich solche Morphismen nach Wahl von Basen von V und W immer durch Matrizen darstellen lassen, und dass man bei geeigneter Wahl dieser Basen auch immer erreichen kann, dass diese Matrizen eine sehr einfache Form haben (siehe Satz 16.42).

Wir wollen uns jetzt den sehr häufig vorkommenden Spezialfall der sogenannten *Endomorphismen* ansehen, bei denen der Startraum V derselbe ist wie der Zielraum W, so dass wir einen Vektor $x \in V$ direkt mit seinem Bildvektor $f(x) \in W$ vergleichen können. Um dies auch anhand der Koordinaten von x und $f(x)$ bezüglich der gewählten Basen machen zu können, werden wir bei der Untersuchung von Endomorphismen in der Regel voraussetzen, dass wir in V und W die gleiche Basis gewählt haben.

Schon ohne irgendeine Rechnung sollte klar sein, dass wir in diesem Fall wohl keine so einfache Abbildungsmatrix wie in Satz 16.42 erhalten können, da wir jetzt ja nur noch *eine* Basis frei wählen dürfen und damit viel weniger Möglichkeiten haben, das Resultat zu manipulieren. In der Tat wird dieses Problem einer möglichst einfachen Abbildungsmatrix durch die verringerten Wahlmöglichkeiten so kompliziert, dass wir es erst am Ende des nächsten Kapitels vollständig lösen können.

19.A Ähnliche Matrizen

Um den Unterschied zwischen linearen Abbildungen mit verschiedenem und gleichem Start- und Zielraum deutlich zu sehen, erinnern wir uns zunächst noch einmal kurz an unsere Resultate über Abbildungsmatrizen aus Kapitel 16.

Bemerkung 19.1 (Abbildungsmatrizen linearer Abbildungen)**.** Es seien V und W endlich erzeugte Vektorräume der Dimensionen n bzw. m mit gewählten Basen $B = (x_1, \dots, x_n)$ und $C = (y_1, \dots, y_m)$. Weiterhin sei $f\colon V \to W$ eine lineare Abbildung. Dann wissen wir bereits:

(a) Wir können f durch die Abbildungsmatrix
$$A_f^{B,C} = (\,\Phi_C(f(x_1)) \mid \cdots \mid \Phi_C(f(x_n))\,)$$
beschreiben (siehe Satz 16.23), wobei $\Phi_C\colon W \to K^m$ die Koordinatenabbildung ist (siehe Definition 16.22). Die Spalten dieser Matrix sind also genau die Koordinatenvektoren bzgl. C der Bilder der Basisvektoren von B. Diese Zuordnung einer Matrix $A_f^{B,C}$ zu einem Morphismus f liefert einen Isomorphismus zwischen dem Raum $\mathrm{Hom}(V,W)$ aller derartigen Morphismen und dem Raum $\mathrm{Mat}(m \times n, K)$ aller Matrizen der passenden Größe.

(b) Sind B' und C' weitere Basen von V bzw. W, so ändert sich die zugehörige Abbildungsmatrix nach Satz 16.32 (a) gemäß
$$A_f^{B',C'} = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B},$$
wobei $A^{C,C'} \in \mathrm{GL}(m,K)$ und $A^{B',B} \in \mathrm{GL}(n,K)$ die Basiswechselmatrizen aus Definition 16.28 sind, es ist also z. B. $A^{C,C'} = (\,\Phi_{C'}(y_1) \mid \cdots \mid \Phi_{C'}(y_m)\,)$.

(c) Umgekehrt sind invertierbare Matrizen auch immer Basiswechselmatrizen. Zwei Matrizen $A, A' \in \mathrm{Mat}(m \times n, K)$ beschreiben damit genau dann dieselbe lineare Abbildung (nur bezüglich evtl. verschiedener Basen), wenn es invertierbare Matrizen S und T gibt mit $A' = SAT$

(siehe Satz 16.32). Wir haben A und A' in diesem Fall *äquivalent* zueinander genannt (siehe Definition 16.36).

(d) Nach Satz 16.42 kann man die Basen B und C stets so wählen, dass die Abbildungsmatrix $A_f^{B,C}$ von f die besonders einfache „Normalform“

$$A_f^{B,C} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

mit $r = \operatorname{rk} f$ besitzt. In der Sprechweise von (c) bedeutet dies einfach, dass jede Matrix äquivalent zu einer Matrix in einer solchen Normalform ist.

Wie schon angekündigt wollen wir nun untersuchen, was sich an diesen Eigenschaften ändert, wenn Start- und Zielraum der betrachteten Morphismen übereinstimmen und wir in beiden Räumen auch die gleiche Basis wählen.

Definition 19.2 (Endomorphismen)**.** Es seien K ein Körper und V ein K-Vektorraum. Ein Morphismus $f\colon V \to V$ von V in sich heißt **Endomorphismus** von V. Der Vektorraum aller Endomorphismen von V wird mit $\operatorname{End}(V) = \operatorname{Hom}(V,V)$ bezeichnet.

Bemerkung 19.3 (Abbildungsmatrizen von Endomorphismen)**.** Wir betrachten einen endlichdimensionalen Vektorraum V mit einer Basis $B = (x_1,\dots,x_n)$. Die Eigenschaften aus Bemerkung 19.1 ändern sich dann für einen Endomorphismus $f\colon V \to V$ wie folgt:

(a) Die *Abbildungsmatrix* von f wie in Bemerkung 19.1 (a) mit Start- und Zielbasis B bezeichnen wir kurz mit

$$A_f^B := A_f^{B,B} = (\,\Phi_B(f(x_1)) \mid \cdots \mid \Phi_B(f(x_n))\,),$$

so dass die Zuordnung $f \mapsto A_f^B$ nun also einen Isomorphismus von $\operatorname{End}(V)$ in den Raum $\operatorname{Mat}(n \times n, K)$ der quadratischen $n \times n$-Matrizen liefert.

(b) Die Transformationsformel aus Bemerkung 19.1 (b) wird dementsprechend zu

$$A_f^{B'} = A^{B,B'} \cdot A_f^B \cdot A^{B',B}$$

für eine zweite Basis B' von V. Weil nach Lemma 16.31 (a) aber $A^{B,B'} = (A^{B',B})^{-1}$ gilt, können wir dies schreiben als

$$A_f^{B'} = T^{-1} A_f^B T \quad \text{mit} \quad T = A^{B',B}.$$

(c) Da invertierbare Matrizen immer Basiswechselmatrizen sind, beschreiben zwei quadratische Matrizen $A, A' \in \operatorname{Mat}(n \times n, K)$ genau dann denselben Endomorphismus (nur bezüglich evtl. einer anderen Basis), wenn es eine invertierbare Matrix T gibt mit $A' = T^{-1}AT$. Wir definieren daher:

Definition 19.4 (Ähnliche Matrizen)**.** Zwei quadratische Matrizen $A, A' \in \operatorname{Mat}(n \times n, K)$ derselben Größe heißen **ähnlich** zueinander, wenn es eine invertierbare Matrix $T \in \operatorname{GL}(n,K)$ gibt, so dass $A' = T^{-1}AT$.

Nach Bemerkung 19.1 (c) und 19.3 (c) sind A und A' also genau dann ähnlich zueinander, wenn sie denselben Endomorphismus bezüglich einer evtl. anderen Basis beschreiben.

Bemerkung 19.5.

(a) Wie im Fall der Äquivalenz von Matrizen (siehe Bemerkung 16.37 (a)) prüft man auch hier leicht nach, dass die Ähnlichkeit von Matrizen eine Äquivalenzrelation ist (siehe Definition 2.27).

(b) Zwei ähnliche Matrizen sind offensichtlich auch immer äquivalent zueinander. Beachte, dass die Wahl der Begriffe hier etwas ungeschickt ist, denn im normalen Sprachgebrauch klingt „Äquivalenz“ ja wie eine stärkere Bedingung als „Ähnlichkeit“ – in Wirklichkeit ist es aber genau umgekehrt.

Die Übertragung von Teil (d) der Bemerkung 19.1, also die Frage nach einer „Normalform von Matrizen bezüglich Ähnlichkeit“, führt auf die sogenannte Jordansche Normalform, die wir im nächsten Kapitel ausführlich untersuchen werden. In diesem Abschnitt wollen wir uns zunächst einmal damit begnügen, ein paar Eigenschaften quadratischer Matrizen anzugeben, die bei ähnlichen Matrizen gleich sind. Neben dem Rang (der bei ähnlichen Matrizen aufgrund von Lemma 16.41 natürlich übereinstimmt), sind dies die uns schon bekannte Determinante sowie die sogenannte Spur, die wir jetzt kurz einführen und untersuchen wollen.

Definition 19.6 (Spur einer Matrix)**.** Ist $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, so heißt die Summe ihrer Einträge auf der Diagonalen von links oben nach rechts unten

$$\mathrm{Spur}\, A := a_{1,1} + \cdots + a_{n,n}$$

die **Spur** von A.

Lemma 19.7. *Für alle $A, B \in \mathrm{Mat}(n \times n, K)$ gilt* $\mathrm{Spur}(AB) = \mathrm{Spur}(BA)$.

Beweis. Es seien $A = (a_{i,j})_{i,j}$ und $B = (b_{i,j})_{i,j}$. Dann gilt nach Definition 16.5 des Matrixprodukts

$$\mathrm{Spur}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} b_{j,i} \quad \text{und} \quad \mathrm{Spur}(BA) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{i,j} a_{j,i}.$$

Dies ist aber derselbe Ausdruck (unter Vertauschung der Bezeichnung der Summationsvariablen $i \leftrightarrow j$). □

Lemma 19.8 (Invarianz von Determinante und Spur unter Ähnlichkeit)**.** *Sind $A, A' \in \mathrm{Mat}(n \times n, K)$ zwei zueinander ähnliche Matrizen, so gilt*

(a) $\det A' = \det A$;

(b) $\mathrm{Spur}\, A' = \mathrm{Spur}\, A$.

Beweis. Es sei $T \in \mathrm{GL}(n, K)$ mit $A' = T^{-1}AT$.

(a) Mit Satz 18.6 folgt

$$\det A' = \det(T^{-1}AT) = (\det T)^{-1} \cdot \det A \cdot \det T = \det A.$$

(b) Es gilt

$$\mathrm{Spur}\, A' = \mathrm{Spur}(T^{-1}A \cdot T) \overset{19.7}{=} \mathrm{Spur}(T \cdot T^{-1}A) = \mathrm{Spur}(EA) = \mathrm{Spur}\, A. \qquad \square$$

Beispiel 19.9. Für die Matrix

$$A = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

gilt $\det A = 4 \cdot (-2) - (-1) \cdot 5 = -3$ und $\mathrm{Spur}\, A = 4 - 2 = 2$. Betrachten wir nun die invertierbare Matrix

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{so berechnet man leicht} \quad T^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad A' := T^{-1}AT = \begin{pmatrix} -1 & 0 \\ 5 & 3 \end{pmatrix}.$$

Offensichtlich ist in Übereinstimmung mit Lemma 19.8 auch hier $\det A' = -3$ und $\mathrm{Spur}\, A' = 2$. Diese Invarianz von Determinante und Spur bedeutet auch bereits, dass A sicher nicht zu einer Matrix in einer Normalform wie in Bemerkung 19.1 (d) ähnlich sein kann, da alle diese Matrizen ja Determinante 0 oder 1 haben.

Die Invarianz der Determinante und Spur einer Matrix hat auch noch eine weitere wichtige Konsequenz: Sie bedeutet, dass wir diese Konzepte nicht nur für Matrizen, sondern auch für Endomorphismen definieren können:

Definition 19.10 (Determinante und Spur von Endomorphismen)**.** Es sei $f\colon V \to V$ ein Endomorphismus eines endlich erzeugten Vektorraums V. Dann definieren wir die **Determinante** und **Spur** von f als

$$\det f := \det A_f^B \qquad \text{und} \qquad \operatorname{Spur} f := \operatorname{Spur} A_f^B$$

für eine beliebige Basis B von V. Beachte, dass dies wohldefiniert ist, weil eine andere Basiswahl nach Bemerkung 19.3 (b) zu einer ähnlichen Abbildungsmatrix und somit nach Lemma 19.8 zur gleichen Determinante und Spur führen würde.

19.B Eigenwerte

Wir wollen nun ein weiteres sehr wichtiges Konzept einführen, das für quadratische Matrizen genauso wie für Endomorphismen definiert werden kann und auf der Idee basiert, dass wir bei gleichem Start- und Zielraum einen Vektor in der Startmenge mit seinem Bildvektor vergleichen können.

Definition 19.11 (Eigenwerte, Eigenvektoren und Eigenräume)**.**

(a) Es sei $A \in \operatorname{Mat}(n \times n, K)$ eine quadratische Matrix. Eine Zahl $\lambda \in K$ heißt **Eigenwert** von A, wenn es einen Vektor $x \in K^n \backslash \{0\}$ gibt mit $Ax = \lambda x$. In diesem Fall heißt x ein zum Eigenwert λ gehöriger **Eigenvektor**. Die Menge

$$\operatorname{Eig}(A, \lambda) := \{x \in K^n : Ax = \lambda x\}$$

wird der **Eigenraum** von A zum Eigenwert λ genannt.

(b) Es sei $f\colon V \to V$ ein Endomorphismus eines Vektorraums V. In diesem Fall heißt $\lambda \in K$ ein **Eigenwert** und $x \in V \backslash \{0\}$ ein zugehöriger **Eigenvektor** von f, wenn $f(x) = \lambda x$. Auch hier ist der **Eigenraum** von f zu λ die Menge

$$\operatorname{Eig}(f, \lambda) := \{x \in V : f(x) = \lambda x\}.$$

Bemerkung 19.12.

(a) Der Eigenraum (einer Matrix oder Endomorphismus) ist also einfach die Menge aller Eigenvektoren – mit Ausnahme des Nullvektors, der nach Definition 19.11 *nie* ein Eigenvektor ist, aber in *jedem* Eigenraum enthalten ist. Ein Vorteil dieser Konvention ist zum Beispiel, dass der Eigenraum $\operatorname{Eig}(A, \lambda)$ einer Matrix A stets ein Untervektorraum von K^n ist, denn er lässt sich wegen

$$\operatorname{Eig}(A, \lambda) = \{x \in K^n : Ax = \lambda x\} = \{x \in K^n : (\lambda E - A)x = 0\} = \operatorname{Ker}(\lambda E - A)$$

als Kern einer Matrix schreiben, wobei E wie üblich die Einheitsmatrix (der Größe $n \times n$) bezeichnet. Diese alternative Beschreibung $\operatorname{Eig}(A, \lambda) = \operatorname{Ker}(\lambda E - A)$ werden wir im Folgenden oft benutzen. Natürlich können wir auch genauso gut $\operatorname{Eig}(A, \lambda) = \operatorname{Ker}(A - \lambda E)$ schreiben, da eine Matrix und ihr Negatives stets denselben Kern haben. Ein wichtiger Spezialfall hiervon ist der des Eigenwerts 0: Hier ist $\operatorname{Eig}(A, 0) = \operatorname{Ker} A$ einfach der Kern der gegebenen Matrix.

(b) Analog zu (a) ist $\operatorname{Eig}(f, \lambda) = \operatorname{Ker}(\lambda \operatorname{id} - f)$ für einen Endomorphismus $f\colon V \to V$.

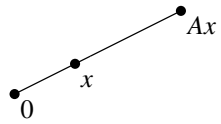
(c) Ist V ein endlich-dimensionaler Vektorraum, so hängen die Eigenwerte und Eigenvektoren eines Endomorphismus $f\colon V \to V$ wie erwartet mit denen der Abbildungsmatrix A_f^B von f bezüglich einer Basis B von V zusammen: Nach Satz 16.23 ist $\Phi_B(f(x)) = A_f^B \cdot \Phi_B(x)$ für alle $x \in V$, und damit

$$\begin{aligned} f(x) = \lambda x \;&\Leftrightarrow\; \Phi_B(f(x)) = \lambda \Phi_B(x) \\ &\Leftrightarrow\; A_f^B \cdot \Phi_B(x) = \lambda\, \Phi_B(x). \end{aligned}$$

Also ist $x \in V$ genau dann ein Eigenvektor zum Eigenwert λ von f, wenn der zugehörige Koordinatenvektor $\Phi_B(x)$ ein Eigenvektor zum selben Eigenwert λ der Abbildungsmatrix A_f^B ist. Insbesondere haben f und A_f^B damit die gleichen Eigenwerte.

41

Beispiel 19.13 (Geometrische Deutung von Eigenvektoren)**.** Nach Definition ist ein Vektor $x \neq 0$ genau dann ein Eigenvektor einer quadratischen Matrix A zum Eigenwert λ, wenn x durch die zu A gehörige lineare Abbildung wie im Bild rechts nur um einen Faktor λ gestreckt wird. Besonders einfach zu sehen ist dies im Fall von Drehungen in der Ebene: Für einen gegebenen Winkel $\alpha \in [0, 2\pi)$ beschreibt die Abbildung



$$f\colon \mathbb{C} \to \mathbb{C},\ x \mapsto \mathrm{e}^{\mathrm{i}\alpha} x,$$

also in Polarkoordinaten $f(r\mathrm{e}^{\mathrm{i}\varphi}) = r\mathrm{e}^{\mathrm{i}(\varphi+\alpha)}$, nach Satz 9.26 gerade die Drehung der komplexen Zahlenebene um den Ursprung um den Winkel α. Identifizieren wir $\mathbb{C}$ mit $\mathbb{R}^2$, indem wir eine komplexe Zahl $x \in \mathbb{C}$ gemäß $x = x_1 + \mathrm{i}x_2$ mit $x_1, x_2 \in \mathbb{R}$ in Real- und Imaginärteil aufteilen, so können wir f auch schreiben als

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ x_1 + \mathrm{i}x_2 \mapsto (\cos\alpha + \mathrm{i}\sin\alpha)(x_1 + \mathrm{i}x_2) = (\cos\alpha \cdot x_1 - \sin\alpha \cdot x_2) + \mathrm{i}(\sin\alpha \cdot x_1 + \cos\alpha \cdot x_2),$$

bzw. in Vektorschreibweise als die lineare Abbildung $f = f_A$ zur sogenannten *Drehmatrix*

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R}).$$

Für eine solche Drehung sehen wir nun auch ohne Rechnung:

(a) Für $\alpha = 0$ ist $A = E$, also $x \mapsto f(x) = Ax = x$ die Identität. Damit ist 1 der einzige Eigenwert von f bzw. A, und $\mathrm{Eig}(f, 1) = \mathrm{Eig}(A, 1) = \mathbb{R}^2$.

(b) Für $\alpha = \pi$ ist $A = -E$, die Abbildung $x \mapsto f(x) = Ax = -x$ spiegelt alle Vektoren am Ursprung. Also ist in diesem Fall -1 der einzige Eigenwert von f bzw. A, und $\mathrm{Eig}(f, -1) = \mathrm{Eig}(A, -1) = \mathbb{R}^2$.

(c) Für alle anderen Drehwinkel wird kein Vektor $x \in \mathbb{R}^2 \setminus \{0\}$ auf ein Vielfaches von sich abgebildet. Die Abbildung f bzw. die Matrix A hat dann also keine Eigenvektoren und damit auch keine Eigenwerte.

Beispiel 19.14 (Eigenvektoren in Funktionenräumen)**.** Ein wichtiges Beispiel von (unendlich-dimensionalen) Vektorräumen, die wir bisher noch nicht untersucht haben, erhalten wir, wenn wir die lineare Algebra mit der Analysis kombinieren und Räume stetiger bzw. differenzierbarer Funktionen betrachten. So ist z. B. für eine Teilmenge $D \subset \mathbb{R}$ (ohne isolierte Punkte) und $n \in \mathbb{N} \cup \{\infty\}$ die in Definition 11.8 bereits eingeführte Menge $C^n(D)$ aller n-mal stetig differenzierbaren Funktionen von D nach $\mathbb{R}$ ein Untervektorraum von $\mathrm{Abb}(D, \mathbb{R})$, da mit $f, g\colon D \to \mathbb{R}$ auch $f + g$ und λf für $\lambda \in \mathbb{R}$ wieder n-mal stetig differenzierbar sind.

Betrachten wir nun z. B. die Ableitungsabbildung

$$f\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}),\ \varphi \mapsto \varphi'$$

(die nach den Rechenregeln für Ableitungen aus Satz 10.8 linear ist), so können wir auch für diesen Endomorphismus nach Eigenwerten und Eigenvektoren suchen. In der Tat sehen wir hier auch schnell, dass jede reelle Zahl $\lambda \in \mathbb{R}$ ein Eigenwert von f ist, denn für die Funktion $\varphi\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \mathrm{e}^{\lambda x}$ gilt ja $\varphi'(x) = \lambda\,\mathrm{e}^{\lambda x} = \lambda\varphi(x)$ für alle $x \in \mathbb{R}$, und damit $f(\varphi) = \lambda\varphi$.

Wir wollen nun aber wieder zum endlich-dimensionalen Fall zurückkehren und als Erstes untersuchen, wie man im Allgemeinen von einer gegebenen Matrix die Eigenwerte und Eigenvektoren berechnen kann. Dies ist sehr einfach:

Lemma 19.15 (Berechnung von Eigenwerten)**.** *Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix. Eine Zahl $\lambda \in K$ ist genau dann ein Eigenwert von A, wenn* $\det(\lambda E - A) = 0$.

Beweis. Für $\lambda \in K$ gilt

$$\begin{aligned}
\lambda \text{ Eigenwert von } A &\Leftrightarrow \text{ es gibt ein } x \in K^n \setminus \{0\} \text{ mit } (\lambda E - A)x = 0 \\
&\Leftrightarrow \dim \operatorname{Ker}(\lambda E - A) > 0 \\
&\Leftrightarrow \dim \operatorname{Im}(\lambda E - A) < n && \text{(Bemerkung 16.16 (a))} \\
&\Leftrightarrow \lambda E - A \text{ nicht invertierbar} && \text{(Bemerkung 16.39 (b))} \\
&\Leftrightarrow \det(\lambda E - A) = 0. && \text{(Satz 18.6 (b))}
\end{aligned}$$

□

Beispiel 19.16. Die Eigenwerte der reellen Matrix

$$A = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix}$$

aus Beispiel 19.9 sind nach Lemma 19.15 die Lösungen der Gleichung

$$0 = \det(\lambda E - A) = \det \begin{pmatrix} \lambda - 4 & 1 \\ -5 & \lambda + 2 \end{pmatrix} = (\lambda - 4)(\lambda + 2) + 5 = \lambda^2 - 2\lambda - 3 = (\lambda + 1)(\lambda - 3),$$

also $\lambda_1 = -1$ und $\lambda_2 = 3$.

Die Eigenräume bzw. Eigenvektoren ergeben sich nun für jeden dieser Eigenwerte λ als Lösung des linearen Gleichungssystems $(\lambda E - A)x = 0$, also indem man in die oben bereits berechnete Matrix $\lambda E - A$ den entsprechenden Eigenwert einsetzt und mit dem Gauß-Verfahren 17.11 den Kern dieser Matrix bestimmt. So erhalten wir z. B. den Eigenraum $\operatorname{Eig}(A, -1)$ als Lösung von

$$0 = (-1E - A)x = \begin{pmatrix} -5 & 1 \\ -5 & 1 \end{pmatrix} \cdot x, \quad \text{d. h. es ist} \quad \operatorname{Eig}(A, -1) = \operatorname{Lin}\begin{pmatrix} 1 \\ 5 \end{pmatrix}.$$

Genauso ergibt sich der Eigenraum $\operatorname{Eig}(A, 3)$ als Lösung von

$$0 = (3E - A)x = \begin{pmatrix} -1 & 1 \\ -5 & 5 \end{pmatrix} \cdot x, \quad \text{also} \quad \operatorname{Eig}(A, 3) = \operatorname{Lin}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Die Eigenvektoren sind jeweils die Vektoren dieser Eigenräume mit Ausnahme des Nullvektors.

Beachte, dass wir hier bei der Rechnung eine gute Kontrolle haben: Wenn wir mit Lemma 19.15 einen Eigenwert λ berechnet haben und nach Einsetzen dieses Wertes in das Gleichungssystem $(\lambda E - A)x = 0$ als Lösung aber nur den Nullvektor, also keinen Eigenvektor herausbekommen, dann müssen wir irgendwo einen Rechenfehler gemacht haben.

Im Beispiel dieser reellen 2×2-Matrix A haben wir für den Ausdruck $\det(\lambda E - A)$ eine normierte quadratische Polynomfunktion (siehe Definition 3.17) herausbekommen. Wir wollen nun sehen, dass dies ein allgemeines Resultat ist und wir für eine $n \times n$-Matrix die Nullstellen einer Polynomfunktion vom Grad n berechnen müssen. Damit wir diesen Nullstellen dann auch Vielfachheiten zuordnen können, sollten wir jedoch zunächst voraussetzen, dass der Grundkörper K unendlich viele Elemente hat und wir damit Polynomfunktionen mit Polynomen gleichsetzen können, d. h. einen Koeffizientenvergleich zur Verfügung haben (siehe Lemma 3.21, Notation 3.22 und Satz 3.24). Wir vereinbaren also:

Im Folgenden (in Kapitel 19 und 20) sei K stets ein Körper mit unendlich vielen Elementen.

Satz und Definition 19.17 (Charakteristisches Polynom)**.** *Für eine Matrix $A \in \operatorname{Mat}(n \times n, K)$ ist*

$$\chi_A(t) := \det(tE - A)$$

ein normiertes Polynom vom Grad n. Man nennt es das ***charakteristische Polynom*** *von A.*

Die Eigenwerte von A sind nach Lemma 19.15 also genau die Nullstellen des charakteristischen Polynoms χ_A.

Beweis. Wir zeigen zunächst mit Induktion über $k = 1, \dots, n$, dass die Determinante jeder $k \times k$-Teilmatrix $B(t) = (b_{i,j}(t))_{i,j}$ von $tE - A$, die durch Auswahl von k Zeilen und Spalten entsteht, ein Polynom vom Grad höchstens k ist.

$k = 1$: In diesem Fall ist die Behauptung offensichtlich, da alle Einträge von $tE - A$ Polynome vom Grad höchstens 1 sind.

$k \to k+1$: Die Matrix $B(t)$ habe nun die Größe $(k+1) \times (k+1)$. Nach der Laplace-Entwicklung aus Satz 18.15 gilt

$$\det B(t) = \sum_{i=1}^{k+1} (-1)^{i+1} b_{i,1}(t) \det B'_{i,1}(t),$$

wobei $B'_{i,1}(t)$ die $k \times k$-Streichungsmatrix aus Definition 18.11 bezeichnet, bei der aus $B(t)$ die Zeile i und Spalte 1 herausgestrichen wurde. Da hier $b_{i,1}(t)$ als Eintrag von $tE - A$ ein Polynom vom Grad höchstens 1 ist und $\det B'_{i,1}(t)$ nach Induktionsvoraussetzung maximal Grad k hat, ist $\det B(t)$ wie behauptet ein Polynom vom Grad höchstens $k+1$.

Für $k = n$ erhalten wir also, dass $\chi_A(t) = c_n t^n + c_{n-1} t^{n-1} + \cdots + c_1 t + c_0$ (mit $c_0, \ldots, c_n \in K$) ein Polynom vom Grad höchstens n ist.

Den Leitkoeffizienten c_n können wir schließlich bestimmen, indem wir den Ausdruck $t^n \chi_A(\frac{1}{t})$ berechnen: Dies ist nämlich wieder ein Polynom, in dem wir dann $t = 0$ einsetzen können und

$$(t^n \chi_A(\tfrac{1}{t}))(0) = (c_n + c_{n-1} t + \cdots + c_1 t^{n-1} + c_0 t^n)(0) = c_n$$

erhalten. Damit folgt

$$c_n = (t^n \det(\tfrac{1}{t} E - A))(0) \overset{(*)}{=} (\det(E - tA))(0) = \det E = 1,$$

wobei sich $(*)$ aus der Multilinearität der Determinante gemäß Definition 18.2 (a) ergibt: Multipliziert man jede Zeile der $n \times n$-Matrix $\frac{1}{t} E - A$ mit t, um daraus $E - tA$ zu erhalten, so multipliziert sich die Determinante dadurch um einen Faktor t^n. □

Wir wollen den Eigenwerten einer quadratischen Matrix jetzt Vielfachheiten zuordnen. Hierfür gibt es zwei grundlegend verschiedene Möglichkeiten.

Definition 19.18 (Vielfachheiten von Eigenwerten)**.** Es seien $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix und $\lambda \in K$.

(a) Die Vielfachheit von λ als Nullstelle im charakteristischen Polynom χ_A (siehe Satz 3.24) heißt die **algebraische** oder **arithmetische Vielfachheit** (oder auch **Multiplizität**) $\mu_{\mathrm{a}}(A, \lambda) \in \mathbb{N}$ von λ in A.

(b) Die Dimension des Eigenraums $\mathrm{Eig}(A, \lambda)$ heißt die **geometrische Vielfachheit** (oder auch **Multiplizität**) $\mu_{\mathrm{g}}(A, \lambda) \in \mathbb{N}$ von λ in A.

Bemerkung 19.19. Für ein $\lambda \in K$ gilt genau dann $\mu_{\mathrm{a}}(A, \lambda) > 0$, wenn λ eine Nullstelle von χ_A ist, also nach Lemma 19.15 genau dann, wenn λ ein Eigenwert von A ist. Ebenso gilt natürlich (unmittelbar nach Definition), dass genau dann $\mu_{\mathrm{g}}(A, \lambda) > 0$ ist, wenn λ ein Eigenwert von A ist.

Beispiel 19.20. Beide Eigenwerte der Matrix

$$A = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

aus Beispiel 19.16 haben algebraische Vielfachheit 1 (da sie einfache Nullstellen des charakteristischen Polynoms $\chi_A(t) = (t-3)^1 (t+1)^1$ sind) und geometrische Vielfachheit 1 (da wir ihre Eigenräume dort als eindimensional erkannt haben).

Wie wir in Beispiel 19.37 (b) noch sehen werden, müssen die algebraische und geometrische Vielfachheit eines Eigenwerts einer Matrix im Allgemeinen nicht übereinstimmen. Wir werden dieses Phänomen und seine Konsequenzen noch ausführlich im nächsten Abschnitt sowie in Kapitel 20 untersuchen. Für den Moment können wir aber schon einmal festhalten, dass das charakteristische Polynom sowie die Vielfachheiten von Eigenwerten invariant unter Ähnlichkeit von Matrizen sind und sich diese Konzepte daher analog zu Definition 19.10 auch für Endomorphismen einführen lassen.

Lemma 19.21 (Invarianz von χ_A, $\mu_{\mathrm{a}}(A,\lambda)$ und $\mu_{\mathrm{g}}(A,\lambda)$)**.** *Sind $A,A' \in \mathrm{Mat}(n\times n,K)$ zwei ähnliche Matrizen, so gilt:*

(a) $\chi_{A'} = \chi_A$;

(b) $\mu_{\mathrm{a}}(A',\lambda) = \mu_{\mathrm{a}}(A,\lambda)$ *für alle* $\lambda \in K$;

(c) $\mu_{\mathrm{g}}(A',\lambda) = \mu_{\mathrm{g}}(A,\lambda)$ *für alle* $\lambda \in K$.

Beweis. Es sei $T \in \mathrm{GL}(n,K)$ mit $A' = T^{-1}AT$.

(a) Nach dem Produktsatz 18.6 für Determinanten gilt für alle $t \in K$

$$\begin{aligned}\chi_{A'}(t) &= \det(tE - T^{-1}AT) = \det(T^{-1}(tE-A)T) = \frac{1}{\det T}\cdot\det(tE-A)\cdot\det T\\ &= \chi_A(t).\end{aligned}$$

(b) Da sich die algebraische Vielfachheit aus dem charakteristischen Polynom ergibt, folgt dies unmittelbar aus (a).

(c) Für alle $x \in K^n$ gilt

$$x \in \mathrm{Eig}(A',\lambda) \;\Leftrightarrow\; T^{-1}ATx = \lambda x \;\Leftrightarrow\; ATx = \lambda Tx \;\Leftrightarrow\; Tx \in \mathrm{Eig}(A,\lambda).$$

Die Abbildung $\mathrm{Eig}(A',\lambda) \to \mathrm{Eig}(A,\lambda),\ x \mapsto Tx$ ist also ein Isomorphismus (mit Umkehrabbildung $x \mapsto T^{-1}x$). Da isomorphe Vektorräume nach Satz 14.22 (c) die gleiche Dimension haben, folgt also wie behauptet $\mu_{\mathrm{g}}(A,\lambda) = \mu_{\mathrm{g}}(A',\lambda)$. □

Definition 19.22 (χ_f, $\mu_{\mathrm{a}}(f,\lambda)$ und $\mu_{\mathrm{g}}(f,\lambda)$ für Endomorphismen)**.** Für einen Endomorphismus $f\colon V \to V$ eines endlich erzeugten Vektorraums V und $\lambda \in K$ definieren wir

$$\begin{aligned}\text{das } \textbf{charakteristische Polynom} \text{ als} \quad & \chi_f := \chi_{A_f^B},\\ \text{die } \textbf{algebraische} \text{ oder } \textbf{arithmetische Vielfachheit} \text{ von } \lambda \text{ als} \quad & \mu_{\mathrm{a}}(f,\lambda) := \mu_{\mathrm{a}}(A_f^B,\lambda),\\ \text{und die } \textbf{geometrische Vielfachheit} \text{ von } \lambda \text{ als} \quad & \mu_{\mathrm{g}}(f,\lambda) := \mu_{\mathrm{g}}(A_f^B,\lambda)\end{aligned}$$

für eine beliebige Basis B von V. Da verschiedene Basiswahlen zu ähnlichen Abbildungsmatrizen führen, ist dies nach Lemma 19.21 wohldefiniert.

Auch die weiteren Konzepte und Ergebnisse in diesem und dem folgenden Kapitel werden auf ähnliche Art jeweils eine Version für Matrizen und eine für Endomorphismen haben, die letztlich äquivalent zueinander sind. Um Schreibarbeit zu sparen, werden wir in vielen Fällen jedoch jeweils nur die Matrixversion angeben.

Bemerkung 19.23 (Charakteristisches Polynom für Hörer der „Algebraischen Strukturen“)**.** Falls ihr die Vorlesung „Algebraische Strukturen“ bereits gehört habt und damit den dort eingeführten Polynomring $K[t]$ kennt [G, Kapitel 9], könnt ihr das charakteristische Polynom χ_A einer Matrix $A \in \mathrm{Mat}(n\times n,K)$ in der Tat auch ohne unsere oben gemachte Voraussetzung $|K| = \infty$ als Polynom (und nicht nur als Polynomfunktion) definieren und damit auch in diesem Fall Eigenwerten algebraische Vielfachheiten zuordnen. Dazu müsst ihr $tE - A$ als Matrix mit Einträgen in $K[t]$ auffassen und die Determinante auch über diesem Polynomring als Abbildung $\det\colon \mathrm{Mat}(n\times n,K[t]) \to K[t]$ konstruieren (was problemlos möglich ist, da hierfür keine Divisionen benötigt werden). Damit wird dann $\chi_A(t) = \det(tE-A)$ wie gewünscht ein Element des Polynomrings $K[t]$.

Verwendet man diese Konstruktion des charakteristischen Polynoms und damit auch der algebraischen Vielfachheiten von Eigenwerten, so sind alle Resultate in diesem und dem nächsten Kapitel ohne irgendwelche Änderungen auch ohne unsere Voraussetzung $|K| = \infty$, also für beliebige Grundkörper gültig.

Aufgabe 19.24. Zeige für jede reelle quadratische Matrix $A \in \mathrm{Mat}(n\times n,\mathbb{R})$:

(a) Ist A invertierbar und λ ein Eigenwert von A, so ist $\frac{1}{\lambda}$ ein Eigenwert von A^{-1}, und es gilt $\mu_{\mathrm{g}}(A^{-1},\frac{1}{\lambda}) = \mu_{\mathrm{g}}(A,\lambda)$ und $\mu_{\mathrm{a}}(A^{-1},\frac{1}{\lambda}) = \mu_{\mathrm{a}}(A,\lambda)$.

(b) Ist jeder Vektor in K^n ein Eigenvektor von A, so ist A ein Vielfaches der Einheitsmatrix.

Aufgabe 19.25. Es seien $f\colon V \to V$ ein Endomorphismus eines endlich-dimensionalen Vektorraums V und U ein Unterraum mit $f(U) \subset U$. Nach Aufgabe 15.21 gibt es dann zugehörige lineare Abbildungen $g\colon U \to U$, $x \mapsto f(x)$ und $h\colon V/U \to V/U$, $\overline{x} \mapsto \overline{f(x)}$. Man zeige:

(a) Ergänzen wir eine Basis $B' = (x_1, \ldots, x_k)$ von U zu einer Basis $B = (x_1, \ldots, x_n)$ von V, so dass $B'' = (\overline{x_{k+1}}, \ldots, \overline{x_n})$ nach Bemerkung 15.19 also eine Basis von V/U ist, so ist die Abbildungsmatrix von f bezüglich B von der Form

$$A_f^B = \left(\begin{array}{c|c} A_g^{B'} & * \\ \hline 0 & A_h^{B''} \end{array}\right).$$

(b) Es gilt $\chi_f = \chi_g \cdot \chi_h$.

Aufgabe 19.26. Es seien $A, B \in \mathrm{Mat}(n \times n, K)$. Zeige, dass die Matrizen AB und BA das gleiche charakteristische Polynom haben.

(Hinweis: Zeige die Aussage zunächst, wenn eine der Matrizen invertierbar ist, dann wenn eine der Matrizen von der Form

$$\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

ist, und schließlich im allgemeinen Fall.)

19.C Diagonalisierbarkeit

Nach den Vorarbeiten im letzten Abschnitt wollen wir uns nun wieder unserem ursprünglichen Problem widmen, zu einer gegebenen Matrix eine möglichst einfache ähnliche Matrix bzw. zu einem gegebenen Endomorphismus eine möglichst einfache Abbildungsmatrix zu finden. Das folgende einfache aber zentrale Lemma zeigt dabei deutlich, warum diese Frage unmittelbar mit Eigenwerten und Eigenvektoren zusammenhängt. Da es so wichtig ist, formulieren wir es sowohl in der Sprache der ähnlichen Matrizen als auch in der Sprache der Endomorphismen, obwohl diese beiden Versionen eigentlich dieselbe Aussage darstellen.

Lemma 19.27 (Abbildungsmatrizen und Eigenvektoren)**.**

(a) *Es seien $A \in \mathrm{Mat}(n \times n, K)$, $T \in \mathrm{GL}(n, K)$, sowie $A' = T^{-1}AT$, so dass A und A' also ähnlich zueinander sind. Ferner seien $k \in \{1, \ldots, n\}$ und $\lambda \in K$.*

Dann ist die k-te Spalte von A' genau dann gleich λe_k (hat also in der k-ten Zeile den Eintrag λ und sonst nur Nullen), wenn die k-te Spalte der Transformationsmatrix T ein Eigenvektor von A zum Eigenwert λ ist.

(b) *Es sei $f\colon V \to V$ ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V mit Basis $B = (x_1, \ldots, x_n)$. Ferner seien wieder $k \in \{1, \ldots, n\}$ und $\lambda \in K$.*

Dann ist die k-te Spalte der Abbildungsmatrix A_f^B genau dann gleich λe_k, wenn der k-te Basisvektor x_k ein Eigenvektor von f zum Eigenwert λ ist.

Beweis.

(a) Die k-te Spalte von A' ist gleich $A'e_k = T^{-1}ATe_k$. Diese ist also genau dann gleich λe_k, wenn

$$T^{-1}ATe_k = \lambda e_k \overset{T\cdot}{\iff} A(Te_k) = \lambda(Te_k)$$

gilt, also wenn die k-te Spalte Te_k von T ein Eigenvektor von A zum Eigenwert λ ist.

(b) Nach Bemerkung 19.1 (a) ist die k-te Spalte von A_f^B gleich $\Phi_B(f(x_k))$. Da Φ_B ein Isomorphismus ist, ist dieser Ausdruck genau dann gleich $\lambda e_k = \Phi_B(\lambda x_k)$, wenn $f(x_k) = \lambda x_k$ gilt, also x_k ein Eigenvektor von f zum Eigenwert λ ist. □

Beispiel 19.28. Es sei noch einmal

$$A = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

die Matrix aus Beispiel 19.16. Wir hatten dort bereits gesehen, dass $\begin{pmatrix} 1 \\ 5 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ Eigenvektoren von A zu den Eigenwerten -1 bzw. 3 sind. Offensichtlich sind diese beiden Vektoren linear unabhängig. Schreiben wir sie also als Spalten in eine Matrix

$$T = \begin{pmatrix} 1 & 1 \\ 5 & 1 \end{pmatrix},$$

so ist diese Matrix invertierbar, und Lemma 19.27 (a) sagt uns ohne weitere Rechnung, dass die beiden Spalten der transformierten Matrix $T^{-1}AT$ genau $-1\,e_1$ und $3\,e_2$ sind, d. h. dass

$$T^{-1}AT = \begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix}$$

gilt. Natürlich könnte man dies nun durch explizite Berechnung der inversen Matrix T^{-1} und des Matrixprodukts $T^{-1}AT$ auch direkt überprüfen.

42

Da in diesem Beispiel in jeder Spalte von T ein Eigenvektor steht, haben wir also gemäß Lemma 19.27 (a) eine transformierte Matrix erhalten, deren k-te Spalte für alle k ein Vielfaches von e_k ist, die also nur auf der Diagonalen Einträge ungleich 0 hat. Dieses Prinzip wollen wir nun allgemein festhalten.

Definition 19.29 (Diagonalmatrizen und Diagonalisierbarkeit)**.**

(a) Eine quadratische Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, K)$ heißt **Diagonalmatrix**, wenn $a_{i,j} = 0$ für alle $i \neq j$ gilt, also wenn A höchstens auf der Diagonalen Einträge ungleich Null hat. Um Platz zu sparen, schreiben wir eine solche Diagonalmatrix oft als $\mathrm{diag}(a_{1,1}, \dots, a_{n,n})$.

(b) Eine quadratische Matrix heißt **diagonalisierbar**, wenn sie ähnlich zu einer Diagonalmatrix ist. Ein Endomorphismus $f\colon V \to V$ eines endlich erzeugten Vektorraums V heißt **diagonalisierbar**, wenn es eine Basis B von V gibt, so dass die zugehörige Abbildungsmatrix A_f^B eine Diagonalmatrix ist.

Folgerung 19.30 (Diagonalisierbarkeit = Basis aus Eigenvektoren)**.** *Eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ ist genau dann diagonalisierbar, wenn es eine Basis $(x_1, \dots, x_n)$ von K^n aus Eigenvektoren von A gibt.*

Sind in diesem Fall $\lambda_1, \dots, \lambda_n$ die zugehörigen Eigenwerte von $x_1, \dots, x_n$, so gilt

$$T^{-1}AT = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

mit $T = (x_1 \mid \cdots \mid x_n)$.

Beweis. Beachte, dass die Matrix T nach Bemerkung 16.39 genau dann invertierbar ist, wenn sie Rang n hat, also wenn ihre Spalten linear unabhängig sind und damit eine Basis von K^n bilden. Da die Matrix $T^{-1}AT$ genau dann eine Diagonalmatrix ist, wenn ihre k-te Spalte für alle $k = 1, \dots, n$ ein Vielfaches von e_k ist, ergibt sich die Aussage der Folgerung damit unmittelbar aus Lemma 19.27 (a), angewendet auf jede Spalte von T. □

Bemerkung 19.31. Natürlich erhalten wir auch hier wieder eine analoge Aussage für Endomorphismen mit Lemma 19.27 (b): Ein Endomorphismus $f\colon V \to V$ eines endlich-dimensionalen Vektorraums ist genau dann diagonalisierbar, wenn es eine Basis $B = (x_1, \dots, x_n)$ aus Eigenvektoren von V gibt, und in diesem Fall ist dann $A_f^B = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, wobei $\lambda_1, \dots, \lambda_n$ die zu $x_1, \dots, x_n$ gehörigen Eigenwerte sind.

Beispiel 19.32. Die Matrix aus Beispiel 19.28 ist diagonalisierbar. Eine allgemeine Drehmatrix wie in Beispiel 19.13 (c) ist dagegen nicht diagonalisierbar, da sie keine Eigenvektoren (und damit insbesondere keine Basis aus Eigenvektoren) besitzt.

Als Erstes wollen wir nun zeigen, dass für diagonalisierbare Matrizen eine Diagonalmatrix mit beliebigen Einträgen auf der Diagonalen wirklich die einfachste mögliche ähnliche Matrix ist, also dass keine zwei solchen Diagonalmatrizen mehr ähnlich zueinander sein können. Wir können eine solche Diagonalform in diesem Fall also als Normalform bezüglich Ähnlichkeit auffassen.

Lemma 19.33. *Zwei Diagonalmatrizen sind genau dann ähnlich zueinander, wenn ihre Diagonaleinträge bis auf die Reihenfolge übereinstimmen.*

Beweis. Wir betrachten zwei Diagonalmatrizen $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ und $B = \operatorname{diag}(\mu_1, \dots, \mu_n)$.

„$\Rightarrow$“: Es sei A ähnlich zu B. Beachte, dass

$$\chi_A(t) = \det \operatorname{diag}(t - \lambda_1, \dots, t - \lambda_n) = (t - \lambda_1) \cdots (t - \lambda_n)$$

und analog $\chi_B(t) = (t - \mu_1) \cdots (t - \mu_n)$. Nach Lemma 19.21 (a) gilt nun aber $\chi_A = \chi_B$, und wegen der Eindeutigkeit der Zerlegung von Polynomen in Faktoren aus Satz 3.24 bedeutet dies gerade, dass die Zahlen $\lambda_1, \dots, \lambda_n$ bis auf die Reihenfolge mit $\mu_1, \dots, \mu_n$ übereinstimmen.

„$\Leftarrow$“: Es seien nun $\lambda_1, \dots, \lambda_n$ bis auf die Reihenfolge gleich $\mu_1, \dots, \mu_n$. Wegen der Diagonalform von A sind die Einheitsvektoren $e_1, \dots, e_n$ offensichtlich Eigenvektoren von A zu den Eigenwerten $\lambda_1, \dots, \lambda_n$. Es seien nun $x_1, \dots, x_n$ diese Einheitsvektoren so durchnummeriert, dass $Ax_i = \mu_i x_i$ für alle $i = 1, \dots, n$. Mit $T = (x_1 \mid \cdots \mid x_n)$ ist dann $T^{-1}AT = \operatorname{diag}(\mu_1 \mid \cdots \mid \mu_n) = B$ nach Lemma 19.27 (a), d. h. A und B sind ähnlich zueinander. □

Folgerung 19.34. *Zwei diagonalisierbare Matrizen sind genau dann ähnlich zueinander, wenn ihre charakteristischen Polynome übereinstimmen.*

Beweis. Es seien $A, B \in \operatorname{Mat}(n \times n, K)$ diagonalisierbar, mit A ähnlich zu $A' = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ und B ähnlich zu $B' = \operatorname{diag}(\mu_1, \dots, \mu_n)$.

„$\Rightarrow$“: Sind A und B ähnlich, so gilt $\chi_A = \chi_B$ nach Lemma 19.21 (a).

„$\Leftarrow$“: Es sei nun $\chi_A = \chi_B$, nach Lemma 19.21 (a) also auch

$$(t - \lambda_1) \cdots (t - \lambda_n) = \chi_{A'}(t) = \chi_A(t) = \chi_B(t) = \chi_{B'}(t) = (t - \mu_1) \cdots (t - \mu_n).$$

Also stimmen $\lambda_1, \dots, \lambda_n$ und $\mu_1, \dots, \mu_n$ nach Satz 3.24 bis auf die Reihenfolge überein. Damit ist A' nach Lemma 19.33 ähnlich zu B', und folglich auch A zu B. □

Im Rest dieses Kapitels wollen wir nun noch ein einfaches Kriterium für die Diagonalisierbarkeit einer Matrix $A \in \operatorname{Mat}(n \times n, K)$ herleiten. Nach Folgerung 19.30 müssen wir ja herausfinden, ob es eine Basis von K^n aus Eigenvektoren von A gibt. Um eine solche Basis zu finden, werden wir offensichtlich zunächst einmal aus jedem Eigenraum $\operatorname{Eig}(A, \lambda)$ maximal viele linear unabhängige Vektoren wählen, also $\mu_g(A, \lambda)$ Stück. Wir müssen uns nun fragen:

(A) Wie viele Vektoren bekommen wir auf diese Art insgesamt? (Wir bräuchten n, um eine Basis von K^n erhalten zu können.)

(B) Sind *alle diese Vektoren zusammen* wirklich linear unabhängig? (Wir wissen zunächst nur, dass die ausgewählten Vektoren innerhalb eines Eigenraums linear unabhängig sind.)

Wir werden im Folgenden sehen, dass (A) wirklich eine Bedingung an die Matrix A stellt (siehe Folgerung 19.36), während (B) für jede Matrix erfüllt ist (siehe Folgerung 19.39).

Um die Bedingung (A) genauer zu untersuchen, benötigen wir zunächst einmal einen wichtigen Zusammenhang zwischen der geometrischen und algebraischen Vielfachheit eines Eigenwerts.

Satz 19.35. *Für jede Matrix $A \in \operatorname{Mat}(n \times n, K)$ und alle $\lambda \in K$ gilt $\mu_g(A, \lambda) \leq \mu_a(A, \lambda)$.*

Beweis. Es sei $k = \mu_g(A, \lambda)$. Wir wählen eine Basis $(x_1, \dots, x_k)$ von $\operatorname{Eig}(A, \lambda)$, ergänzen diese zu einer Basis $(x_1, \dots, x_n)$, und schreiben diese Vektoren als Spalten in eine Matrix $T = (x_1 \mid \cdots \mid x_n)$.

Nach Lemma 19.27 (a) ist die i-te Spalte von $A' := T^{-1}AT$ dann gleich λe_i für alle $i = 1, \dots, k$, d. h. es ist

$$A' = \left(\begin{array}{c|c} \lambda E_k & * \\ \hline 0 & B \end{array}\right)$$

für eine Matrix $B \in \mathrm{Mat}((n-k)\times(n-k),K)$. Damit folgt nach Lemma 19.21 (a) und Aufgabe 18.24

$$\chi_A(t) = \chi_{A'}(t) = \det\left(\begin{array}{c|c} (t-\lambda)E_k & * \\ \hline 0 & tE-B \end{array}\right) = (t-\lambda)^k \chi_B(t),$$

d. h. die Vielfachheit $\mu_\mathrm{a}(A,\lambda)$ von λ als Nullstelle in χ_A ist mindestens gleich $k = \mu_\mathrm{g}(A,\lambda)$. □

Folgerung 19.36 (Notwendige Kriterien für Diagonalisierbarkeit)**.** *Ist eine Matrix $A \in \mathrm{Mat}(n\times n,K)$ diagonalisierbar, so muss gelten:*

(a) *Das charakteristische Polynom χ_A zerfällt in Linearfaktoren.*

(b) *Für alle Eigenwerte λ von A gilt $\mu_\mathrm{g}(A,\lambda) = \mu_\mathrm{a}(A,\lambda)$.*

Beweis. Nach Satz 3.24 können wir das charakteristische Polynom von A schreiben als

$$\chi_A(t) = g(t)\cdot(t-\lambda_1)^{\mu_\mathrm{a}(A,\lambda_1)}\cdots(t-\lambda_k)^{\mu_\mathrm{a}(A,\lambda_k)}, \tag{1}$$

wobei g ein Polynom ohne Nullstellen ist und $\lambda_1,\dots,\lambda_k$ die verschiedenen Eigenwerte von A sind. Da χ_A nach Satz 19.17 Grad n hat, folgt daraus insbesondere

$$\mu_\mathrm{a}(A,\lambda_1)+\dots+\mu_\mathrm{a}(A,\lambda_k) \le n. \tag{2}$$

Um nun gemäß Folgerung 19.30 eine Basis von K^n aus Eigenvektoren von A zu finden, können wir aus jedem Eigenraum $\mathrm{Eig}(A,\lambda_i)$ für $i=1,\dots,k$ maximal $\mu_\mathrm{g}(A,\lambda_i) = \dim\mathrm{Eig}(A,\lambda_i)$ Vektoren wählen. Die Anzahl der so insgesamt gefundenen Vektoren ist also

$$\sum_{i=1}^k \mu_\mathrm{g}(A,\lambda_i) \overset{19.35}{\le} \sum_{i=1}^k \mu_\mathrm{a}(A,\lambda_i) \overset{(2)}{\le} n.$$

Da wir für eine Basis von K^n aber n Vektoren benötigen, muss hier im Fall einer diagonalisierbaren Matrix an beiden Stellen die Gleichheit gelten. An der ersten Stelle ergibt dies genau die Bedingung $\mu_\mathrm{g}(A,\lambda_i) = \mu_\mathrm{a}(A,\lambda_i)$ für alle i; an der zweiten erhalten wir, dass das Polynom g in (1) Grad 0 haben und χ_A damit in Linearfaktoren zerfallen muss. □

Beispiel 19.37.

(a) Die Matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \in \mathrm{Mat}(2\times 2,\mathbb{R})$$

hat das charakteristische Polynom

$$\chi_A(t) = \det\begin{pmatrix} t & 1 \\ -1 & t \end{pmatrix} = t^2+1,$$

das über $\mathbb{R}$ nicht in Linearfaktoren zerfällt. Also ist A nach Folgerung 19.36 (a) nicht diagonalisierbar. In der Tat ist A auch genau die Drehmatrix um den Winkel $\frac{\pi}{2}$, von der wir in Beispiel 19.32 bereits gesehen haben, dass sie nicht diagonalisierbar ist.

(b) Die Matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \in \mathrm{Mat}(2\times 2,\mathbb{R})$$

hat das charakteristische Polynom

$$\chi_A(t) = \det\begin{pmatrix} t & -1 \\ 0 & t \end{pmatrix} = t^2$$

mit doppelter Nullstelle 0. Also ist 0 ein Eigenwert von A mit $\mu_\mathrm{a}(A,0) = 2$. Der zugehörige Eigenraum

$$\mathrm{Eig}(A,0) = \mathrm{Ker}(A) = \mathrm{Lin}(e_1)$$

ist aber nur eindimensional. Es ist also $\mu_\mathrm{g}(A,0) = 1 < \mu_\mathrm{a}(A,0)$, und damit ist A nach Folgerung 19.36 (b) nicht diagonalisierbar.

Beachte, dass dieses Beispiel auch zeigt, dass Folgerung 19.34 für beliebige Matrizen falsch ist: Auch die Nullmatrix hat das charakteristische Polynom $\chi_0(t) = \det(tE) = t^2$, aber A ist sicher nicht zur Nullmatrix ähnlich, da es kein $T \in \mathrm{GL}(2,\mathbb{R})$ geben kann mit $T^{-1}\cdot 0\cdot T = A$.

Wenn die Bedingungen aus Folgerung 19.36 erfüllt sind, haben wir jetzt gesehen, dass wir insgesamt wirklich n Vektoren erhalten, wenn wir Basen aller Eigenräume einer Matrix $A \in \mathrm{Mat}(n \times n, K)$ berechnen. Wir müssen jetzt nur noch zeigen, dass diese Vektoren wirklich auch zusammen genommen linear unabhängig sind, damit sie dann eine Basis von K^n bilden.

Satz 19.38 (Lineare Unabhängigkeit von Eigenvektoren)**.** *Es seien $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix und $x_1, \ldots, x_k$ Eigenvektoren von A zu verschiedenen Eigenwerten. Dann sind $x_1, \ldots, x_k$ linear unabhängig.*

Beweis. Wir beweisen den Satz mit Induktion über k. Für $k = 1$ ist die Aussage klar, denn ein Eigenvektor ist nach Definition immer ungleich dem Nullvektor und damit linear unabhängig.

Für den Induktionsschritt $k \to k+1$ nehmen wir nun an, dass wir die Aussage des Satzes für k Vektoren schon gezeigt haben. Es seien nun $x_1, \ldots, x_{k+1}$ Eigenvektoren zu verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_{k+1}$ von A. Angenommen, wir haben eine Linearkombination des Nullvektors

$$\mu_1 x_1 + \cdots + \mu_{k+1} x_{k+1} = 0 \tag{$*$}$$

für gewisse $\mu_1, \ldots, \mu_{k+1} \in K$. Multiplizieren wir diese Gleichung von links mit der Matrix A, so erhalten wir wegen $Ax_i = \lambda_i x_i$ für alle i

$$\mu_1 \lambda_1 x_1 + \cdots + \mu_{k+1} \lambda_{k+1} x_{k+1} = 0.$$

Multiplizieren wir dieselbe Gleichung stattdessen mit dem Skalar λ_{k+1}, so ergibt sich hingegen

$$\mu_1 \lambda_{k+1} x_1 + \cdots + \mu_{k+1} \lambda_{k+1} x_{k+1} = 0.$$

Wir subtrahieren nun die letzten beiden Gleichungen voneinander und erhalten so

$$\mu_1 (\lambda_1 - \lambda_{k+1}) x_1 + \cdots + \mu_k (\lambda_k - \lambda_{k+1}) x_k = 0.$$

Dies ist eine Linearkombination des Nullvektors von k Eigenvektoren zu verschiedenen Eigenwerten. Nach Induktionsvoraussetzung müssen also alle Vorfaktoren Null sein, d. h. es ist

$$\mu_1 (\lambda_1 - \lambda_{k+1}) = \cdots = \mu_k (\lambda_k - \lambda_{k+1}) = 0.$$

Da aber alle Eigenwerte $\lambda_1, \ldots, \lambda_{k+1}$ verschieden sind, folgt daraus $\mu_1 = \cdots = \mu_k = 0$, und damit eingesetzt in $(*)$ natürlich auch $\mu_{k+1} = 0$. Also sind $x_1, \ldots, x_{k+1}$ linear unabhängig. □

Folgerung 19.39 (Die Summe von Eigenräumen ist direkt)**.** *Es seien $A \in \mathrm{Mat}(n \times n, K)$ und $\lambda_1, \ldots, \lambda_k$ verschiedene Eigenwerte von A. Dann ist die Summe*

$$\mathrm{Eig}(A, \lambda_1) + \cdots + \mathrm{Eig}(A, \lambda_k)$$

direkt (siehe Definition 15.2).

Beweis. Dies ist letztlich nur eine Umformulierung von Satz 19.38. Haben wir nämlich zwei Darstellungen eines Vektors in der gegebenen Summe

$$x_1 + \cdots + x_k = y_1 + \cdots + y_k$$

mit $x_i, y_i \in \mathrm{Eig}(A, \lambda_i)$ für alle $i = 1, \ldots, k$, so bedeutet dies ja

$$\underbrace{x_1 - y_1}_{\in \mathrm{Eig}(A, \lambda_1)} + \cdots + \underbrace{x_k - y_k}_{\in \mathrm{Eig}(A, \lambda_k)} = 0.$$

Nun ist der Eigenraum $\mathrm{Eig}(A, \lambda_i)$ für alle $i = 1, \ldots, k$ nach Definition 19.11 (b) aber gerade der Nullvektor zusammen mit allen Eigenvektoren zum Eigenwert λ_i. Wäre in der obigen Gleichung also einer der Vektoren $x_i - y_i$ nicht gleich dem Nullvektor, so wäre er ein Eigenvektor zum Eigenwert λ_i, und wir hätten somit eine nicht-triviale Linearkombination des Nullvektors aus Eigenvektoren von A zu verschiedenen Eigenwerten – im Widerspruch zu Satz 19.38. Also gilt $x_i = y_i$ für alle $i = 1, \ldots, k$, d. h. die Summe der Eigenräume ist direkt. □

43

Wir können nun unsere Ergebnisse zur Diagonalisierbarkeit zusammenfassen:

Folgerung 19.40 (Diagonalisierbarkeit)**.** *Für eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ sind äquivalent:*

(a) *A ist diagonalisierbar.*

(b) *Das charakteristische Polynom χ_A zerfällt in Linearfaktoren, und es gilt $\mu_g(A,\lambda) = \mu_a(A,\lambda)$ für alle Eigenwerte λ von A.*

(c) *Sind $\lambda_1,\dots,\lambda_k$ die verschiedenen Eigenwerte von A, so ist $\operatorname{Eig}(A,\lambda_1)\oplus\dots\oplus\operatorname{Eig}(A,\lambda_k) = K^n$.*

Bestimmt man in diesem Fall eine Basis jedes Eigenraums und nimmt alle diese Vektoren zusammen, so erhält man eine Basis von K^n. Schreibt man diese Vektoren als Spalten in eine Matrix T, so ist $T^{-1}AT$ eine Diagonalmatrix mit den zugehörigen Eigenwerten auf der Diagonalen.

Beweis.

„(a) $\Rightarrow$ (b)“: Dies ist genau Folgerung 19.36.

„(b) $\Rightarrow$ (c)“: Nach Folgerung 19.39 ist die Summe der Eigenräume direkt. Also folgt aus der Voraussetzung mit Lemma 15.3

$$\dim(\operatorname{Eig}(A,\lambda_1)\oplus\dots\oplus\operatorname{Eig}(A,\lambda_k)) = \sum_{i=1}^k \dim\operatorname{Eig}(A,\lambda_i) = \sum_{i=1}^k \mu_g(A,\lambda_i) = \sum_{i=1}^k \mu_a(A,\lambda_i) = n,$$

und damit $\operatorname{Eig}(A,\lambda_1)\oplus\dots\oplus\operatorname{Eig}(A,\lambda_k) = K^n$.

„(c) $\Rightarrow$ (a)“: Ist $\operatorname{Eig}(A,\lambda_1)\oplus\dots\oplus\operatorname{Eig}(A,\lambda_k) = K^n$, so erhalten wir nach Aufgabe 15.35 eine Basis von K^n, indem wir Basen der einzelnen Eigenräume zusammen nehmen. Da diese Basis dann natürlich aus Eigenvektoren zu den zugehörigen Eigenwerten besteht, folgt die zu zeigende Aussage damit aus Folgerung 19.30. □

Bemerkung 19.41.

(a) Ist ein Eigenwert λ eine einfache Nullstelle des charakteristischen Polynoms χ_A einer quadratischen Matrix A, ist also $\mu_a(A,\lambda) = 1$, so folgt aus Satz 19.35 für diesen Eigenwert auch unmittelbar $\mu_g(A,\lambda) = 1$. Im Kriterium (b) von Folgerung 19.40 muss man die Bedingung $\mu_g(A,\lambda) = \mu_a(A,\lambda)$ für diesen Eigenwert also nicht mehr überprüfen. Zerfällt χ_A also z. B. in paarweise verschiedene Linearfaktoren, so ist A damit stets diagonalisierbar.

Im Fall des Körpers $K = \mathbb{C}$ zerfällt außerdem *jedes* Polynom nach dem Fundamentalsatz der Algebra (siehe Satz 6.11) in Linearfaktoren. Hat ein solches Polynom „zufällig gewählte Koeffizienten“, so können wir erwarten, dass diese Linearfaktoren auch alle verschieden sind. Ohne dass wir dies hier zu einer mathematisch exakten Aussage machen wollen, können wir anschaulich also sagen, dass eine „zufällig gewählte komplexe Matrix“ diagonalisierbar ist.

(b) Wie bei unseren vorherigen Ergebnissen in diesem Kapitel gibt es natürlich auch zu Folgerung 19.40 eine analoge Version für Endomorphismen, deren Formulierung (und Beweis) offensichtlich sein sollte.

Beispiel 19.42. Wir betrachten die Matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \in \operatorname{Mat}(3\times 3,\mathbb{R}).$$

Das charakteristische Polynom von A berechnet sich als Determinante einer unteren Dreiecksmatrix nach Folgerung 18.18 zu

$$\chi_A(t) = \det\begin{pmatrix} t-1 & 0 & 0 \\ 0 & t-1 & 0 \\ -1 & 1 & t \end{pmatrix} = (t-1)^2 t.$$

Es zerfällt also in Linearfaktoren und hat als Nullstellen die beiden Eigenwerte 1 (mit $\mu_a(A,1) = 2$) und 0 (mit $\mu_a(A,0) = 1$). Um herauszufinden, ob A diagonalisierbar ist, müssen wir nach Folgerung

19.40 und Bemerkung 19.41 also nur noch überprüfen, ob $\mu_g(A,1) = 2$ gilt. Dazu berechnen wir den Eigenraum $\mathrm{Eig}(A,1)$ durch das Gleichungssystem

$$(1 \cdot E - A)x = 0, \quad \text{also} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 1 \end{pmatrix} \cdot x = 0.$$

Dieser Eigenraum ist zweidimensional; eine Basis bilden z. B. die Vektoren $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Also ist $\mu_g(A,1) = \mu_a(A,1) = 2$, und A ist nach Folgerung 19.40 damit diagonalisierbar.

Um eine Matrix T zu bestimmen, so dass $T^{-1}AT$ eine Diagonalmatrix ist, müssen wir auch vom anderen Eigenraum $\mathrm{Eig}(A,0)$ eine Basis bestimmen: Man sieht hier sofort, dass der dritte Einheitsvektor ein Element von $\mathrm{Eig}(A,0) = \mathrm{Ker}\,A$ und somit eine Basis dieses eindimensionalen Eigenraums ist. Wir brauchen nach Folgerung 19.40 also nur noch die drei gefundenen Vektoren als Spalten in die Matrix T zu schreiben, und erhalten somit

$$T^{-1}AT = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{mit} \quad T = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

(was man natürlich durch direktes Nachrechnen auch überprüfen könnte).

Aufgabe 19.43. Es seien

$$A = \begin{pmatrix} 3 & 2 & -2 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{pmatrix} \in \mathrm{Mat}(3 \times 3, \mathbb{R}).$$

(a) Bestimme die Eigenwerte und Eigenvektoren von A und B mit ihren algebraischen und geometrischen Vielfachheiten.

(b) Welche dieser Matrizen ist / sind diagonalisierbar? Im Fall der Diagonalisierbarkeit bestimme man jeweils eine Matrix $T \in \mathrm{GL}(3,\mathbb{R})$, so dass $T^{-1}AT$ bzw. $T^{-1}BT$ eine Diagonalmatrix ist.

(c) Sind A und B ähnlich zueinander?

(d) Wie kann man mit Hilfe von (a) auf einfache Art eine allgemeine Formel für die Potenzen A^n für $n \in \mathbb{N}$ bestimmen?

Aufgabe 19.44. Bestimme die Eigenwerte und Eigenvektoren der linearen Abbildung

$$f\colon \mathrm{Mat}(2 \times 2, \mathbb{R}) \to \mathrm{Mat}(2 \times 2, \mathbb{R}), \quad A \mapsto A^{\mathsf{T}}.$$

Ist f diagonalisierbar? Falls ja, bestimme man eine Basis B von $\mathrm{Mat}(2 \times 2, \mathbb{R})$, so dass die zugehörige Abbildungsmatrix A_f^B von f eine Diagonalmatrix ist, und gebe diese Abbildungsmatrix an.

Aufgabe 19.45.

(a) Es seien $V = C^0(\mathbb{R})$ der $\mathbb{R}$-Vektorraum aller stetigen Funktionen auf $\mathbb{R}$ und $f\colon V \to V$ die lineare Abbildung mit $f(\varphi)(x) = \varphi(x+1)$. Bestimme alle Eigenwerte von f.

(b) Es seien V der $\mathbb{R}$-Vektorraum aller Polynomfunktionen auf $\mathbb{R}$ und $f\colon V \to V$ wieder die lineare Abbildung mit $f(\varphi)(x) = \varphi(x+1)$. Bestimme alle Eigenwerte von f.

Aufgabe 19.46.

(a) Ist die Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$, deren Einträge alle gleich 1 sind, diagonalisierbar? Falls ja, zu welcher Diagonalmatrix ist A ähnlich?

(b) Es seien V ein K-Vektorraum mit Basis $(x_1, \ldots, x_n)$ und $f\colon V \to V$ der Endomorphismus mit

$$f(x_i) = x_{i+1} \quad \text{für } i = 1, \ldots, n-1 \quad \text{und} \quad f(x_n) = x_1,$$

der also die Basisvektoren zyklisch permutiert.

Untersuche f in den Fällen $K = \mathbb{R}$ und $K = \mathbb{C}$ in Abhängigkeit von n auf Diagonalisierbarkeit, und gib ggfs. an, welche Diagonalmatrix bei geeigneter Basiswahl die Abbildungsmatrix von f sein kann.

Aufgabe 19.47. Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine reelle quadratische Matrix.

(a) Man zeige: A ist genau dann diagonalisierbar, wenn A^T diagonalisierbar ist.

(b) Es sei $\chi_A(t) = t^3 - 4t^2 + 3t$. Berechne χ_{A^2}.

Aufgabe 19.48. Zu einem Endomorphismus $f\colon V \to V$ eines endlich erzeugten $\mathbb{R}$-Vektorraums gebe es $x, y \in V \setminus \{0\}$ mit $f(x) = y$ und $f(y) = -x$.

Zeige, dass f dann nicht diagonalisierbar ist.

Aufgabe 19.49. Es seien $f, g\colon V \to V$ zwei diagonalisierbare Endomorphismen eines endlichdimensionalen K-Vektorraums V, so dass $f \circ g = g \circ f$. Wir bezeichnen die (verschiedenen) Eigenwerte von f und g mit $\lambda_1, \dots, \lambda_k$ bzw. $\mu_1, \dots, \mu_l$. Man zeige:

(a) $g(\mathrm{Eig}(f, \lambda_i)) \subset \mathrm{Eig}(f, \lambda_i)$ für alle $i = 1, \dots, k$;

(b) $\mathrm{Eig}(g, \mu_j) = (\mathrm{Eig}(f, \lambda_1) \cap \mathrm{Eig}(g, \mu_j)) \oplus \dots \oplus (\mathrm{Eig}(f, \lambda_k) \cap \mathrm{Eig}(g, \mu_j))$ für alle $j = 1, \dots, l$;

(c) es gibt eine Basis B von V, so dass sowohl A_f^B als auch A_g^B eine Diagonalmatrix ist (d. h. f und g sind mit der gleichen Basis diagonalisierbar).

20. Die Jordansche Normalform

Im letzten Kapitel haben wir bereits ausführlich untersucht, wie man zu einer gegebenen quadratischen Matrix $A \in \mathrm{Mat}(n \times n, K)$ eine möglichst einfache ähnliche Matrix finden kann (bzw. wie man zu einem Endomorphismus $f\colon V \to V$ eines endlich-dimensionalen Vektorraums V eine Basis B von V finden kann, so dass die zugehörige Abbildungsmatrix A_f^B möglichst einfach wird). Wir haben dabei in Folgerung 19.40 gesehen, dass A unter den Bedingungen

(a) χ_A zerfällt in Linearfaktoren, und

(b) $\mu_{\mathrm{g}}(A,\lambda) = \mu_{\mathrm{a}}(A,\lambda)$ für alle Eigenwerte λ

diagonalisierbar, also zu einer Diagonalmatrix ähnlich ist, und dass eine solche Diagonalform in diesem Fall auch das bestmögliche Ergebnis ist (siehe Lemma 19.33). Wir wollen uns in diesem Kapitel nun anschauen, was wir erreichen können, wenn eine der beiden obigen Bedingungen nicht erfüllt ist (und A nach Folgerung 19.40 damit nicht diagonalisierbar ist).

Die Bedingung (a), dass das charakteristische Polynom in Linearfaktoren zerfällt, ist dabei eher harmlos. Ist dies nämlich z. B. bei einer reellen Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ nicht der Fall, so können wir A einfach als *komplexe* Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ auffassen. Nach dem Fundamentalsatz der Algebra (siehe Satz 6.11) zerfällt χ_A dann zumindest über $\mathbb{C}$ in Linearfaktoren, und somit können wir wieder eine Diagonalform erreichen (wenn auch mit komplexen Eigenwerten und Eigenvektoren). So war z. B. die Matrix A aus Beispiel 19.37 (a) mit charakteristischem Polynom $\chi_A(t) = t^2 + 1$ zwar über $\mathbb{R}$ nicht diagonalisierbar, über $\mathbb{C}$ ist aber $\chi_A(t) = (t+\mathrm{i})(t-\mathrm{i})$ und A damit diagonalisierbar nach Bemerkung 19.41 (a).

Wie in der Vorlesung „Einführung in die Algebra“ des zweiten Studienjahres gezeigt wird, lässt sich dieser Trick nicht nur für reelle Matrizen, sondern über jedem Grundkörper K anwenden: Man kann stets zu einem größeren Körper übergehen, in dem das charakteristische Polynom dann in Linearfaktoren zerfällt. Das Problem eines nicht zerfallenden charakteristischen Polynoms ist also recht einfach zu beheben, und daher wollen wir hier auch nicht mehr weiter darauf eingehen. Wir werden in diesem Kapitel in der Regel also Matrizen betrachten, deren charakteristisches Polynom zwar in Linearfaktoren zerfällt, für die aber i. A. nicht die Bedingung (b) oben gilt, und wollen wiederum nach einer möglichst einfachen ähnlichen Matrix suchen. Nach Satz 19.35 bedeutet dies dann $\mu_{\mathrm{g}}(A,\lambda) < \mu_{\mathrm{a}}(A,\lambda)$ für einen Eigenwert λ von A.

20.A Haupträume

Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Nach Lemma 19.27 (a) ist A genau dann diagonalisierbar, wenn es eine Basis von K^n aus Eigenvektoren von A gibt – diese können wir dann als Spalten in eine Matrix $T \in \mathrm{GL}(n,K)$ schreiben, und erhalten für $T^{-1}AT$ eine Diagonalform.

Ist nun $\mu_{\mathrm{g}}(A,\lambda) < \mu_{\mathrm{a}}(A,\lambda)$ für mindestens einen Eigenwert λ von A, so haben wir im Beweis von Folgerung 19.36 gezeigt, dass wir nicht genug linear unabhängige Eigenvektoren für eine solche Basis finden. Unsere Strategie wird es in diesem Fall daher sein, den Begriff der Eigenvektoren etwas zu verallgemeinern, so dass einerseits von diesen verallgemeinerten Eigenvektoren stets genügend viele existieren, um aus ihnen eine Basis von K^n zusammenstellen und in eine invertierbare Matrix T schreiben zu können, und andererseits die transformierte Matrix $T^{-1}AT$ auch im Fall einer Basis aus derartigen verallgemeinerten Eigenvektoren noch recht einfach ist. Die Idee hierfür ist, statt Eigenvektoren, also Vektoren $x \neq 0$ mit $(\lambda E - A)x = 0$, nun solche zu betrachten, für die zumindest $(\lambda E - A)^r x = 0$ für irgendein $r \in \mathbb{N}$ gilt.

Konstruktion 20.1 (Haupträume)**.** Es seien $A \in \mathrm{Mat}(n \times n, K)$ und $\lambda \in K$.

(a) Für $r \in \mathbb{N}$ heißt

$$H_r(A,\lambda) := \operatorname{Ker}((\lambda E - A)^r) \quad \leq K^n$$

der **verallgemeinerte Eigenraum** der Stufe r von A zu λ.

Offensichtlich ist $H_0(A,\lambda) = \operatorname{Ker} E = \{0\}$ und $H_1(A,\lambda) = \operatorname{Eig}(A,\lambda)$.

(b) Die verallgemeinerten Eigenräume bilden eine aufsteigende Kette

$$\{0\} = H_0(A,\lambda) \leq H_1(A,\lambda) \leq H_2(A,\lambda) \leq \cdots$$

von Unterräumen von K^n: Ist nämlich $x \in H_r(A,\lambda)$ für ein $r \in \mathbb{N}$, also $(\lambda E - A)^r x = 0$, so ist natürlich auch $(\lambda E - A)^{r+1} x = (\lambda E - A)(\lambda E - A)^r x = 0$ und damit $x \in H_{r+1}(A,\lambda)$. Die Dimensionen der verallgemeinerten Eigenräume sind also monoton wachsend, können aber natürlich nicht größer als die Dimension n des umgebenden Raumes werden. Es muss also ein $r \in \mathbb{N}$ geben mit

$$H_r(A,\lambda) = H_{r+1}(A,\lambda) = H_{r+2}(A,\lambda) = \cdots \quad =: H(A,\lambda).$$

Dieser Unterraum $H(A,\lambda)$, den wir uns also als „Grenzwert der verallgemeinerten Eigenräume $H_r(A,\lambda)$ für $r \to \infty$“ vorstellen können, heißt der **Hauptraum** von A zu λ.

(c) Für $r \in \mathbb{N}_{>0}$ nennt man die Vektoren $x \in H_r(A,\lambda) \backslash H_{r-1}(A,\lambda)$, für die also $(\lambda E - A)^r x = 0$ und $(\lambda E - A)^{r-1} x \neq 0$ gilt, **Hauptvektoren** oder **verallgemeinerte Eigenvektoren** der Stufe r von A zum Eigenwert λ.

Der Hauptraum besteht also aus dem Nullvektor zusammen mit den Hauptvektoren einer beliebigen Stufe. Die gewöhnlichen Eigenvektoren aus Definition 19.11 (a) sind gerade die Hauptvektoren der Stufe 1, nämlich die Elemente von $H_1(A,\lambda) \backslash H_0(A,\lambda) = \operatorname{Eig}(A,\lambda) \backslash \{0\}$.

Bemerkung 20.2. Es seien $A \in \operatorname{Mat}(n \times n, K)$, $\lambda \in K$ und $r \in \mathbb{N}$ mit $H_r(A,\lambda) = H_{r+1}(A,\lambda)$. Dann gilt auch $H_{r+1}(A,\lambda) = H_{r+2}(A,\lambda)$, denn für alle $x \in H_{r+2}(A,\lambda)$ folgt dann

$$\begin{aligned} & (\lambda E - A)^{r+1} (\lambda E - A) x = 0 \\ \Rightarrow\ & (\lambda E - A) x \in H_{r+1}(A,\lambda) = H_r(A,\lambda) \\ \Rightarrow\ & (\lambda E - A)^r (\lambda E - A) x = 0 \end{aligned}$$

und damit auch $x \in H_{r+1}(A,\lambda)$. Mit Induktion ergibt sich daraus natürlich sofort, dass aus der Gleichheit $H_r(A,\lambda) = H_{r+1}(A,\lambda)$ bereits

$$H_r(A,\lambda) = H_{r+1}(A,\lambda) = H_{r+2}(A,\lambda) = \cdots \quad = H(A,\lambda)$$

folgt. Gilt in der aufsteigenden Kette von verallgemeinerten Eigenräumen aus Konstruktion 20.1 (b) also an irgendeiner Stelle die Gleichheit, so ist dies bereits der Hauptraum zu dem betrachteten Eigenwert.

Für $r = 0$ bedeutet dies insbesondere: Ist $\lambda \in K$ kein Eigenwert von A, gibt es also keine Eigenvektoren zu λ und ist damit $H_1(A,\lambda) = \operatorname{Eig}(A,\lambda) = \{0\} = H_0(A,\lambda)$, so ist auch $H(A,\lambda) = \{0\}$, d. h. dann gibt es auch keine verallgemeinerten Eigenvektoren zu λ. Anders formuliert verallgemeinern wir also nur den Begriff der Eigen*vektoren*, aber nicht den der Eigen*werte*.

Bemerkung 20.3 (Haupträume von Endomorphismen)**.** Natürlich gibt es zu Konstruktion 20.1 und Bemerkung 20.2 auch wieder eine entsprechende Version für Endomorphismen $f\colon V \to V$ eines endlich erzeugten Vektorraums V: Analog zu Bemerkung 19.12 (b) sind die verallgemeinerten Eigenräume in diesem Fall $H_r(f,\lambda) = \operatorname{Ker}((\lambda \operatorname{id} - f)^r)$, wobei die Potenz hier für die r-fache Verkettung von $\lambda \operatorname{id} - f$ mit sich selbst steht.

Beispiel 20.4. Die reelle Matrix

$$A = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

hat nach Folgerung 18.18 das charakteristische Polynom $\chi_A(t) = t^2(t-2)$. Wir wollen zum Eigenwert $\lambda = 0$ den Hauptraum $H(A,0)$ bestimmen. Dazu berechnen wir gemäß Bemerkung 20.2 der Reihe nach für $r = 1,2,3,\ldots$ die Räume $H_r(A,0)$, bis zwei von ihnen übereinstimmen:

$$H_1(A,0) = \operatorname{Ker}(0E-A) = \operatorname{Ker}\begin{pmatrix} 0 & -2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix} = \operatorname{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right),$$

$$H_2(A,0) = \operatorname{Ker}(0E-A)^2 = \operatorname{Ker}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \operatorname{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right),$$

$$H_3(A,0) = \operatorname{Ker}(0E-A)^3 = \operatorname{Ker}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -8 \end{pmatrix} = \operatorname{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right).$$

Also ist der gesuchte Hauptraum $H(A,0) = H_2(A,0)$ der von den ersten beiden Einheitsvektoren aufgespannte Unterraum von $\mathbb{R}^3$.

Wir sehen an diesem Beispiel also bereits, dass es mehr Hauptvektoren als Eigenvektoren geben kann, denn $\operatorname{Eig}(A,0) = H_1(A,0)$ ist hier ja ein echter Unterraum von $H(A,0) = H_2(A,0)$. In der Tat führt dies dazu, dass die Standardbasis eine Basis von $\mathbb{R}^3$ aus Hauptvektoren von A ist (die ersten beiden Einheitsvektoren sind ja Hauptvektoren zum Eigenwert 0, der dritte ist offensichtlich ein Eigenvektor zum Eigenwert 2), wohingegen wir wegen $\mu_g(A,0) = \mu_g(A,2) = 1$ keine Basis von $\mathbb{R}^3$ aus Eigenvektoren von A finden können und A damit nicht diagonalisierbar ist.

Wir wollen nun zeigen, dass wir in der Tat zu jeder Matrix $A \in \operatorname{Mat}(n \times n, K)$ mit zerfallendem charakteristischen Polynom eine Basis aus Hauptvektoren finden können. Analog zu Abschnitt 19.C sind dies gerade die beiden Aussagen:

(A) Wir finden in jedem Hauptraum genügend viele linear unabhängige Vektoren, d. h. es ist $\dim H(A,\lambda) = \mu_a(A,\lambda)$ für alle Eigenwerte λ.

(B) Alle diese Vektoren zusammen genommen sind linear unabhängig, d. h. die Summe der Haupträume ist direkt.

Für den Beweis dieser Aussagen in den nächsten beiden Sätzen benötigen wir zuerst zwei kleine Bemerkungen.

Bemerkung 20.5. Es sei $A \in \operatorname{Mat}(n \times n, K)$.

(a) (Kommutativität von Matrixprodukten aus A und E) Natürlich sind Matrixprodukte im Allgemeinen nicht kommutativ, d. h. in der Regel ist $AB \neq BA$ für eine weitere Matrix $B \in \operatorname{Mat}(n \times n, K)$. Jedoch gilt hier offensichtlich die Gleichheit, falls $B = A$ oder $B = E$ ist. Dementsprechend sind auch Produkte von Matrizen kommutativ, die aus A und E zusammengesetzt sind, wie z. B.

$$(\lambda E - A) \cdot A = A \cdot (\lambda E - A)$$

(in der Tat sind beide Seiten gleich $\lambda A - A^2$). Wir werden diese Kommutativität im Folgenden oft benutzen, ohne jedes Mal wieder darauf hinzuweisen.

(b) (Invarianz der Haupträume) Die Abbildung $x \mapsto Ax$ lässt sich auf jeden verallgemeinerten Eigenraum $H_r(A,\lambda)$ (und damit auch auf $H(A,\lambda)$) zu einem Endomorphismus

$$H_r(A,\lambda) \to H_r(A,\lambda),\ x \mapsto Ax$$

einschränken: Ist nämlich $x \in H_r(A,\lambda)$, also $(\lambda E - A)^r x = 0$, so gilt nach (a) auch

$$(\lambda E - A)^r Ax = A\,(\lambda E - A)^r x = 0$$

und damit $Ax \in H_r(A,\lambda)$. Analog zu Aufgabe 15.21 sagt man hierfür auch, dass die verallgemeinerten Eigenräume und der Hauptraum A-*invariant* sind.

Satz 20.6 (Hauptraumdimension)**.** *Es seien $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix mit zerfallendem charakteristischen Polynom und $\lambda \in K$. Dann gilt:*

(a) *Für alle $\mu \in K$ mit $\mu \neq \lambda$ ist $H(A,\lambda) \cap H(A,\mu) = \{0\}$.*

(b) $\dim H(A,\lambda) = \mu_{\mathrm{a}}(A,\lambda)$.

Beweis. Es sei $r \in \mathbb{N}$, so dass $H(A,\lambda) = H_r(A,\lambda)$. Nach Bemerkung 20.5 (b) können wir die Abbildung $f\colon K^n \to K^n$, $x \mapsto Ax$ auf den Hauptraum $H(A,\lambda)$ einschränken. Es gibt nach Aufgabe 15.21 also lineare Abbildungen

$$g\colon H(A,\lambda) \to H(A,\lambda),\ x \mapsto Ax \quad \text{und} \quad h\colon K^n/H(A,\lambda) \to K^n/H(A,\lambda),\ \overline{x} \mapsto \overline{Ax}.$$

Für die Eigenwerte dieser beiden Abbildungen gilt:

(1) Kein $\mu \neq \lambda$ ist ein Eigenwert von g: Ist $x \in H(A,\lambda)$ mit $Ax = \mu x$, so wissen wir einerseits $(\lambda E - A)^r x = 0$ und andererseits $(\lambda E - A)x = (\lambda - \mu)x$, und damit

$$0 = (\lambda E - A)^r x = (\lambda - \mu)^r x,$$

was wegen $\mu \neq \lambda$ nur für $x = 0$ möglich ist. Also hat g keine Eigenvektoren zu μ, und damit nach Bemerkung 20.2 auch keine verallgemeinerten Eigenvektoren zu μ. Mit anderen
44 Worten ist $H(A,\lambda) \cap H(A,\mu) = H(g,\mu) = \{0\}$, was (a) zeigt.

(2) Umgekehrt ist λ kein Eigenwert von h: Ist $\overline{x} \in K^n/H(A,\lambda)$ mit $\overline{Ax} = \lambda\overline{x}$, so folgt

$$\begin{aligned} & \overline{(\lambda E - A)x} = \overline{0} \\ \Rightarrow\ & (\lambda E - A)x \in H_r(A,\lambda) \\ \Rightarrow\ & (\lambda E - A)^r(\lambda E - A)x = 0 \\ \Rightarrow\ & x \in H_{r+1}(A,\lambda) = H(A,\lambda) \end{aligned}$$

und damit $\overline{x} = \overline{0}$. Also besitzt h keine Eigenvektoren zu λ.

Nach Aufgabe 19.25 ist nun $\chi_A = \chi_f = \chi_g \cdot \chi_h$, und dieses Polynom zerfällt nach Voraussetzung in Linearfaktoren. Aus χ_g können nach (1) aber ausschließlich Linearfaktoren $t - \lambda$ kommen, aus χ_h dagegen nach (2) kein Linearfaktor $t - \lambda$. Die algebraische Vielfachheit $\mu_{\mathrm{a}}(A,\lambda)$, also die Anzahl der Linearfaktoren $t - \lambda$ in χ_A, ist damit genau gleich dem Grad von χ_g, also der Dimension $\dim H(A,\lambda)$ des Startraums von g. Damit ist auch (b) gezeigt. □

Satz 20.7 (**Hauptraumzerlegung**)**.** *Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Sind $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von A, so gilt*

$$H(A,\lambda_1) \oplus \cdots \oplus H(A,\lambda_k) = K^n.$$

Beweis. Es seien $r_1, \dots, r_k \in \mathbb{N}$, so dass $H(A,\lambda_i) = H_{r_i}(A,\lambda_i)$ für $i = 1, \dots, k$. Wir zeigen nun mit Induktion über $l = 1, \dots, k$, dass die Summe $H(A,\lambda_1) + \cdots + H(A,\lambda_l)$ direkt ist. Für den Induktionsanfang $l = 1$ ist dabei nichts zu zeigen.

Für den Induktionsschritt $l \to l+1$ seien

$$x_1 + \cdots + x_{l+1} = y_1 + \cdots + y_{l+1} \tag{1}$$

zwei Darstellungen desselben Vektors in K^n mit $x_i, y_i \in H(A,\lambda_i)$ für alle $i = 1, \dots, l+1$. Multiplizieren wir diese Gleichung von links mit $(\lambda_{l+1}E - A)^{r_{l+1}}$, so fallen x_{l+1} und y_{l+1} weg, und wir erhalten

$$(\lambda_{l+1}E - A)^{r_{l+1}}(x_1 - y_1) + \cdots + (\lambda_{l+1}E - A)^{r_{l+1}}(x_l - y_l) = 0. \tag{2}$$

Nun ist aber $(\lambda_{l+1}E - A)^{r_{l+1}}(x_i - y_i) \in H(A,\lambda_i)$ für alle i, denn nach Bemerkung 20.5 (a) ist

$$(\lambda_i E - A)^{r_i}(\lambda_{l+1}E - A)^{r_{l+1}}(x_i - y_i) = (\lambda_{l+1}E - A)^{r_{l+1}} \underbrace{(\lambda_i E - A)^{r_i}(x_i - y_i)}_{=0} = 0.$$

Da die Summe $H(A,\lambda_1) + \cdots + H(A,\lambda_l)$ nach Induktionsvoraussetzung direkt ist, ergibt sich aus (2) also

$$(\lambda_{l+1}E - A)^{r_{l+1}}(x_i - y_i) = 0 \quad \text{für alle } i = 1, \dots, l.$$

Damit ist $x_i - y_i \in H(A,\lambda_i) \cap H(A,\lambda_{l+1})$, mit Satz 20.6 (a) also $x_i - y_i = 0$ und somit $x_i = y_i$ für $i = 1,\dots,l$. Einsetzen in (1) zeigt dann auch $x_{l+1} = y_{l+1}$, und damit ist die Summe der Haupträume direkt.

Da χ_A in Linearfaktoren zerfällt, ist nun aber auch

$$\dim(H(A,\lambda_1) \oplus \cdots \oplus H(A,\lambda_k)) \overset{15.3}{=} \sum_{i=1}^{k} \dim H(A,\lambda_i) \overset{20.6\,(b)}{=} \sum_{i=1}^{k} \mu_a(A,\lambda_i) = n$$

und damit $H(A,\lambda_1) \oplus \cdots \oplus H(A,\lambda_k) = K^n$. □

Bemerkung 20.8. Wollen wir den Hauptraum $H(A,\lambda)$ einer Matrix A zu einem Eigenwert λ bestimmen, indem wir nacheinander die verallgemeinerten Eigenräume $H_1(A,\lambda), H_2(A,\lambda), H_3(A,\lambda), \dots$ berechnen, so folgt aus Satz 20.6 (b) insbesondere, dass wir mit der Berechnung dieser Räume aufhören können, sobald einer von ihnen die Dimension $\mu_a(A,\lambda)$ hat – und nicht erst wie in Beispiel 20.4, wenn zwei von ihnen gleich sind. In der Tat hätten wir in diesem Beispiel damit $H_3(A,0)$ nicht mehr berechnen müssen, da dort bereits $\dim H_2(A,0) = \mu_a(A,0) = 2$ war und somit $H(A,0) = H_2(A,0)$ gelten musste.

20.B Jordandiagramme

Wir wollen nun endlich unser Problem lösen, zu einer quadratischen Matrix mit zerfallendem charakteristischen Polynom eine möglichst einfache ähnliche Matrix zu finden. Dazu haben wir im letzten Abschnitt gesehen, dass sich diese Frage auf die Untersuchung der einzelnen Haupträume reduzieren lässt. Wir fassen unser Ergebnis hier noch einmal in der jetzt benötigten Form zusammen.

Bemerkung 20.9 (Hauptraumzerlegung)**.** Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Sind $\lambda_1,\dots,\lambda_k$ die verschiedenen Eigenwerte von A, so haben wir in Satz 20.7 gesehen, dass

$$H(A,\lambda_1) \oplus \cdots \oplus H(A,\lambda_k) = K^n.$$

Wählen wir also Basen $B_1,\dots,B_k$ dieser Haupträume, so erhalten wir nach Aufgabe 15.35 mit allen diesen Vektoren zusammen genommen eine Basis B von K^n. Damit können wir diese Vektoren in die Spalten einer invertierbaren Matrix T schreiben.

Wegen der Invarianz der Haupträume gemäß Bemerkung 20.5 (b) können wir nun die Abbildung $f\colon K^n \to K^n$, $x \mapsto Ax$ auf die Haupträume einschränken zu

$$f_i\colon H(A,\lambda_i) \to H(A,\lambda_i),\ x \mapsto Ax$$

für $i = 1,\dots,k$. Die Matrix A_f^B (die nach Bemerkung 19.3 (b) gleich $T^{-1}AT$ ist) kann nach Satz 16.23 in den zu B_i gehörigen Spalten also auch nur in den entsprechenden Zeilen Einträge ungleich 0 haben, und ist dort gleich der Abbildungsmatrix $A_{f_i}^{B_i}$. Wir erhalten für diese Matrix damit

$$A_f^B = T^{-1}AT = \begin{pmatrix} A_{f_1}^{B_1} & & 0 \\ & \ddots & \\ 0 & & A_{f_k}^{B_k} \end{pmatrix}$$

Man sagt, dass A_f^B eine *Blockdiagonalmatrix* mit den Blöcken $A_{f_1}^{B_1},\dots,A_{f_k}^{B_k}$ ist.

Um unsere Suche nach einer möglichst einfachen zu A ähnlichen Matrix zu beenden, müssen wir jetzt also nur noch die Basen B_i der einzelnen Haupträume $H(A,\lambda_i)$ für $i = 1,\dots,k$ so wählen, dass die Abbildungsmatrizen $A_{f_i}^{B_i}$ der auf die Haupträume eingeschränkten Abbildung möglichst einfach werden. Hierfür ist das folgende Konzept der Jordandiagramme sehr nützlich, da es die etwas komplizierte Berechnung von B_i auf grafische Art gut veranschaulicht.

Definition 20.10 (Jordandiagramme)**.** Es sei $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Wir wählen einen festen Eigenwert λ von A und setzen zur Abkürzung der Schreibweise $H_r := H_r(A, \lambda)$ für alle $r \in \mathbb{N}$.

Ein **Jordandiagramm** von A zum Eigenwert λ ist dann (wie im Bild unten dargestellt) ein in linksbündigen Zeilen angeordnetes Diagramm von Kästchen, mit von oben nach unten (nicht notwendig streng) monoton fallenden Zeilenlängen, und zusammen mit einem Vektor in K^n in jedem dieser Kästchen, so dass die folgenden beiden Bedingungen erfüllt sind:

(a) (*Zeilenbedingung*) Für alle $r > 0$ bilden die Vektoren der ersten r Zeilen eine Basis von H_r.

(b) (*Spaltenbedingung*) Steht ein Vektor y in einer Spalte des Diagramms unmittelbar über dem Vektor x, so ist $y = (A - \lambda E)x$ (im Diagramm unten ist also z. B. $x_2 = (A - \lambda E)x_3$).

Bemerkung 20.11. Nach der Zeilenbedingung aus Definition 20.10 (a) steht in der ersten Zeile eines Jordandiagramms eine Basis von $H_1 = \mathrm{Eig}(A, \lambda)$, im gesamten Diagramm eine Basis von $H(A, \lambda)$. Die Anzahl der Kästchen in der ersten Zeile eines Jordandiagramms (und damit auch die Anzahl der Spalten) ist also gleich $\dim \mathrm{Eig}(A, \lambda) = \mu_{\mathrm{g}}(A, \lambda)$, die Anzahl der Kästchen im gesamten Diagramm nach Satz 20.6 (b) gleich $\dim H(A, \lambda) = \mu_{\mathrm{a}}(A, \lambda)$.

Bemerkung 20.12 (Alternative Zeilenbedingung)**.** Oft ist es nützlich, die Zeilenbedingung aus Definition 20.10 (a) äquivalent umzuformulieren zu:

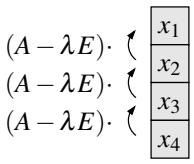
> *(Alternative Zeilenbedingung)* Für alle $r > 0$ liegen die Vektoren der Zeile r in H_r, und ihre Klassen bilden eine Basis von H_r / H_{r-1}.

Diese alternative Bedingung ist nämlich nach Bemerkung 15.19 äquivalent dazu, dass die Vektoren der Zeile r eine Basis von H_{r-1} zu einer Basis von H_r ergänzen – und dies ist ja gerade die ursprüngliche Zeilenbedingung.

Insbesondere heißt dies auch, dass die Vektoren einer Zeile r in $H_r \backslash H_{r-1}$ liegen, also Hauptvektoren der Stufe r sind.

Wir werden in Satz 20.14 noch beweisen, dass Jordandiagramme stets existieren und auch einfach berechenbar sind. Zuerst wollen wir aber sehen, warum gerade eine Basiswahl wie in den Kästchen eines Jordandiagramms zu einer besonders einfachen Abbildungsmatrix führt.

Konstruktion 20.13 (Abbildungsmatrizen aus Jordandiagrammen)**.** Wir betrachten zunächst eine Spalte der Länge m in einem Jordandiagramm zu einer Matrix $A \in \mathrm{Mat}(n \times n, K)$ zum Eigenwert λ; die Vektoren von oben nach unten gelesen seien $x_1, \dots, x_m$ (im Bild rechts ist eine solche Situation für den Fall $m = 4$ dargestellt).



Wegen $x_1 \in H_1$ ist dann $Ax_1 = \lambda x_1$, während für alle $i = 2, \dots, m$ nach der Spaltenbedingung (b) aus Definition 20.10

$$(A - \lambda E)x_i = x_{i-1}, \quad \text{also} \quad Ax_i = \lambda x_i + x_{i-1}$$

gilt. Der Unterraum $U = \mathrm{Lin}(x_1, \dots, x_m)$ ist damit im Sinne von Bemerkung 20.5 (b) A-invariant, d. h. die Abbildung $K^n \to K^n$, $x \mapsto Ax$ lässt sich zu einer Abbildung $U \to U$, $x \mapsto Ax$ einschränken, und die Abbildungsmatrix dieser eingeschränkten Abbildung ist bezüglich der Basis $(x_1, \dots, x_m)$ von U nach Satz 16.23 gleich

$$J_m(\lambda) := \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in \mathrm{Mat}(m \times m, K).$$

Wir nennen diese $m \times m$-Matrix $J_m(\lambda)$, bei der alle Einträge auf der Diagonalen gleich λ, die unmittelbar darüber gleich 1, und alle anderen gleich 0 sind, einen **Jordanblock** (oder manchmal auch ein **Jordankästchen**) der Größe m zum Eigenwert λ.

Wir gehen nun zum gesamten Jordandiagramm zu λ über, nehmen also alle Spalten zusammen. Wählen wir als Basis von $H(A,\lambda)$ die Vektoren im Diagramm, spaltenweise von oben nach unten gelesen, so erhalten wir für die Abbildungsmatrix zu $x \mapsto Ax$ auf $H(A,\lambda)$ analog zu Bemerkung 20.9 eine Blockdiagonalform mit Jordanblöcken wie oben beschrieben, also

$$\begin{pmatrix} J_{m_1}(\lambda) & & 0 \\ & \ddots & \\ 0 & & J_{m_l}(\lambda) \end{pmatrix},$$

wobei jeder Jordanblock einer Spalte im Jordandiagramm entspricht. Die Größen $m_1,\dots,m_l$ der Jordanblöcke sind dabei genau die Längen der Spalten im Jordandiagramm in der gewählten Reihenfolge. Dies ist die einfache Form der Abbildungsmatrix, nach der wir gesucht haben.

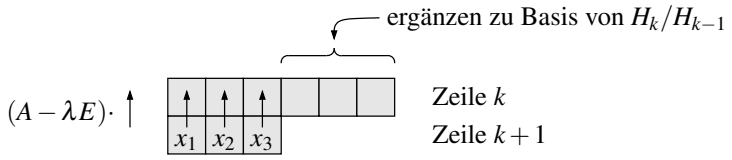
Wir haben nun also gesehen, dass man mit Hilfe von Jordandiagrammen Hauptraumbasen bestimmen kann, die letztlich zu sehr einfachen Abbildungsmatrizen führen. Im nächsten Satz wollen wir daher zeigen, dass solche Jordandiagramme auch wirklich existieren. Der Beweis des Satzes gibt gleichzeitig auch ein einfaches konstruktives Verfahren zur Berechnung eines solchen Diagramms.

Satz 20.14 (Existenz bzw. Berechnung von Jordandiagrammen)**.** *Es seien $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, und λ ein Eigenwert von A. Dann gibt es ein Jordandiagramm von A zum Eigenwert λ.*

Beweis. Wie oben setzen wir wieder $H_k := H_k(A,\lambda)$ für alle $k \in \mathbb{N}$. Es sei $r \in \mathbb{N}$ die kleinste Zahl mit $H_r(A,\lambda) = H(A,\lambda)$, also die benötigte Anzahl Zeilen im Jordandiagramm. Wir konstruieren das gesuchte Diagramm nun nach folgendem Verfahren zeilenweise von unten nach oben:

(a) In die letzte Zeile r schreiben wir beliebige Vektoren in H_r, deren Klassen wie in Bemerkung 20.12 eine Basis von H_r/H_{r-1} bilden – z. B. indem wir wie in Bemerkung 15.19 eine Basis von H_{r-1} zu einer Basis von H_r ergänzen und die hinzugenommenen Vektoren in die letzte Zeile des Diagramms schreiben. (Die dafür bestimmte Basis von H_{r-1} wird dabei im Folgenden nicht mehr benötigt.)

(b) Um für $k = r-1,\dots,1$ die Zeile k aus Zeile $k+1$ zu konstruieren, schreiben wir wie im Bild unten dargestellt zunächst unmittelbar über die Vektoren $x_1,\dots,x_m$ der Zeile $k+1$ die Vektoren $(A-\lambda E)x_1,\dots,(A-\lambda E)x_m$ in Zeile k, und ergänzen diese Vektoren dann so, dass die Klassen der Vektoren in Zeile k insgesamt eine Basis von H_k/H_{k-1} bilden.



Nach Konstruktion sind im ausgefüllten Diagramm dann überall die alternative Zeilenbedingung aus Bemerkung 20.12 und die Spaltenbedingung aus Definition 20.10 (b) erfüllt. Außerdem müssen die Zeilenlängen dann natürlich von oben nach unten monoton fallend sein. Um sicherzustellen, dass Schritt (b) immer funktioniert (wenn das Diagramm unterhalb dieser Zeile schon korrekt ausgefüllt ist), müssen wir aber noch zwei Dinge überprüfen: 45

- Die Vektoren $(A-\lambda E)x_i$ liegen wirklich in H_k für alle $i = 1,\dots,m$: Da x_i in Zeile $k+1$ steht, ist $x_i \in H_{k+1}$, also $(\lambda E - A)^{k+1}x_i = 0$ bzw. $(A-\lambda E)^{k+1}x_i = 0$. Dies bedeutet aber auch $(A-\lambda E)^k(A-\lambda E)x_i = 0$, und damit $(A-\lambda E)x_i \in H_k$.

- Die Klassen der Vektoren $(A-\lambda E)x_1,\dots,(A-\lambda E)x_m$ sind linear unabhängig in H_k/H_{k-1}, so dass wir sie zu einer Basis dieses Raumes ergänzen können: Für $\mu_1,\dots,\mu_m \in K$ gilt

$$\begin{aligned}
& \mu_1\,\overline{(A-\lambda E)x_1}+\dots+\mu_m\,\overline{(A-\lambda E)x_m}=\overline{0} \quad \in H_k/H_{k-1} \\
\Rightarrow\ & \mu_1\,(A-\lambda E)x_1+\dots+\mu_m\,(A-\lambda E)x_m \in H_{k-1} && \text{(Bemerkung 15.14)} \\
\Rightarrow\ & (A-\lambda E)^{k-1}(A-\lambda E)(\mu_1 x_1+\dots+\mu_m x_m)=0 && \text{(Definition von } H_{k-1}) \\
\Rightarrow\ & \mu_1 x_1+\dots+\mu_m x_m \in H_k && \text{(Definition von } H_k) \\
\Rightarrow\ & \mu_1\overline{x_1}+\dots+\mu_m\overline{x_m}=\overline{0} \quad \in H_{k+1}/H_k && \text{(Bemerkung 15.14)} \\
\Rightarrow\ & \mu_1=\dots=\mu_m=0,
\end{aligned}$$

da die Vektoren $\overline{x_1},\dots,\overline{x_m}$ eine Basis von H_{k+1}/H_k bilden und damit in diesem Quotientenraum linear unabhängig sind. □

Ein konkretes Beispiel für dieses Verfahren werden wir in Beispiel 20.18 noch angeben. Zunächst aber wollen wir unsere Ergebnisse zusammenfassen und damit das Hauptergebnis dieses Kapitels zeigen.

Folgerung 20.15 (Jordansche Normalform)**.** *Es sei $A \in \mathrm{Mat}(n\times n,K)$ eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Dann ist A ähnlich zu einer Blockdiagonalmatrix der Form*

$$J=\begin{pmatrix} J_{m_1}(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J_{m_k}(\lambda_k) \end{pmatrix},$$

wobei die $J_{m_1}(\lambda_1),\dots,J_{m_k}(\lambda_k)$ Jordanblöcke wie in Konstruktion 20.13 für gewisse $m_1,\dots,m_k \in \mathbb{N}_{>0}$ und nicht notwendig verschiedene Eigenwerte $\lambda_1,\dots,\lambda_k$ von A sind, d. h. für $i=1,\dots,k$ ist

$$J_{m_i}(\lambda_i)=\begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} \in \mathrm{Mat}(m_i\times m_i,K).$$

*Man nennt eine solche Matrix J die **Jordanform** oder **Jordansche Normalform** von A (wir werden in Aufgabe 20.22 noch sehen, dass sie bis auf die Reihenfolge der Blöcke eindeutig bestimmt ist).*

*Ist $T=(x_1 \mid \dots \mid x_n) \in \mathrm{GL}(n,K)$ eine Matrix mit $J=T^{-1}AT$, so dass J also die Abbildungsmatrix von $K^n \to K^n,\ x \mapsto Ax$ bezüglich der Basis $B=(x_1,\dots,x_n)$ ist, so heißt B eine **Jordanbasis** von A.*

Beweis. Auch der Beweis dieser Folgerung ist konstruktiv: Man bestimme zunächst mit Hilfe des charakteristischen Polynoms χ_A die Eigenwerte von A, und dann zu jedem dieser Eigenwerte ein Jordandiagramm mit dem Verfahren aus Satz 20.14. Aus jedem Diagramm erhalten wir eine Basis des entsprechenden Hauptraums, und so nach der Hauptraumzerlegung aus Satz 20.7 insgesamt eine Basis von K^n. Dies ist dann eine Jordanbasis: Lesen wir die Basisvektoren in den Diagrammen spaltenweise von oben nach unten und schreiben sie in die Matrix T, so ist $T^{-1}AT$ nach Bemerkung 20.9 eine Blockdiagonalmatrix mit einem Block für jeden Eigenwert λ, wobei nach Konstruktion 20.13 jeder dieser Blöcke selbst wieder eine Blockdiagonalform aus Jordanblöcken zu λ ist. □

Bemerkung 20.16 (Bestimmung der Jordanform ohne Jordanbasis)**.** Möchte man zu einer Matrix A (mit in Linearfaktoren zerfallendem charakteristischen Polynom) nur die Jordanform, aber keine Jordanbasis bestimmen, so ist das Verfahren hierfür sehr viel einfacher. Weil wir genau einen Jordanblock der Größe m zum Eigenwert λ für jede Spalte der Länge m im Jordandiagramm zu λ bekommen, benötigen wir zur Bestimmung der Jordanform nämlich nur die äußere Form der Jordandiagramme, aber nicht die in ihnen stehenden Vektoren. Da die äußere Form eines solchen Diagramms zum Eigenwert λ nach Bemerkung 20.12 aber wiederum dadurch eindeutig bestimmt ist, dass Zeile r für alle r die Länge $\dim H_r(A,\lambda)-\dim H_{r-1}(A,\lambda)$ hat, genügt also die Information der Dimensionen aller Räume $H_r(A,\lambda)$, um die Jordanform zu bestimmen.

In vielen Fällen reicht sogar noch viel weniger Information zur Bestimmung der Jordanform aus. So haben z. B. die algebraische und geometrische Vielfachheit eines Eigenwerts eine direkte Interpretation in den Jordandiagrammen und damit auch in der Jordanform:

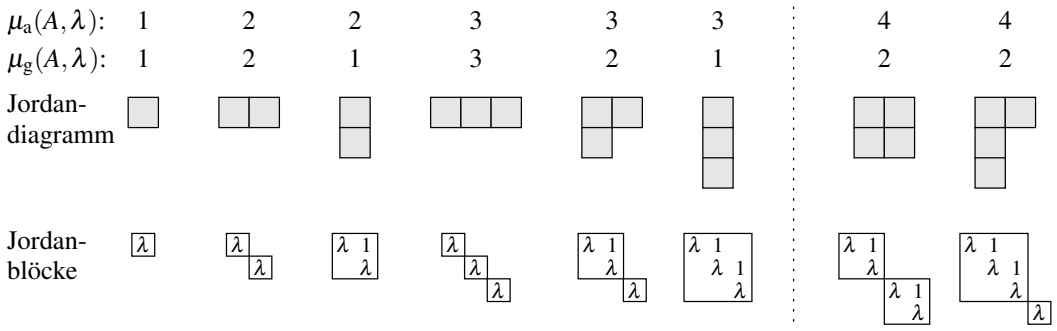
(a) Die algebraische Vielfachheit $\mu_{\mathrm{a}}(A,\lambda)$ ist nach Bemerkung 20.11 gleich der Anzahl der Kästchen im Jordandiagramm zu λ. Natürlich ist dies dann gleichzeitig auch die Summe der Längen aller Spalten in diesem Diagramm, und damit gleich der Summe der Größen aller Jordanblöcke in der Jordanform von A.

(b) Die geometrische Vielfachheit $\mu_{\mathrm{g}}(A,\lambda)$ ist nach Definition die Dimension des Eigenraums $\mathrm{Eig}(A,\lambda) = H_1(A,\lambda)$ und damit nach der Zeilenbedingung aus Definition 20.10 (a) die Länge der ersten Zeile des Jordandiagramms zu λ. Da die Zeilenlängen in einem Jordandiagramm von oben nach unten monoton fallend sind, ist dies aber dasselbe wie die Anzahl der Spalten im Jordandiagramm und damit wie die Anzahl der Jordanblöcke zum Eigenwert λ in der Jordanform von A.

Wir halten also fest:

$$\boxed{\begin{aligned}\mu_{\mathrm{a}}(A,\lambda) &= \text{Anzahl der Kästchen im Jordandiagramm zum Eigenwert } \lambda\\ &= \text{Summe der Größen der Jordanblöcke zum Eigenwert } \lambda \text{ in der Jordanform}\\ \mu_{\mathrm{g}}(A,\lambda) &= \text{Anzahl der Spalten im Jordandiagramm zum Eigenwert } \lambda\\ &= \text{Anzahl der Jordanblöcke zum Eigenwert } \lambda \text{ in der Jordanform}\end{aligned}}$$

Im Fall $\mu_{\mathrm{a}}(A,\lambda) \leq 3$ reichen diese beiden Zahlen in der Tat bereits aus, um das Jordandiagramm und damit auch die von diesem Eigenwert kommenden Jordanblöcke in der Jordanform eindeutig zu bestimmen: Im Bild unten links sind alle möglichen Jordandiagramme mit höchstens 3 Kästchen (und die sich daraus ergebenden Jordanblöcke) angegeben, und diese unterscheiden sich alle in der Anzahl ihrer Kästchen oder Spalten, also durch die algebraische oder geometrische Vielfachheit des Eigenwerts. Erst im Fall $\mu_{\mathrm{a}}(A,\lambda) = 4$ gibt es zum ersten Mal zwei Jordandiagramme, die sich durch diese beiden Zahlen nicht unterscheiden lassen, nämlich die beiden Diagramme unten rechts mit 4 Kästchen und 2 Spalten – und dementsprechend auch zwei verschiedene mögliche Jordanformen. Möchte man zwischen diesen beiden Fällen unterscheiden, muss man auch noch $\dim H_2(A,\lambda)$ berechnen: Diese Zahl ist nach Definition 20.10 ja gleich der Anzahl der Kästchen in den ersten beiden Zeilen des Jordandiagramms, in den beiden Fällen im Bild unten rechts also 4 bzw. 3.



Bemerkung 20.17.

(a) Natürlich gibt es einen zu Folgerung 20.15 analogen Satz auch wieder für Endomorphismen $f\colon V \to V$ eines endlich erzeugten Vektorraums V: Zerfällt χ_f in Linearfaktoren, so gibt es eine Jordanbasis B von V für f – also eine Basis, so dass die zugehörige Abbildungsmatrix A_f^B in Jordanform ist.

(b) Beachte, dass der Fall von diagonalisierbaren Matrizen, also Folgerung 19.40, im Satz über die Jordansche Normalform enthalten ist: Ist A eine quadratische Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt, so ergibt sich aus Folgerung 20.15 und Bemerkung

20.16

$$\begin{aligned} A \text{ diagonalisierbar} &\Leftrightarrow \text{alle Jordanblöcke haben die Größe 1} \\ &\Leftrightarrow \text{alle Spalten der Jordandiagramme haben die Länge 1} \\ &\Leftrightarrow \text{alle Jordandiagramme haben genauso viele Kästchen wie Spalten} \\ &\Leftrightarrow \mu_g(A,\lambda) = \mu_a(A,\lambda) \text{ für alle Eigenwerte } \lambda, \end{aligned}$$

und damit Folgerung 19.40.

Beispiel 20.18. Es sei

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -4 & 4 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 3 \end{pmatrix} \in \mathrm{Mat}(4\times 4,\mathbb{R}).$$

Das charakteristische Polynom von A berechnet sich wohl am einfachsten durch Laplace-Entwicklung gemäß Satz 18.15, zunächst nach der letzten Spalte und dann nach der letzten Zeile:

$$\begin{aligned} \chi_A(t) &= \det \begin{pmatrix} t & -1 & -1 & 0 \\ 4 & t-4 & -2 & 0 \\ 0 & 0 & t-2 & 0 \\ -1 & 0 & -1 & t-3 \end{pmatrix} = (t-3)\cdot \det \begin{pmatrix} t & -1 & -1 \\ 4 & t-4 & -2 \\ 0 & 0 & t-2 \end{pmatrix} \\ &= (t-3)(t-2)\cdot \det \begin{pmatrix} t & -1 \\ 4 & t-4 \end{pmatrix} = (t-3)(t-2)(t^2-4t+4) = (t-3)(t-2)^3. \end{aligned}$$

Da χ_A in Linearfaktoren zerfällt, besitzt A nach Folgerung 20.15 also eine Jordanform.

Wollen wir diese berechnen, müssen wir die Jordandiagramme zu den beiden Eigenwerten $\lambda_1 = 2$ und $\lambda_2 = 3$ bestimmen. Für $\lambda_2 = 3$ ist dies einfach: Aus $\mu_a(A,3) = 1$ folgt natürlich bereits, dass das zugehörige Jordandiagramm aus nur einem Kästchen besteht und die Jordanform von A daher genau einen Jordanblock der Größe 1 zum Eigenwert 3 hat. Für den Eigenwert $\lambda_1 = 2$ hingegen ist $\mu_a(A,2) = 3$, und daher gibt es für das Jordandiagramm noch die drei im Bild von Bemerkung 20.16 aufgelisteten Möglichkeiten, die drei Kästchen im Jordandiagramm anzuordnen. Um zu entscheiden, welcher dieser Fälle hier vorliegt, berechnen wir die geometrische Vielfachheit $\mu_g(A,2)$: Mit dem Gauß-Verfahren ergibt sich

$$H_1(A,2) = \mathrm{Ker}(2E-A) = \mathrm{Ker}\begin{pmatrix} 2 & -1 & -1 & 0 \\ 4 & -2 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & -1 \end{pmatrix} = \mathrm{Lin}\left(\underbrace{\begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}}_{=:x_1}, \underbrace{\begin{pmatrix} 1 \\ 3 \\ -1 \\ 0 \end{pmatrix}}_{=:x_2} \right) \qquad (*)$$

und damit $\mu_g(A,2) = 2$. Das Jordandiagramm zum Eigenwert 2 hat gemäß Bemerkung 20.16 also 3 Kästchen und 2 Spalten und muss damit wie im Bild am Ende dieses Beispiels in der ersten Spalte zwei und in der zweiten Spalte ein Kästchen haben. Zum Eigenwert 2 gibt es also zwei Jordanblöcke, deren Größen gerade diese Spaltenlängen 2 bzw. 1 sind. Die Jordanform von A ist damit

$$J = \left(\begin{array}{cc|cc} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{array}\right).$$

Wollen wir auch noch eine Jordanbasis bestimmen, müssen wir die beiden Jordandiagramme zu den Eigenwerten 2 und 3 noch mit Vektoren füllen. Für den Eigenwert $\lambda_2 = 3$ ist dies wieder einfach: Da der vierte Einheitsvektor offensichtlich ein Eigenvektor zu diesem Eigenwert ist, können wir ihn wie im Bild unten in das eine Kästchen des Jordandiagramms schreiben. Um auch das Jordandiagramm für $\lambda_1 = 2$ zu füllen, verwenden wir das Verfahren aus Satz 20.14 und beginnen also in der zweiten Zeile des Diagramms, wofür wir eine Basis für den eindimensionalen Raum $H_2(A,2)/H_1(A,2)$ brauchen. Wir müssen also erst einmal den verallgemeinerten Eigenraum $H_2(A,2)$ bestimmen. Dies

können wir durch explizite Berechnung von $H_2(A,2) = \operatorname{Ker}(2E-A)^2$ mit dem Gauß-Verfahren machen und erhalten

$$H_2(A,2) = \operatorname{Ker}(2E-A)^2 = \operatorname{Ker}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 2 & 1 \end{pmatrix} = \operatorname{Lin}\left(\underbrace{\begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix}}_{=x_1}, \underbrace{\begin{pmatrix} 1 \\ 3 \\ -1 \\ 0 \end{pmatrix}}_{=x_2}, \underbrace{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}}_{=:x_3}\right)$$

Offensichtlich ist x_3 ein Vektor, der eine Basis von $H_1(A,2)$ zu einer von $H_2(A,2)$ ergänzt. Wir schreiben ihn also in das untere Kästchen des Jordandiagramms. Direkt darüber gehört nun nach dem Verfahren aus Satz 20.14 der Vektor

$$(A-2E)\cdot x_3 = \begin{pmatrix} -2 & 1 & 1 & 0 \\ -4 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 0 \\ 1 \end{pmatrix}.$$

Beachte, dass dieser Vektor nach $(*)$ tatsächlich in $H_1(A,2)$ liegt (dies ist eine gute Kontrolle der Rechnung – wir hätten irgendwo einen Rechenfehler gemacht, wenn dies nicht so wäre). Für das rechte Kästchen des Jordandiagramms müssen wir ihn noch zu einer Basis von $H_1(A,2)$ ergänzen, nach $(*)$ z. B. durch x_2.

Insgesamt haben wir also die folgenden Jordandiagramme erhalten:

$\begin{pmatrix} -1 \\ -2 \\ 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 3 \\ -1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$
$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$	$\lambda_1 = 2$	$\lambda_2 = 3$

Für eine Jordanbasis müssen wir die Vektoren in diesen Kästchen nun nur noch spaltenweise von oben nach unten lesen bzw. sie in dieser Reihenfolge als Spalten in die Transformationsmatrix T schreiben. Dabei müssen wir die Spalten der Diagramme in der gleichen Reihenfolge durchgehen, in der wir oben die Jordanblöcke in J angeordnet haben:

$$T := \begin{pmatrix} -1 & 1 & 1 & 0 \\ -2 & 1 & 3 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Damit gilt dann $T^{-1}AT = J$ nach Folgerung 20.15 (was wir durch direkte Berechnung von T^{-1} und $T^{-1}AT$ natürlich auch explizit überprüfen könnten).

Aufgabe 20.19. Berechne die Jordanschen Normalformen der reellen Matrizen

$$A = \begin{pmatrix} -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & -1 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 & 4 \end{pmatrix}.$$

Für die Matrix A bestimme man dabei zusätzlich eine Jordanbasis; für die Matrix B hingegen versuche man, mit möglichst wenig Rechenaufwand lediglich die Jordanform zu ermitteln.

Aufgabe 20.20. Für gegebenes $n \in \mathbb{N}$ sei V der Vektorraum aller reellen Polynome vom Grad höchstens n. Berechne die Jordanform der linearen Abbildung $f\colon V \to V$ mit $f(\varphi)(x) = \varphi(x+1)$.

Aufgabe 20.21 (Anwendung der Jordanform auf Systeme von Differentialgleichungen)**.** In dieser Aufgabe wollen wir ein vermutlich recht unerwartetes Beispiel für die Anwendung der Jordanform aus dem Bereich der Analysis geben. Das Ziel soll es sein, reelle differenzierbare Funktionen $f_1, f_2, f_3 \colon \mathbb{R} \to \mathbb{R}$ zu bestimmen, so dass $f_1(0) = f_2(0) = f_3(0) = 1$ und

$$\begin{aligned} f_1' &= f_2 + 2f_3 \\ f_2' &= f_1 + f_2 + 3f_3 \\ f_3' &= -f_1 - f_3 \end{aligned}$$

(an jeder Stelle $x \in \mathbb{R}$) gilt, wobei f_i' wie üblich die Ableitung von f_i bezeichnet.

Derartige Systeme von sogenannten *Differentialgleichungen*, die also in jedem Punkt die Änderung von Funktionen durch die Funktionswerte selbst ausdrücken, kommen in der Praxis überall vor. Ausgehend von einem Anfangswert (hier bei $x = 0$) kann man mit ihnen die gesamten Funktionen f_i rekonstruieren.

Zur Lösung schreibe man die gegebenen Gleichungen in Matrixform $f' = A \cdot f$ mit $A \in \mathrm{Mat}(3 \times 3, \mathbb{R})$ und bringe A in Jordanform, d. h. bestimme eine Matrix $T \in \mathrm{GL}(3, \mathbb{R})$, so dass $T^{-1}AT = J$ eine Jordanmatrix ist. Wenn ihr die Gleichungen dann umschreibt in Gleichungen für $g = T^{-1}f$ (also $f = Tg$), sollte sich dieses neue Differentialgleichungssystem für g (mit den passenden Werten bei $x = 0$) leicht lösen lassen.

Aufgabe 20.22 (Eindeutigkeit der Jordanform)**.** Wir wollen nun die bereits in Folgerung 20.15 behauptete Eindeutigkeit der Jordanform (bis auf die Reihenfolge der Jordanblöcke) beweisen. Man zeige dazu:

(a) Sind A und B zwei ähnliche Matrizen, so gilt $\dim H_r(A, \lambda) = \dim H_r(B, \lambda)$ für alle $r \in \mathbb{N}$ und $\lambda \in K$.

(b) Ist A eine Matrix in Jordanscher Normalform, λ ein Eigenwert von A und $k \in \mathbb{N}_{>0}$, so ist die Anzahl der Jordanblöcke der Größe r zum Eigenwert λ in A genau

$$2 \dim H_r(A, \lambda) - \dim H_{r-1}(A, \lambda) - \dim H_{r+1}(A, \lambda).$$

(c) Zwei Matrizen in Jordanscher Normalform sind genau dann ähnlich zueinander, wenn sie aus den gleichen Jordanblöcken, nur evtl. in anderer Reihenfolge bestehen.

Zum Abschluss dieses Abschnitts wollen wir schließlich noch ein paar Aufgaben betrachten, in denen man die Nützlichkeit der einfachen Matrixdarstellung der Jordanform in theoretischen Problemen sieht. Ihnen allen ist gemeinsam, dass man die zu zeigende Aussage für Jordanblöcke bzw. Matrizen in Jordanform recht einfach sehen und sie dann mit Hilfe einer Ähnlichkeitstransformation problemlos auf beliebige Matrizen (mit zerfallendem charakteristischen Polynom) übertragen kann – während ein direkter Beweis für beliebige Matrizen deutlich schwieriger wäre.

Aufgabe 20.23. Beweise, dass jede quadratische komplexe Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ zu ihrer transponierten Matrix A^T ähnlich ist.

Aufgabe 20.24. Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{C})$. Zeige, dass $\dim\{B \in \mathrm{Mat}(n \times n, \mathbb{C}) : AB = BA\} \geq n$.

Aufgabe 20.25. Eine quadratische Matrix $N \in \mathrm{Mat}(n \times n, K)$ heißt *nilpotent*, wenn es ein $k \in \mathbb{N}$ gibt mit $N^k = 0$.

(a) Zeige, dass sich jede komplexe Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ als Summe $A = D + N$ schreiben lässt, wobei D diagonalisierbar und N nilpotent ist, sowie $DN = ND$ gilt.

(b) Berechne mit Hilfe von (a) für alle $n \in \mathbb{N}$ die Potenzen

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}^n \quad \text{und} \quad \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix}^n$$

in $\mathrm{Mat}(2 \times 2, \mathbb{R})$ durch eine direkte Rechnung (also ohne erst durch Ausprobieren eine Formel zu raten und sie danach durch Induktion zu beweisen).

(c) Wir definieren eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ rekursiv durch

$$a_0 = a_1 = 1 \quad \text{und} \quad a_{n+2} = 4a_{n+1} - 4a_n \quad \text{für alle } n \in \mathbb{N}.$$

Bestimme eine explizite Formel für alle a_n.

(Hinweis: Die Rekursionsgleichung ist offensichtlich äquivalent zu

$$\begin{pmatrix} a_{n+2} \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 4 & -4 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} \quad \text{für alle } n \in \mathbb{N}.)$$

20.C Minimalpolynome

Als eine interessante Anwendung der Jordanschen Normalform wollen wir in diesem Abschnitt Polynomausdrücke in Matrizen betrachten. Hier sind zunächst einmal zwei Beispiele dafür.

Beispiel 20.26.

(a) Wenn wir im Polynom $p(t) = (t-\lambda)^r$ „für t die Matrix A einsetzen“, erhalten wir $(A-\lambda E)^r$. Diese Matrix haben wir in diesem Kapitel bereits oft betrachtet, denn ihr Kern ist ja gerade der verallgemeinerte Eigenraum $H_r(A,\lambda)$.

(b) Es sei A eine quadratische Matrix mit $A^2 = A$, also eine „Nullstelle des Polynoms $t^2 - t$“. Wir können uns fragen, was wir dann über A aussagen können, also ob wir die quadratische Gleichung $t^2 - t = 0$ „im Matrizenraum lösen“ können. In der Tat werden wir die allgemeine Matrixlösung dieser Gleichung in Beispiel 20.36 angeben können. Wir können hier aber schon einmal am Beispiel der reellen Matrix

$$A = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} \quad \text{mit} \quad A^2 - A = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

sehen, dass wir (im Gegensatz zur Lösung der Gleichung $t^2 - t = t(t-1) = 0$ in K) hier nicht nur die „offensichtlichen“ Lösungen $A = 0$ und $A = E$ erhalten werden.

Als Erstes sollten wir aber das Einsetzen einer Matrix in ein Polynom exakt definieren.

Definition 20.27 (Polynomausdrücke in Matrizen)**.** Es seien $p(t) = c_k t^k + \cdots + c_1 t + c_0$ ein Polynom mit Koeffizienten in K und $A \in \mathrm{Mat}(n \times n, K)$. Dann setzen wir

$$p(A) := c_k A^k + \cdots + c_1 A + c_0 E \quad \in \mathrm{Mat}(n \times n, K).$$

46

Im Rest dieses Abschnitts wollen wir der Einfachheit halber annehmen, dass unser Grundkörper K gleich $\mathbb{C}$ ist, so dass jedes Polynom $p \neq 0$ nach dem Fundamentalsatz 6.11 der Algebra in Linearfaktoren zerfällt und somit als $p(t) = c(t-\lambda_1)\cdots(t-\lambda_k)$ für $c, \lambda_1, \ldots, \lambda_k \in \mathbb{C}$ mit $c \neq 0$ geschrieben werden kann. Einsetzen einer Matrix ergibt dann $p(A) = c(A-\lambda_1 E)\cdots(A-\lambda_k E)$. Da wir jede reelle Matrix auch als komplexe auffassen können, ist diese Beschränkung auf $\mathbb{C}$ für uns kein wesentliches Problem. Wie schon in der Einleitung zu diesem Kapitel erwähnt, kann man in der Tat sogar zu jedem Körper einen größeren finden, in dem Polynome immer in Linearfaktoren zerfallen. Mit einer solchen Aussage würden sich die Ergebnisse, die wir jetzt zeigen werden, dann auch auf beliebige Körper übertragen lassen.

Wir wollen nun zu einer gegebenen Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ alle komplexen Polynome p finden, für die $p(A) = 0$ gilt. Es wird sich dabei herausstellen, dass es unter diesen Polynomen ein *eindeutiges* normiertes Polynom mit minimalem Grad gibt, das wir dann das Minimalpolynom von A nennen werden.

Für die konkrete Berechnung, ob $p(A) = 0$ gilt, ist es natürlich nützlich, wenn die Matrix A eine möglichst einfache Form hat. Daher werden wir diese Frage zunächst für Jordanblöcke untersuchen (wo sie einfach zu beantworten ist), dann für Matrizen in Jordanform, und schließlich mit Hilfe von Ähnlichkeitstransformationen für beliebige Matrizen.

Lemma 20.28. *Es seien $p \neq 0$ ein komplexes Polynom und $A = J_m(\lambda) \in \mathrm{Mat}(m \times m, \mathbb{C})$ ein Jordanblock wie in Konstruktion 20.13.*

Dann gilt $p(A) = 0$ genau dann, wenn λ in p eine Nullstelle der Vielfachheit mindestens m ist.

Beweis. Wir können p als $p(t) = c(t-\lambda_1)^{a_1} \cdots (t-\lambda_k)^{a_k}$ mit $c \in \mathbb{C}\backslash\{0\}$, $a_1, \dots, a_k \in \mathbb{N}$ und verschiedenen $\lambda_1, \dots, \lambda_k$ faktorisieren, so dass also

$$p(A) = c\,(A-\lambda_1 E)^{a_1} \cdots (A-\lambda_k E)^{a_k}. \tag{1}$$

Ein einzelner Faktor in $p(A)$ hat dabei die Form

$$A - \lambda_i E = \begin{pmatrix} \lambda-\lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda-\lambda_i \end{pmatrix}. \tag{2}$$

Für alle i mit $\lambda_i \neq \lambda$ hat diese Matrix Determinante $(\lambda-\lambda_i)^m \neq 0$ und ist damit invertierbar. In (1) kann also höchstens dann $p(A) = 0$ gelten, wenn $\lambda_i = \lambda$ für ein i gilt; und in diesem Fall ist $p(A) = 0$ äquivalent zu $(A-\lambda_i E)^{a_i} = 0$. Die Matrix $A - \lambda_i E$ hat dann die Form (2) mit Nullen auf der Diagonalen, und bildet damit die Einheitsvektoren $e_1, \dots, e_m$ auf

$$(A-\lambda_i E)\,e_j = \begin{cases} e_{j-1} & \text{für } j > 1, \\ 0 & \text{für } j = 1 \end{cases}$$

ab, verschiebt also jeden Einheitsvektor auf den vorherigen und bildet e_1 auf 0 ab. Die a_1-fache Anwendung $(A-\lambda_i E)^{a_1}$ dieser Operation bildet damit e_j ab auf e_{j-a_1} für $a_1 < j$ und auf 0 sonst. Sie ist also genau dann gleich 0, bildet also alle Einheitsvektoren (insbesondere e_m) auf 0 ab, wenn $a_1 \geq m$ gilt, also wie behauptet wenn die Vielfachheit der Nullstelle λ in p mindestens m ist. □

Bemerkung 20.29. Es seien p ein komplexes Polynom und $A \in \mathrm{Mat}(n \times n, \mathbb{C})$.

(a) ($p(A)$ für Blockdiagonalmatrizen) Ist A in Blockdiagonalform mit quadratischen Blöcken $A_1, \dots, A_k$, so gilt aufgrund der Blockmultiplikation aus Aufgabe 16.9

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix} \quad \Rightarrow \quad A^i = \begin{pmatrix} A_1^i & & 0 \\ & \ddots & \\ 0 & & A_k^i \end{pmatrix} \quad \text{für alle } i \in \mathbb{N},$$

und damit auch für den Ausdruck $p(A)$, der ja eine Linearkombination solcher Potenzen ist,

$$p(A) = \begin{pmatrix} p(A_1) & & 0 \\ & \ddots & \\ 0 & & p(A_k) \end{pmatrix}.$$

Also ist $p(A) = 0$ genau dann, wenn $p(A_j) = 0$ für alle $j = 1, \dots, k$.

(b) ($p(A)$ für ähnliche Matrizen) Ist $T \in \mathrm{GL}(n, \mathbb{C})$ und damit $A' := T^{-1}AT$ ähnlich zu A, so gilt $p(A') = 0$ genau dann, wenn $p(A) = 0$: Es ist nämlich

$$(A')^i = \underbrace{(T^{-1}AT)(T^{-1}AT)\cdots(T^{-1}AT)}_{i\text{-mal}} = T^{-1}A^iT$$

für alle $i \in \mathbb{N}$, da sich die Matrixprodukte TT^{-1} hier in der Mitte herauskürzen, und damit für $p(t) = c_n t^n + \cdots c_1 t + c_0$

$$p(A') = T^{-1}\,(c_n A^n + \cdots c_1 A + c_0 E)\,T = T^{-1}\,p(A)\,T,$$

was wegen der Invertierbarkeit von T genau dann gleich 0 ist, wenn $p(A) = 0$ gilt.

Folgerung und Definition 20.30 (Minimalpolynom)**.** *Es seien $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte einer Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$. Ferner sei a_i für $i = 1, \dots, k$ die maximale Größe eines Jordanblocks zum Eigenwert λ_i in der Jordanform von A.*

(a) *Es gibt ein eindeutiges normiertes Polynom p_A minimalen Grades mit $p_A(A) = 0$, nämlich*

$$p_A(t) = (t-\lambda_1)^{a_1} \cdots (t-\lambda_k)^{a_k}.$$

*Man nennt p_A das **Minimalpolynom** von A.*

(b) *Für ein beliebiges Polynom p gilt*

$$p(A) = 0 \quad \Leftrightarrow \quad \text{es gibt ein Polynom } q \text{ mit } p = q \cdot p_A$$

(man sagt dafür auch, dass p ein Vielfaches von p_A bzw. p_A ein Teiler von p ist).

Beweis. Nach Folgerung 20.15 und Bemerkung 20.29 (b) können wir annehmen, dass A eine Matrix in Jordanform ist. Ist $p \neq 0$ dann ein beliebiges komplexes Polynom, so ist nach Bemerkung 20.29 (a) genau dann $p(A) = 0$, wenn p ausgewertet an jedem Jordanblock von A gleich 0 ist, was wiederum nach Lemma 20.28 genau dann der Fall ist, wenn die Nullstellenordnung von allen λ_i in p mindestens so groß ist wie jeder Jordanblock zu λ_i, also mindestens so groß wie a_i. Die Polynome p mit $p(A) = 0$ sind also genau diejenigen der Form

$$p(t) = q \cdot (t-\lambda_1)^{a_1} \cdots (t-\lambda_k)^{a_k}$$

für ein Polynom q. Dies zeigt bereits beide Teile der Folgerung. □

Bemerkung 20.31.

(a) Nach Folgerung 20.30 (a) sind die Nullstellen des Minimalpolynoms p_A genau alle Eigenwerte von A.

(b) Da jeder Jordanblock der Größe m einer Spalte der Länge m im Jordandiagramm entspricht, ist die maximale Größe eines solchen Blocks genau die Anzahl der Zeilen im Jordandiagramm. Zusammen mit Bemerkung 20.16 können wir also für jeden Eigenwert einer komplexen Matrix als Merkregel für die Jordandiagramme festhalten:

Anzahl der Kästchen im Jordandiagramm = algebraische Vielfachheit
Anzahl der Spalten im Jordandiagramm = geometrische Vielfachheit
Anzahl der Zeilen im Jordandiagramm = Vielfachheit im Minimalpolynom

Beispiel 20.32. Beachte, dass Folgerung 20.30 das Minimalpolynom einer Matrix A auf zwei ganz unterschiedliche Arten beschreibt: als normiertes Polynom kleinsten Grades, das beim Einsetzen von A Null ergibt, und als Polynom, dessen Nullstellenordnung bei jedem λ die maximale Größe eines Jordanblocks zum Eigenwert λ in der Jordanform von A ist. Um die Nützlichkeit beider Beschreibungen zu sehen, betrachten wir noch einmal die reellen Matrizen

$$A = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0 & 1 & 1 & 0 \\ -4 & 4 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 3 \end{pmatrix}.$$

(a) Für die Matrix A haben wir in Beispiel 20.26 (a) bereits gesehen, dass $p(A) = 0$ für $p(t) = t^2 - t = t(t-1)$. Nach Folgerung 20.30 (b) muss dieses Polynom ein Vielfaches des Minimalpolynoms p_A sein. Für p_A kommen also nur die Polynome $t(t-1)$, t und $t-1$ in Frage. Da die letzten beiden Polynome beim Einsetzen von A aber offensichtlich nicht 0 ergeben (es ist $A \neq 0$ und $A - E \neq 0$), ist $p_A(t) = t^2 - t$.

In diesem Fall konnten wir das Minimalpolynom also einfach berechnen, ohne irgendetwas über die Eigenwerte oder die Jordanform von A zu wissen.

(b) Von der Matrix B haben wir in Beispiel 20.18 schon die Jordanform berechnet; sie bestand aus zwei Jordanblöcken zum Eigenwert 2 mit den Größen 1 und 2 sowie einem Jordanblock zum Eigenwert 3 der Größe 1. Mit Folgerung 20.30 (a) können wir daraus also sofort ablesen, dass $p_B(t) = (t-2)^2(t-3)$.

Hier konnten wir durch die Kenntnis der Jordanform also das Minimalpolynom bestimmen, ohne irgendwelche Polynomausdrücke in Matrizen berechnen zu müssen.

Bemerkung 20.33 (Minimalpolynome reeller Matrizen sind reell)**.** Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine reelle Matrix, deren charakteristisches Polynom über $\mathbb{R}$ nicht notwendig in Linearfaktoren zerfällt. In der Konstruktion des Minimalpolynoms in Folgerung 20.30 treten dann evtl. komplexe Eigenwerte von A auf, so dass nicht mehr klar ist, ob p_A ein reelles Polynom ist. In der Tat ist dies aber immer der Fall: Ist $p_A(t) = t^m + c_{m-1}t^{m-1} + \dots + c_1 t + c_0$ (mit zunächst evtl. komplexen Koeffizienten), so gilt für das komplex konjugierte Polynom $\overline{p_A}(t) = t^m + \overline{c_{m-1}}t^{m-1} + \dots + \overline{c_1}t + \overline{c_0}$ wegen $\overline{A} = A$

$$\overline{p_A}(A) = A^m + \overline{c_{m-1}}A^{m-1} + \dots + \overline{c_1}A + \overline{c_0}E = \overline{A^m + c_{m-1}A^{m-1} + \dots + c_1 A + c_0 E} = \overline{p_A(A)} = 0.$$

Da dieses Polynom $\overline{p_A}$ außerdem normiert ist und den gleichen Grad hat wie p_A, muss es nach der Eindeutigkeitsaussage in Folgerung 20.30 (a) bereits das Minimalpolynom sein. Also ist $\overline{p_A} = p_A$, d. h. p_A ist ein reelles Polynom.

Aus unserer Charakterisierung des Minimalpolynoms ergeben sich zwei unmittelbare interessante Folgerungen.

Folgerung 20.34 (**Satz von Cayley-Hamilton**)**.** *Für jede quadratische Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ gilt $\chi_A(A) = 0$.*

Beweis. Das charakteristische Polynom von A hat die Form

$$\chi_A(t) = (t - \lambda_1)^{\mu_a(A,\lambda_1)} \cdots (t - \lambda_k)^{\mu_a(A,\lambda_k)},$$

wobei $\lambda_1, \dots, \lambda_k$ die verschiedenen Eigenwerte von A sind. Für alle i ist aber nun $\mu_a(A, \lambda_i)$ nach Bemerkung 20.31 (b) die Anzahl der Kästchen im Jordandiagramm zu λ_i, und damit sicher mindestens so groß wie die Anzahl der Zeilen darin, also wie die Vielfachheit von λ_i im Minimalpolynom p_A. Wie in Folgerung 20.30 ist χ_A also ein Vielfaches des Minimalpolynoms p_A, und damit ist $\chi_A(A) = 0$. □

Folgerung 20.35 (Diagonalisierbarkeit und Minimalpolynom)**.** *Eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ ist genau dann diagonalisierbar, wenn ihr Minimalpolynom p_A nur einfache Nullstellen hat.*

Beweis. Natürlich ist A genau dann diagonalisierbar, wenn alle Jordanblöcke in der Jordanform von A die Größe 1 haben, das Jordandiagramm zu jedem Eigenwert also nur Spalten der Länge 1 und damit genau eine Zeile besitzt. Nach Bemerkung 20.31 (b) ist dies äquivalent dazu, dass alle Eigenwerte einfache Nullstellen von p_A sind. □

Beispiel 20.36 (Matrizen mit $A^2 = A$)**.** Wir können nun unser Beispiel 20.26 (b) vom Anfang dieses Abschnitts noch einmal aufgreifen und die Frage klären, welche Matrizen $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ die Gleichung $A^2 = A$ erfüllen.

Wie in Beispiel 20.32 (a) ergibt sich auch hier zunächst wieder aus Folgerung 20.30 (b), dass das Polynom $t^2 - t = t(t-1)$ ein Vielfaches des Minimalpolynoms p_A sein muss, also dass für p_A nur $t(t-1)$, t und $t-1$ in Frage kommen. Insbesondere hat p_A damit in jedem Fall nur einfache Nullstellen, so dass A nach Folgerung 20.35 diagonalisierbar sein muss. Außerdem können die Eigenwerte von A nur 0 oder 1 sein, weil sie nach Bemerkung 20.31 (a) Nullstellen von p_A und damit von $t(t-1)$ sein müssen. Es gibt also ein $T \in \mathrm{GL}(n, \mathbb{C})$, so dass $T^{-1}AT$ eine Diagonalmatrix mit Diagonaleinträgen 0 oder 1 ist, d. h. es muss

$$T^{-1}AT = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{und damit} \quad A = T\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)T^{-1} \qquad (*)$$

für ein $r \in \{0, \dots, n\}$ gelten. Wegen

$$A^2 = T\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)T^{-1} \cdot T\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)T^{-1} = T\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)^2 T^{-1} = T\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)T^{-1} = A$$

sind dies auch wirklich Lösungen der gegebenen Matrixgleichung $A^2 = A$, so dass $(*)$ also die allgemeine Lösung dieser Gleichung ist. Beachte, dass dies die „offensichtlichen“ Lösungen $A = 0$ und $A = E$ enthält (für $r = 0$ bzw. $r = n$). Wir sehen hier aber, dass es noch viele weitere Lösungen gibt.

Bemerkung 20.37 (Minimalpolynome für Endomorphismen)**.** Natürlich lassen sich auch die Konzepte und Ergebnisse dieses Abschnitts wieder auf die übliche Art auf Endomorphismen $f\colon V \to V$ eines endlich-dimensionalen komplexen Vektorraums V übertragen: Für ein gegebenes komplexes Polynom $p(t) = c_n t^n + \cdots + c_1 t + c_0$ setzt man

$$p(f) := c_n f^n + \cdots + c_1 f + c_0 \,\mathrm{id} \quad \in \mathrm{End}(V),$$

wobei f^k für $k = 1, \ldots, n$ für die k-fache Verkettung von f mit sich selbst steht. Auch hier erhält man dann ein eindeutig bestimmtes normiertes Polynom p_f minimalen Grades mit $p_f(f) = 0$, das *Minimalpolynom* von f. Es hat die gleiche Charakterisierung durch die Jordanform wie in Folgerung 20.30, und auch der Satz von Cayley-Hamilton aus Folgerung 20.34 und die Charakterisierung der Diagonalisierbarkeit aus Folgerung 20.35 gelten ganz analog.

Aufgabe 20.38. Es seien $A, B \in \mathrm{Mat}(6 \times 6, \mathbb{C})$,so dass gilt:

(a) A hat Rang 6 und erfüllt $A^4 + 3A^2 = 3A^3 + A$, und einer der Eigenräume von A ist zweidimensional.

(b) $\mathrm{Ker}\, B = \mathrm{Im}\, B$.

Bestimme das Minimalpolynom und die Jordansche Normalform dieser beiden Matrizen.

Aufgabe 20.39. Untersuche, ob es eine komplexe 4×4-Matrix A bzw. B gibt mit

$$\text{(a) } A^5 = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \qquad \text{(b) } B^5 + 5B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}.$$

(Es ist nicht notwendig, im Fall der Existenz eine solche Matrix anzugeben.)

Aufgabe 20.40. Es sei $A \in \mathrm{GL}(n, \mathbb{K})$. Zeige, dass es ein Polynom p mit Koeffizienten in $\mathbb{K}$ gibt, so dass $A^{-1} = p(A)$.

Aufgabe 20.41. Man zeige: Ist $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ eine invertierbare Matrix, so dass A^m für ein $m \in \mathbb{N}_{>0}$ diagonalisierbar ist, so ist auch A diagonalisierbar. 47

21. Euklidische und unitäre Räume

Wir wollen uns nun mit einem ganz anderen Thema beschäftigen, nämlich wie man Längen von Vektoren und Winkel zwischen zwei Vektoren berechnen (und überhaupt erst einmal definieren) kann. Zur Motivation betrachten wir dazu zunächst einmal den sehr einfachen Fall des Vektorraums $\mathbb{R}^2$, in dem sich diese beiden Fragen mit Hilfe von Elementargeometrie und Schulmathematik leicht beantworten lassen.

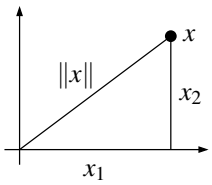
Beispiel 21.1 (Längen und Winkel in $\mathbb{R}^2$)**.** Wie ihr sicher aus der Schule wisst, ist das wesentliche Hilfsmittel für die Längen- und Winkelmessung in $\mathbb{R}^2$ das sogenannte *Skalarprodukt*

$$\langle x,y\rangle = x_1y_1 + x_2y_2 \in \mathbb{R} \qquad \text{für } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},\ y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathbb{R}^2.$$

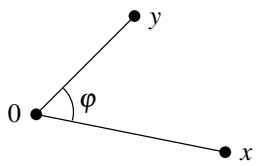
So ergibt sich z. B. wie im Bild unten links dargestellt aus dem Satz des Pythagoras, dass die Länge eines Vektors $x \in \mathbb{R}^2$ durch den Ausdruck

$$\sqrt{x_1^2 + x_2^2} = \sqrt{\langle x,x\rangle}$$

gegeben ist, den wir im Folgenden kurz als $\|x\|$ schreiben werden.



$\|x\| = \sqrt{\langle x,x\rangle}$



$\frac{y}{\|y\|} = e^{i\varphi} \cdot \frac{x}{\|x\|}$

Wollen wir den Winkel φ zwischen zwei Vektoren $x,y \in \mathbb{R}^2 \backslash \{0\}$ wie im Bild oben rechts berechnen, betrachten wir dazu am besten zunächst einmal die Vektoren $\frac{x}{\|x\|}$ und $\frac{y}{\|y\|}$, die in die gleiche Richtung wie x bzw. y zeigen, aber die Länge 1 haben. Fassen wir dann $x = x_1 + ix_2$ und $y = y_1 + iy_2$ als Elemente der komplexen Ebene $\mathbb{C} = \mathbb{R}^2$ auf, so folgt aus den Bemerkungen 6.5 und 9.11, dass

$$\frac{y}{\|y\|} = e^{i\varphi} \cdot \frac{x}{\|x\|}$$

ist, da sich die Winkel bei der komplexen Multiplikation addieren und $e^{i\varphi}$ eine Zahl mit Winkel φ und Betrag 1 ist. Einfache Umformungen in $\mathbb{C}$ ergeben nun wegen $\|x\|^2 = \bar{x}x$

$$\frac{\bar{x}y}{\bar{x}x} \cdot \frac{\|x\|}{\|y\|} = e^{i\varphi} \qquad \Rightarrow \qquad \frac{\bar{x}y}{\|x\| \cdot \|y\|} = \cos\varphi + i\sin\varphi.$$

Wegen $\bar{x}y = (x_1 - ix_2)(y_1 + iy_2) = x_1y_1 + x_2y_2 + i(x_1y_2 - x_2y_1)$ besagt der Realteil dieser Gleichung

$$\frac{x_1y_1 + x_2y_2}{\|x\| \cdot \|y\|} = \cos\varphi,$$

woraus wir mit der obigen Definition des Skalarprodukts folgern, dass

$$\varphi = \arccos \frac{\langle x,y\rangle}{\|x\| \cdot \|y\|}$$

ist (beachte hierbei, dass der Arkuskosinus nur Werte zwischen 0 und π zurückliefert und aufgrund der Symmetrien der Kosinusfunktion damit den *unorientierten* Winkel zwischen x und y ergibt).

Sowohl Längen als auch Winkel lassen sich damit durch das Skalarprodukt ausdrücken. Wenn wir diese beiden Konzepte auch in anderen Vektorräumen definieren wollen, sollten wir den Begriff des Skalarprodukts also auf beliebige Vektorräume verallgemeinern.

Das Problem dabei ist jedoch, dass die Formel $\langle x,y\rangle = x_1y_1 + x_2y_2$ (oder eine entsprechend verallgemeinerte Version für höhere Dimensionen) explizit die Koordinaten der beiden Vektoren x und y benutzt. In einem allgemeinen Vektorraum gäbe es solche Koordinaten aber erst nach Wahl einer Basis – und die Formel würde natürlich auch unterschiedliche Ergebnisse liefern, wenn man die Koordinaten bezüglich verschiedener Basen nehmen würde. Wir schließen daraus, dass es in einem allgemeinen Vektorraum *kein natürlich definiertes* Skalarprodukt gibt, sondern dass ein Skalarprodukt eine *Zusatzstruktur* darstellt, die man zusätzlich zum Vektorraum erst einmal festlegen muss, bevor man mit konkreten Rechnungen anfangen kann.

Wir müssen diese Zusatzstruktur also zunächst erst einmal genauer definieren, d. h. konkret angeben, welche Eigenschaften ein Skalarprodukt haben soll. Klar ist, dass wir zwei Elementen eines K-Vektorraums V ein Element des zugrunde liegenden Körpers K zuordnen wollen, also formal eine Abbildung von $V \times V$ nach K betrachten müssen. Mit solchen Abbildungen wollen wir uns nun zunächst beschäftigen.

21.A Bilinearformen

Die erste wichtige Eigenschaft eines Skalarprodukts ist, dass es linear in beiden Vektoren ist. Derartige Abbildungen bezeichnet man als Bilinearformen.

Definition 21.2 (Bilinearformen)**.** Es sei V ein K-Vektorraum. Eine **Bilinearform** auf V ist eine Abbildung $b\colon V \times V \to K,\ (x,y) \mapsto b(x,y)$, die in beiden Komponenten eine lineare Abbildung gemäß Definition 13.16 ist, d. h. für die für alle $x_1, x_2, x, y_1, y_2, y \in V$ und $\lambda \in K$ die Eigenschaften

$$\begin{aligned} b(x_1+x_2,y) &= b(x_1,y) + b(x_2,y), \\ b(\lambda x,y) &= \lambda\, b(x,y), \\ b(x,y_1+y_2) &= b(x,y_1) + b(x,y_2), \\ b(x,\lambda y) &= \lambda\, b(x,y) \end{aligned}$$

gelten. Wie man leicht nachprüft, ist die Menge aller Bilinearformen auf V ein Unterraum von $\mathrm{Abb}(V \times V, K)$ (siehe Beispiel 13.3 (d)) und damit ein K-Vektorraum. Wir bezeichnen ihn mit $\mathrm{BLF}(V)$.

Beispiel 21.3.

(a) Die Abbildung

$$b\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R},\ \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) \mapsto x_1y_1 + x_1y_2 + x_2y_1 + 4x_2y_2$$

ist offensichtlich eine Bilinearform: Hält man y_1 und y_2 fest, so ist der gegebene Ausdruck eine lineare Abbildung in x_1 und x_2, und umgekehrt. Hingegen ist

$$b\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R},\ \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) \mapsto x_1y_1 + x_1 + y_1$$

keine Bilinearform: Da lineare Abbildungen nach Bemerkung 13.18 (b) stets 0 auf 0 abbilden, kann b wegen

$$b\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = 1$$

bei festgehaltener zweiter Komponente y nicht linear im ersten Argument x sein.

(b) Ist $A \in \mathrm{Mat}(n \times n, K)$ eine quadratische Matrix, so ist

$$b\colon K^n \times K^n \to K,\ b(x,y) = x^{\mathsf{T}} A\, y \qquad (*)$$

nach den Rechenregeln für Matrizen aus Lemma 16.7 eine Bilinearform – beachte, dass das Ergebnis hierbei als Produkt dreier Matrizen der Größen $1 \times n$, $n \times n$ und $n \times 1$ eine 1×1-Matrix, also ein Element von K ist. Ist $A = (a_{i,j})_{i,j}$ und sind $x_1, \dots, x_n$ und $y_1, \dots, y_n$ die Koordinaten von x und y, so ist eine alternative Schreibweise für $(*)$ nach Definition 16.5

$$b(x,y) = (x_1 \ \cdots \ x_n) \cdot \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \sum_{i,j=1}^{n} x_i a_{i,j} y_j.$$

Wir wollen nun sehen, dass man in der Tat sogar jede mögliche Bilinearform auf K^n auf diese Art aus einer eindeutig bestimmten Matrix erhalten kann. Dies besagt der folgende Satz, der völlig analog zu Satz 16.11 über den Zusammenhang zwischen linearen Abbildungen und Matrizen ist, und der damit letztlich besagt, dass Bilinearformen auf K^n und $n \times n$-Matrizen über K „im Prinzip dasselbe“ sind.

Satz und Definition 21.4 (Bilinearformen auf K^n und Matrizen)**.** *Für alle $n \in \mathbb{N}$ ist die Abbildung*

$$\mathrm{Mat}(n \times n, K) \to \mathrm{BLF}(K^n), \quad A \mapsto b_A \qquad \text{mit } b_A(x,y) := x^\mathsf{T} A y$$

ein K-Vektorraumisomorphismus mit Umkehrabbildung

$$\mathrm{BLF}(K^n) \to \mathrm{Mat}(n \times n, K), \quad b \mapsto A_b \qquad \text{mit } A_b := (b(e_i, e_j))_{i,j}.$$

*Man nennt A_b die **Gramsche Matrix** von b.*

Beweis. Nach Beispiel 21.3 (b) ist b_A wirklich eine Bilinearform auf K^n. Wir zeigen, dass die beiden angegebenen Abbildungen invers zueinander sind: Starten wir mit einer Matrix $A = (a_{i,j})_{i,j}$, bilden die zugehörige Bilinearform $b_A(x,y) = x^\mathsf{T} A y$ und dazu wieder die Gramsche Matrix, so erhalten wir

$$A_{b_A} = (b_A(e_i, e_j))_{i,j} = \left(e_i^\mathsf{T} A e_j\right)_{i,j} = (a_{i,j})_{i,j} = A.$$

Beginnen wir umgekehrt mit einer Bilinearform b, nehmen dazu die Gramsche Matrix A_b und bilden dazu wieder die zugehörige Bilinearform, so erhalten wir aufgrund der Bilinearität von b für alle $x, y \in K^n$

$$b_{A_b}(x,y) = x^\mathsf{T} A_b y = \sum_{i,j=1}^{n} x_i b(e_i, e_j) y_j = b\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\right) = b(x,y),$$

wobei $x_1, \dots, x_n$ und $y_1, \dots, y_n$ die Koordinaten von x bzw. y bezeichnen. Also sind die beiden gegebenen Abbildungen wirklich invers zueinander.

Schließlich ist die gegebene Abbildung $A \mapsto b_A$ auch linear, denn für $A, B \in \mathrm{Mat}(n \times n, K)$, $\lambda \in K$ und $x, y \in K^n$ gilt

$$b_{A+B}(x,y) = x^\mathsf{T}(A+B)y = x^\mathsf{T} A y + x^\mathsf{T} B y = b_A(x,y) + b_B(x,y)$$

$$\text{und} \qquad b_{\lambda A}(x,y) = x^\mathsf{T}(\lambda A) y = \lambda x^\mathsf{T} A y = \lambda b_A(x,y). \qquad \square$$

Beispiel 21.5.

(a) Zur Bilinearform

$$b\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, \ \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) \mapsto x_1 y_1 + x_1 y_2 + x_2 y_1 + 4 x_2 y_2 \qquad (*)$$

aus Beispiel 21.3 (a) ist die zugehörige Gramsche Matrix

$$A_b = \begin{pmatrix} b(e_1, e_1) & b(e_1, e_2) \\ b(e_2, e_1) & b(e_2, e_2) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}.$$

Eine alternative Beschreibung von $A_b = (a_{i,j})_{i,j}$ ist offensichtlich, dass $a_{i,j}$ in einer Darstellung der Form $(*)$ von $b(x,y)$ genau der Koeffizient von $x_i y_j$ ist.

Umgekehrt können wir nun nach Satz 21.4 aus dieser Matrix auch die ursprüngliche Bilinearform durch die Formel

$$b(x,y) = (x_1 \ x_2) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_1 y_2 + x_2 y_1 + 4 x_2 y_2$$

zurückgewinnen.

(b) Die Einheitsmatrix $E_n \in \mathrm{Mat}(n \times n, K)$ entspricht in der Korrespondenz aus Satz 21.4 genau der Bilinearform

$$b\colon K^n \times K^n \to K,\ (x,y) \mapsto x^\mathsf{T} y = x_1 y_1 + \cdots + x_n y_n,$$

die wir in Beispiel 21.1 im Fall $K = \mathbb{R}$ und $n = 2$ schon beim gewöhnlichen Skalarprodukt auf $\mathbb{R}^2$ gesehen haben.

Genau wie bei linearen Abbildungen in Satz 16.23 können wir unsere Korrespondenz zwischen Bilinearformen und Matrizen nun unmittelbar von K^n auf beliebige endlich-dimensionale Vektorräume erweitern, indem wir dort eine Basis wählen und mit den Koordinaten bezüglich dieser Basis arbeiten.

Folgerung 21.6 (Bilinearformen auf V und Matrizen)**.** *Es seien V ein endlich-dimensionaler K-Vektorraum sowie $B = (x_1, \ldots, x_n)$ eine Basis von V mit zugehöriger Koordinatenabbildung $\Phi_B\colon V \to K^n$ (siehe Definition 16.22). Dann ist die Abbildung*

$$\mathrm{Mat}(n \times n, K) \to \mathrm{BLF}(V), \quad A \mapsto b_A^B \qquad \text{mit } b_A^B(x,y) := \Phi_B(x)^\mathsf{T} A\, \Phi_B(y)$$

wieder ein K-Vektorraumisomorphismus mit Umkehrabbildung

$$\mathrm{BLF}(V) \to \mathrm{Mat}(n \times n, K), \quad b \mapsto A_b^B \qquad \text{mit } A_b^B := (b(x_i, x_j))_{i,j}.$$

*Wie oben nennt man A_b^B die **Gramsche Matrix** von b bezüglich der Basis B.*

Beweis. Die Abbildung

$$\mathrm{BLF}(K^n) \to \mathrm{BLF}(V), \quad b \mapsto \Big((x,y) \mapsto b(\Phi_B(x), \Phi_B(y)) \Big),$$

die einer Bilinearform b auf K^n die Bilinearform auf V zuordnet, bei der man einfach in b die Koordinatenvektoren der Vektoren aus V einsetzt, ist offensichtlich ein Isomorphismus mit Umkehrabbildung

$$\mathrm{BLF}(V) \to \mathrm{BLF}(K^n), \quad b \mapsto \Big((x,y) \mapsto b(\Phi_B^{-1}(x), \Phi_B^{-1}(y)) \Big).$$

Verketten wir den Isomorphismus aus Satz 21.4 mit dieser Abbildung, erhalten wir also wie behauptet einen Isomorphismus $\mathrm{Mat}(n \times n, K) \to \mathrm{BLF}(K^n) \to \mathrm{BLF}(V)$, der eine Matrix A auf die Bilinearform $(x,y) \mapsto b_A(\Phi_B(x), \Phi_B(y)) = \Phi_B(x)^\mathsf{T} A\, \Phi_B(y)$ abbildet, und dessen Umkehrung einer Bilinearform b auf V die Matrix $(b(\Phi_B^{-1}(e_i), \Phi_B^{-1}(e_j)))_{i,j} = (b(x_i, x_j))_{i,j}$ zuordnet. □

Natürlich hängt die Gramsche Matrix einer Bilinearform wie in Folgerung 21.6 von der gewählten Basis ab. Wie das folgende Lemma zeigt, ist die Transformationsformel bei einem Basiswechsel jedoch eine andere als für Endomorphismen (siehe Bemerkung 19.3 (b)).

Lemma 21.7 (Verhalten von Gramschen Matrizen unter Basiswechsel)**.** *Es seien b eine Bilinearform auf einem endlich-dimensionalen K-Vektorraum V sowie B und B' zwei Basen von V. Dann gilt für die Gramschen Matrizen von b bezüglich B und B'*

$$A_b^{B'} = T^\mathsf{T} A_b^B\, T,$$

wobei $T = A^{B',B}$ die Basiswechselmatrix aus Definition 16.28 ist.

Beweis. Es seien $B = (x_1, \ldots, x_n)$ und $B' = (y_1, \ldots, y_n)$ die gewählten Basen. Nach Definition 16.28 enthält die k-te Spalte von $T = (a_{i,j})_{i,j}$ für $k = 1, \ldots, n$ genau den Koordinatenvektor von y_k bezüglich B, d. h. es gilt

$$y_k = a_{1,k} x_1 + \cdots + a_{n,k} x_n.$$

Damit folgt für die Gramschen Matrizen mit der Formel aus Folgerung 21.6 sofort

$$A_b^{B'} = (b(y_k, y_l))_{k,l} = \left(b\left(\sum_{i=1}^n a_{i,k} x_i, \sum_{j=1}^n a_{j,l} x_j \right) \right)_{k,l} = \left(\sum_{i,j=1}^n a_{i,k}\, b(x_i, x_j)\, a_{j,l} \right)_{k,l} = T^\mathsf{T} A_b^B\, T. \quad \square$$

Bemerkung 21.8 (Matrizen unter Basiswechsel)**.** Bisher hatten wir Matrizen nahezu ausschließlich zur Beschreibung von linearen Abbildungen benutzt. Nach Definition ist eine Matrix aber zunächst einmal nichts weiter als ein rechteckiges Zahlenschema, ohne Vorgabe einer Bedeutung dieser Zahlen. In der Tat haben wir nun gesehen, dass man Matrizen auch noch für ganz andere Dinge verwenden kann, nämlich z. B. zur Darstellung von Bilinearformen.

Ohne weitere Informationen ergibt es daher keinen Sinn zu fragen, wie sich eine Matrix unter einem Basiswechsel transformiert. Die Antwort auf diese Frage hängt nach Bemerkung 19.3 (b) und Lemma 21.7 davon ab, welche Bedeutung die Einträge in der Matrix haben:

Bei einem Basiswechsel mit zugehöriger Basiswechselmatrix T transformiert sich ...
... eine zu einem *Endomorphismus* gehörige Matrix A in die Matrix $T^{-1}AT$,
... eine zu einer *Bilinearform* gehörige Matrix A in die Matrix $T^{\mathsf{T}}AT$.

Wie bei linearen Abbildungen oder Endomorphismen könnten wir uns nun schließlich auch bei Bilinearformen wieder nach einer Normalform fragen: Wie können wir zu einer gegebenen Bilinearform $b \in \mathrm{BLF}(V)$ eine Basis von V so wählen, dass die zugehörige Gramsche Matrix A_b^B möglichst einfach wird? Wir wollen diese Frage hier allerdings nicht in dieser vollen Allgemeinheit beantworten, da wir im Folgenden hauptsächlich an Bilinearformen mit noch weiteren speziellen Eigenschaften interessiert sind. Diese Eigenschaften wollen wir jetzt einführen.

21.B Skalarprodukte

Wir kommen nun zu den in der Einleitung zu diesem Kapitel bereits angekündigten Skalarprodukten. Wir werden sie als Bilinearformen mit den folgenden beiden Eigenschaften definieren, die wir auch gleich wieder analog für Matrizen einführen wollen.

Definition 21.9 (Symmetrie und positive Definitheit)**.** Es seien $b \in \mathrm{BLF}(V)$ eine Bilinearform auf einem K-Vektorraum V und $A \in \mathrm{Mat}(n \times n, K)$.

(a) Die Bilinearform b heißt **symmetrisch**, wenn $b(x,y) = b(y,x)$ für alle $x,y \in V$.

Die Matrix A heißt **symmetrisch**, wenn $A^{\mathsf{T}} = A$.

(b) Es sei nun zusätzlich $K = \mathbb{R}$.

Die Bilinearform b heißt dann **positiv definit**, wenn $b(x,x) > 0$ für alle $x \in V$ mit $x \neq 0$.

Die Matrix A heißt dann **positiv definit**, wenn $x^{\mathsf{T}}Ax > 0$ für alle $x \in \mathbb{R}^n$ mit $x \neq 0$.

Bemerkung 21.10.

(a) Die Bedingung der positiven Definitheit lässt sich offensichtlich nur für einen geordneten Körper (siehe Kapitel 4.B) formulieren. Für uns ist hierbei eigentlich nur der Fall $K = \mathbb{R}$ interessant. Wir werden die Bedingung der positiven Definitheit in Konstruktion 21.19 aber noch etwas abändern, so dass sie dann auch im Fall $K = \mathbb{C}$ anwendbar ist.

(b) Da eine Bilinearform b linear in jedem Eintrag ist, gilt natürlich stets $b(0,0) = 0$. Eine positiv definite Bilinearform auf einem $\mathbb{R}$-Vektorraum V erfüllt damit also immer $b(x,x) \geq 0$ für alle $x \in V$. Diese Bedingung, die wir später für Skalarprodukte fordern werden, wird uns dann sicherstellen, dass wir aus $b(x,x)$ die Wurzel ziehen und so die Länge von x definieren können.

48

In manchen Fällen (siehe Satz 26.20) benötigt man allerdings auch noch die folgenden zur positiven Definitheit analogen Bedingungen: Eine Bilinearform $b \in \mathrm{BLF}(V)$ auf einem reellen Vektorraum V heißt ...

- ... **negativ definit**, wenn $b(x,x) < 0$ für alle $x \in V$ mit $x \neq 0$;
- ... **positiv semidefinit**, wenn $b(x,x) \geq 0$ für alle $x \in V$;
- ... **negativ semidefinit**, wenn $b(x,x) \leq 0$ für alle $x \in V$;

- ... **indefinit**, wenn sie weder positiv noch negativ semidefinit ist, also wenn es $x, y \in V$ gibt mit $b(x,x) > 0$ und $b(y,y) < 0$.

Entsprechende Eigenschaften definiert man natürlich auch für reelle quadratische Matrizen.

Als Erstes wollen wir nun die wohl erwartete Aussage zeigen, dass sich die in Definition 21.9 eingeführten Begriffe für Bilinearformen und Matrizen entsprechen.

Lemma 21.11 (Symmetrie und positive Definitheit bei Bilinearformen und Matrizen)**.** *Es seien b eine Bilinearform auf einem endlich-dimensionalen K-Vektorraum V, B eine Basis von V, und A_b^B wie in Folgerung 21.6 die zugehörige Gramsche Matrix. Dann gilt:*

(a) *Die Bilinearform b ist genau dann symmetrisch, wenn die Matrix A_b^B symmetrisch ist.*

(b) *Im Fall $K = \mathbb{R}$ ist b genau dann positiv definit, wenn A_b^B positiv definit ist.*

Beweis. Es sei $B = (x_1, \dots, x_n)$.

(a) „$\Rightarrow$“: Ist b symmetrisch, so folgt natürlich sofort

$$(A_b^B)^\mathsf{T} = (b(x_j, x_i))_{i,j} = (b(x_i, x_j))_{i,j} = A_b^B.$$

„$\Leftarrow$“: Ist umgekehrt A_b^B symmetrisch, so gilt nach Folgerung 21.6 für alle $x, y \in V$

$$b(x,y) = \Phi_B(x)^\mathsf{T} A_b^B \Phi_B(y) \overset{(*)}{=} \Phi_B(y)^\mathsf{T} (A_b^B)^\mathsf{T} \Phi_B(x) = \Phi_B(y)^\mathsf{T} A_b^B \Phi_B(x) = b(y,x),$$

wobei wir in $(*)$ gemäß Lemma 16.7 (d) die transponierte 1×1-Matrix gebildet haben.

(b) Da die Koordinatenabbildung $\Phi_B \colon V \to \mathbb{R}^n$ ein Isomorphismus ist, ist die Matrix A_b^B genau dann positiv definit, wenn $\Phi_B(x)^\mathsf{T} A_b^B \Phi_B(x) > 0$ für alle $x \in V \setminus \{0\}$ gilt. Nach Folgerung 21.6 bedeutet dies genau $b(x,x) > 0$ für alle $x \in V \setminus \{0\}$, also dass b positiv definit ist. □

Bemerkung 21.12 (Invarianz von Symmetrie und positiver Definitheit)**.** Sind $A, A' \in \mathrm{Mat}(n \times n, K)$ zwei quadratische Matrizen mit $A' = T^\mathsf{T} A T$ für ein $T \in \mathrm{GL}(n, K)$, so besagt Lemma 21.11 insbesondere, dass A' genau dann symmetrisch (bzw. im Fall $K = \mathbb{R}$ positiv definit) ist, wenn dies für A gilt: A' und A beschreiben nach Lemma 21.7 nämlich die gleiche Bilinearform bezüglich zweier evtl. verschiedener Basen, und nach Lemma 21.11 hängt es nur von dieser Bilinearform (aber eben nicht von der gewählten Basis) ab, ob die Matrix symmetrisch bzw. positiv definit ist.

Mit Hilfe der eingeführten Konzepte können wir nun Skalarprodukte auf reellen Vektorräumen definieren.

Definition 21.13 (Skalarprodukte)**.** Es sei V ein $\mathbb{R}$-Vektorraum. Ein **Skalarprodukt** auf V ist eine positiv definite, symmetrische Bilinearform $b \colon V \times V \to \mathbb{R}$. Ein $\mathbb{R}$-Vektorraum V zusammen mit einem Skalarprodukt heißt ein **euklidischer Raum**.

Für $x, y \in V$ schreiben wir statt $b(x,y)$ dann auch $\langle x, y \rangle$. Die (wegen der positiven Definitheit existierende) Zahl

$$\|x\| := \sqrt{\langle x, x \rangle} \quad \in \mathbb{R}_{\geq 0}$$

heißt in Verallgemeinerung von Beispiel 21.1 die **Norm** oder **Länge** von x (bezüglich b). Man nennt einen Vektor $x \in V$ **normiert** (bezüglich b), falls $\|x\| = 1$.

Bemerkung 21.14. Ist V ein endlich erzeugter $\mathbb{R}$-Vektorraum und $B = (x_1, \dots, x_n)$ eine Basis von V, so lässt sich ein Skalarprodukt auf V nach Lemma 21.11 also genau durch eine positiv definite, symmetrische Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ beschreiben bzw. definieren (nämlich durch die Gramsche Matrix des Skalarprodukts bezüglich der Basis B).

Beispiel 21.15.

(a) Für $V = \mathbb{R}^n$ ist die in Beispiel 21.5 (b) bereits betrachtete Bilinearform

$$\langle x, y \rangle := x^\mathsf{T} y = \sum_{i=1}^n x_i y_i$$

ein Skalarprodukt: Die Symmetrie ist offensichtlich, und die positive Definitheit ergibt sich sofort daraus, dass für alle $x \in \mathbb{R}^n$ mit $x \neq 0$ natürlich

$$\langle x,x \rangle = \sum_{i=1}^n x_i^2 > 0$$

gilt. Es ist die direkte Verallgemeinerung von Beispiel 21.1 und wird als das **Standardskalarprodukt** auf $\mathbb{R}^n$ bezeichnet. Bezüglich der Standardbasis ist die Gramsche Matrix dieses Skalarprodukts wie in Beispiel 21.5 (b) offensichtlich genau die Einheitsmatrix E_n.

(b) Die in Beispiel 21.5 (a) bereits betrachtete reelle Matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$ ist positiv definit: Für alle $x \in \mathbb{R}^2$ ist zunächst

$$x^\mathsf{T} A x = x_1^2 + x_1 x_2 + x_2 x_1 + 4x_2^2 = (x_1 + x_2)^2 + 3x_2^2 \geq 0,$$

und die Gleichheit kann hier nur gelten für $x_1 + x_2 = x_2 = 0$, also für $x = 0$. Da A auch symmetrisch ist, ist die durch diese Matrix (bezüglich der Standardbasis) definierte Bilinearform

$$\langle x,y \rangle = x^\mathsf{T} A y = x_1 y_1 + x_1 y_2 + x_2 y_1 + 4x_2 y_2$$

also ein Skalarprodukt auf $\mathbb{R}^2$.

(c) Betrachten wir reelle $m \times n$-Matrizen über ihre Einträge als Vektoren in $\mathbb{R}^{mn}$, so können wir das Standardskalarprodukt zweier Matrizen $A = (a_{i,j})_{i,j}$ und $B = (b_{i,j})_{i,j}$ im Vektorraum $V = \mathrm{Mat}(m \times n, \mathbb{R})$ mit der Definition des Matrixprodukts und der Spur schreiben als

$$\langle A,B \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{i,j} b_{i,j} = \mathrm{Spur}\left(A^\mathsf{T} B\right).$$

Wir werden dies in Zukunft als das Standardskalarprodukt auf Matrizenräumen wählen.

(d) Es sei $V = C^0([0,1])$ der reelle Vektorraum der stetigen Funktionen auf dem Intervall $[0,1]$. Wir behaupten, dass dann die Abbildung

$$\langle f,g \rangle := \int_0^1 f(x) g(x)\, dx \qquad (*)$$

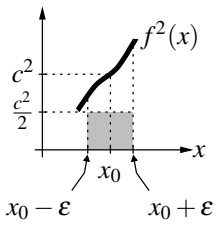
ein Skalarprodukt auf V definiert. Die Bilinearität ergibt sich hierbei einfach aus den Eigenschaften des Integrals, z. B. ist für $f_1, f_2, g \in V$

$$\begin{aligned} \langle f_1 + f_2, g \rangle &= \int_0^1 (f_1(x) + f_2(x)) g(x)\, dx = \int_0^1 f_1(x) g(x)\, dx + \int_0^1 f_2(x) g(x)\, dx \\ &= \langle f_1, g \rangle + \langle f_2, g \rangle . \end{aligned}$$

Die Symmetrie ist natürlich offensichtlich. Für die positive Definitheit bemerken wir zunächst, dass die Funktion f^2 für jedes $f \in V$ überall nicht-negativ ist und damit schon einmal

$$\langle f,f \rangle = \int_0^1 f(x)^2\, dx \geq 0$$

gilt. Ist nun zusätzlich $f \neq 0$ nicht die Nullfunktion, so gibt es also ein $x_0 \in [0,1]$ mit $c := f(x_0) \neq 0$. Da f und damit auch f^2 stetig ist, gibt es dann aber wie im Bild rechts ein $\varepsilon > 0$, so dass $f(x)^2 \geq \frac{c^2}{2}$ für alle $x \in [0,1]$ mit $|x - x_0| < \varepsilon$ ist. Unter dem Graphen von f^2 liegt also sicher ein positives Flächenstück (der Breite 2ε und Höhe $\frac{c^2}{2}$, falls x_0 nicht zu weit am Rand des Intervalls liegt), und somit ist wie gewünscht $\langle f,f \rangle = \int_0^1 f(x)^2\, dx > 0$.



Beachte, dass die Formel $(*)$ ganz analog zu der in (a) ist, wenn wir die Summe über die Koordinaten von $\mathbb{R}^n$ durch ein Integral über alle Punkte im Definitionsintervall $[0,1]$ ersetzen. Auch die Interpretation z. B. der Norm $\|f\| = \sqrt{\langle f,f \rangle}$ ist in dem Sinne analog, dass diese Norm klein ist, wenn die Funktion „nur wenig von der Nullfunktion abweicht“. Man bezeichnet dieses Skalarprodukt daher auch als Standardskalarprodukt auf $C^0([0,1])$. Um es

definieren zu können, ist die Stetigkeit der betrachteten Funktionen entscheidend: Wäre z. B. die unstetige Funktion

$$f\colon [0,1] \to \mathbb{R},\ x \mapsto \begin{cases} 0 & \text{für } x \neq \frac{1}{2}, \\ 1 & \text{für } x = \frac{1}{2} \end{cases}$$

zugelassen, so wäre hier zwar $f \neq 0$, aber trotzdem $\langle f, f\rangle = \int_0^1 f(x)^2\,dx = 0$. Wir hätten in diesem Fall also kein Skalarprodukt (siehe auch Aufgabe 21.24 (c)).

Bemerkung 21.16 (Skalarprodukte auf Unterräumen)**.** Die Einschränkung eines Skalarprodukts auf einen Untervektorraum ist offensichtlich wieder ein Skalarprodukt. So können wir in Beispiel 21.15 z. B. das Standardskalarprodukt aus (a) genauso gut auch auf einem Unterraum von $\mathbb{R}^n$ betrachten, oder das aus (d) auf den Vektorräumen aller differenzierbaren Funktionen oder aller Polynomfunktionen.

Aufgabe 21.17. Es sei $n \in \mathbb{N}_{>0}$. Untersuche, ob die folgenden Abbildungen b Skalarprodukte auf dem reellen Vektorraum V sind:

(a) $V = \mathbb{R}^n$, $b(x,y) = x^\mathsf{T} A y$ mit $A = \begin{pmatrix} 2 & 1 & & & 0 \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ 0 & & & 1 & 2 \end{pmatrix}$;

(b) $V = \mathrm{Mat}(n \times n, \mathbb{R})$, $b(A,B) = \mathrm{Spur}(AB)$.

Aufgabe 21.18 (Duale Vektorräume)**.** Zu einem K-Vektorraum V heißt $V^* := \mathrm{Hom}(V,K)$ sein *Dualraum*.

Es sei nun V ein euklidischer Vektorraum. Man zeige:

(a) Die Abbildung

$$\Phi\colon V \to V^*,\ x \mapsto \varphi_x \qquad \text{mit } \varphi_x(y) := \langle x, y\rangle$$

ist linear und injektiv.

(b) Ist V endlich-dimensional, so ist die Abbildung Φ aus (a) sogar ein Isomorphismus.

(c) Ist $V = C^0([0,1])$ der Raum aller stetigen Funktionen auf dem Intervall $[0,1]$ mit dem Standardskalarprodukt $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$, so ist die Auswerteabbildung am Nullpunkt

$$\varphi\colon V \to \mathbb{R} \quad \text{mit} \quad \varphi(f) = f(0)$$

ein Element von V^*, das nicht im Bild von Φ liegt. In diesem Fall ist die Abbildung Φ aus (a) also nicht surjektiv und damit kein Isomorphismus.

Im Rest dieses Kapitels wollen wir nun die grundlegenden Eigenschaften von euklidischen Räumen untersuchen. Da es in der Praxis öfters einmal vorkommt, werden wir den Begriff des Skalarprodukts aber zunächst noch auf den Fall von komplexen Vektorräumen erweitern.

Konstruktion 21.19 (Skalarprodukte im komplexen Fall)**.** Wollen wir die Definition 21.13 auf einen komplexen Vektorraum übertragen, so haben wir das Problem, dass die Bedingung der positiven Definitheit über $\mathbb{C}$ zunächst einmal keinen Sinn ergibt, da $\mathbb{C}$ kein geordneter Körper ist. Dies hat zur Folge, dass wir die Norm eines Vektors nicht mehr wie gewohnt definieren können: Würden wir wie beim Standardskalarprodukt im Reellen auch in $\mathbb{C}^n$ die Formeln

$$\langle x, y\rangle = \sum_{i=1}^n x_i y_i \quad \text{und} \quad \|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \cdots + x_n^2}$$

verwenden, so müssten wir hier die Wurzel aus einer im Allgemeinen komplexen Zahl $x_1^2 + \cdots + x_n^2$ bilden – was nicht eindeutig möglich ist und auch nicht wie gewünscht zu einer nicht-negativen reellen Zahl als Länge eines Vektors führen würde. Die Lösung dieses Problems besteht darin, im

Skalarprodukt grundsätzlich jede Koordinate *eines* Eintrags komplex zu konjugieren. Es spielt dabei keine Rolle, welchen Eintrag wir dafür wählen – wir werden in dieser Vorlesung immer den ersten nehmen, aber auch die umgekehrte Konvention des zweiten Eintrags ist in der Literatur zu finden. Durch diese komplexe Konjugation erhalten wir z. B. für das Standardskalarprodukt die Formeln

$$\langle x,y\rangle = \sum_{i=1}^{n} \overline{x_i}\, y_i \quad \text{und damit} \quad \|x\| = \sqrt{\langle x,x\rangle} = \sqrt{\overline{x_1}\,x_1 + \dots + \overline{x_n}\,x_n} = \sqrt{|x_1|^2 + \dots + |x_n|^2},$$

was wieder zu einer reellen, nicht-negativen Länge eines Vektors führt.

Mit dem Hintergrund dieser Idee sind die entsprechenden Abänderungen für beliebige Skalarprodukte auf komplexen Vektorräumen relativ offensichtlich. Wir werden die sich daraus ergebenden Definitionen und Resultate im Folgenden nur kurz auflisten; die Beweise dieser Aussagen sind völlig analog zu denen im reellen Fall.

Es sei also V ein $\mathbb{C}$-Vektorraum. Eine **Sesquilinearform** auf V ist eine Abbildung $s\colon V \times V \to \mathbb{C}$, so dass für alle $x_1, x_2, x, y_1, y_2, y \in V$ und $\lambda \in \mathbb{C}$ die Eigenschaften

$$\begin{aligned} s(x_1 + x_2, y) &= s(x_1, y) + s(x_2, y), \\ s(\lambda x, y) &= \overline{\lambda}\, s(x,y), \\ s(x, y_1 + y_2) &= s(x, y_1) + s(x, y_2), \\ s(x, \lambda y) &= \lambda\, s(x,y) \end{aligned}$$

gelten. Der einzige Unterschied zu Bilinearformen besteht also in der komplexen Konjugation des Skalars in der zweiten Zeile oben – und in der Tat kommt der Begriff „Sesquilinearform“ aus dem Lateinischen und bedeutet „eineinhalbfach lineare Form“.

Die Sesquilinearformen auf V bilden einen $\mathbb{C}$-Vektorraum, den wir mit $\mathrm{SLF}(V)$ bezeichnen wollen. Ist V endlich-dimensional und $B = (x_1, \dots, x_n)$ eine Basis von V, so ist $\mathrm{SLF}(V)$ wie in Folgerung 21.6 isomorph zu $\mathrm{Mat}(n \times n, \mathbb{C})$ über die beiden zueinander inversen Isomorphismen

$$\begin{aligned} & \mathrm{Mat}(n\times n,\mathbb{C}) \to \mathrm{SLF}(V), \quad A \mapsto s_A^B && \text{mit } s_A^B(x,y) := \overline{\Phi_B(x)}^{\mathsf{T}} A\, \Phi_B(y) \\ \text{und} \quad & \mathrm{SLF}(V) \to \mathrm{Mat}(n\times n,\mathbb{C}), \quad s \mapsto A_s^B && \text{mit } A_s^B := (s(x_i,x_j))_{i,j}, \end{aligned}$$

wobei der Querstrich über $\Phi_B(x) \in \mathbb{C}^n$ bedeutet, dass jede Komponente des Vektors komplex konjugiert wird. Man bezeichnet A_s^B wieder als die *Gramsche Matrix* von s. Ist B' eine weitere Basis von V, so transformieren sich die Gramschen Matrizen analog zu Lemma 21.7 gemäß

$$A_s^{B'} = \overline{T}^{\mathsf{T}} A_s^B\, T \quad \text{mit } T = A^{B',B},$$

wobei $T = A^{B',B}$ die übliche Basiswechselmatrix ist.

Eine Sesquilinearform s auf einem $\mathbb{C}$-Vektorraum V heißt **hermitesch**, wenn $s(y,x) = \overline{s(x,y)}$ für alle $x, y \in V$. Ist dies der Fall, so erhalten wir daraus insbesondere mit $y = x$

$$s(x,x) = \overline{s(x,x)}, \quad \text{also} \quad s(x,x) \in \mathbb{R}$$

für alle $x \in V$. Wir können daher fragen, ob diese reelle Zahl immer nicht-negativ ist, und nennen eine hermitesche Sesquilinearform s analog zum reellen Fall **positiv definit**, wenn $s(x,x) > 0$ für alle $x \in V$ mit $x \neq 0$.

Entsprechend heißt eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ **hermitesch**, wenn $\overline{A}^{\mathsf{T}} = A$. In diesem Fall ist $\overline{x}^{\mathsf{T}} A x \in \mathbb{R}$ für alle $x \in \mathbb{C}^n$, denn es ist

$$\overline{\overline{x}^{\mathsf{T}} A x} = x^{\mathsf{T}} \overline{A}\, \overline{x} = (x^{\mathsf{T}} \overline{A}\, \overline{x})^{\mathsf{T}} = \overline{x}^{\mathsf{T}} \overline{A}^{\mathsf{T}} x = \overline{x}^{\mathsf{T}} A x$$

nach Lemma 16.7 (d). Gilt sogar $\overline{x}^{\mathsf{T}} A x > 0$ für alle $x \in \mathbb{C}^n$ mit $x \neq 0$, so heißt A **positiv definit**. Wie in Lemma 21.11 ist eine Sesquilinearform auf einem endlich-dimensionalen Vektorraum genau dann hermitesch, bzw. eine hermitesche Sesquilinearform genau dann positiv definit, wenn ihre Gramsche Matrix zu einer beliebigen Basis diese Eigenschaft besitzt. Die anderen Definitheitsbegriffe aus Bemerkung 21.10 (b) definiert man natürlich sowohl für hermitesche Sesquilinearformen als auch für hermitesche Matrizen analog.

Ein **Skalarprodukt** auf einem komplexen Vektorraum V ist nun eine positiv definite, hermitesche Sesquilinearform s auf V, die wir dann typischerweise als $\langle x,y\rangle := s(x,y)$ schreiben. Ein $\mathbb{C}$-Vektorraum zusammen mit einem Skalarprodukt heißt ein **unitärer Raum**. Für $x \in V$ nennen wir in diesem Fall wieder

$$\|x\| := \sqrt{\langle x,x\rangle} \quad \in \mathbb{R}_{\geq 0}$$

die *Norm* (oder *Länge*) von x.

Das **Standardskalarprodukt** auf $\mathbb{C}^n$ ist dasjenige, das bezüglich der Standardbasis der Einheitsmatrix entspricht, also

$$\langle x,y\rangle = \overline{x}^{\mathsf{T}} y = \overline{x_1}\, y_1 + \cdots + \overline{x_n}\, y_n,$$

wobei $x_1,\ldots,x_n$ und $y_1,\ldots,y_n$ die Koordinaten von x bzw. y sind. Beachte, dass die Norm eines Vektors $x \in \mathbb{C}^n$ in diesem Fall dann

$$\|x\| = \sqrt{\langle x,x\rangle} = \sqrt{|x_1|^2 + \cdots + |x_n|^2} = \sqrt{(\operatorname{Re} x_1)^2 + (\operatorname{Im} x_1)^2 + \cdots + (\operatorname{Re} x_n)^2 + (\operatorname{Im} x_n)^2}$$

und damit gleich der Norm dieses Vektors in $\mathbb{R}^{2n}$ bezüglich des Standardskalarprodukts ist. Das Standardskalarprodukt für komplexe Matrizen $A,B \in \operatorname{Mat}(m \times n, \mathbb{C})$ ist $\langle A,B\rangle = \operatorname{Spur}\left(\overline{A}^{\mathsf{T}} B\right)$ analog zu Beispiel 21.15 (c).

Im Folgenden wollen wir den reellen und komplexen Fall in der Regel zusammen behandeln. Wie in der Analysis schreiben wir daher wieder $\mathbb{K}$ für einen der Körper $\mathbb{R}$ oder $\mathbb{C}$, und sprechen von einem *Vektorraum mit Skalarprodukt*, wenn wir einen euklidischen bzw. unitären Raum meinen. Die Formeln werden wir dabei wie in Konstruktion 21.19 mit der komplexen Konjugation schreiben, so dass sie dann für $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$ gleichermaßen gelten – im reellen Fall ist die komplexe Konjugation dann zwar unnötig, schadet aber natürlich auch nicht.

Als erstes Resultat über Skalarprodukte wollen wir nun eine sehr wichtige Ungleichung beweisen.

Satz 21.20 (Cauchy-Schwarz-Ungleichung). *In jedem Vektorraum V mit Skalarprodukt gilt*

$$|\langle x,y\rangle| \leq \|x\| \cdot \|y\|$$

für alle $x,y \in V$, wobei die Gleichheit genau dann gilt, wenn x und y linear abhängig sind.

Beweis. Für $y = 0$ ist die Aussage des Satzes offensichtlich, denn dann sind beide Seiten gleich Null und die Vektoren linear abhängig. Andernfalls setzen wir $\lambda := \frac{\langle y,x\rangle}{\langle y,y\rangle}$ (und damit $\overline{\lambda} = \frac{\langle x,y\rangle}{\langle y,y\rangle}$ wegen der Symmetrie bzw. Hermitizität des Skalarprodukts) und folgern aus der positiven Definitheit, dass

$$\begin{aligned}
0 &\leq \langle x - \lambda y, x - \lambda y\rangle \qquad (*)\\
&= \langle x,x\rangle - \lambda\,\langle x,y\rangle - \overline{\lambda}\,\langle y,x\rangle + \lambda\,\overline{\lambda}\,\langle y,y\rangle\\
&= \langle x,x\rangle - \frac{\langle y,x\rangle\,\langle x,y\rangle}{\langle y,y\rangle} - \frac{\langle x,y\rangle\,\langle y,x\rangle}{\langle y,y\rangle} + \frac{\langle x,y\rangle\,\langle y,x\rangle}{\langle y,y\rangle}\\
&= \langle x,x\rangle - \frac{\langle x,y\rangle\,\langle y,x\rangle}{\langle y,y\rangle}\\
&= \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}
\end{aligned}$$

und damit $|\langle x,y\rangle|^2 \leq \|x\|^2 \cdot \|y\|^2$ gilt. Wurzelziehen liefert nun die behauptete Ungleichung.

Gilt in dieser Ungleichung sogar die Gleichheit, gilt also die Gleichheit in $(*)$, so ergibt sich aus der positiven Definitheit des Skalarprodukts sofort $x - \lambda y = 0$, d. h. x und y sind linear abhängig. Sind umgekehrt x und y linear abhängig, gilt also $x = \mu y$ für ein $\mu \in \mathbb{K}$, so ist $\langle y,x\rangle = \langle y,\mu y\rangle = \mu\,\langle y,y\rangle$ und damit $\lambda = \frac{\langle y,x\rangle}{\langle y,y\rangle} = \mu$, so dass in $(*)$ oben $x - \lambda y = x - \mu y = 0$ ist und damit sowohl dort als auch in der Cauchy-Schwarz-Ungleichung sogar die Gleichheit gilt. □

49

Mit der Cauchy-Schwarz-Ungleichung können wir die folgenden Eigenschaften der Norm herleiten:

Satz 21.21 (Eigenschaften der Norm)**.** *In jedem Vektorraum V mit Skalarprodukt gilt für alle $x,y \in V$ und $\lambda \in \mathbb{K}$:*

(a) $\|\lambda x\| = |\lambda| \cdot \|x\|$*;*

(b) $\|x\| > 0$ *für alle* $x \neq 0$*;*

(c) $\|x+y\| \leq \|x\| + \|y\|$ ***(Dreiecksungleichung)****.*

Beweis.

(a) Es ist $\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\overline{\lambda}\lambda \langle x,x\rangle} = \sqrt{|\lambda|^2 \langle x,x\rangle} = |\lambda| \cdot \|x\|$.

(b) folgt sofort aus der positiven Definitheit des Skalarprodukts.

(c) Es gilt

$$\begin{aligned}
\|x+y\|^2 &= \langle x+y, x+y\rangle \\
&= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\
&= \|x\|^2 + \langle x,y\rangle + \overline{\langle x,y\rangle} + \|y\|^2 \\
&= \|x\|^2 + 2\,\mathrm{Re}\,\langle x,y\rangle + \|y\|^2 && \text{(wegen } \mathrm{Re}\, z = \tfrac{1}{2}(z+\overline{z}) \text{ für alle } z \in \mathbb{C}) \\
&\leq \|x\|^2 + 2\,|\langle x,y\rangle| + \|y\|^2 && \text{(wegen } \mathrm{Re}\, z \leq |z| \text{ für alle } z \in \mathbb{C}) \\
&\leq \|x\|^2 + 2\,\|x\| \cdot \|y\| + \|y\|^2 && \text{(Satz 21.20)} \\
&= (\|x\| + \|y\|)^2,
\end{aligned}$$

woraus durch Wurzelziehen die Behauptung folgt. □

Bemerkung 21.22 (Normen in der linearen Algebra und Analysis)**.** In der Analysis werden wir später normierte Vektorräume definieren als reelle oder komplexe Vektorräume mit einer reellwertigen „Normabbildung“ $\|\cdot\|$, die die drei Eigenschaften aus Satz 21.21 erfüllt (siehe Definition 23.1). In diesem Sinne sind Vektorräume mit Skalarprodukt also immer normierte Vektorräume aus der Sicht der Analysis. Wir werden allerdings sehen, dass es auch sehr viele Normen gibt, die nicht von einem Skalarprodukt kommen, z. B. in $\mathbb{R}^2$ die Summennorm $\|x\| = |x_1| + |x_2|$ oder die Maximumsnorm $\|x\| = \max\{|x_1|, |x_2|\}$.

Wir kommen nun zum Winkel zwischen zwei Vektoren. Man definiert ihn in der Regel nur im reellen Fall und orientiert sich dabei an der geometrischen Deutung aus Beispiel 21.1 im Fall des Standardskalarprodukts.

Konstruktion 21.23 (Winkel)**.** Es seien V ein euklidischer Raum und $x,y \in V\backslash\{0\}$. Nach der Cauchy-Schwarz-Ungleichung aus Satz 21.20 ist dann

$$\frac{|\langle x,y\rangle|}{\|x\| \cdot \|y\|} \leq 1, \qquad \text{also} \qquad -1 \leq \frac{\langle x,y\rangle}{\|x\| \cdot \|y\|} \leq 1.$$

Die Zahl

$$\varphi = \arccos \frac{\langle x,y\rangle}{\|x\| \cdot \|y\|} \quad \in [0,\pi]$$

ist daher wohldefiniert; wir nennen sie in Analogie zu Beispiel 21.1 den (unorientierten) **Winkel** zwischen x und y.

Aufgabe 21.24 (Skalarprodukte aus positiv semidefiniten Formen)**.** Zu einer symmetrischen Bilinearform b auf einem reellen Vektorraum V sei $U_b = \{x \in V : b(x,x) = 0\}$. Man zeige:

(a) U_b ist im Allgemeinen kein Unterraum von V.

(b) Ist b jedoch positiv semidefinit, so ist U_b ein Unterraum, und $\overline{b}(\overline{x},\overline{y}) := b(x,y)$ ist ein wohldefiniertes Skalarprodukt auf V/U_b.

(c) Was ergibt sich aus dieser Konstruktion, wenn V der Vektorraum aller stückweise stetigen Funktionen auf einem Intervall $[a,b]$ und $b(f,g) = \int_a^b f(x)g(x)\,dx$ ist? Wie kann man sich in diesem Fall die Elemente von U_b und V/U_b anschaulich vorstellen?

Aufgabe 21.25. Es sei $f\colon V \to V$ ein Endomorphismus eines $\mathbb{K}$-Vektorraums mit Skalarprodukt, so dass $\langle f(x), x\rangle = 0$ für alle $x \in V$. Man zeige:

(a) Ist V ein unitärer Raum (also $\mathbb{K} = \mathbb{C}$), so ist f die Nullabbildung.

(b) Ist V ein euklidischer Raum (also $\mathbb{K} = \mathbb{R}$), so ist f nicht notwendig die Nullabbildung.

21.C Orthogonalität

Der mit Abstand wichtigste Fall der Winkeldefinition ist derjenige, in dem die beiden betrachteten Vektoren „senkrecht aufeinander stehen“, also dieser Winkel gleich $\frac{\pi}{2}$ und nach Konstruktion 21.23 damit das Skalarprodukt gleich 0 ist. Im Gegensatz zu allgemeinen Winkeln können wir diesen Spezialfall auch wieder sowohl für reelle als auch für komplexe Vektorräume definieren.

Definition 21.26 (Orthogonale Vektoren)**.** Es sei V ein Vektorraum mit Skalarprodukt.

(a) Zwei Vektoren $x, y \in V$ heißen **orthogonal** bzw. **senkrecht** zueinander (in Zeichen: $x \perp y$), wenn $\langle x, y\rangle = 0$ gilt. Für $x \in V$ und einen Unterraum $U \leq V$ schreiben wir kurz $x \perp U$, falls $x \perp y$ für alle $y \in U$.

(b) Eine Familie $B = (x_i)_{i\in I}$ in V heißt **orthogonal**, wenn $x_i \perp x_j$ für alle $i, j \in I$ mit $i \neq j$ gilt, also wenn die gegebenen Vektoren paarweise zueinander senkrecht sind. Gilt zusätzlich $\|x_i\| = 1$ für alle $i \in I$, so nennt man B **orthonormal**.

(c) Eine **Orthogonalbasis** bzw. **Orthonormalbasis** von V ist eine orthogonale bzw. orthonormale Familie, die gleichzeitig eine Basis von V ist.

Beispiel 21.27.

(a) Nach Definition ist der Nullvektor orthogonal zu jedem anderen Vektor.

(b) Die Standardbasis von $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ ist natürlich eine Orthonormalbasis bezüglich des Standardskalarprodukts.

Bemerkung 21.28.

(a) Die Orthogonalitätsrelation ist symmetrisch: Gilt $x \perp y$ für zwei Vektoren x und y in einem Vektorraum mit Skalarprodukt, ist also $\langle x, y\rangle = 0$, so folgt auch $\langle y, x\rangle = \overline{\langle x, y\rangle} = \overline{0} = 0$ und damit $y \perp x$.

(b) Es sei U ein Unterraum eines Vektorraums V mit Skalarprodukt. Sind dann $x \in V$ und $(x_1, \ldots, x_n)$ eine Basis von U, so genügt es für die Bedingung $x \perp U$ zu überprüfen, dass $x \perp x_i$ für alle $i = 1, \ldots, n$: In diesem Fall gilt nämlich $\langle x, x_i\rangle = 0$ für alle $i = 1, \ldots, n$, also auch

$$\langle x, \lambda_1 x_1 + \cdots + \lambda_n x_n\rangle = \lambda_1 \langle x, x_1\rangle + \cdots + \lambda_n \langle x, x_n\rangle = 0$$

für alle $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$, und damit $\langle x, y\rangle = 0$ für alle $y \in U$.

Eine erste sehr nützliche Eigenschaft der Orthogonalität einer Familie ist, dass daraus bereits ihre lineare Unabhängigkeit folgt:

Lemma 21.29. *In einem Vektorraum mit Skalarprodukt ist jede orthogonale Familie, die nicht den Nullvektor enthält, linear unabhängig.*

Insbesondere ist sie damit also stets eine Orthogonalbasis des von ihr aufgespannten Unterraums.

Beweis. Es sei $B = (x_i)_{i\in I}$ orthogonal mit $x_i \neq 0$ für alle $i \in I$. Weiterhin sei $\sum_{i\in I} \lambda_i x_i = 0$ eine Linearkombination des Nullvektors (wobei nur endlich viele $\lambda_i \in \mathbb{K}$ ungleich 0 sind, siehe Definition 14.1 (b)). Bilden wir dann das Skalarprodukt dieser Gleichung mit einem x_k für $k \in I$, so folgt

$$0 = \left\langle x_k, \sum_{i\in I} \lambda_i x_i \right\rangle = \sum_{i\in I} \lambda_i \langle x_k, x_i\rangle = \lambda_k \|x_k\|^2,$$

weil B orthogonal ist und damit $\langle x_k, x_i\rangle$ für alle $i \neq k$ gilt. Da weiterhin nach Voraussetzung $x_k \neq 0$ und damit aufgrund der positiven Definitheit $\|x_k\| \neq 0$ ist, folgt $\lambda_k = 0$. Dies gilt aber für alle k, d. h. die ursprüngliche Linearkombination ist trivial, und B damit linear unabhängig. □

Wir wollen nun sehen, dass Orthonormalbasen in jedem (endlich-dimensionalen) Vektorraum mit Skalarprodukt existieren und zu besonders schönen Eigenschaften führen. In der Tat gilt sogar noch mehr, nämlich eine „Basisergänzungseigenschaft" analog zu Folgerung 14.16 für Orthonormalbasen. Um dies zu beweisen, brauchen wir die folgende Konstruktion der orthogonalen Projektion auf einen Unterraum.

Satz und Definition 21.30 (Orthogonale Projektionen)**.** *Es sei U ein endlich-dimensionaler Unterraum eines Vektorraums V mit Skalarprodukt. Ferner sei $(x_1, \dots, x_n)$ eine Orthonormalbasis von U (wir werden in Satz 21.32 noch sehen, dass eine solche Orthonormalbasis immer existiert). Dann gilt:*

(a) *Es gibt genau eine lineare Abbildung $f\colon V \to U$ mit $x - f(x) \perp U$ für alle $x \in V$, nämlich*

$$f(x) = \sum_{i=1}^{n} \langle x_i, x\rangle x_i.$$

*Man nennt sie die **orthogonale Projektion** von V auf U.*

(b) *Die orthogonale Projektion $f(x)$ von x aus (a) ist der eindeutig bestimmte Punkt in U mit kleinstem Abstand zu x, d. h. es gilt $\|x - y\| > \|x - f(x)\|$ für alle $y \in U$ mit $y \neq f(x)$.*

Beweis.

(a) Für ein zunächst festes $x \in V$ seien $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ die Koordinaten des Vektors $f(x) \in U$ bezüglich der gegebenen Basis $(x_1, \dots, x_n)$, also $f(x) = \lambda_1 x_1 + \dots + \lambda_n x_n$. Dann gilt

$$\begin{aligned} x - f(x) \perp U &\overset{21.28\,(b)}{\Leftrightarrow} \langle x_i, x - f(x)\rangle = 0 \text{ für alle } i = 1, \dots, n \\ &\Leftrightarrow \langle x_i, x - \lambda_1 x_1 - \dots - \lambda_n x_n\rangle = 0 \text{ für alle } i = 1, \dots, n \\ &\overset{(*)}{\Leftrightarrow} \langle x_i, x\rangle - \lambda_i = 0 \text{ für alle } i = 1, \dots, n \\ &\Leftrightarrow f(x) = \langle x_1, x\rangle x_1 + \dots + \langle x_n, x\rangle x_n, \end{aligned}$$

wobei wir in $(*)$ wieder die Orthonormalität der Basis verwendet haben. Da diese Vorschrift offensichtlich linear in x ist, ist das Lemma damit gezeigt.

(b) Es sei $y \in U$ mit $y \neq f(x)$, also $y = f(x) + u$ für ein $u \in U \setminus \{0\}$. Dann gilt

$$\begin{aligned} \|x - y\|^2 &= \|x - f(x) - u\|^2 = \langle x - f(x) - u, x - f(x) - u\rangle \\ &= \underbrace{\langle x - f(x), x - f(x)\rangle}_{=\|x - f(x)\|^2} + \underbrace{\langle u, u\rangle}_{=\|u\|^2} + \underbrace{\langle x - f(x), -u\rangle + \langle -u, x - f(x)\rangle}_{=0 \text{ wegen } x - f(x) \perp U}. \end{aligned}$$

Wegen der positiven Definitheit des Skalarprodukts ist aber $\|u\| > 0$, und damit wie behauptet $\|x - y\|^2 > \|x - f(x)\|^2$. □

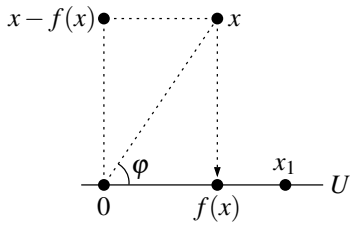
Bemerkung 21.31 (Geometrische Deutung orthogonaler Projektionen)**.** Orthogonale Projektionen haben eine einfache geometrische Deutung, die im Bild unten rechts für einen eindimensionalen Unterraum $U = \mathrm{Lin}(x_1)$ von $V = \mathbb{R}^2$ mit $\|x_1\| = 1$ dargestellt ist.

Die Abbildung konstruiert in diesem Fall zu einem $x \in V$ das Lot auf den Unterraum U; der so entstehende Punkt ist die orthogonale Projektion $f(x)$. Nach Konstruktion 21.23 ist dann nämlich $f(x) = \lambda x_1$ mit

$$\lambda = \|x\| \cos\varphi = \|x\| \cdot \frac{\langle x_1, x\rangle}{\|x_1\| \cdot \|x\|} = \langle x_1, x\rangle,$$

und damit wie in Satz 21.30

$$f(x) = \langle x_1, x\rangle x_1.$$



Beachte, dass $f(x)$ wie in Satz 21.30 auch anschaulich im Bild der Punkt in U ist, der am nächsten an x liegt, und dass die Differenz $x - f(x)$ senkrecht zu x_1 ist. Wir können also einen Vektor konstruieren, der auf (allen Vektoren von) U senkrecht steht, indem wir von einem beliebigen Vektor $x \notin U$ seine orthogonale Projektion auf U abziehen. Dies ist die Grundidee des folgenden Verfahrens zur Bestimmung von Orthonormalbasen.

Satz 21.32 (Gram-Schmidtsches Orthonormalisierungsverfahren). *Es seien V ein endlich-dimensionaler Vektorraum mit Skalarprodukt und $U \leq V$ ein Untervektorraum. Dann lässt sich jede Orthonormalbasis von U zu einer Orthonormalbasis von V ergänzen.*

Insbesondere besitzt also jeder endlich-dimensionale Vektorraum mit Skalarprodukt eine Orthonormalbasis.

Beweis. Der Beweis dieses Satzes ist konstruktiv und erlaubt daher auch eine einfache (iterative) Konstruktion solcher Orthonormalbasen.

Es sei $(x_1, \ldots, x_k)$ die gegebene Orthonormalbasis von U. Ist bereits $U = V$, so sind wir natürlich fertig. Ansonsten führen wir die folgenden Schritte aus:

- Wähle einen beliebigen Vektor $x \in V \setminus U$.
- Wir subtrahieren von x die orthogonale Projektion von x auf U und erhalten so nach Satz 21.30 (a) den auf allen $x_1, \ldots, x_k$ senkrecht stehenden Vektor
$$y_{k+1} := x - \sum_{i=1}^{k} \langle x_i, x\rangle x_i.$$
- Wegen $x \notin U = \mathrm{Lin}(x_1, \ldots, x_k)$ ist $y_{k+1} \neq 0$. Wir können y_{k+1} also normieren zu
$$x_{k+1} := \frac{y_{k+1}}{\|y_{k+1}\|}.$$
Damit ist $(x_1, \ldots, x_{k+1})$ eine orthonormale Familie, die also nach Lemma 21.29 eine Orthonormalbasis eines Unterraums $U' = \mathrm{Lin}(x_1, \ldots, x_{k+1}) \supsetneq U$ ist.

Ist jetzt $U' = V$, so sind wir fertig. Ansonsten setzen wir das obige Verfahren iterativ mit dem neuen Unterraum U' fort, bis wir genügend Vektoren gefunden haben. □

Beispiel 21.33. Wir wollen eine Orthonormalbasis von $\mathbb{R}^2$ für das Skalarprodukt

$$\langle x, y\rangle = x^\mathsf{T} \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} y$$

aus Beispiel 21.15 (b) bestimmen. Beachte zunächst, dass die Standardbasis (e_1, e_2) *keine* Orthonormalbasis ist, denn es ist z. B.

$$\langle e_1, e_2\rangle = (1 \;\; 0) \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1 \neq 0.$$

Wir wenden also das Gram-Schmidtsche Orthonormalisierungsverfahren aus Satz 21.32 an. Den ersten Vektor können wir dabei (ungleich 0) beliebig wählen, z. B. $y_1 = e_1$. Wegen $\langle e_1, e_1\rangle = 1$, also $\|e_1\| = 1$, erhalten wir nach Normieren mit den Notationen wie im Beweis des Satzes

$$x_1 = \frac{y_1}{\|y_1\|} = \frac{e_1}{1} = e_1.$$

Für den zweiten Vektor starten wir mit einem beliebigen Element von $\mathbb{R}^2 \setminus \operatorname{Lin}(x_1)$, z. B. $x = e_2$, und subtrahieren von ihm die orthogonale Projektion auf $\operatorname{Lin}(x_1)$:

$$y_2 = e_2 - \langle x_1, e_2 \rangle x_1 = e_2 - \langle e_1, e_2 \rangle e_1 = e_2 - 1\, e_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Wegen

$$\|y_2\|^2 = \langle y_2, y_2 \rangle = \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 3$$

erhalten wir nach Normieren also die Orthonormalbasis

$$\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right)$$

von $\mathbb{R}^2$ bezüglich des gegebenen Skalarprodukts.

Bemerkung 21.34 (Gram-Schmidt als Normalformensatz)**.** Es sei V ein endlich erzeugter Vektorraum mit Skalarprodukt $\langle x, y \rangle = b(x, y)$. Ist dann $B = (x_1, \dots, x_n)$ gemäß Satz 21.32 eine Orthonormalbasis von V, so ist die Gramsche Matrix von b bezüglich B nach Folgerung 21.6 gerade

$$A_b^B = \left(\langle x_i, x_j \rangle \right)_{i,j} = E_n,$$

denn es ist ja $\langle x_i, x_j \rangle$ gleich 1 für $i = j$ und 0 für $i \neq j$. In Analogie zu den Normalformaussagen aus Satz 16.42 und Folgerung 20.15 kann man die Existenz von Orthonormalbasen damit auch so auffassen, dass es zu jedem Skalarprodukt eine Basis gibt, bezüglich der die Gramsche Matrix die Einheitsmatrix ist, also diese sehr einfache Form hat.

Da sich Gramsche Matrizen unter Basiswechsel wie in Lemma 21.7 bzw. Konstruktion 21.19 verhalten, ist die entsprechende Aussage in Matrixform gemäß Lemma 21.11 also, dass es zu jeder positiv definiten, symmetrischen (im Fall $\mathbb{K} = \mathbb{R}$) bzw. hermiteschen (im Fall $\mathbb{K} = \mathbb{C}$) Matrix $A \in \operatorname{Mat}(n \times n, \mathbb{K})$ eine invertierbare Matrix $T \in \operatorname{GL}(n, \mathbb{K})$ gibt mit $\overline{T}^\mathsf{T} A T = E$: Man muss in die Spalten von T einfach eine Orthonormalbasis von $\mathbb{K}^n$ bezüglich des Skalarprodukts $\langle x, y \rangle = \overline{x}^\mathsf{T} A y$ schreiben. Für das Skalarprodukt und die Orthonormalbasis aus Beispiel 21.33 ergibt sich also z. B.

$$T^\mathsf{T} A T = E \quad \text{mit} \quad A = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix} \quad \text{und} \quad T = \begin{pmatrix} 1 & -\frac{1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix},$$

was man durch direkte Berechnung des Matrixprodukts auch sofort bestätigen kann.

Eine Verallgemeinerung dieser Aussage auf nicht notwendig positiv definite symmetrische Bilinearformen bzw. hermitesche Sesquilinearformen werden wir in Satz 22.36 bzw. Bemerkung 22.37 (b) kennenlernen.

50

Folgerung 21.35 (Determinante hermitescher und positiv definiter Matrizen)**.** *Für jede hermitesche Matrix* $A \in \operatorname{Mat}(n \times n, \mathbb{C})$ *gilt:*

(a) $\det A \in \mathbb{R}$.

(b) *Ist A zusätzlich positiv definit, so ist sogar* $\det A > 0$.

Beweis.

(a) Ist $\overline{A}^\mathsf{T} = A$, so folgt $\overline{\det A} = \det \overline{A} = \det \overline{A}^\mathsf{T} = \det A$ und damit $\det A \in \mathbb{R}$.

(b) Nach Bemerkung 21.34 gibt es eine invertierbare Matrix T mit $\overline{T}^\mathsf{T} A T = E$. Damit ist

$$1 = \det E = \det \overline{T}^\mathsf{T} \cdot \det A \cdot \det T = \det A \cdot \det \overline{T} \cdot \det T = \det A \cdot |\det T|^2,$$

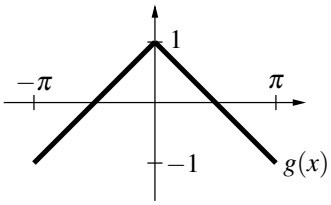
woraus $\det A > 0$ folgt. □

Aufgabe 21.36. Es seien $n \in \mathbb{N}$ und V_n der Vektorraum aller reellen Polynome vom Grad höchstens n.

(a) Für welche $m \in \mathbb{N}$ definiert $\langle f, g \rangle := \sum_{i=0}^{m} f(i)\, g(i)$ ein Skalarprodukt auf V_n?

(b) Berechne für dieses Skalarprodukt eine Orthonormalbasis von V_2 im Fall $m = n = 2$.

Aufgabe 21.37. Es sei $V = C^0([-\pi,\pi])$ der reelle Vektorraum aller stetigen Funktionen auf dem Intervall $[-\pi,\pi]$ mit dem üblichen Skalarprodukt $\langle f,g\rangle := \int_{-\pi}^{\pi} f(x)g(x)\,dx$ wie in Beispiel 21.15 (d). Wir betrachten darin das Element $g \in V$ definiert durch $g(x) := 1 - \frac{2}{\pi}|x|$.



Für $n \in \mathbb{N}_{>0}$ seien weiterhin $f_n \in V$ mit $f_n(x) := \frac{1}{\sqrt{\pi}}\cos nx$, und $U_n := \mathrm{Lin}(f_1,\dots,f_n) \le V$.

(a) Zeige, dass $(f_1,\dots,f_n)$ für alle n eine Orthonormalbasis von U_n ist.

(b) Berechne für alle $n \in \mathbb{N}_{>0}$ die orthogonale Projektion g_n von g auf U_n (also die Funktion in U_n, die nach Satz 21.30 (b) von g den kleinsten Abstand hat, d. h. sie am besten approximiert).

(c) Zeichne die Funktionen g_n für kleine n mit einem Computer und vergleiche sie mit der ursprünglichen Funktion g.

Wir wollen nun noch drei interessante Anwendungen von Orthonormalbasen betrachten. Die erste betrifft Abbildungsmatrizen bzw. Basiswechselmatrizen: Bisher mussten wir zur Berechnung solcher Matrizen wie in Bemerkung 19.1 Koordinatendarstellungen von Vektoren berechnen und die dabei erhaltenen Koeffizienten in die Spalten einer Matrix schreiben. Dies erfordert die Lösung linearer Gleichungssysteme und ist daher recht rechenaufwändig. In einem Vektorraum mit Skalarprodukt ist die Berechnung solcher Matrizen bezüglich Orthonormalbasen dagegen deutlich einfacher:

Lemma 21.38 (Abbildungsmatrizen zu Orthonormalbasen)**.** *Es seien V ein endlich-dimensionaler Vektorraum mit Skalarprodukt und $B = (x_1,\dots,x_n)$ eine Orthonormalbasis von V.*

(a) *Für alle $x \in V$ gilt*

$$x = \langle x_1,x\rangle x_1 + \cdots + \langle x_n,x\rangle x_n.$$

Mit anderen Worten ist in der Koordinatendarstellung $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$ bezüglich B also $\lambda_i = \langle x_i,x\rangle$ für alle $i = 1,\dots,n$.

(b) *Ist $f\colon V \to V$ ein Endomorphismus, so ist die Abbildungsmatrix von f bezüglich B gleich $A_f^B = \big(\langle x_i, f(x_j)\rangle\big)_{i,j}$.*

(c) *Ist $B' = (y_1,\dots,y_n)$ eine weitere Basis von V, so ist die zugehörige Basiswechselmatrix gleich $A^{B',B} = \big(\langle x_i,y_j\rangle\big)_{i,j}$.*

Beweis.

(a) Dies folgt unmittelbar aus Satz 21.30 (a) für $U = V$, da die orthogonale Projektion von V auf sich selbst offensichtlich die Identität ist, in der Notation des Satzes also $f(x) = x$ gilt.

(b) Nach Satz 16.23 ist der Eintrag in Zeile i und Spalte j von A_f^B genau die i-te Koordinate von $f(x_j)$ bezüglich B, nach (a) also $\langle x_i, f(x_j)\rangle$.

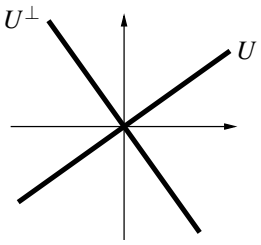
(c) Nach Definition 16.28 ist der Eintrag in Zeile i und Spalte j von $A^{B',B}$ genau die i-te Koordinate von y_j bezüglich B, nach (a) also $\langle x_i,y_j\rangle$. □

In der zweiten Anwendung von Orthonormalbasen geht es um Komplemente von Unterräumen in endlich erzeugten Vektorräumen: Wir wissen ja nach Beispiel 15.9 und Satz 15.10 bereits, dass solche Komplemente zwar immer existieren, aber in der Regel nicht eindeutig bestimmt sind. Falls im zugrundeliegenden Vektorraum aber ein Skalarprodukt gegeben ist, wollen wir jetzt sehen, dass es zu jedem Unterraum ein besonderes Komplement gibt – nämlich das sogenannte *orthogonale Komplement* – das immer eindeutig bestimmt ist.

Definition 21.39 (Orthogonales Komplement)**.** Es seien V ein Vektorraum mit Skalarprodukt und $U \leq V$ ein Unterraum von V. Dann nennen wir die Menge

$$\begin{aligned} U^\perp &:= \{x \in V : x \perp U\} \\ &= \{x \in V : \langle x, y\rangle = 0 \text{ für alle } y \in U\} \end{aligned}$$

das **orthogonale Komplement** von U. (Das Bild rechts illustriert dies für den Vektorraum $\mathbb{R}^2$ mit dem Standardskalarprodukt.)



Satz 21.40. *Es seien V ein endlich-dimensionaler Vektorraum mit Skalarprodukt und $U \leq V$ ein Unterraum. Dann ist das orthogonale Komplement $U^\perp$ ein Komplement von U im Sinne von Definition 15.7, d. h. es gilt $V = U \oplus U^\perp$.*

Beweis. Nach Satz 21.32 können wir eine Orthonormalbasis $(x_1, \dots, x_k)$ von U finden und zu einer Orthonormalbasis $(x_1, \dots, x_n)$ von V ergänzen. Nach Bemerkung 15.11 ist $U' := \operatorname{Lin}(x_{k+1}, \dots, x_n)$ dann ein Komplement von U. Es genügt also zu zeigen, dass $U^\perp = U'$ gilt: Für einen Vektor $x \in V$ mit der Darstellung $x = \lambda_1 x_1 + \dots + \lambda_n x_n$ gilt

$$\begin{aligned} x \in U^\perp &\overset{21.28\,(b)}{\Leftrightarrow} \langle x_i, \lambda_1 x_1 + \dots + \lambda_n x_n\rangle = 0 \text{ für alle } i = 1, \dots, k \\ &\Leftrightarrow \lambda_1 \langle x_i, x_1\rangle + \dots + \lambda_n \langle x_i, x_n\rangle = 0 \text{ für alle } i = 1, \dots, k \\ &\Leftrightarrow \lambda_i \|x_i\|^2 = 0 \text{ für alle } i = 1, \dots, k \\ &\Leftrightarrow \lambda_i = 0 \text{ für alle } i = 1, \dots, k \\ &\Leftrightarrow x \in U'. \end{aligned}$$

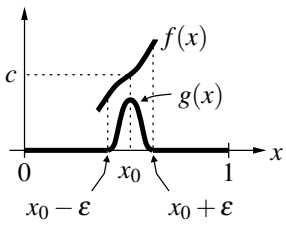
□

Bemerkung 21.41. Der Beweis von Satz 21.40 zeigt auch, wie man das orthogonale Komplement eines Unterraums U in V berechnen kann: Man ergänzt eine Orthonormalbasis $(x_1, \dots, x_k)$ von U zu einer Orthonormalbasis $(x_1, \dots, x_n)$ von V; die dabei neu hinzu genommenen Vektoren $(x_{k+1}, \dots, x_n)$ bilden dann eine Orthonormalbasis von $U^\perp$.

Da in diesem Fall natürlich auch $x_1, \dots, x_k$ die Vektoren $x_{k+1}, \dots, x_n$ zu einer Orthonormalbasis von V ergänzen, folgt aus dieser Beschreibung auch sofort, dass $(U^\perp)^\perp = U$.

Beispiel 21.42 (Orthogonale Komplemente in unendlich-dimensionalen Vektorräumen)**.** Als „abschreckendes Beispiel" dafür, dass Satz 21.40 nicht so selbstverständlich ist, wie er vielleicht scheint, wollen wir kurz zeigen, dass dieser Satz für unendlich-dimensionale Vektorräume im Allgemeinen *falsch* ist, dass das orthogonale Komplement dann also nicht unbedingt immer ein Komplement im Sinne von Definition 15.7 ist. Dazu sei $V = C^0([0,1])$ der Vektorraum aller stetigen Funktionen auf dem Intervall $[0,1]$ mit dem in Beispiel 21.15 (d) betrachteten Skalarprodukt $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. Wir betrachten hierin den Unterraum $U \subset V$ aller *differenzierbaren* Funktionen; offensichtlich ist $U \subsetneq V$. Wir werden aber zeigen, dass $U^\perp = \{0\}$ ist, so dass also insbesondere $U + U^\perp = U \neq V$, d. h. $U^\perp$ kein Komplement von U ist.

Dazu sei $f \in V \setminus \{0\}$ beliebig; wir werden zeigen, dass $f \notin U^\perp$. Wegen $f \neq 0$ gibt es ein $x_0 \in [0,1]$ mit $f(x_0) =: c \neq 0$. Wir können ohne Einschränkung annehmen, dass $c > 0$ ist. Da f stetig ist, gibt es ein $\varepsilon > 0$, so dass $f(x) \geq \frac{c}{2}$ für alle x mit $|x - x_0| < \varepsilon$ ist. Es sei nun $g \in U$ eine differenzierbare Funktion wie im Bild. Dann ist offensichtlich $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx \geq \frac{c}{2}\int_0^1 g(x)\,dx > 0$, denn die Funktion g ist überall nicht negativ und hat außerdem eine nicht verschwindende Fläche unter ihrem Graphen.



Wir haben damit also ein $g \in U$ mit $\langle f, g\rangle \neq 0$ gefunden. Damit ist $f \notin U^\perp$. Da f beliebig war, folgt also wie behauptet $U^\perp = \{0\}$.

Unsere letzte Anwendung der Existenz von Orthonormalbasen in diesem Kapitel ist das folgende einfache Kriterium für die positive Definitheit einer Matrix.

Satz 21.43 (Hurwitz-Kriterium). *Es sei $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, \mathbb{K})$ eine symmetrische (im Fall $\mathbb{K} = \mathbb{R}$) bzw. hermitesche (im Fall $\mathbb{K} = \mathbb{C}$) Matrix. Für $k = 1, \dots, n$ bezeichnen wir mit $A_k = (a_{i,j})_{i,j=1,\dots,k} \in \mathrm{Mat}(k \times k, \mathbb{K})$ die Matrizen, die man erhält, wenn man von A nur die ersten k Zeilen und Spalten betrachtet. (Beachte, dass ihre Determinanten nach Folgerung 21.35 (a) in jedem Fall reell sind.)*

Dann ist A genau dann positiv definit, wenn $\det A_k > 0$ für alle $k = 1, \dots, n$.

Beweis. Es sei b die zu A gehörige symmetrische Bilinearform bzw. hermitesche Sesquilinearform auf $\mathbb{K}^n$, also $b(x,y) = \overline{x}^\mathsf{T} A y$ für alle $x, y \in \mathbb{K}^n$.

„$\Rightarrow$“ Ist A positiv definit und b damit ein Skalarprodukt, so ist nach Bemerkung 21.16 auch die Einschränkung von b auf den Unterraum $\mathrm{Lin}(e_1, \dots, e_k)$ ein Skalarprodukt. Die zugehörige Gramsche Matrix $(b(e_i, e_j))_{i,j=1,\dots,k} = A_k$ ist also ebenfalls positiv definit und hat damit nach Folgerung 21.35 (b) eine positive Determinante.

„$\Leftarrow$“ Wir zeigen die Aussage mit Induktion über n, der Induktionsanfang für $n = 1$ ist trivial.

Für den Induktionsschritt $n \to n+1$ bemerken wir zunächst, dass nach Annahme insbesondere $\det A_k > 0$ für $k = 1, \dots, n$ gilt. Damit ist A_n nach Induktionsvoraussetzung positiv definit, und wir können nach Satz 21.32 eine Orthonormalbasis $(x_1, \dots, x_n)$ für das zugehörige Skalarprodukt finden, das gerade die Einschränkung von b auf $\mathrm{Lin}(e_1, \dots, e_n)$ ist. Bilden wir nun wie im Gram-Schmidtschen Orthonormalisierungsverfahren in Satz 21.32 den Vektor

$$x_{n+1} := e_{n+1} - \sum_{i=1}^{n} b(x_i, e_{n+1})\, x_i \quad \notin \mathrm{Lin}(x_1, \dots, x_n),$$

so gilt wieder

$$b(x_j, x_{n+1}) = b(x_j, e_{n+1}) - \sum_{i=1}^{n} b(x_i, e_{n+1}) b(x_j, x_i) = b(x_j, e_{n+1}) - b(x_j, e_{n+1}) = 0$$

für alle $j = 1, \dots, n$, da $(x_1, \dots, x_n)$ orthonormal ist. Damit hat die Gramsche Matrix von b bezüglich der Basis $B = (x_1, \dots, x_{n+1})$ die Form

$$A_b^B = \overline{T}^\mathsf{T} A T = (b(x_i, x_j))_{i,j} = \left(\begin{array}{c|c} E_n & 0 \\ \hline 0 & b(x_{n+1}, x_{n+1}) \end{array}\right),$$

wobei wir wie üblich $T = (\, x_1 \mid \cdots \mid x_{n+1} \,)$ gesetzt haben. Insbesondere folgt damit

$$b(x_{n+1}, x_{n+1}) = \det A_b^B = \det\left(\overline{T}^\mathsf{T} A T\right) = \overline{\det T} \cdot \det A \cdot \det T = |\det T|^2 \cdot \det A > 0.$$

Ist nun $x = \lambda_1 x_1 + \dots + \lambda_{n+1} x_{n+1} \neq 0$ beliebig, so ergibt sich daraus

$$b(x,x) = \sum_{i,j=1}^{n+1} \overline{\lambda_i} \lambda_j b(x_i, x_j) = |\lambda_1|^2 + \dots + |\lambda_n|^2 + \underbrace{b(x_{n+1}, x_{n+1})}_{>0} |\lambda_{n+1}|^2 > 0.$$

Also ist b und damit auch A positiv definit. □

Bemerkung 21.44. Die Determinanten von quadratischen Teilmatrizen einer gegebenen Matrix A werden oft auch *Minoren* von A genannt, die in Satz 21.43 auftretenden Determinanten $\det A_k$ bezeichnet man als *Hauptminoren* von A. Das Hurwitz-Kriterium ist daher auch unter dem Namen *Hauptminorenkriterium* bekannt.

Beispiel 21.45. Wir betrachten noch ein letztes Mal die symmetrische reelle Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$$

aus Beispiel 21.15 (b), die wir dort schon durch eine direkte Rechnung als positiv definit erkannt haben. Mit dem Hurwitz-Kriterium folgt dies (in der Notationen des Satzes) nun auch einfacher aus

$$\det A_1 = \det(1) = 1 > 0 \qquad \text{und} \qquad \det A_2 = \det A = 3 > 0.$$

Aufgabe 21.46. Für ein $\lambda \in \mathbb{R}$ sei $A = \begin{pmatrix} \lambda & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 7 \end{pmatrix}$.

(a) Für welche λ ist A positiv definit? Für welche λ ist A negativ definit?

(b) Wir betrachten nun $V = \mathbb{R}^3$ als euklidischen Vektorraum mit dem Skalarprodukt b_A für den Wert $\lambda = 1$. Berechne zu $U = \mathrm{Lin}(e_1) \leq V$ eine Orthonormalbasis des orthogonalen Komplements $U^\perp$.

22. Endomorphismen euklidischer und unitärer Räume

Im letzten Kapitel haben wir ausführlich Vektorräume studiert, auf denen die Zusatzstruktur eines Skalarprodukts gegeben ist. Wir wollen nun sehen, welche Vorteile und Vereinfachungen uns diese Zusatzstruktur bei der Untersuchung von Endomorphismen gibt, wenn diese in gewissem Sinne mit dem gegebenen Skalarprodukt verträglich sind. Das zentrale Resultat dieses Kapitels wird schließlich der sogenannte Spektralsatz in Abschnitt 22.C sein, der die Diagonalisierbarkeit solcher Endomorphismen garantiert und – wie wir sehen werden – so universell ist, dass er bei geschickter Anwendung auch Aussagen über Bilinearformen bzw. Sesquilinearformen und in Abschnitt 22.D sogar über lineare Abbildungen mit unterschiedlichem Start- und Zielraum machen kann.

22.A Orthogonale und unitäre Abbildungen

Die natürlichste Verträglichkeitsbedingung zwischen Morphismen und Skalarprodukten ist vermutlich die folgende.

Definition 22.1 (Orthogonale und unitäre Abbildungen und Matrizen)**.**

(a) Ein Endomorphismus $f\colon V \to V$ eines Vektorraums V mit Skalarprodukt heißt **orthogonal** (im Fall $\mathbb{K} = \mathbb{R}$) bzw. **unitär** (im Fall $\mathbb{K} = \mathbb{C}$), wenn

$$\langle f(x), f(y)\rangle = \langle x, y\rangle$$

für alle $x, y \in V$ gilt.

(b) Eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ heißt **orthogonal**, wenn $A^{\mathsf{T}}A = E$ gilt, also wenn A invertierbar ist mit $A^{-1} = A^{\mathsf{T}}$. Wir bezeichnen die Menge aller orthogonalen $n \times n$-Matrizen mit $\mathrm{O}(n) \subset \mathrm{Mat}(n \times n, \mathbb{R})$.

Eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ heißt **unitär**, wenn $\overline{A}^{\mathsf{T}}A = E$ gilt, also wenn A invertierbar ist mit $A^{-1} = \overline{A}^{\mathsf{T}}$. Die Menge aller unitären $n \times n$-Matrizen wird mit $\mathrm{U}(n) \subset \mathrm{Mat}(n \times n, \mathbb{C})$ bezeichnet.

Wie üblich schreiben wir diese Bedingung im Folgenden oft in beiden Fällen als $\overline{A}^{\mathsf{T}}A = E$.

Bemerkung 22.2.

(a) Ist $B = (x_1, \ldots, x_n)$ eine Basis von V, so genügt es, die Bedingung eines orthogonalen bzw. unitären Morphismus $f\colon V \to V$ für alle Paare von Basisvektoren zu überprüfen: Ist nämlich $\big\langle f(x_i), f(x_j)\big\rangle = \big\langle x_i, x_j\big\rangle$ für alle $i, j = 1, \ldots, n$, so gilt wegen der Linearität von f und der Bilinearität bzw. Sesquilinearität des Skalarprodukts auch für alle $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$ und $y = \mu_1 x_1 + \cdots + \mu_n x_n$

$$\langle f(x), f(y)\rangle = \sum_{i,j=1}^{n} \overline{\lambda_i}\mu_j \left\langle f(x_i), f(x_j)\right\rangle = \sum_{i,j=1}^{n} \overline{\lambda_i}\mu_j \left\langle x_i, x_j\right\rangle = \langle x, y\rangle\,.$$

(b) Eine Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, \mathbb{K})$ ist genau dann orthogonal bzw. unitär, wenn die Spalten von A eine Orthonormalbasis bezüglich des Standardskalarprodukts bilden: Nach Definition des Matrixprodukts ist nämlich

$$\overline{A}^{\mathsf{T}}A = \left(\sum_{j=1}^{n} \overline{a_{j,i}}\, a_{j,k}\right)_{i,k}.$$

Der Ausdruck $\sum_{j=1}^{n} \overline{a_{j,i}}\, a_{j,k}$ ist aber genau das Standardskalarprodukt der i-ten mit der k-ten Spalte von A. Daher bilden diese Spalten genau dann eine Orthonormalbasis bezüglich des

Standardskalarprodukts, wenn dieser Ausdruck gleich 1 für $i = k$ und 0 für $i \neq k$ ist, also wenn $\overline{A}^{\mathsf{T}} A = E$ ist.

Es wäre also vermutlich konsequenter, eine reelle Matrix A mit $A^{\mathsf{T}} A = E$ ortho*normal* statt ortho*gonal* zu nennen. Die Bezeichnung „orthogonale Matrix" ist in der Literatur aber so üblich, dass wir hier nicht davon abweichen wollen.

51

Bemerkung 22.3 (Geometrische Deutung orthogonaler Abbildungen)**.** Nach Definition 22.1 erhalten orthogonale und unitäre Abbildungen Skalarprodukte, und damit auch Längen, Orthogonalität und (im reellen Fall) Winkel zwischen zwei Vektoren. Über $\mathbb{R}$ kann man sie sich daher als Drehungen, Spiegelungen und Kombinationen davon vorstellen (siehe auch Beispiel 22.6 und Aufgabe 22.11).

Wir wollen nun als Erstes zeigen, dass die oben eingeführten Bedingungen für orthogonale und unitäre Matrizen wie erwartet denen der zugehörigen Endomorphismen entsprechen, sofern es sich um Abbildungsmatrizen bezüglich einer Orthonormalbasis handelt. Die Interpretation als Drehungen bzw. Spiegelungen gilt damit also auch für orthogonale Matrizen.

Lemma 22.4. *Es sei V ein endlich-dimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt und B eine Orthonormalbasis von V. Dann gilt:*

(a) *Ein Endomorphismus $f\colon V \to V$ ist genau dann orthogonal bzw. unitär, wenn A_f^B orthogonal bzw. unitär ist.*

(b) *Eine weitere Basis B' von V ist genau dann auch eine Orthonormalbasis, wenn $A^{B',B}$ orthogonal bzw. unitär ist.*

Beweis. Es seien $B = (x_1, \dots, x_n)$ und $A = \big(\langle x_i, y_j \rangle\big)_{i,j}$ für Vektoren $y_1, \dots, y_n \in \mathbb{K}^n$. Dann gilt

$$\begin{aligned}
\overline{A}^{\mathsf{T}} A &= \left(\sum_{j=1}^n \overline{\langle x_j, y_i \rangle} \langle x_j, y_k \rangle \right)_{i,k} && \text{(Definition 16.5)} \\
&= \left(\left\langle \sum_{j=1}^n \langle x_j, y_i \rangle x_j, y_k \right\rangle \right)_{i,k} && \text{(Sesquilinearität des Skalarprodukts)} \\
&= \big(\langle y_i, y_k \rangle\big)_{i,k}. && \text{(Lemma 21.38 (a))}
\end{aligned}$$

(a) Setzen wir $y_i = f(x_i)$ für alle i und eine lineare Abbildung $f\colon V \to V$, so ist $A = A_f^B$ nach Lemma 21.38 (b). Damit ist diese Abbildungsmatrix genau dann orthogonal bzw. unitär, wenn $\big(\langle f(x_i), f(x_k) \rangle\big)_{i,k} = E = \big(\langle x_i, x_k \rangle\big)_{i,k}$ ist, also nach Bemerkung 22.2 (a) wenn f orthogonal bzw. unitär ist.

(b) Setzen wir $B' = (y_1, \dots, y_n)$, so ist $A = A^{B',B}$ nach Lemma 21.38 (c). Damit ist diese Basiswechselmatrix genau dann orthogonal bzw. unitär, wenn $\big(\langle y_i, y_k \rangle\big)_{i,k} = E$ ist, also wenn B' eine Orthonormalbasis ist. □

Bemerkung 22.5. Gemäß Definition 22.1 (b) sind orthogonale bzw. unitäre Matrizen invertierbar. Mit Lemma 22.4 (a) bedeutet dies also, dass orthogonale bzw. unitäre Endomorphismen eines endlich erzeugten Vektorraums mit Skalarprodukt stets Isomorphismen sind.

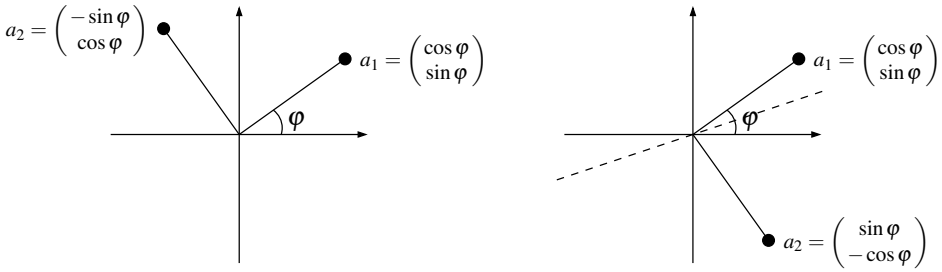
Beispiel 22.6 (O(2))**.** Nach Bemerkung 22.2 (b) ist eine Matrix $A = (a_1 \,|\, a_2) \in \mathrm{Mat}(2 \times 2, \mathbb{R})$ mit $a_1, a_2 \in \mathbb{R}^2$ genau dann orthogonal, wenn bezüglich des Standardskalarprodukts

$$\|a_1\| = 1, \quad \|a_2\| = 1 \quad \text{und} \quad a_1 \perp a_2$$

gilt. Also liegt a_1 auf dem Einheitskreis und lässt sich nach der Polarkoordinatendarstellung aus Satz 9.26 damit als

$$a_1 = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$$

für ein $\varphi \in \mathbb{R}$ schreiben, und a_2 entsteht aus a_1 durch eine positive oder negative Drehung um $\frac{\pi}{2}$. Wir erhalten also die folgenden beiden Möglichkeiten:



(a) $A = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ (b) $A = \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}$

In beiden Fällen werden die Einheitsvektoren e_1 und e_2 durch A auf a_1 und a_2 abgebildet. In (a) haben wir geometrisch also eine Drehung um den Winkel φ, in (b) eine Spiegelung an der gestrichelt eingezeichneten Winkelhalbierenden zwischen e_1 und a_1. Damit lassen sich alle Elemente von $\mathrm{O}(2)$ in der Tat wie erwartet durch Drehungen und Spiegelungen beschreiben.

Lemma 22.7 ($\mathrm{O}(n)$ und $\mathrm{U}(n)$ als Gruppen)**.** *Sind $A, B \in \mathrm{Mat}(n \times n, \mathbb{K})$ orthogonal bzw. unitär, so auch AB und A^{-1}. Insbesondere sind $\mathrm{O}(n)$ und $\mathrm{U}(n)$ also Gruppen (siehe Definition 3.1) mit der Matrixmultiplikation als Verknüpfung; man nennt sie die* ***orthogonalen*** *bzw.* ***unitären Gruppen****.*

Beweis. Es gelte $A^{-1} = \overline{A}^{\mathsf{T}}$ und $B^{-1} = \overline{B}^{\mathsf{T}}$. Dann ist

$$(AB)^{-1} = B^{-1}A^{-1} = \overline{B}^{\mathsf{T}}\overline{A}^{\mathsf{T}} = \overline{AB}^{\mathsf{T}}$$

$$\text{und} \quad (A^{-1})^{-1} = \left(\overline{A}^{\mathsf{T}}\right)^{-1} = \left(\overline{A}^{-1}\right)^{\mathsf{T}} = \overline{A^{-1}}^{\mathsf{T}}$$

nach Lemma 16.7 (d) und Lemma 16.20, d. h. AB und A^{-1} sind ebenfalls orthogonal bzw. unitär. □

Lemma 22.8 (Determinante und Eigenwerte orthogonaler und unitärer Matrizen)**.** *Für jede orthogonale oder unitäre Matrix A gilt:*

(a) $|\det A| = 1$.

(b) *Ist λ ein Eigenwert von A, so ist $|\lambda| = 1$.*

Beweis.

(a) Aus $\overline{A}^{\mathsf{T}}A = E$ ergibt sich sofort $1 = \det E = \det\overline{A}^{\mathsf{T}} \cdot \det A = \overline{\det A} \cdot \det A = |\det A|^2$.

(b) Gilt $Ax = \lambda x$ für ein $x \in \mathbb{K}^n \backslash \{0\}$, so ist auch $\overline{A}\overline{x} = \overline{\lambda}\overline{x}$, nach Transponieren also $\overline{x}^{\mathsf{T}}\overline{A}^{\mathsf{T}} = \overline{\lambda}\overline{x}^{\mathsf{T}}$. Zusammen erhalten wir mit $\overline{A}^{\mathsf{T}}A = E$ so

$$\overline{x}^{\mathsf{T}}x = \overline{x}^{\mathsf{T}}\overline{A}^{\mathsf{T}}Ax = \overline{\lambda}\overline{x}^{\mathsf{T}} \cdot \lambda x = |\lambda|^2 \cdot \overline{x}^{\mathsf{T}}x,$$

wegen $\overline{x}^{\mathsf{T}}x \neq 0$ also $|\lambda|^2 = 1$. □

Bemerkung 22.9. Nach Lemma 22.4 (a) gelten die Aussagen aus Lemma 22.8 genauso für orthogonale bzw. unitäre Endomorphismen endlich erzeugter Vektorräume mit Skalarprodukt.

Bemerkung 22.10 (Spezielle orthogonale und unitäre Matrizen)**.** Bei einer orthogonalen Matrix können die Determinante und die Eigenwerte nach Lemma 22.8 nur 1 und -1 sein. Für eine komplexe unitäre Matrix ist dagegen jede Zahl auf dem komplexen Einheitskreis möglich. In beiden Fällen spielen aber diejenigen Matrizen, deren Determinante gleich 1 ist, eine große Rolle. Man definiert daher:

(a) Eine orthogonale Matrix $A \in \mathrm{O}(n)$ heißt *spezielle orthogonale Matrix*, wenn $\det A = 1$ gilt. Die Menge der speziellen orthogonalen $n \times n$-Matrizen wird mit $\mathrm{SO}(n)$ bezeichnet.

(b) Eine unitäre Matrix $A \in \mathrm{U}(n)$ heißt *spezielle unitäre Matrix*, wenn $\det A = 1$ gilt. Man bezeichnet die Menge der speziellen unitären $n \times n$-Matrizen mit $\mathrm{SU}(n)$.

Man sieht leicht, dass auch $\mathrm{SO}(n)$ und $\mathrm{SU}(n)$ Gruppen bezüglich der Matrixmultiplikation sind, denn wenn $\det A = \det B = 1$ gilt, so ist nach Satz 18.6 ja auch $\det(AB) = \det A^{-1} = 1$ – sind also A und B speziell orthogonal bzw. unitär, so auch AB und A^{-1}.

Wir hatten in Bemerkung 22.3 ja schon gesehen, dass man sich die Elemente von $\mathrm{O}(n)$ als Drehungen bzw. Spiegelungen vorstellen kann. Dabei gilt:

- Drehungen entsprechen den Elementen von $\mathrm{SO}(n)$, da diese kontinuierlich aus der identischen Abbildung erzeugt werden können und ihre Determinante daher nicht von 1 auf -1 wechseln kann.
- Spiegelungen sind bezüglich einer geeigneten Orthonormalbasis $(x_1,\dots,x_n)$ von der Form $x_1 \mapsto -x_1$ und $x_i \mapsto x_i$ für $i \geq 2$, so dass die zugehörige Abbildungsmatrix $\mathrm{diag}(-1,1,\dots,1)$ ist und damit Determinante -1 hat. Damit entsprechen Spiegelungen Elementen von $\mathrm{O}(n) \setminus \mathrm{SO}(n)$.

Aufgabe 22.11 ($\mathrm{O}(3)$ und $\mathrm{O}(4)$)**.** Man zeige:

(a) Ist $A \in \mathrm{O}(3)$, so gibt es ein $\varphi \in \mathbb{R}$ und $T \in \mathrm{O}(3)$ mit

$$T^{-1}AT = \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

(b) Ist $A \in \mathrm{O}(4)$, so gibt es im Allgemeinen *kein* $\varphi \in \mathbb{R}$ und $T \in \mathrm{O}(4)$ mit

$$T^{-1}AT = \begin{pmatrix} \pm 1 & 0 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & -\sin\varphi \\ 0 & 0 & \sin\varphi & \cos\varphi \end{pmatrix}.$$

(Hinweis: Untersuche, ob A einen Eigenwert besitzen muss.)

Was bedeutet das Ergebnis geometrisch?

Aufgabe 22.12. Es sei f ein Endomorphismus eines euklidischen Vektorraums V. Man zeige: Gilt $\|f(x)\| = \|x\|$ für alle $x \in V$, erhält f also Längen von Vektoren, dann ist f bereits orthogonal.

Aufgabe 22.13. Es seien V ein euklidischer Vektorraum und $v \in V$ mit $\|v\| = 1$. Wir betrachten den Endomorphismus $f\colon V \to V$, $x \mapsto x - 2\,\langle v,x\rangle\, v$.

(a) Zeige, dass f eine orthogonale Abbildung ist.

(b) Berechne das charakteristische Polynom χ_f.

(c) Wie kann man f geometrisch beschreiben?

Aufgabe 22.14 (*QR*-Zerlegung)**.** Zeige, dass sich jede reelle invertierbare Matrix $A \in \mathrm{GL}(n,\mathbb{R})$ eindeutig in der Form $A = QR$ schreiben lässt, wobei $Q \in \mathrm{O}(n)$ eine orthogonale Matrix und R eine obere Dreiecksmatrix mit positiven Diagonaleinträgen ist.

(Hinweis: Bezeichne die Spaltenvektoren von A bzw. Q mit $a_1,\dots,a_n \in \mathbb{R}^n$ bzw. $x_1,\dots,x_n \in \mathbb{R}^n$ und schreibe die Matrixgleichung $A = QR$ in Gleichungen für diese Spaltenvektoren um.)

Diese Zerlegung wird in der Vorlesung „Einführung in die Numerik“ im zweiten Studienjahr noch eine wichtige Rolle spielen. Möchte man ein Gleichungssystem $Ax = b$ lösen, so ist dies mit der Zerlegung $A = QR$ wie oben natürlich äquivalent zum System $Rx = Q^{\mathsf{T}}b$, das sich wegen der oberen Diagonalform von R einfach durch Rückwärtseinsetzen lösen lässt. In der Regel ist dieses Verfahren für große Gleichungssysteme schneller und unanfälliger für eine Verstärkung von Rundungsfehlern als das euch bekannte Gauß-Verfahren.

22.B Selbstadjungierte Abbildungen

In diesem Abschnitt wollen wir eine weitere Art der Verträglichkeit eines Endomorphismus mit einem Skalarprodukt behandeln. Obwohl sie im Gegensatz zur Bedingung einer orthogonalen bzw. unitären Abbildung keine einfache geometrische Interpretation besitzt, ist sie dennoch in der Praxis sehr wichtig – vor allem, da sie einfach an der Abbildungsmatrix abzulesen ist. Um sie einzuführen, brauchen wir das folgende Konzept der adjungierten Abbildungen.

Satz und Definition 22.15 (Adjungierte Abbildung)**.** *Es sei $f\colon V \to V$ ein Endomorphismus eines endlich-dimensionalen Vektorraums V mit Skalarprodukt. Dann gilt:*

(a) *Es gibt genau einen Endomorphismus $f^*\colon V \to V$ mit*

$$\langle f(x), y\rangle = \langle x, f^*(y)\rangle$$

für alle $x, y \in V$. Man nennt f^* die zu f ***adjungierte Abbildung****.*

(b) *Ist B eine Orthonormalbasis von V, so gilt für die Abbildungsmatrix von f^* bezüglich B*

$$A^B_{f^*} = \overline{A^B_f}^{\mathsf{T}}.$$

Beweis. Es seien $B = (x_1, \dots, x_n)$ eine Orthonormalbasis von V und $g\colon V \to V$ ein Endomorphismus. Beachte zunächst, dass mit demselben Argument wie in Bemerkung 22.2 (a) genau dann

$$\langle f(x), y\rangle = \langle x, g(y)\rangle \tag{$*$}$$

für alle $x, y \in V$ ist, wenn $\big\langle f(x_i), x_j\big\rangle = \big\langle x_i, g(x_j)\big\rangle$ für alle $i, j = 1, \dots, n$ gilt. Wegen der Hermitizität des Skalarprodukts ist diese Bedingung nun aber äquivalent zu

$$\big(\overline{\langle x_j, f(x_i)\rangle}\big)_{i,j} = \big(\langle x_i, g(x_j)\rangle\big)_{i,j},$$

nach Lemma 21.38 (b) also zu $\overline{A^B_f}^{\mathsf{T}} = A^B_g$. Also gibt es genau einen solchen Endomorphismus g, nämlich denjenigen, der bezüglich der Basis B zur Abbildungsmatrix $\overline{A^B_f}^{\mathsf{T}}$ gehört. Dies zeigt bereits beide Teile des Satzes. □

Bemerkung 22.16. Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$.

(a) Aufgrund von Satz 22.15 (b) nennt man $\overline{A}^{\mathsf{T}}$ manchmal auch die zu A *adjungierte Matrix* und schreibt sie als A^*. Bezüglich einer Orthonormalbasis entsprechen sich dann also die Konzepte der adjungierten Abbildung und der adjungierten Matrix.

(b) Da offensichtlich $\overline{\overline{A}^{\mathsf{T}}}^{\mathsf{T}} = A$ gilt, ist nach Satz 22.15 (b) auch $(f^*)^* = f$ für jeden Endomorphismus f eines endlich erzeugten $\mathbb{K}$-Vektorraums mit Skalarprodukt. Dies bedeutet, dass nicht nur

$$\langle f(x), y\rangle = \langle x, f^*(y)\rangle, \quad \text{sondern auch} \quad \langle f^*(x), y\rangle = \langle x, (f^*)^*(y)\rangle = \langle x, f(y)\rangle$$

für alle $x, y \in V$ gilt: Man kann in einem Skalarprodukt stets die Abbildung f auf der einen Seite durch f^* auf der anderen Seite ersetzen (und umgekehrt).

Beispiel 22.17 (Adjungierte Abbildung einer orthogonalen bzw. unitären Abbildung)**.** Ist eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ orthogonal bzw. unitär, also $\overline{A}^{\mathsf{T}} A = E$, so ist $\overline{A}^{\mathsf{T}} = A^{-1}$ die adjungierte Matrix zu A. Für eine orthogonale oder unitäre Abbildung f eines endlich-dimensionalen Vektorraums mit Skalarprodukt ist dementsprechend also $f^* = f^{-1}$.

Die für diesen Abschnitt angekündigte Verträglichkeitsbedingung eines Endomorphismus mit einem Skalarprodukt besteht nun einfach darin, dass die adjungierte Abbildung mit der ursprünglichen übereinstimmt.

Definition 22.18 (Selbstadjungierte Abbildungen)**.** Ein Endomorphismus $f\colon V \to V$ eines endlichdimensionalen Vektorraums V mit Skalarprodukt heißt **selbstadjungiert**, wenn $f^* = f$, also

$$\langle f(x), y\rangle = \langle x, f(y)\rangle$$

für alle $x, y \in V$ gilt.

Bemerkung 22.19 (Selbstadjungierte Matrizen)**.** Gemäß Satz 22.15 (b) ist ein Endomorphismus $f\colon V \to V$ genau dann selbstadjungiert, wenn die Abbildungsmatrix A_f^B bezüglich einer Orthonormalbasis B von V symmetrisch bzw. hermitesch ist, also $\overline{A_f^B}^\mathsf{T} = A_f^B$ gilt. Man bezeichnet symmetrische bzw. hermitesche Matrizen daher manchmal auch als *selbstadjungiert*.

Beispiel 22.20 (Selbstadjungiertheit orthogonaler Projektionen)**.** Es seien U ein Unterraum eines endlich erzeugten Vektorraums V mit Skalarprodukt und $f\colon V \to V$ die orthogonale Projektion auf U wie in Definition 21.30. Ist $(x_1, \dots, x_k)$ eine Orthonormalbasis von U, die wir zu einer Orthonormalbasis $B = (x_1, \dots, x_n)$ von V ergänzen (siehe Satz 21.32), so ist also

$$f(x_i) = x_i \text{ für } i = 1 \dots, k \quad \text{und} \quad f(x_i) = 0 \text{ für } i = k+1, \dots, n,$$

und damit $A_f^B = \operatorname{diag}(1, \dots, 1, 0, \dots, 0)$ (mit k Einträgen 1 und $n-k$ Einträgen 0) nach Satz 16.23. Da diese Matrix hermitesch ist, ist f nach Bemerkung 22.19 selbstadjungiert.

Beachte jedoch, dass orthogonale Projektionen (auch wenn man es vom Namen her anders vermuten könnte) nach Bemerkung 22.5 keine orthogonalen Abbildungen sind, da sie ja in der Regel nicht invertierbar sind, wie man an der obigen Form der Abbildungsmatrix ablesen kann.

Lemma 22.21 (Determinante und Eigenwerte hermitescher Matrizen)**.** *Für jede hermitesche Matrix $A \in \operatorname{Mat}(n \times n, \mathbb{C})$ gilt:*

(a) $\det A \in \mathbb{R}$.

(b) *Ist λ ein Eigenwert von A, so ist $\lambda \in \mathbb{R}$.*

Beweis.

(a) Dies hatten wir bereits in Folgerung 21.35 (a) gesehen.

(b) Gilt $Ax = \lambda x$ für ein $x \in \mathbb{K}^n \backslash \{0\}$, so auch $\overline{A}\overline{x} = \overline{\lambda}\overline{x}$, und damit $\overline{x}^\mathsf{T}\overline{A}^\mathsf{T} = \overline{\lambda}\overline{x}^\mathsf{T}$. Also folgt

$$\lambda \overline{x}^\mathsf{T} x = \overline{x}^\mathsf{T} A x = \overline{x}^\mathsf{T}\overline{A}^\mathsf{T} x = \overline{\lambda}\overline{x}^\mathsf{T} x,$$

nach Division durch $\overline{x}^\mathsf{T} x \neq 0$ also $\lambda = \overline{\lambda}$ und damit $\lambda \in \mathbb{R}$. □

Um die weiteren Eigenschaften von orthogonalen, unitären und selbstadjungierten Abbildungen zu untersuchen, wenden wir nun einen kleinen Trick an: Es stellt sich heraus, dass alle diese Abbildungen Spezialfälle einer noch etwas allgemeineren Klasse von Abbildungen, den sogenannten *normalen Abbildungen*, sind, und dass man viele gemeinsame Eigenschaften orthogonaler, unitärer und selbstadjungierter Abbildungen auch für diese normalen Abbildungen zeigen kann – mit einem einzigen Beweis, der dann alle bisher betrachteten Fälle abdeckt.

Definition 22.22 (Normale Endomorphismen und Matrizen)**.**

(a) Ein Endomorphismus $f\colon V \to V$ eines endlich erzeugten Vektorraums V mit Skalarprodukt heißt **normal**, wenn $f^* \circ f = f \circ f^*$ gilt.

(b) Eine Matrix $A \in \operatorname{Mat}(n \times n, \mathbb{K})$ heißt **normal**, wenn $\overline{A}^\mathsf{T} A = A \overline{A}^\mathsf{T}$ gilt.

Bemerkung 22.23.

(a) Da die Verkettung von Morphismen der Matrixmultiplikation entspricht, ist ein Endomorphismus nach Satz 22.15 (b) genau dann normal, wenn seine Abbildungsmatrix bezüglich einer beliebigen Orthonormalbasis normal ist.

(b) Der Begriff eines normalen Endomorphismus ist ebenso wie der eines selbstadjungierten Endomorphismus ein theoretisches Konzept, für das es keine einfache geometrische Interpretation gibt. Auch sollte der Name „normal“ nicht darüber hinwegtäuschen, dass es für einen Endomorphismus durchaus nicht normal ist, normal zu sein: Die „meisten“ Endomorphismen sind dies nicht. Die speziellen Morphismen, die wir in diesem Kapitel betrachten wollen, haben diese Eigenschaft jedoch:

Beispiel 22.24.

(a) Eine Diagonalmatrix $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ ist normal: Wegen $\overline{A}^{\mathsf{T}} = \operatorname{diag}(\overline{\lambda_1}, \dots, \overline{\lambda_n})$ ist dann nämlich $\overline{A}^{\mathsf{T}} A = \operatorname{diag}(|\lambda_1|^2, \dots, |\lambda_n|^2) = A\overline{A}^{\mathsf{T}}$.

Beachte jedoch, dass diagonalisierbare Matrizen nicht notwendig normal sind: Für die diagonalisierbare Matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{ist} \quad A^{\mathsf{T}} A = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}, \quad \text{aber} \quad A A^{\mathsf{T}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

(b) Ist $f\colon V \to V$ ein orthogonaler bzw. unitärer Endomorphismus eines endlich-dimensionalen Vektorraums V mit Skalarprodukt, so gilt $f^* = f^{-1}$ nach Beispiel 22.17, und daher auch

$$f^* \circ f = f^{-1} \circ f = \mathrm{id}_V = f \circ f^{-1} = f \circ f^*.$$

Ist f selbstadjungiert, also $f^* = f$, so gilt natürlich trivialerweise ebenfalls $f^* \circ f = f \circ f^*$.

Orthogonale, unitäre und selbstadjungierte Endomorphismen (und damit auch Matrizen) sind also normal.

52

Wir werden daher jetzt solche normalen Endomorphismen und Matrizen untersuchen – und können dann jedes Ergebnis dazu auf einen beliebigen der vorherigen Fälle von orthogonalen, unitären oder selbstadjungierten Endomorphismen anwenden. Wie immer beginnen wir mit den grundlegenden Eigenschaften solcher Morphismen.

Satz 22.25 (Eigenschaften normaler Endomorphismen)**.** *Es seien V ein endlich erzeugter Vektorraum mit Skalarprodukt und $f\colon V \to V$ ein normaler Endomorphismus. Dann gilt:*

(a) *Für alle $x \in V$ ist $\|f^*(x)\| = \|f(x)\|$. Insbesondere ist also $f^*(x) = 0$ genau dann, wenn $f(x) = 0$, d. h. es gilt $\operatorname{Ker} f^* = \operatorname{Ker} f$.*

(b) *Ist x ein Eigenvektor von f zum Eigenwert λ, so ist x auch ein Eigenvektor von f^* zum Eigenwert $\overline{\lambda}$.*

(c) *Eigenvektoren von f zu verschiedenen Eigenwerten stehen senkrecht aufeinander.*

Beweis.

(a) Für $x \in V$ ist

$$\|f^*(x)\|^2 = \langle f^*(x), f^*(x)\rangle = \langle x, f(f^*(x))\rangle \overset{(*)}{=} \langle x, f^*(f(x))\rangle = \langle f(x), f(x)\rangle = \|f(x)\|^2,$$

wobei wir in $(*)$ die Normalität von f ausgenutzt haben.

(b) Es sei $f(x) = \lambda x$. Dann ist

$$\begin{aligned}
\|f^*(x) - \overline{\lambda} x\|^2 &= \left\langle f^*(x) - \overline{\lambda} x, f^*(x) - \overline{\lambda} x \right\rangle \\
&= \langle f^*(x), f^*(x)\rangle - \lambda \langle x, f^*(x)\rangle - \overline{\lambda} \langle f^*(x), x\rangle + \lambda\overline{\lambda} \langle x, x\rangle \\
&\overset{(a)}{=} \langle f(x), f(x)\rangle - \lambda \langle f(x), x\rangle - \overline{\lambda} \langle x, f(x)\rangle + \lambda\overline{\lambda} \langle x, x\rangle \\
&= \langle f(x) - \lambda x, f(x) - \lambda x\rangle \\
&= 0
\end{aligned}$$

und damit $f^*(x) - \overline{\lambda} x = 0$.

(c) Sind $x,y \in V$ und $f(x) = \lambda x$ sowie $f(y) = \mu y$ mit $\lambda \neq \mu$, so folgt

$$\lambda \langle x,y\rangle = \left\langle \overline{\lambda} x, y \right\rangle \overset{\text{(b)}}{=} \langle f^*(x), y\rangle = \langle x, f(y)\rangle = \langle x, \mu y\rangle = \mu \langle x,y\rangle,$$

also $(\lambda - \mu)\langle x,y\rangle = 0$. Wegen $\lambda \neq \mu$ ergibt sich daraus sofort $\langle x,y\rangle = 0$. □

22.C Der Spektralsatz

Wir kommen nun zum Hauptresultat dieses Kapitels. Es besagt, dass normale Endomorphismen mit zerfallendem charakteristischen Polynom stets diagonalisierbar sind, und zwar sogar mit einer Orthonormalbasis. Dieser Satz wird der Spektralsatz genannt, da die Menge aller Eigenwerte eines Endomorphismus oft als das Spektrum dieses Endomorphismus bezeichnet wird und dieser Satz durch die Diagonalisierbarkeit eine Aussage über diese Eigenwerte und ihre Vielfachheiten macht.

Satz 22.26 (Spektralsatz). *Für einen Endomorphismus f eines endlich-dimensionalen Vektorraums V mit Skalarprodukt sind äquivalent:*

(a) *f ist normal und χ_f zerfällt in Linearfaktoren;*

(b) *V besitzt eine Orthonormalbasis aus Eigenvektoren von f.*

Insbesondere ist f dann nach Folgerung 19.30 also diagonalisierbar; man sagt für (b) auch, dass f ***orthogonal*** *bzw.* ***unitär diagonalisierbar*** *ist.*

Beweis.

(a) $\Rightarrow$ (b): Wir zeigen diese Richtung mit Induktion über $n := \dim V$; der Fall $n = 0$ ist trivial. Für den Induktionsschritt $n-1 \to n$ sei nun also $n > 0$.

Da χ_f nach Voraussetzung in Linearfaktoren zerfällt, gibt es einen Eigenwert λ von f mit zugehörigem Eigenvektor x_1. Durch Normieren können wir diesen so wählen, dass $\|x_1\| = 1$. Es sei $U = \operatorname{Lin}(x_1)$.

Wir ergänzen nun x_1 zu einer Orthonormalbasis $B = (x_1, x_2', \dots, x_n')$ von V, so dass nach Bemerkung 21.41 also $B' = (x_2', \dots, x_n')$ eine Orthonormalbasis von $U^\perp$ ist. Unser Ziel ist es, die Induktionsvoraussetzung auf die Einschränkung von f auf $U^\perp$ anzuwenden. Dazu müssen wir die folgenden Dinge überprüfen:

- $f(U^\perp) \subset U^\perp$, d. h. f lässt sich zu einem Endomorphismus $f|_{U^\perp} \colon U^\perp \to U^\perp$ einschränken: Für alle $x \in U^\perp$ gilt

$$\begin{aligned} \langle f(x), x_1\rangle &= \langle x, f^*(x_1)\rangle && \text{(Satz 22.15 (a))} \\ &= \left\langle x, \overline{\lambda} x_1 \right\rangle && \text{(Satz 22.25 (b), da } f \text{ normal)} \\ &= \overline{\lambda}\, \langle x, x_1\rangle \\ &= 0 && \text{(wegen } x \in U^\perp\text{)}, \end{aligned}$$

  und damit auch $f(x) \in U^\perp$.

- Die Einschränkung $f' := f|_{U^\perp}$ erfüllt die Voraussetzung (a) des Satzes: Für die Abbildungsmatrix A_f^B gilt

$$A_f^B = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & A_{f'}^{B'} \end{array}\right)$$

  (wobei gemäß Satz 16.23 die erste Spalte aus $f(x_1) = \lambda x_1$ und die Nullen neben λ aus $f(U^\perp) \subset U^\perp$ folgen). Da f normal ist, ist nun nach Bemerkung 22.23 (a)

$$\overline{A_f^B}^{\mathsf{T}} A_f^B = A_f^B \overline{A_f^B}^{\mathsf{T}} \quad \Rightarrow \quad \left(\begin{array}{c|c} |\lambda|^2 & 0 \\ \hline 0 & \overline{A_{f'}^{B'}}^{\mathsf{T}} A_{f'}^{B'} \end{array}\right) = \left(\begin{array}{c|c} |\lambda|^2 & 0 \\ \hline 0 & A_{f'}^{B'} \overline{A_{f'}^{B'}}^{\mathsf{T}} \end{array}\right),$$

also $A_{f'}^{B'}$ und damit f' normal; und wegen $\chi_f(t) = (t-\lambda)\,\chi_{f'}(t)$ zerfällt mit χ_f auch $\chi_{f'}$ in Linearfaktoren.

Nach der Induktionsvoraussetzung gibt es nun also eine Orthonormalbasis $(x_2,\dots,x_n)$ von $U^\perp$ aus Eigenvektoren von f (bzw. f'), so dass wie gewünscht $(x_1,\dots,x_n)$ eine Orthonormalbasis von V aus Eigenvektoren von f ist.

(b) $\Rightarrow$ (a): Es sei B eine Orthonormalbasis von V aus Eigenvektoren von f. Dann ist die Abbildungsmatrix $A = A_f^B$ nach Folgerung 19.30 eine Diagonalmatrix. Damit ist diese Matrix normal nach Beispiel 22.24 (a) und hat ein zerfallendes charakteristisches Polynom nach Folgerung 19.36 (a). □

Da der Spektralsatz in der Praxis sehr häufig gebraucht wird, schreiben wir ihn hier auch noch einmal in Matrixform auf:

Folgerung 22.27 (Spektralsatz in Matrixform)**.** *Für eine quadratische Matrix $A \in \mathrm{Mat}(n\times n,\mathbb{K})$ sind äquivalent:*

(a) *A ist normal und χ_A zerfällt in Linearfaktoren;*

(b) *es gibt eine orthogonale bzw. unitäre Matrix T, so dass*

$$T^{-1}AT = \overline{T}^{\mathsf{T}}AT = \mathrm{diag}(\lambda_1,\dots,\lambda_n)$$

eine Diagonalmatrix (mit den Eigenwerten $\lambda_1,\dots,\lambda_n$ von A auf der Diagonalen) ist.

*Man sagt für (b) wieder, dass A **orthogonal** bzw. **unitär diagonalisierbar** ist.*

Beweis. Es sei $f = f_A\colon \mathbb{K}^n \to \mathbb{K}^n$, $x \mapsto Ax$ die zu A gehörige lineare Abbildung, so dass umgekehrt also A die Abbildungsmatrix von f bezüglich der Standardbasis von $\mathbb{K}^n$ ist. Wir versehen $\mathbb{K}^n$ mit dem Standardskalarprodukt.

Da die Standardbasis eine Orthonormalbasis für das Standardskalarprodukt ist, ist A nach Bemerkung 22.23 (a) genau dann normal mit in Linearfaktoren zerfallendem charakteristischen Polynom, wenn dies für f gilt. Nach dem Spektralsatz 22.26 ist dies wiederum genau dann der Fall, wenn $\mathbb{K}^n$ eine Orthonormalbasis $(x_1,\dots,x_n)$ aus Eigenvektoren von f bzw. A besitzt. Mit $T = (x_1 \mid \cdots \mid x_n)$ und $\lambda_1,\dots,\lambda_n$ den Eigenwerten zu $x_1,\dots,x_n$ ist diese Aussage nach Folgerung 19.30 und Lemma 22.4 (b) aber genau äquivalent dazu, dass T orthogonal bzw. unitär ist und $T^{-1}AT$ die Diagonalmatrix mit Einträgen $\lambda_1,\dots,\lambda_n$ ist. □

Beispiel 22.28. Die für uns wichtigsten Spezialfälle von Folgerung 22.27 sind:

(a) $\mathbb{K} = \mathbb{C}$ und A ist hermitesch oder unitär: Dann ist A nach Bemerkung 22.24 (b) normal, und wegen des Fundamentalsatzes der Algebra aus Satz 6.11 zerfällt χ_A in Linearfaktoren. Also ist A dann nach Folgerung 22.27 unitär diagonalisierbar.

(b) $\mathbb{K} = \mathbb{R}$ und A ist symmetrisch (also selbstadjungiert): Wie in (a) ist A auch dann wieder normal. Außerdem können wir A auch als komplexe hermitesche Matrix auffassen; das charakteristische Polynom zerfällt dann zumindest über $\mathbb{C}$ wieder in Linearfaktoren, d. h. es ist $\chi_A(t) = (t-\lambda_1)\cdots(t-\lambda_n)$ mit den Eigenwerten $\lambda_1,\dots,\lambda_n \in \mathbb{C}$ von A. Diese Eigenwerte sind nach Lemma 22.21 (b) aber reell, und damit zerfällt χ_A sogar über $\mathbb{R}$ in Linearfaktoren. Wir erhalten so den sicher wichtigsten Fall des Spektralsatzes:

$$\boxed{\text{Symmetrische reelle Matrizen sind stets orthogonal diagonalisierbar.}}$$

(c) $\mathbb{K} = \mathbb{R}$ und A ist orthogonal: Dann muss χ_A nicht notwendig in Linearfaktoren zerfallen – wie z. B. bei einer Drehung in $\mathbb{R}^2$ wie in Beispiel 19.13 (c). Nach Folgerung 19.40 ist A dann also auch nicht notwendig diagonalisierbar.

Beispiel 22.29. Die Berechnung der Matrix T in Folgerung 22.27 bzw. der Orthonormalbasis in Satz 22.26 ist mit unserem Vorwissen nicht mehr weiter kompliziert: Wir berechnen mit dem Verfahren von Gram-Schmidt aus Satz 21.32 Orthonormalbasen aller Eigenräume – besonders einfach ist dies natürlich für alle eindimensionalen Eigenräume, weil wir für diese nur jeweils einen normierten Eigenvektor benötigen. Nach Satz 22.25 (c) erhalten wir auf diese Art eine orthonormale Familie, und wegen der aus dem Spektralsatz folgenden Diagonalisierbarkeit sogar eine Orthonormalbasis. Als konkretes Beispiel sei

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R}).$$

Die Matrix A ist offensichtlich reell symmetrisch und damit nach Beispiel 22.28 (b) orthogonal diagonalisierbar. Wie üblich (siehe Beispiel 19.16) berechnen wir dazu zuerst die Eigenwerte und -räume von A. Die Rechnung zeigt, dass A zwei Eigenwerte besitzt, und zwar

- $\lambda_1 = 3$ mit Eigenraum $\mathrm{Eig}(A,3) = \mathrm{Lin}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, also normiertem Eigenvektor $\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, und
- $\lambda_1 = 1$ mit Eigenraum $\mathrm{Eig}(A,1) = \mathrm{Lin}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, also normiertem Eigenvektor $\frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

(wie man durch Rückwärtseinsetzen auch leicht bestätigen kann). Beachte, dass wir dabei für die Bestimmung der Orthonormalbasen das Standardskalarprodukt nehmen müssen und *nicht* die durch A bestimmte Bilinearform $b(x,y) = x^\mathsf{T} A y$. Dies wird aus dem Beweis von Folgerung 22.27 deutlich; man sieht es aber auch schon daran, dass wir nach Bemerkung 22.2 (b) für eine orthogonale Matrix ja Spaltenvektoren brauchen, die *bezüglich des Standardskalarprodukts* orthonormal sind.

Schreiben wir diese Vektoren nun wie üblich als Spalten in eine Matrix, so erhalten wir also die orthogonale Matrix

$$T = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \in \mathrm{O}(2)$$

(wie man leicht sieht, bilden die Spalten in der Tat eine Orthonormalbasis von $\mathbb{R}^2$ bezüglich des Standardskalarprodukts), und es gilt dann

$$T^{-1} A T = T^\mathsf{T} A T = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix},$$

wie man auch leicht durch direkte Matrixmultiplikation (bzw. Inversenbildung) überprüfen kann.

Aufgabe 22.30 (Wurzelziehen für Matrizen)**.** Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine symmetrische Matrix. Man zeige:

(a) Ist A positiv semidefinit, so gibt es eine eindeutig bestimmte symmetrische positiv semidefinite Matrix $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ mit $B^2 = A$. Man nennt sie die *Wurzel* aus A.

(b) Ist A nicht positiv semidefinit, so kann es zwar eine Matrix $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ mit $B^2 = A$ geben, aber keine symmetrische.

Aufgabe 22.31. Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$. Man zeige:

(a) Ist A symmetrisch und nilpotent (siehe Aufgabe 20.25), so ist $A = 0$.

(b) Ist A diagonalisierbar und $A^\mathsf{T} = -A$, so ist $A = 0$.

Aufgabe 22.32. Es sei $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine symmetrische Matrix. Zeige, dass es eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ gibt mit $A^3 + A = B$.

In seiner ursprünglichen Form macht der Spektralsatz 22.26 eine Aussage über Endomorphismen. Wir wollen nun sehen, dass man ihn jedoch ebenso gewinnbringend auf Bilinearformen bzw. Sesquilinearformen anwenden kann – denn obwohl sich deren Gramsche Matrizen nach Bemerkung 21.8 zunächst einmal anders transformieren als die Abbildungsmatrizen von Endomorphismen (nämlich mit $A \mapsto \overline{T}^\mathsf{T} A T$ statt mit $A \mapsto T^{-1} A T$), stimmen diese beiden Transformationsregeln für eine orthogonale bzw. unitäre Matrix T ja überein.

Als Erstes wollen wir eine in der Praxis besonders wichtige Charakterisierung positiv definiter Matrizen (bzw. Bilinear- oder Sesquilinearformen) mit Hilfe von Eigenwerten angeben. Wir stellen sie im Folgenden noch einmal mit dem uns bereits aus Satz 21.43 bekannten Hurwitz-Kriterium zusammen und erweitern diese Aussagen auch auf negativ definite und indefinite Matrizen, da wir dies für die spätere Anwendung auf Extremwertuntersuchungen benötigen werden (siehe Satz 26.20).

Satz 22.33 (Kriterien für die Definitheit von Matrizen)**.** *Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ eine symmetrische (für $\mathbb{K} = \mathbb{R}$) bzw. hermitesche (für $\mathbb{K} = \mathbb{C}$) Matrix. Ferner sei $A_k \in \mathrm{Mat}(k \times k, \mathbb{K})$ für $k = 1, \dots, n$ die Matrix, die aus den ersten k Zeilen und Spalten von A besteht. Dann gilt:*

(a) ***(Eigenwertkriterium)***

A ist genau dann positiv definit / negativ definit / positiv semidefinit / negativ semidefinit, wenn $\lambda > 0$ / $\lambda < 0$ / $\lambda \geq 0$ / $\lambda \leq 0$ für jeden (nach Lemma 22.21 (b) reellen) Eigenwert λ von A gilt.

A ist genau dann indefinit, wenn A mindestens einen positiven und einen negativen Eigenwert besitzt.

(b) ***(Hurwitz-Kriterium)***

A ist genau dann positiv definit, wenn $\det A_k > 0$ für alle k.

A ist genau dann negativ definit, wenn $\det A_k > 0$ für alle geraden und $\det A_k < 0$ für alle ungeraden k.

Ist $\det A \neq 0$, so ist A genau dann positiv bzw. negativ semidefinit, wenn A positiv bzw. negativ definit ist.

Ist $\det A \neq 0$, so ist A genau dann indefinit, wenn A weder positiv noch negativ definit ist, also wenn die Vorzeichenfolge von $\det A_k$ für $k = 1, \dots, n$ weder $(+, +, +, \dots)$ noch $(-, +, -, +, \dots)$ ist.

Beweis.

(a) Nach der Folgerung 22.27 aus dem Spektralsatz gibt es eine orthogonale bzw. unitäre Matrix T mit $\overline{T}^{\mathsf{T}} A T = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, wobei $\lambda_1, \dots, \lambda_n$ die Eigenwerte von A sind. Mit der invertierbaren Transformation $y := \overline{T}^{\mathsf{T}} x = T^{-1} x$, d. h. $x = Ty$, ist A also genau dann positiv definit, d. h. $\overline{x}^{\mathsf{T}} A x > 0$ für alle $x \in \mathbb{K}^n \backslash \{0\}$, wenn

$$\overline{y}^{\mathsf{T}} \overline{T}^{\mathsf{T}} A T y > 0 \quad \Leftrightarrow \quad \overline{y}^{\mathsf{T}} \mathrm{diag}(\lambda_1, \dots, \lambda_n) y > 0 \quad \Leftrightarrow \quad \lambda_1 |y_1|^2 + \dots + \lambda_n |y_n|^2 > 0$$

für alle $y \neq 0$ gilt. Dies ist offensichtlich der Fall, wenn alle Eigenwerte $\lambda_1, \dots, \lambda_n$ positiv sind; umgekehrt folgt aus dieser Ungleichung durch Einsetzen des Einheitsvektors e_i aber auch sofort $\lambda_i > 0$ für alle i.

Die Aussagen zur negativen Definitheit und Semidefinitheit zeigt man analog. Die Matrix A ist indefinit, wenn sie weder negativ noch positiv semidefinit ist, also wenn sie mindestens einen positiven und einen negativen Eigenwert hat.

(b) Der Fall der positiven Definitheit ist genau Satz 21.43. Weiterhin ist A genau dann negativ definit, wenn $\overline{x}^{\mathsf{T}} A x < 0$ und damit $\overline{x}^{\mathsf{T}} (-A) x > 0$ für alle $x \in \mathbb{K}^n \backslash \{0\}$ gilt, also genau dann, wenn $-A$ positiv definit ist. Anwenden von Satz 21.43 auf $-A$ ergibt in diesem Fall also wegen $\det(-A_k) = (-1)^k \det A_k$ die Behauptung.

Ist nun $\det A \neq 0$, so ist A invertierbar, d. h. es ist $\mathrm{Eig}(A, 0) = \mathrm{Ker} A = \{0\}$ und damit 0 kein Eigenwert von A. Nach (a) ist A damit genau dann positiv bzw. negativ semidefinit, wenn A positiv bzw. negativ definit ist, und indefinit, wenn dies beides nicht der Fall ist. □

Bemerkung 22.34. Zur Bestimmung der Definitheitseigenschaften einer symmetrischen bzw. hermiteschen Matrix A mit Satz 22.33 ist das Hurwitz-Kriterium oftmals geeigneter, da die Berechnung von Determinanten einfacher ist als die aller Eigenwerte. Es liefert aber nicht in jedem Fall eine

Entscheidung: Ist $\det A = 0$, so lässt sich mit dem Hurwitz-Kriterium in der Regel keine allgemeine Aussage treffen. Dies zeigt das Beispiel der drei Matrizen

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

die nach Satz 22.33 (a) positiv semidefinit, negativ semidefinit bzw. indefinit sind, für die aber alle Untermatrizen der ersten $k = 1,2,3$ Zeilen und Spalten die Determinante 0 haben.

Aufgabe 22.35. Es sei $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine symmetrische reelle Matrix mit

$$a_{i,i} > \sum_{j \neq i} |a_{i,j}|$$

für alle $i = 1, \dots, n$. Beweise, dass A dann positiv definit ist – symmetrische Matrizen, deren Diagonaleinträge im Vergleich zu den anderen „groß genug“ sind, sind also positiv definit.

53

Als weitere Anwendung des Spektralsatzes auf symmetrische Bilinearformen bzw. Sesquilinearformen können wir den folgenden „Normalformensatz“ zeigen, der die Aussage von Bemerkung 21.34 auf den nicht notwendig positiv definiten Fall verallgemeinert:

Satz 22.36 (Trägheitssatz von Sylvester). *Es seien V ein endlich erzeugter $\mathbb{K}$-Vektorraum und b eine symmetrische Bilinearform (für $\mathbb{K} = \mathbb{R}$) bzw. hermitesche Sesquilinearform (für $\mathbb{K} = \mathbb{C}$). Dann gibt es eine Basis B von V, bezüglich der die Gramsche Matrix von b die einfache Form*

$$A_b^B = \mathrm{diag}(\underbrace{1, \dots, 1}_{k}, \underbrace{-1, \dots, -1}_{l}, 0, \dots, 0) = \left(\begin{array}{c|c|c} E_k & & 0 \\ \hline & -E_l & \\ \hline 0 & & 0 \end{array}\right)$$

hat. Dabei ist die Anzahl k bzw. l der Einträge 1 bzw. -1 auf der Diagonalen durch b bereits eindeutig bestimmt, und zwar ist k bzw. l gleich

(a) *der maximalen Dimension eines Unterraums von V, auf dem b positiv bzw. negativ definit ist; und*

(b) *der (mit Vielfachheiten gezählten) Anzahl der positiven bzw. negativen Eigenwerte einer beliebigen Gramschen Matrix zu b.*

Beweis. Wir teilen den Beweis in zwei Teile:

Teil 1: Als Erstes zeigen wir die Eindeutigkeit von k und l und den Ausdruck aus (a). Dazu betrachten wir eine beliebige Basis $B = (x_1, \dots, x_n)$ von V, für die $A_b^B = (b(x_i, x_j))_{i,j}$ die im Satz angegebene Form hat. Dann ist $U_+ := \mathrm{Lin}(x_1, \dots, x_k)$ sicher ein k-dimensionaler Unterraum von V, auf dem b positiv definit ist, denn die Gramsche Matrix der Einschränkung von b auf U_+ ist ja gerade die positiv definite Matrix E_k. Genauso sieht man natürlich, dass $U_- := \mathrm{Lin}(x_{k+1}, \dots, x_n)$ ein $(n-k)$-dimensionaler Unterraum ist, auf dem b negativ semidefinit ist.

Ist nun andererseits $U \leq V$ ein beliebiger Unterraum, auf dem b positiv definit ist, so ist mit dem eben gefundenen U_- sicher $U \cap U_- = \{0\}$, denn für ein $x \in U \cap U_-$ mit $x \neq 0$ ergäbe sich aus $b(x,x) > 0$ wegen $x \in U$ und $b(x,x) \leq 0$ wegen $x \in U_-$ sofort ein Widerspruch. Mit der Dimensionsformel aus Folgerung 15.31 erhalten wir also

$$n \geq \dim(U + U_-) = \dim U + \dim U_- = \dim U + n - k$$

und damit $\dim U \leq k$. Also ist k wirklich die maximale Dimension eines Unterraums von V, auf dem b positiv definit ist. Analog zeigt man natürlich die entsprechende Aussage für l. Wir haben damit also den Ausdruck für k und l in (a) bewiesen, und somit insbesondere auch die Eindeutigkeit der im Satz angegebenen Matrixdarstellung.

Teil 2: Wir zeigen nun (konstruktiv) die Existenz einer Basis B wie in der Behauptung und dabei den Ausdruck aus (b). Dazu sei zunächst B' eine beliebige Basis von V. Nach Lemma 21.11 bzw.

Konstruktion 21.19 ist mit b auch die Gramsche Matrix $A := A_b^{B'}$ symmetrisch bzw. hermitesch. Wir gehen nun in zwei Schritten vor:

- Schritt 1: Drehung auf Diagonalform.

  Aus dem Spektralsatz folgt wie in Beispiel 22.28 die Existenz einer orthogonalen bzw. unitären Matrix T, so dass
  $$T^{-1}AT = \overline{T}^{\mathsf{T}}AT = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$
  eine Diagonalmatrix mit den (nach Lemma 22.21 (b) reellen) Eigenwerten $\lambda_1, \dots, \lambda_n$ von A ist. Dabei wählen wir die Reihenfolge der Spalten von T und damit der Diagonaleinträge der Matrix so, dass die Eigenwerte $\lambda_1, \dots, \lambda_k$ positiv, $\lambda_{k+1}, \dots, \lambda_{k+l}$ negativ, und $\lambda_{k+l+1}, \dots, \lambda_n$ gleich 0 sind.

- Schritt 2: Koordinatenstreckung auf Normalform.

  Wir setzen nun
  $$S = \overline{S}^{\mathsf{T}} = \operatorname{diag}\left(\frac{1}{\sqrt{|\lambda_1|}}, \dots, \frac{1}{\sqrt{|\lambda_{k+l}|}}, 1, \dots, 1\right) \in \mathrm{GL}(n, \mathbb{K}).$$
  Dann ist
  $$\overline{TS}^{\mathsf{T}}ATS = \overline{S}^{\mathsf{T}}\overline{T}^{\mathsf{T}}ATS = \overline{S}^{\mathsf{T}} \cdot \operatorname{diag}(\lambda_1, \dots, \lambda_n) \cdot S = \operatorname{diag}(1, \dots, 1, -1, \dots, -1, 0, \dots, 0).$$

Nach Satz 16.32 (b) gibt es nun aber eine Basis B, so dass die Basiswechselmatrix $A^{B,B'}$ gleich TS ist. Mit der Transformationsregel für Bilinearformen aus Lemma 21.7 (bzw. für Sesquilinearformen aus Konstruktion 21.19) hat die Matrix A_b^B dann die gewünschte Form, wobei k und l wie in (b) sind. □

Bemerkung 22.37.

(a) Im Fall einer positiv definiten symmetrischen Bilinearform bzw. hermiteschen Sesquilinearform ist natürlich $k = n$ und $l = 0$ in Satz 22.36, so dass A_b^B die Einheitsmatrix ist. Diese Aussage ist genau die uns schon bekannte Existenz von Orthonormalbasen zu Skalarprodukten wie in Bemerkung 21.34.

(b) Natürlich gibt es auch vom Trägheitssatz 22.36 eine Matrixform: Für jede reelle symmetrische bzw. komplexe hermitesche Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ gibt es eine invertierbare Matrix T, so dass $\overline{T}^{\mathsf{T}}AT$ die in Satz 22.36 angegebene Gestalt hat. Dabei sind auch hier dann die Anzahlen k und l der Diagonaleinträge 1 bzw. -1 eindeutig bestimmt und gleich der maximalen Dimension eines Unterraums von $\mathbb{K}^n$, auf dem A positiv bzw. negativ definit ist, sowie gleich der Anzahl der positiven bzw. negativen Eigenwerte von $\overline{S}^{\mathsf{T}}AS$ für eine beliebige invertierbare Matrix $S \in \mathrm{GL}(n, \mathbb{K})$. Mit den uns bekannten Entsprechungen zwischen Bilinearformen bzw. Sesquilinearformen und Matrizen ergeben sich diese Aussagen unmittelbar aus Satz 22.36 angewendet auf $b(x, y) = \overline{x}^{\mathsf{T}}Ay$.

  Insbesondere bedeutet dies, dass die Matrizen A und $\overline{S}^{\mathsf{T}}AS$ für alle $S \in \mathrm{GL}(n, \mathbb{K})$ die gleiche Anzahl positiver (und analog negativer) Eigenwerte haben – obwohl die Eigenwerte selbst ja nicht übereinstimmen, da A und $\overline{S}^{\mathsf{T}}AS$ im Allgemeinen nicht ähnlich zueinander sind. Diese Aussage, die aus unserem Satz 22.36 folgt, wird in der Literatur auch oft Trägheitssatz von Sylvester genannt. Aus ihr leitet sich auch der Name „Trägheitssatz“ ab: Der Satz zeigt, dass sich die Anzahlen der positiven und negativen Eigenwerte einer hermiteschen Matrix unter Transformationen der Form $A \mapsto \overline{S}^{\mathsf{T}}AS$ nicht ändern, sich also in diesem Sinne „träge“ verhalten.

Die in Teil 2 vom Beweis des Trägheitssatzes 22.36 konstruierte Transformation, um eine gegebene reelle symmetrische Bilinearform in die dort angegebene Normalform zu bringen, hat auch eine einfache geometrische Bedeutung. Der Einfachheit halber betrachten wir diese zunächst im positiv definiten Fall, also wenn alle Diagonaleinträge der transformierten Matrix gleich 1 sind.

Beispiel 22.38 (Hauptachsentransformation). Es sei b ein Skalarprodukt auf $\mathbb{R}^n$ mit Gramscher Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$. Um dieses Skalarprodukt zu visualisieren, können wir die nach Definition 21.13 zugehörige Norm $\|x\| = \sqrt{b(x,x)} = \sqrt{x^\mathsf{T} A x}$ betrachten und die Menge

$$K := \{x \in \mathbb{R}^n : \|x\| = 1\} = \{x \in \mathbb{R}^n : x^\mathsf{T} A x = 1\}$$

aller Punkte mit Norm 1 zeichnen, also den „Rand der Einheitskugel" bezüglich b. Um K geometrisch zu beschreiben, führen wir die zwei Schritte in Teil 2 des Beweises von Satz 22.36 durch:

- Drehung auf Diagonalform: Wir finden nach dem Spektralsatz eine orthogonale Matrix T, so dass $T^\mathsf{T} A T = \mathrm{diag}(\lambda_1, \dots, \lambda_n) =: D$ eine Diagonalmatrix ist. Als Eigenwerte von A sind diese $\lambda_1, \dots, \lambda_n$ nach Satz 22.33 (a) positiv.

  Wir machen nun die Koordinatentransformation $y := T^{-1}x = T^\mathsf{T} x$, also $x = Ty$, die wir uns wegen $T \in \mathrm{O}(n)$ als Drehung (bzw. Spiegelung) in $\mathbb{R}^n$ vorstellen können. Mit diesen neuen Koordinaten ist

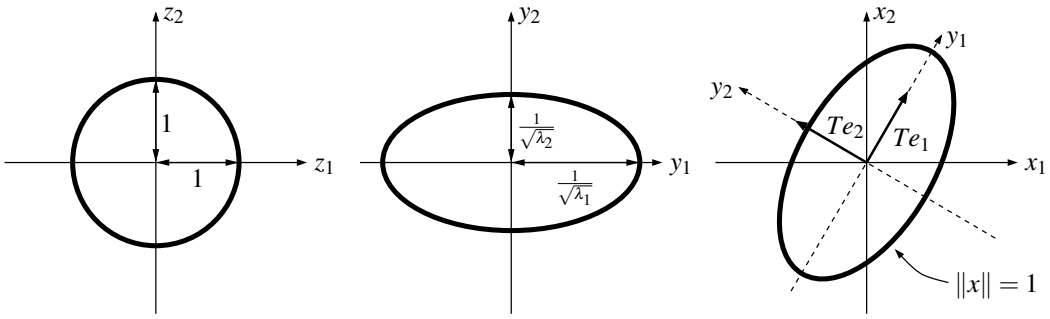
  $$x^\mathsf{T} A x = 1 \quad \Leftrightarrow \quad y^\mathsf{T} T^\mathsf{T} A T y = 1 \quad \Leftrightarrow \quad y^\mathsf{T} D y = 1 \quad \Leftrightarrow \quad \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 = 1.$$

- Koordinatenstreckung auf Normalform: Wegen $\lambda_i > 0$ für alle i können wir nun noch die weitere Koordinatentransformation $z_i := \sqrt{\lambda_i}\, y_i$, also $y_i = \frac{1}{\sqrt{\lambda_i}} z_i$ für $i = 1, \dots, n$ durchführen, die einer Streckung der Koordinatenachsen (mit unterschiedlichen Streckfaktoren) entspricht. In diesen neuen Koordinaten ist nun einfach

  $$x^\mathsf{T} A x = 1 \quad \Leftrightarrow \quad \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 = 1 \quad \Leftrightarrow \quad z_1^2 + \dots + z_n^2 = 1,$$

  d. h. hier bekommen wir nun den Rand der „gewöhnlichen Einheitskugel".

Unsere ursprüngliche Menge K ensteht also aus dem Rand der normalen Einheitskugel (in den Koordinaten z_i), indem wir zuerst die einzelnen Koordinaten strecken (Übergang von den z_i zu den y_i) und das resultierende Ellipsoid dann im Raum drehen (Übergang von den y_i zu den x_i). Die gesuchte Menge K ist also wie im folgenden Bild für $n = 2$ dargestellt ein im Raum gedrehtes Ellipsoid.



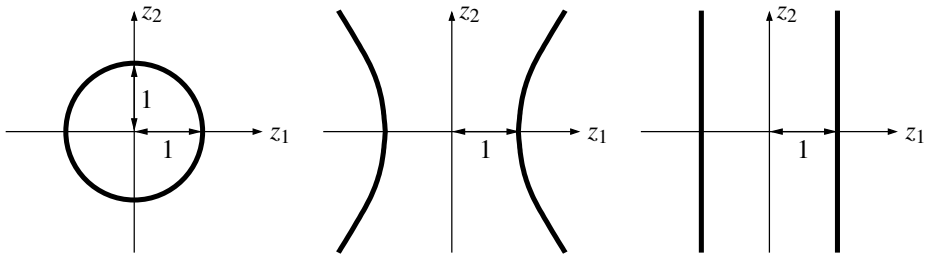
Wie im Bild angedeutet sind die charakteristischen Merkmale dieses Ellipsoids:

- Seine Radien sind gerade $\frac{1}{\sqrt{\lambda_i}}$, denn dies sind die y_i-Werte, die $z_i = 1$ entsprechen;
- Seine Symmetrieachsen werden aufgespannt von den x-Vektoren, die im y-Koordinatensystem den Einheitsvektoren entsprechen – wegen $x = Ty$ also von den Spalten $Te_1, \dots, Te_n$ von T und damit genau von den Eigenvektoren von A.

Die Symmetrieachsen eines Ellipsoids werden auch als seine **Hauptachsen** bezeichnet – daher der Name Hauptachsentransformation.

Bemerkung 22.39. Untersuchen wir die Menge $K = \{x \in \mathbb{R}^n : x^\mathsf{T} A x = 1\}$ wie in Beispiel 22.38 für eine symmetrische Matrix A, die nicht mehr notwendig positiv definit ist, so können wir immer noch das Verfahren aus Teil 2 des Beweises von Satz 22.36 anwenden, erhalten jedoch am Ende in den Koordinaten z_i eine quadratische Gleichung, deren Koeffizienten 1, -1 und 0 sein können (je nachdem, wie viele Eigenwerte von A positiv, negativ bzw. 0 sind). Bis auf Permutation dieser Variablen erhalten wir so z. B. für $n = 2$ als Möglichkeiten für ein nicht-leeres K wie im folgenden Bild

neben einem Kreis (bei zwei positiven Eigenwerten) eine *Hyperbel* (bei einem positiven und einem negativen Eigenwert) und ein *Geradenpaar* (bei einem positiven Eigenwert und einem Eigenwert 0):



$z_1^2 + z_2^2 = 1$ Kreis | $z_1^2 - z_2^2 = 1$ Hyperbel | $z_1^2 = 1$ Geradenpaar

Wie in Beispiel 22.38 ist die ursprüngliche Menge K dann eine in den Koordinatenrichtungen gestreckte und anschließend gedrehte Variante dieser Bilder. Hat A keinen positiven Eigenwert, so ist $K = \emptyset$, da eine Summe von Quadraten mit nicht-positiven Vorfaktoren niemals 1 ergeben kann. Für $n > 2$ gibt es natürlich entsprechend mehr qualitativ verschiedene Möglichkeiten für K.

Aufgabe 22.40. Die symmetrische reelle Matrix $A = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 7 & 4 \\ -2 & 4 & 7 \end{pmatrix}$ hat genau die beiden Eigenwerte 3 und 12. Da alle Eigenwerte positiv sind, ist A nach Satz 22.33 (b) also positiv definit und bestimmt somit ein Skalarprodukt auf $\mathbb{R}^3$.

(a) Berechne eine orthogonale Matrix T, so dass $T^{-1}AT$ eine Diagonalmatrix ist.

(b) Bestimme mit einer Hauptachsentransformation die Punkte der Menge $\{x \in \mathbb{R}^3 : x^\mathsf{T} A x = 3\}$, die vom Ursprung (bezüglich des Standardskalarprodukts) den kleinsten Abstand haben.

Aufgabe 22.41. Man zeige:

(a) Ist $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ normal, so ist A genau dann hermitesch, wenn alle Eigenwerte von A reell sind.

(b) Sind $A, B \in \mathrm{Mat}(n \times n, \mathbb{R})$ symmetrisch mit $x^\mathsf{T} A x \leq x^\mathsf{T} B x$ für alle $x \in \mathbb{R}^n$, so hat B mindestens so viele positive Eigenwerte wie A (mit ihrer jeweiligen algebraischen Vielfachheit gezählt).

Aufgabe 22.42. Es sei b eine Bilinearform auf einem endlich-dimensionalen Vektorraum V über einem Körper K, in dem $1 + 1 \neq 0$ gilt. Wir nennen b *antisymmetrisch*, wenn $b(x,y) = -b(y,x)$ für alle $x, y \in V$ gilt. Eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ heißt *antisymmetrisch*, wenn $A^\mathsf{T} = -A$.

(a) Zeige, dass b genau dann antisymmetrisch ist, wenn A_b^B für eine beliebige Basis B von V antisymmetrisch ist.

(b) Man zeige mit Induktion über $n = \dim V$: Ist b antisymmetrisch, so ist A_b^B für eine geeignete Basis B von V von der Form

$$A_b^B = \begin{pmatrix} \boxed{I} & & & 0 \\ & \ddots & & \\ & & \boxed{I} & \\ 0 & & & 0 \end{pmatrix}$$

mit $I = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, ist also eine Blockdiagonalmatrix mit einer gewissen Anzahl k von Blöcken I (mit $0 \leq 2k \leq n$) und $n - 2k$ anschließenden Nullzeilen und -spalten.

(c) Zeige, dass die Determinante jeder ganzzahligen antisymmetrischen Matrix eine Quadratzahl ist.

22.D Die Singulärwertzerlegung

Als Abschluss der linearen Algebra wollen wir nun noch sehen, dass die Eigenwerttheorie und der Spektralsatz überraschenderweise sogar auf beliebige lineare Abbildungen mit unterschiedlichem Start- und Zielraum angewendet werden können – und damit eine Brücke zurück zu Kapitel 16 schlagen, in dem wir derartige Abbildungen untersucht haben.

Dazu erinnern wir uns an unseren ersten Normalformensatz 16.42: Zu einem Morphismus $f\colon V \to W$ zwischen endlich-dimensionalen Vektorräumen V und W gibt es stets Basen B und C von V bzw. W, so dass die zugehörige Abbildungsmatrix die besonders einfache Form

$$A_f^{B,C} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

hat. Wir suchen nun nach einer analogen Aussage, bei der V und W Vektorräume mit Skalarprodukt sind und wir nur *Orthonormalbasen* B und C zulassen wollen. Da uns dies mehr einschränkt, erwarten wir natürlich, dass die zugehörige Normalform dann nicht mehr ganz so einfach sein wird. Wir werden allerdings sehen, dass es genügt, statt der Einsen auf der „Diagonalen" der Abbildungsmatrix beliebige positive reelle Zahlen zuzulassen. Daher führen wir für derartige Matrizen zunächst eine einfache Notation ein:

Notation 22.43 (Nichtquadratische Diagonalmatrizen)**.** Im Folgenden wollen wir auch eine nicht notwendig quadratische Matrix $D = (d_{i,j})_{i,j} \in \operatorname{Mat}(m \times n, K)$ als *Diagonalmatrix* bezeichnen, wenn $d_{i,j} = 0$ für alle $i \neq j$. Analog zu Definition 19.29 schreiben wir für $\lambda_1, \dots, \lambda_r \in K$ mit $r \leq \min(m,n)$

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_r) := \left(\begin{array}{ccc|c} \lambda_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \lambda_r & \\ \hline & 0 & & 0 \end{array}\right) \in \operatorname{Mat}(m \times n, K)$$

(wobei die Größe der Matrix aus dieser Schreibweise nicht ersichtlich ist und aus dem Zusammenhang klar sein muss). Oft verlangen wir dabei, dass $\lambda_1, \dots, \lambda_r$ ungleich 0 sind, in diesem Fall ist dann natürlich $r = \operatorname{rk} D$.

Beachte, dass auch diese Diagonalmatrizen wie erwartet transponiert und multipliziert werden können: Ist $D = \operatorname{diag}(\lambda_1, \dots, \lambda_r) \in \operatorname{Mat}(m \times n, K)$, so ist z. B. auch

$$D^\mathsf{T} = \operatorname{diag}(\lambda_1, \dots, \lambda_r) \in \operatorname{Mat}(n \times m, K) \quad \text{und} \quad D^\mathsf{T} D = \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2) \in \operatorname{Mat}(n \times n, K).$$

Damit können wir nun den angekündigten Normalformensatz beweisen, der in der Literatur unter dem Namen *Singulärwertzerlegung* bekannt ist. Für den Beweis ist es praktisch, zunächst die Matrixform dieses Satzes zu betrachten.

Satz 22.44 (Singulärwertzerlegung)**.** *Zu jeder Matrix $A \in \operatorname{Mat}(m \times n, \mathbb{K})$ gibt es orthogonale bzw. unitäre Matrizen $S \in \operatorname{Mat}(m \times m, \mathbb{K})$ und $T \in \operatorname{Mat}(n \times n, \mathbb{K})$, so dass*

$$S^{-1} A T = \overline{S}^\mathsf{T} A T = \operatorname{diag}(\lambda_1, \dots, \lambda_r) =: D \quad \in \operatorname{Mat}(m \times n, \mathbb{K})$$

eine Diagonalmatrix mit reellen positiven $\lambda_1, \dots, \lambda_r$ und $r = \operatorname{rk} A$ ist.

Die $\lambda_1, \dots, \lambda_r$ sind dabei bis auf die Reihenfolge eindeutig bestimmt und heißen die ***Singulärwerte*** *von A. Man nennt die Produktdarstellung $A = S D T^{-1} = S D \overline{T}^\mathsf{T}$ die* ***Singulärwertzerlegung*** *von A.*

54

Beweis. Wir zeigen zunächst die Eindeutigkeit der Singulärwerte: Sind S und T beliebige orthogonale bzw. unitäre Matrizen der passenden Größen, so dass $\overline{S}^\mathsf{T} A T = \operatorname{diag}(\lambda_1, \dots, \lambda_r) =: D$ eine

Diagonalmatrix mit reellen positiven $\lambda_1, \dots, \lambda_r$ ist, so folgt durch Transponieren und Konjugieren auch $\overline{T}^\mathsf{T}\overline{A}^\mathsf{T} S = \overline{D}^\mathsf{T} = D^\mathsf{T}$, und damit

$$T^{-1}\overline{A}^\mathsf{T} A T = \overline{T}^\mathsf{T}\overline{A}^\mathsf{T} A T = \overline{T}^\mathsf{T}\overline{A}^\mathsf{T} S \overline{S}^\mathsf{T} A T = D^\mathsf{T} D = \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2) \quad \in \operatorname{Mat}(n \times n, \mathbb{K}).$$

Die Matrix $\overline{A}^\mathsf{T} A$ ist also ähnlich zur Diagonalmatrix $D^\mathsf{T} D$ und hat damit dieselben Eigenwerte. Die $\lambda_1^2, \dots, \lambda_r^2$ (und damit auch $\lambda_1, \dots, \lambda_r$) sind daher bis auf ihre Reihenfolge eindeutig bestimmt, nämlich als die Eigenwerte von $\overline{A}^\mathsf{T} A$ ungleich 0.

Der Existenzbeweis der Singulärwertzerlegung ist konstruktiv und erfolgt in zwei Schritten:

(a) Bestimmung von T.

Die Matrix $\overline{A}^\mathsf{T} A \in \operatorname{Mat}(n \times n, \mathbb{K})$ ist symmetrisch bzw. hermitesch, da $\overline{\overline{A}^\mathsf{T} A}^\mathsf{T} = \overline{A}^\mathsf{T} A$. Nach dem Spektralsatz gibt es wie in Beispiel 22.28 also eine orthogonale bzw. unitäre Matrix T, so dass $\overline{T}^\mathsf{T}\overline{A}^\mathsf{T} A T$ eine Diagonalmatrix ist.

Nun ist $A^\mathsf{T} A$ aber auch positiv semidefinit, denn es gilt $\overline{x}^\mathsf{T}\overline{A}^\mathsf{T} A x = \overline{Ax}^\mathsf{T} A x = \|Ax\|^2 \geq 0$ für alle $x \in \mathbb{K}^n$, wobei $\|\cdot\|$ die Norm des Standardskalarprodukts ist. Die Einträge der Diagonalmatrix $\overline{T}^\mathsf{T}\overline{A}^\mathsf{T} A T$ sind also nicht-negativ, und damit können wir (nach geeigneter Anordnung der Spalten von T)

$$\overline{T}^\mathsf{T}\overline{A}^\mathsf{T} A T = \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2)$$

mit reellen $\lambda_1, \dots, \lambda_r > 0$ für ein $r \leq n$ schreiben.

(b) Bestimmung von S.

Für alle $i \leq r$ setzen wir

$$s_i := \frac{1}{\lambda_i} A T e_i \quad \in \mathbb{K}^m.$$

Diese Vektoren sind bezüglich des Standardskalarprodukts orthonormal, denn für alle $i, j \leq r$ gilt

$$\overline{s_i}^\mathsf{T} s_j = \frac{1}{\lambda_i \lambda_j} \overline{e_i}^\mathsf{T} \overline{T}^\mathsf{T} \overline{A}^\mathsf{T} A T e_j \overset{\text{(a)}}{=} \frac{1}{\lambda_i \lambda_j} \overline{e_i}^\mathsf{T} \cdot \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2) \cdot e_j = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Wir können sie nach Satz 21.32 also zu einer Orthonormalbasis von $\mathbb{K}^m$ ergänzen (insbesondere ist damit auch $r \leq m$) und daraus die orthogonale bzw. unitäre Matrix $S := (s_1 \mid \cdots \mid s_m)$ bilden.

Wir setzen nun $D := \operatorname{diag}(\lambda_1, \dots, \lambda_r) \in \operatorname{Mat}(m \times n, \mathbb{K})$ und behaupten, dass damit wie gewünscht $\overline{S}^\mathsf{T} A T = D$, also $A T = S D \in \operatorname{Mat}(m \times n, \mathbb{K})$ gilt. Die Gleichheit dieser Matrizen überprüfen wir spaltenweise, d. h. wir zeigen $A T e_i = S D e_i$ für alle $i = 1, \dots, n$:

- Für $i \leq r$ gilt $A T e_i = s_i \lambda_i = S D e_i$ nach (b).
- Für $i > r$ ist

$$\|A T e_i\|^2 = \overline{e_i}^\mathsf{T} \overline{T}^\mathsf{T} \overline{A}^\mathsf{T} A T e_i \overset{\text{(a)}}{=} \overline{e_i}^\mathsf{T} \cdot \operatorname{diag}(\lambda_1^2, \dots, \lambda_r^2) \cdot e_i = 0$$

und damit $A T e_i = 0 = S D e_i$, da die i-te Spalte $D e_i$ von D eine Nullspalte ist.

Mit $\overline{S}^\mathsf{T} A T = D$ sind A und D nun insbesondere auch äquivalent, so dass $r = \operatorname{rk} D = \operatorname{rk} A$ nach Lemma 16.41 gilt. Damit ist alles gezeigt. □

Bemerkung 22.45 (Singulärwertzerlegung für Morphismen)**.** Wie am Anfang dieses Abschnitts erwähnt gibt es natürlich auch eine Variante der Singulärwertzerlegung für Morphismen: Ist $f\colon V \to W$ eine lineare Abbildung zwischen zwei endlich-dimensionalen Vektorräumen V und W mit Skalarprodukt, so gibt es Orthonormalbasen B und C von V bzw. W, so dass die Abbildungsmatrix $A_f^{B,C}$ diagonal mit reellen nicht-negativen Diagonaleinträgen ist. Dies ergibt sich unmittelbar aus Satz 22.44 angewendet auf eine Abbildungsmatrix von f zu beliebigen Orthonormalbasen.

Beispiel 22.46.

(a) Ist $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ quadratisch, symmetrisch bzw. hermitesch und positiv semidefinit, so finden wir nach dem Spektralsatz wie in Beispiel 22.28 bereits eine orthogonale bzw. unitäre Matrix T, so dass $\overline{T}^{\mathsf{T}} A T$ eine Diagonalmatrix ist, deren Einträge nach Satz 22.33 (a) in $\mathbb{R}_{\geq 0}$ liegen. Wir können in Satz 22.44 dann also $S = T$ wählen, und die Singulärwerte von A sind genau die positiven Eigenwerte von A.

(b) Wir wollen eine Singulärwertzerlegung der Matrix

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \in \mathrm{Mat}(3 \times 2, \mathbb{R})$$

finden. Um wie im Beweis von Satz 22.44 zunächst T zu bestimmen, müssen wir die symmetrische, positiv semidefinite Matrix

$$A^{\mathsf{T}} A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

nach dem Spektralsatz orthogonal diagonalisieren. Dies haben wir bereits in Beispiel 22.29 getan: Mit

$$T = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad \in \mathrm{O}(2) \quad \text{ist} \quad T^{\mathsf{T}} A^{\mathsf{T}} A T = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} =: \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix}.$$

Die Singulärwerte von A sind also $\lambda_1 = \sqrt{3}$ und $\lambda_2 = \sqrt{1} = 1$. Um nun auch S zu bestimmen, setzen wir

$$s_1 := \frac{1}{\lambda_1} A T e_1 = \frac{1}{\sqrt{3}} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{6}} \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\text{und} \qquad s_2 := \frac{1}{\lambda_2} A T e_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Beachte, dass diese beiden Vektoren in der Tat wie in der Konstruktion von Satz 22.44 bezüglich des Standardskalarprodukts orthonormal sind. Wir ergänzen sie leicht (z. B. mit dem Gram-Schmidtschen Orthonormalisierungsverfahren aus Satz 21.32, falls man das Ergebnis nicht bereits sieht) mit

$$s_3 := \frac{1}{\sqrt{3}} \cdot \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

zu einer Orthonormalbasis von $\mathbb{R}^3$. Nach Satz 22.44 gilt dann mit der orthogonalen Matrix $S := (s_1 \,|\, s_2 \,|\, s_3) \in \mathrm{O}(3)$ also

$$S^{-1} A T = S^{\mathsf{T}} A T = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Beispiel 22.47 (Approximation von Matrizen)**.** Eine interessante praktische Anwendung der Singulärwertzerlegung liegt im Bereich der Approximation von Matrizen durch Matrizen von kleinem Rang. Dazu betrachten wir einmal eine (große) reelle Matrix $A = (a_{i,k})_{i,k} \in \mathrm{Mat}(m \times n, \mathbb{R})$ in ihrer Singulärwertzerlegung $A = SDT^{\mathsf{T}}$, so dass für alle $i = 1, \ldots, m$ und $k = 1, \ldots, n$ nach Definition 16.5 der Matrixmultiplikation also

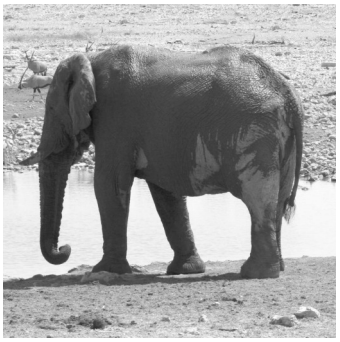
$$a_{i,k} = \sum_{j=1}^{r} s_{i,j} \lambda_j t_{k,j} \tag{$*$}$$

mit $S = (s_{i,j})_{i,j} \in \mathrm{O}(m)$, $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_r) \in \mathrm{Mat}(m \times n, \mathbb{R})$ und $T = (t_{j,k})_{j,k} \in \mathrm{O}(n)$ gilt, wobei wir die Singulärwerte so anordnen, dass $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ (mit $r = \mathrm{rk} A$). Beachte dabei, dass in

$(*)$ nur die ersten r Spalten von S und T benötigt werden, da in der Summe stets $j \leq r$ ist. Wir können A also exakt rekonstruieren, wenn wir nur diese Teile von S und T sowie die Diagonaleinträge von D kennen, was insgesamt

$$rm + rn + r = r(m+n+1)$$

reelle Zahlen sind. Wenn der Rang r von A klein ist, können dies deutlich weniger Zahlen sein als wenn wir uns die mn Einträge von A direkt merken würden. Matrizen von kleinem Rang lassen sich so also z. B. in einem Computer sehr platzsparend abspeichern.



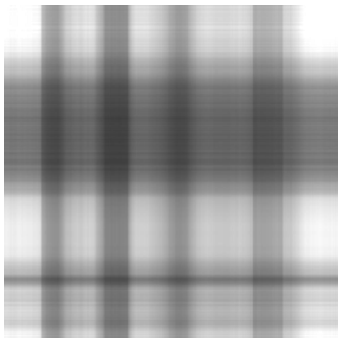
Für beliebige Matrizen, deren Rang in der Regel nicht klein ist, hilft dies natürlich erst einmal nicht weiter. Ist es aber akzeptabel, die Werte in der Matrix nur näherungsweise abzuspeichern – wie etwa in dem Foto rechts, das mit seinen Helligkeitswerten eine reelle Matrix der Größe 1000×1000 darstellt – so können wir sehr einfach eine Näherung des Ausdrucks $(*)$ bilden, indem wir ein $r' \leq r$ wählen und in der Summe nur die größten (also „wichtigsten“) r' Singulärwerte $\lambda_1, \ldots, \lambda_{r'}$ berücksichtigen, d. h. $A = SDT^\mathsf{T}$ durch die Matrix

$$A' = SD'T^\mathsf{T} \quad \text{mit} \quad D' = \operatorname{diag}(\lambda_1, \ldots, \lambda_{r'})$$

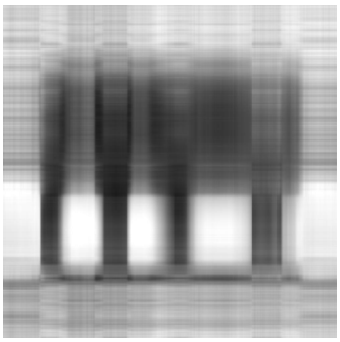
vom Rang r' approximieren, die sich dann wieder wie oben platzsparend abspeichern lässt. Wir können den Fehler, den wir dabei machen, auch genau berechnen: Bezüglich des Standardskalarprodukts auf $\operatorname{Mat}(m \times n, \mathbb{R}) = \mathbb{R}^{mn}$ wie in Beispiel 21.15 (c) ist

$$\begin{aligned}
\|A - A'\|^2 &= \left\|S(D-D')T^\mathsf{T}\right\|^2 \\
&= \operatorname{Spur}\big(T(D-D')^\mathsf{T} \underbrace{S^\mathsf{T} \cdot S}_{=E}(D-D')T^\mathsf{T}\big) \\
&\overset{19.7}{=} \operatorname{Spur}\big((D-D')^\mathsf{T}(D-D')\underbrace{T^\mathsf{T}T}_{=E}\big) \\
&= \sum_{j>r'} \lambda_j^2 .
\end{aligned}$$

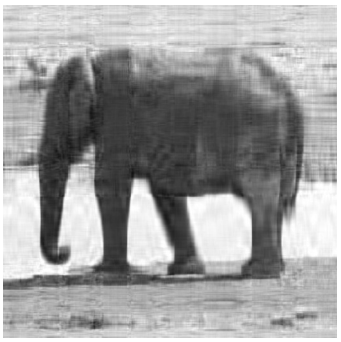
Da die weggelassenen Singulärwerte λ_j für $j > r'$ ja die kleinsten sind, ist diese Näherung also wirklich recht gut. In der Tat kann man zeigen, dass sie im Sinne dieser Norm die beste Approximation von A durch eine Matrix vom Rang höchstens r' ist. Für das obige Foto sind drei dieser Näherungen für (sehr) kleine Werte von r' unten dargestellt.



$r' = 1$



$r' = 3$



$r' = 20$

Die ursprüngliche Matrix $A \in \operatorname{Mat}(1000 \times 1000, \mathbb{R})$ hat wie erwartet maximalen Rang $r = 1000$, ihre Singulärwerte fallen allerdings sehr schnell ab: Schon der erste im mittleren Näherungsfoto weggelassene Singulärwert λ_4 ist nur noch etwa $\frac{1}{10}$-mal so groß wie λ_1, und nahezu die Hälfte aller

Singulärwerte ist kleiner als $\frac{1}{1000} \cdot \lambda_1$, so dass deren Vernachlässigung praktisch nicht mehr erkennbar ist. Das rechte Näherungsfoto belegt weniger als 5 % des Speicherplatzes der ursprünglichen Matrix A. Beachte auch, dass man dem linken Näherungsfoto „ansieht", dass es Rang 1 hat: Alle Spalten der Matrix sind ein Vielfaches desselben Vektors, haben also von oben nach unten die gleiche Helligkeitsverteilung und sind nur insgesamt heller oder dunkler – wodurch das deutlich sichtbare Streifenmuster entsteht.

Aufgabe 22.48 (Pseudoinverse Matrizen)**.**

(a) Zeige mit Hilfe der Singulärwertzerlegung, dass es zu jeder Matrix $A \in \mathrm{Mat}(m \times n, \mathbb{R})$ eine eindeutig bestimmte Matrix $B \in \mathrm{Mat}(n \times m, \mathbb{R})$ gibt, so dass

$$ABA = A \quad \text{und} \quad BAB = B$$

gilt und AB sowie BA symmetrisch sind.

Ist A quadratisch und invertierbar, so ist dann offensichtlich $B = A^{-1}$. Für eine allgemeine Matrix A, die nicht notwendig quadratisch ist bzw. vollen Rang hat, nennt man B daher die *pseudoinverse Matrix* zu A.

(b) Berechne für die Matrix

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

eine Singulärwertzerlegung und die pseudoinverse Matrix wie in (a).

Aufgabe 22.49. Es seien $n \in \mathbb{N}_{>0}$ und $A \in \mathrm{Mat}(n \times n, \mathbb{R})$. Man zeige:

(a) Die Matrix A kann geschrieben werden als $A = QB$, wobei $Q \in \mathrm{O}(n)$ eine orthogonale und $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine positiv semidefinite symmetrische Matrix ist.

(b) In der Zerlegung aus (a) ist die Matrix B eindeutig bestimmt. Ist darüber hinaus A invertierbar, so ist auch Q eindeutig bestimmt, und B ist sogar positiv definit.

Wie kann man diese Aussage geometrisch interpretieren?

Aufgabe 22.50. Zu einer Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, \mathbb{K})$ mit $A \neq 0$ bezeichne $\sigma_A \in \mathbb{R}_{>0}$ ihren größten Singulärwert. Man zeige bezüglich der Normen zu den Standardskalarprodukten auf $\mathbb{K}^m$, $\mathbb{K}^n$ und $\mathrm{Mat}(m \times n, \mathbb{K})$:

(a) $\sigma_A = \max\left\{ \frac{\|Ax\|}{\|x\|} : x \in \mathbb{K}^n \backslash \{0\} \right\}$;

(b) $\sigma_A \leq \|A\|$.

Für welche Matrizen gilt hierbei die Gleichheit?

Grundlagen der Mathematik 2: Analysis

23. Topologische Grundbegriffe

Wir haben nun unser Studium der linearen Algebra beendet und wenden uns wieder der Analysis zu. Unser Ziel für den Rest dieser Vorlesung wird es sein, die in den Kapiteln 5 bis 12 entwickelte Theorie für Funktionen in einer reellen Variablen – insbesondere die Differential- und Integralrechnung – auf den mehrdimensionalen Fall zu übertragen, also analoge Resultate z. B. auch für Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$ zu finden. Da es bekanntlich die grundlegende Idee der Differentialrechnung ist, beliebige Funktionen durch lineare zu approximieren, werden unsere Ergebnisse zu linearen Abbildungen, die wir in den vorangegangenen Kapiteln zur linearen Algebra erzielt haben, dabei sehr nützlich sein.

23.A Normierte und metrische Räume

Erinnern wir uns an die eindimensionale Analysis zurück: Der zentrale Begriff, mit dem wir damals in Kapitel 5 begonnen haben, war der des Grenzwerts einer Folge. Wir haben dabei eine (reelle oder komplexe) Zahlenfolge $(a_n)_{n\in\mathbb{N}}$ konvergent gegen ein $a \in \mathbb{K}$ genannt, wenn in jeder ε-Umgebung

$$U_\varepsilon(a) = \{x \in \mathbb{K} : |x-a| < \varepsilon\} \tag{$*$}$$

von a fast alle Folgenglieder a_n liegen (siehe Definition 5.1 und Bemerkung 5.2). Wenn wir diese Definition auf den höherdimensionalen Fall übertragen wollen, brauchen wir dazu offensichtlich eine Verallgemeinerung der Betragsfunktion, mit der wir in $(*)$ den Abstand von x zu a messen konnten. In der Tat haben wir so etwas in Definition 21.13 bereits kennengelernt: die Norm bzw. Länge eines Vektors in einem Vektorraum mit Skalarprodukt. Wir wollen daher zuerst diesen Normbegriff genauer untersuchen. Wie schon in Bemerkung 21.22 erwähnt ist in der Analysis das Konzept einer Norm auf einem $\mathbb{K}$-Vektorraum sehr allgemein und erlaubt nicht nur die Normen von Skalarprodukten, die wir in Abschnitt 21.B kennengelernt haben. Vielmehr ist eine allgemeine Norm definiert als eine Abbildung, die jedem Vektor eine reelle Zahl zuordnet und die erwarteten Eigenschaften erfüllt – nämlich genau diejenigen, die wir in Satz 21.21 für Normen zu Skalarprodukten bereits bewiesen haben.

Definition 23.1 (Normen und normierte Räume)**.** Es sei V ein $\mathbb{K}$-Vektorraum. Wir nennen eine Abbildung $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$, $x \mapsto \|x\|$ eine **Norm** auf V, wenn für alle $x, y \in V$ und $\lambda \in \mathbb{K}$ gilt:

(a) $\|\lambda x\| = |\lambda| \cdot \|x\|$;

(b) $\|x\| > 0$ für alle $x \neq 0$;

(c) $\|x+y\| \leq \|x\| + \|y\|$ (**Dreiecksungleichung**).

Ein $\mathbb{K}$-Vektorraum V zusammen mit einer Norm $\|\cdot\|$ wird als **normierter Raum** bezeichnet. Wir schreiben ihn als $(V, \|\cdot\|)$, oder auch einfach nur als V, wenn die betrachtete Norm aus dem Zusammenhang klar ist.

Bemerkung 23.2. In jedem normierten Raum ist die Norm des Nullvektors nach Definition 23.1 (a) gleich $\|0_V\| = \|0_\mathbb{K} \cdot 0_V\| = 0_\mathbb{K} \cdot \|0_V\| = 0$. Eine Norm nimmt also nur reelle nicht-negative Werte an, und ist genau dann gleich 0, wenn der Vektor der Nullvektor ist.

Beispiel 23.3.

(a) Jeder $\mathbb{K}$-Vektorraum V mit Skalarprodukt, also jeder euklidische oder unitäre Raum, ist mit der Vorschrift $\|x\| := \sqrt{\langle x, x\rangle}$ ein normierter Raum – dies haben wir in Satz 21.21 bewiesen.

Wenn wir nichts anderes angeben, werden wir in Zukunft jeden Vektorraum mit Skalarprodukt auf diese Art als normierten Raum ansehen.

Um Normen anschaulich darzustellen, zeichnet man in der Regel die sogenannte Einheitskugel $\{x \in V : \|x\| \leq 1\}$, also die Menge aller Vektoren, die bezüglich dieser Norm die Länge höchstens 1 haben. Wir haben bei der Hauptachsentransformation in Beispiel 22.38 gesehen, dass diese Einheitskugel im Fall einer von einem Skalarprodukt bestimmten Norm ein gedrehtes Ellipsoid mit Mittelpunkt im Ursprung ist. Im Fall eines reellen zweidimensionalen Vektorraums ist dies im Bild unten dargestellt.

Der wichtigste Fall ist dabei natürlich $V = \mathbb{K}^n$ mit dem Standardskalarprodukt. Wenn wir nichts Gegenteiliges angeben, wollen wir $\mathbb{K}^n$ in Zukunft immer mit der hieraus resultierenden Norm

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

als normierten Raum betrachten. Man nennt dies die **euklidische Norm**. Im Fall $\mathbb{K} = \mathbb{R}$ ist
55 die Einheitskugel $\{x \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 \leq 1\}$ dann natürlich die „gewöhnliche“ Kugel.

(b) Für $V = \mathbb{K}^n$ ist die **Maximumsnorm** definiert als

$$\|x\| = \max(|x_1|, \ldots, |x_n|).$$

In der Tat ist dies eine Norm: Die Eigenschaften (a) und (b) aus Definition 23.1 sind offensichtlich, und die Dreiecksungleichung folgt sofort aus der in $\mathbb{K}$, denn es ist

$$\|x+y\| = |x_i + y_i| \leq |x_i| + |y_i| \leq \|x\| + \|y\|,$$

wobei $i \in \{1, \ldots, n\}$ einen Index bezeichnet, für den in $\|x+y\| = \max(|x_1+y_1|, \ldots, |x_n+y_n|)$ das Maximum angenommen wird. Die zugehörige Einheitskugel

$$\{x \in \mathbb{K}^n : \max(|x_1|, \ldots, |x_n|) \leq 1\} = \{x \in \mathbb{K}^n : |x_i| \leq 1 \text{ für alle } i = 1, \ldots, n\}$$

ist in diesem Fall ein achsenparalleler Würfel; er ist wieder im Bild unten eingezeichnet. Da wir hier kein Ellipsoid erhalten, sehen wir also auch schon, dass die Maximumsnorm keine Norm sein kann, die von einem Skalarprodukt kommt (siehe auch Aufgabe 23.6). Das Konzept eines normierten Raumes lässt also allgemeinere Normen zu als die, die wir in Abschnitt 21.B kennengelernt haben.

(c) Für $V = \mathbb{K}^n$ definieren wir die **Summennorm** durch

$$\|x\| = |x_1| + \cdots + |x_n|.$$

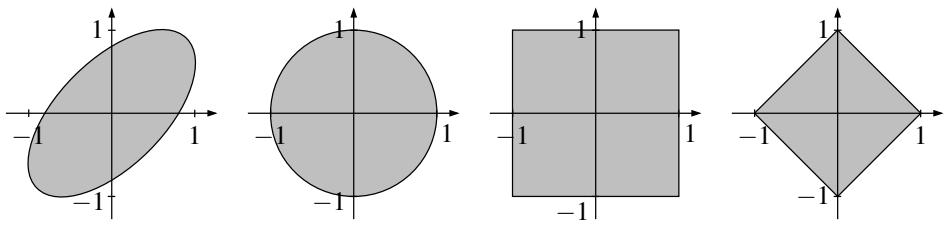
Auch hier sind die Bedingungen (a) und (b) in Definition 23.1 wieder offensichtlich, und die Dreiecksungleichung ergibt sich aus

$$\|x+y\| = |x_1+y_1| + \cdots + |x_n+y_n| \leq |x_1| + |y_1| + \cdots + |x_n| + |y_n| = \|x\| + \|y\|.$$

Die zugehörige Einheitskugel ist für $n = 2$ wieder im folgenden Bild dargestellt. Man kann sich ihre Form leicht dadurch überlegen, dass sie wegen der Beträge in der Definition der Norm symmetrisch zu beiden Achsen sein muss und im positiven Quadranten $\mathbb{R}^2_{\geq 0}$ durch die Ungleichung $x_1 + x_2 \leq 1$ gegeben ist.

Die Einheitskugel in $\mathbb{R}^2$ bezüglich der ...

Norm zu einem allgemeinen Skalarprodukt	euklidischen Norm	Maximumsnorm	Summennorm



(d) Es sei wieder $V = \mathbb{K}^n$. Für eine reelle Zahl $p \geq 1$ kann man zeigen, dass durch

$$\|x\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p}$$

eine Norm definiert wird, die sogenannte **p-Norm**. In der Tat sind die ersten beiden Eigenschaften von Definition 23.1 auch hier wieder klar, die Dreiecksungleichung ist jedoch etwas aufwändiger als für unsere bisher betrachteten Normen nachzurechnen (und benötigt die Bedingung $p \geq 1$). Da wir diese allgemeinen p-Normen in unserer Vorlesung nicht weiter benötigen, verzichten wir hier auf den Beweis. Ihr könnt ihn z. B. in [Fo1, Kapitel 16, Satz 8] finden.

Wir erwähnen die p-Norm hier nur deshalb, weil sie eine Verallgemeinerung unserer oben betrachteten Normen ist und sich daraus auch deren übliche Bezeichnungsweise ableitet: Offensichtlich ist die 1-Norm gerade die Summennorm und die 2-Norm die euklidische Norm. Für $p \to \infty$ hingegen ergibt sich die Maximumsnorm: Ist nämlich $x \in \mathbb{K}^n$ ein Vektor, für den ohne Beschränkung der Allgemeinheit das Maximum in $\|x\| = \max(|x_1|, \ldots, |x_n|)$ für $|x_1|$ angenommen wird, so ist

$$\lim_{p\to\infty} \|x\|_p = |x_1| \cdot \lim_{p\to\infty} \sqrt[p]{1 + \frac{|x_2|^p}{|x_1|^p} + \cdots + \frac{|x_n|^p}{|x_1|^p}} = |x_1| = \|x\|,$$

da der Ausdruck unter der Wurzel zwischen 1 und n liegt und der Grenzwert für $p \to \infty$ somit durch $\sqrt[p]{1} \to 1$ und $\sqrt[p]{n} \to 1$ eingeschachtelt wird. Für $x \in \mathbb{K}^n$ bezeichnet man daher ...

- die Summennorm mit $\|x\|_1 := |x_1| + \cdots + |x_n|$,
- die euklidische Norm mit $\|x\|_2 := \sqrt{|x_1|^2 + \cdots + |x_n|^2}$,
- die Maximumsnorm mit $\|x\|_\infty := \max(|x_1|, \ldots, |x_n|)$.

(e) Alle oben betrachteten Normen gibt es analog auch auf dem Vektorraum $C^0([a,b])$ aller stetigen reellen Funktionen auf einem abgeschlossenen Intervall $[a,b]$: So ist z. B. für eine solche Funktion f

$$\|f\|_1 := \int_a^b |f(x)|\,dx, \quad \|f\|_2 := \sqrt{\int_a^b f(x)^2\,dx}, \quad \|f\|_\infty := \max\{|f(x)| : x \in [a,b]\}$$

(das Maximum existiert nach Satz 8.25, und die Normeigenschaften beweist man analog zu den Fällen oben bzw. zu Beispiel 21.15 (d)). Die Maximumsnorm $\|f\|_\infty$ ist dabei die, die wir in Aufgabe 8.43 bereits in einem etwas allgemeineren Fall als *Supremumsnorm* kennengelernt haben.

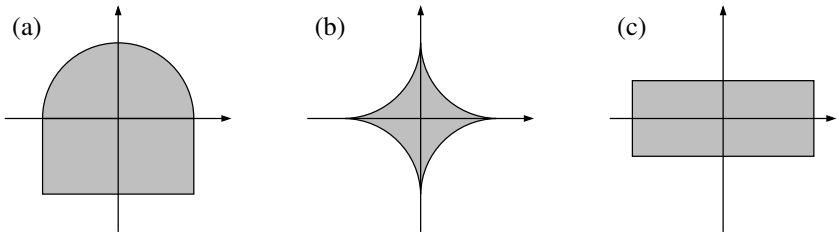
(f) Auf den Matrizenräumen $\mathrm{Mat}(m \times n, \mathbb{K})$ heißt die zum Standardskalarprodukt aus Beispiel 21.15 (c) gehörige Norm

$$\|A\| = \sqrt{\mathrm{Spur}\left(\overline{A}^{\mathsf{T}} A\right)} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^2} \quad \text{für } A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, \mathbb{K})$$

die **Frobenius-Norm**. Wir werden $\mathrm{Mat}(m \times n, \mathbb{K})$ im Folgenden standardmäßig mit dieser Norm versehen, sofern wir nichts Gegenteiliges angeben. Beachte aber, dass diese Norm *nicht* mit $\|A\|_2$ bezeichnet wird, weil dieser Name bereits für eine andere Matrixnorm reserviert ist, die wir auch später in Definition 24.37 und Beispiel 24.39 (a) noch kennenlernen werden.

Bemerkung 23.4 (Normen auf Unterräumen)**.** Genau wie bei Skalarprodukten in Bemerkung 21.16 kann man offensichtlich auch eine Norm auf einem $\mathbb{K}$-Vektorraum V auf einen Unterraum $U \leq V$ einschränken, um auch U zu einem normierten Raum zu machen.

Aufgabe 23.5. Welche der folgenden Bilder können Einheitskugeln einer Norm auf $\mathbb{R}^2$ sein?



Aufgabe 23.6. Man zeige:

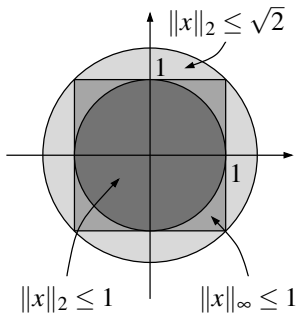
(a) In jedem Vektorraum V mit Skalarprodukt gilt die *Parallelogrammgleichung*
$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2) \quad \text{für alle } x,y \in V.$$

(b) Die Maximumsnorm auf $\mathbb{K}^n$ für $n \geq 2$ erfüllt die Parallelogrammgleichung nicht, und kann damit nach (a) also nicht die Norm zu einem Skalarprodukt sein.

Bemerkung 23.7 (Vergleich von Normen)**.** In Beispiel 23.3 haben wir viele verschiedene Normen gesehen. Wie hängen diese miteinander zusammen? Oft ist es so, dass man Normen gegeneinander abschätzen kann: So gelten z. B. für die Maximumsnorm und die euklidische Norm auf $\mathbb{K}^n$ die Ungleichungen
$$\|x\|_\infty = \max(|x_1|,\dots,|x_n|) = |x_i| \leq \sqrt{|x_1|^2 + \dots + |x_n|^2} = \|x\|_2$$
$$\text{und} \quad \|x\|_2 = \sqrt{|x_1|^2 + \dots + |x_n|^2} \leq \sqrt{|x_i|^2 + \dots + |x_i|^2} = \sqrt{n} \cdot |x_i| = \sqrt{n} \cdot \|x\|_\infty,$$
wobei $i \in \{1,\dots,n\}$ ein Index mit maximalem Wert von $|x_i|$ ist. Es gilt also für alle $x \in \mathbb{K}^n$
$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \cdot \|x\|_\infty,$$
d. h. wir haben die Folgerungen:

(a) Wenn $\|x\|_2 \leq 1$ gilt, so ist $\|x\|_\infty \leq 1$.

(b) Wenn $\|x\|_\infty \leq 1$ gilt, so ist $\|x\|_2 \leq \sqrt{n}$.



Dies kann man sich auch gut anschaulich vorstellen: Betrachten wir zu einer Norm $\|\cdot\|$ Kugeln $\{x \in \mathbb{K}^n : \|x\| \leq r\}$ mit einem anderen Radius als 1, so sind diese wegen der Linearitätseigenschaft aus Definition 23.1 (a) einfach nur skalierte Versionen der Einheitskugel, und die Ergebnisse (a) und (b) oben besagen genau, dass solche Kugeln wie im Bild rechts ineinander liegen: In jedem Kreis liegt noch ein Quadrat, und in jedem Quadrat liegt noch ein Kreis (mit geeigneten Radien).

Wir werden später (z. B. in Lemma 23.17 und Bemerkung 23.18) noch sehen, dass dies dazu führt, dass sich die beiden betrachteten Normen in vielerlei Hinsicht gleich verhalten. Man definiert daher:

Definition 23.8 (Äquivalente Normen)**.** Zwei Normen $\|\cdot\|$ und $\|\cdot\|'$ auf einem $\mathbb{K}$-Vektorraum V heißen **äquivalent**, wenn es Konstanten $a,b \in \mathbb{R}_{>0}$ gibt mit
$$a\,\|x\| \leq \|x\|' \leq b\,\|x\| \quad \text{für alle } x \in V.$$
Man prüft leicht nach, dass dies in der Tat eine Äquivalenzrelation auf der Menge aller Normen auf V ist.

Beispiel 23.9.

(a) Ist $\|\cdot\|$ eine beliebige Norm auf einem $\mathbb{K}$-Vektorraum V, so können wir daraus für ein gegebenes $\lambda \in \mathbb{R}_{>0}$ eine skalierte Variante $\|\cdot\|'$ mit
$$\|x\|' := \lambda \cdot \|x\| \quad \text{für alle } x \in V$$
konstruieren, die ebenfalls eine Norm ist. Diese neue Norm $\|\cdot\|'$ ist natürlich zur ursprünglichen äquivalent (wobei in Definition 23.8 mit $a := b := \lambda$ sogar die Gleichheit gilt).

(b) Wie wir in Bemerkung 23.7 gesehen haben, ist die Maximumsnorm auf $\mathbb{K}^n$ zur euklidischen Norm äquivalent.

In der Tat werden wir in Satz 24.36 noch sehen, dass sogar *alle* Normen auf $\mathbb{K}^n$ zueinander äquivalent sind (was sehr praktisch ist, weil wir ab dann für die meisten Anwendungen die uns jeweils am einfachsten erscheinende Norm benutzen können, ohne etwas am Ergebnis zu ändern). Für den Moment bedeutet dies aber erst einmal nur, dass wir für ein Beispiel nicht-äquivalenter Normen zwangsläufig einen unendlich-dimensionalen Vektorraum betrachten müssen:

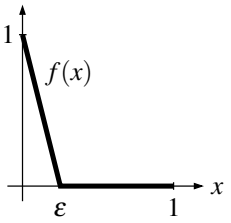
(c) Es sei $V = C^0([0,1])$ der $\mathbb{R}$-Vektorraum der stetigen Funktionen auf $[0,1]$ mit den beiden Normen

$$\|f\|_1 = \int_0^1 |f(x)|\,dx \quad \text{und} \quad \|f\|_\infty = \max\{|f(x)| : x \in [0,1]\}$$

aus Beispiel 23.3 (e). Angenommen, es gäbe eine Konstante $a \in \mathbb{R}_{>0}$ mit $a\|f\|_\infty \leq \|f\|_1$ für alle $f \in V$. Mit $\varepsilon := \min(a,1)$ erhalten wir dann aber für die rechts abgebildete Funktion $f \in V$

$$\|f\|_1 = \frac{\varepsilon}{2} \leq \frac{a}{2} < a \quad \text{und} \quad \|f\|_\infty = 1,$$

also im Widerspruch zur Annahme $\|f\|_1 < a\|f\|_\infty$. Es kann daher keine solche Konstante a geben, d. h. $\|\cdot\|_1$ und $\|\cdot\|_\infty$ sind nicht äquivalent auf $C^0([0,1])$.



Bevor wir zur angekündigten Anwendung des Längen- bzw. Abstandsbegriffs in der Analysis kommen, wollen wir das Konzept eines normierten Raumes aber erst noch etwas verallgemeinern. Eine recht große Einschränkung ist es in der Praxis nämlich, dass normierte Räume nach Definition stets Vektorräume (über $\mathbb{R}$ oder $\mathbb{C}$) sein müssen. Um nur Abstände zwischen Punkten sinnvoll definieren zu können, benötigt man aber keine Vektorraumstruktur. Diese Idee führt zum folgenden Begriff eines metrischen Raumes, der lediglich den Abstand zweier Punkte, aber nicht die Länge eines Vektors definiert.

Definition 23.10 (Metriken und metrische Räume)**.** Es sei M eine Menge. Man nennt eine Abbildung $d\colon M \times M \to \mathbb{R}_{\geq 0}$, $(x,y) \mapsto d(x,y)$ (die man sich als *Abstandsfunktion* zwischen zwei Punkten vorstellen sollte; der hierfür übliche Buchstabe d steht dabei für *Distanz*) eine **Metrik** auf M, wenn für alle $x,y,z \in M$ gilt:

(a) $d(x,y) = d(y,x)$ (**Symmetrie**);

(b) $d(x,y) = 0$ genau dann, wenn $x = y$;

(c) $d(x,z) \leq d(x,y) + d(y,z)$ (**Dreiecksungleichung**).

Eine Menge M zusammen mit einer Metrik d heißt **metrischer Raum** und wird manchmal auch als (M,d) geschrieben.

Lemma 23.11. *Jeder normierte Raum $(V, \|\cdot\|)$ ist mit der Abstandsfunktion $d(x,y) := \|x-y\|$ ein metrischer Raum.*

Wenn wir nichts anderes spezifizieren, werden wir in Zukunft daher jeden normierten Raum auf diese Art als metrischen Raum auffassen.

Beweis. Wir überprüfen die Eigenschaften aus Definition 23.10: Für alle $x,y,z \in V$ folgt nach Definition 23.1

- aus Teil (a): $d(x,y) = \|x-y\| = \|(-1)(y-x)\| = |-1| \cdot \|y-x\| = d(y,x)$;
- aus Teil (b): $d(x,y) = \|x-y\| = 0$ genau dann, wenn $x-y=0$, also $x=y$;
- aus Teil (c): $d(x,z) = \|x-z\| = \|x-y+y-z\| \leq \|x-y\| + \|y-z\| = d(x,y) + d(y,z)$. $\square$

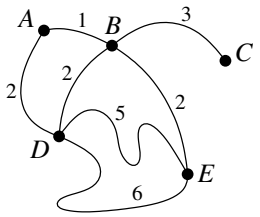
Beispiel 23.12.

(a) Ist (M,d) ein metrischer Raum, so ist jede Teilmenge $X \subset M$ mit der eingeschränkten Metrik $d|_{X\times X}$ ebenfalls wieder ein metrischer Raum. Standardmäßig werden wir in Zukunft jede Teilmenge eines metrischen Raumes auf diese Art wieder als metrischen Raum betrachten. Fassen wir alle unsere Konventionen zusammen, so werden wir im Folgenden also jede Teilmenge $X \subset \mathbb{K}^n$ als metrischen Raum mit der *euklidischen Metrik*

$$d(x,y) = \|x-y\|_2 = \sqrt{|x_1-y_1|^2+\cdots+|x_n-y_n|^2}$$

auffassen, sofern wir nichts Gegenteiliges angeben.

(b) Wir betrachten die „Landkarte“ mit 5 Städten A,B,C,D,E und Verbindungsstraßen wie im Bild rechts, wobei die (positiven) Zahlen an den Straßen deren Längen angeben sollen – man nennt ein solches Diagramm auch einen (zusammenhängenden) *gewichteten Graphen*. Für die Menge $M = \{A,B,C,D,E\}$ und $x,y \in M$ sei nun $d(x,y)$ die Länge eines kürzesten Weges von x nach y. So ist z. B. $d(D,E) = 4$, da von D nach E der Weg über B der kürzeste ist und eine Gesamtlänge von 4 hat.



Man sieht leicht ein, dass diese Abstandsfunktion d dann eine Metrik auf M definiert: Die Eigenschaften (a) und (b) aus Definition 23.10 sind offensichtlich, und (c) folgt aus der einfachen Tatsache, dass $d(x,y)+d(y,z)$ ja die Länge eines kürzesten Weges von x über y nach z ist und diese natürlich mindestens gleich der Länge $d(x,z)$ eines kürzesten Weges von x nach z ist, bei dem man nicht notwendig über y laufen muss.

Metrische Räume dieser Art spielen z. B. in der Optimierung eine große Rolle. In unserer Vorlesung werden sie nicht weiter vorkommen, sondern sollen hier nur als Beispiel dafür dienen, dass metrische Räume sehr allgemeine Objekte sind und auch „ganz anders aussehen“ können als normierte Vektorräume.

(c) Auf jeder Menge M ist

$$d(x,y) := \begin{cases} 0 & \text{für } x = y, \\ 1 & \text{für } x \neq y \end{cases}$$

offensichtlich eine Metrik, die sogenannte **diskrete Metrik**.

23.B Konvergenz in metrischen Räumen

Wie bereits angekündigt können wir nun analog zu Definition 5.1 Grenzwerte von Folgen in metrischen Räumen (und damit auch in normierten Räumen bzw. Vektorräumen mit Skalarprodukt) definieren.

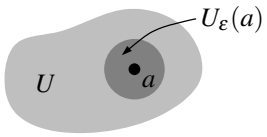
Definition 23.13 (Kugeln, Umgebungen und Grenzwerte)**.** Es seien M ein metrischer Raum und $a \in M$.

(a) Zu $r \in \mathbb{R}_{\geq 0}$ heißt

$$U_r(a) := \{x \in M : d(x,a) < r\} \quad \text{die } \textbf{offene Kugel}\text{, und}$$
$$K_r(a) := \{x \in M : d(x,a) \leq r\} \quad \text{die } \textbf{abgeschlossene Kugel}$$

um a mit Radius r.

(b) Eine Teilmenge $U \subset M$ heißt **Umgebung** von a, wenn es wie im Bild rechts ein $\varepsilon > 0$ gibt mit $U_\varepsilon(a) \subset U$. Insbesondere ist $U_\varepsilon(a)$ also selbst eine Umgebung von a. Man nennt sie oft auch die **ε-Umgebung** von a.



(c) Es sei $(a_n)_{n\in\mathbb{N}}$ eine Folge in M. Dann heißt a **Grenzwert** von $(a_n)_n$, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : d(a_n, a) < \varepsilon.$$

Wie im Fall von Folgen in $\mathbb{K}$ werden wir gleich in Lemma 23.15 sehen, dass ein solcher Grenzwert a eindeutig ist, sofern er existiert. Wir können ihn dann also *den* Grenzwert der Folge nennen und schreiben wie gewohnt $\lim_{n\to\infty} a_n = a$ oder „$a_n \to a$ für $n \to \infty$“. Existiert dieser Grenzwert, so heißt die Folge $(a_n)_n$ **konvergent**, andernfalls **divergent**.

Bemerkung 23.14.

(a) Es gibt mehrere äquivalente Umformulierungen der obigen Grenzwertdefinition. Am nützlichsten ist vermutlich das Kriterium

$$\lim_{n\to\infty} a_n = a \quad \Leftrightarrow \quad \lim_{n\to\infty} d(a_n, a) = 0,$$

das wegen der Positivität der Metrik unmittelbar durch Vergleich mit Definition 5.1 (b) folgt und den Grenzwert $\lim_{n\to\infty} a_n$ im metrischen Raum M damit direkt auf einen Grenzwert $\lim_{n\to\infty} d(a_n, a)$ in $\mathbb{R}$ zurückführt. Eine andere äquivalente Formulierung von Definition 23.13 (c) mit Hilfe des Umgebungsbegriffs und der Notation „fast alle“ für „alle bis auf endlich viele“ ist offensichtlich genau wie in Bemerkung 5.2

$$\lim_{n\to\infty} a_n = a \quad \Leftrightarrow \quad \text{In jeder } \varepsilon\text{-Umgebung von } a \text{ liegen fast alle Folgenglieder } a_n.$$

Dies kann man schließlich noch umformulieren als

$$\lim_{n\to\infty} a_n = a \quad \Leftrightarrow \quad \text{In jeder Umgebung von } a \text{ liegen fast alle Folgenglieder } a_n.$$

Liegen nämlich in jeder Umgebung fast alle Folgenglieder, so natürlich insbesondere auch in jeder ε-Umgebung. Enthält umgekehrt jede ε-Umgebung von a fast alle Folgenglieder, so auch jede Umgebung von a, da eine solche ja nach Definition noch eine ε-Umgebung von a enthält.

(b) Wendet man Definition 23.13 auf einen normierten Raum V an, so ist nach Lemma 23.11 also $d(x,y) = \|x-y\|$. In diesem Fall ist demnach z. B. $U_r(a) = \{x \in V : \|x-a\| < r\}$, und eine Folge $(a_n)_n$ in V konvergiert genau dann gegen $a \in V$, wenn $\lim_{n\to\infty} \|a_n - a\| = 0$.

Lemma 23.15 (Eindeutigkeit des Grenzwerts)**.** *In einem metrischen Raum hat jede Folge höchstens einen Grenzwert.*

Beweis. Würde eine Folge $(a_n)_n$ in einem metrischen Raum M gegen zwei Punkte $a \neq b$ konvergieren, so müsste mit $\varepsilon := \frac{d(a,b)}{2}$ sowohl $d(a_n, a) < \varepsilon$ als auch $d(a_n, b) < \varepsilon$ für fast alle n gelten. Damit wäre nach der Dreiecksungleichung aber auch für fast alle n

$$d(a,b) \leq d(a, a_n) + d(a_n, b) < \varepsilon + \varepsilon = d(a,b),$$

was ein Widerspruch ist. □

Beispiel 23.16. Wir betrachten die Folge $(a_n)_{n\in\mathbb{N}_{>0}}$ in $\mathbb{R}^2$ mit $a_n = \begin{pmatrix} \frac{1}{n} \\ \frac{1}{n^2} \end{pmatrix}$ für alle n.

(a) Versehen wir $\mathbb{R}^2$ mit der euklidischen Norm, so konvergiert die Folge nach Bemerkung 23.14 (b) gegen $0 \in \mathbb{R}^2$, denn es ist

$$\lim_{n\to\infty} \|a_n - 0\|_2 = \lim_{n\to\infty} \sqrt{\frac{1}{n^2} + \frac{1}{n^4}} = 0.$$

(b) Auch mit der Maximumsnorm aus Beispiel 23.3 (b) konvergiert die Folge gegen 0, denn

$$\lim_{n\to\infty} \|a_n - 0\|_\infty = \lim_{n\to\infty} \max\left(\frac{1}{n}, \frac{1}{n^2}\right) = \lim_{n\to\infty} \frac{1}{n} = 0.$$

(c) In der diskreten Metrik aus Beispiel 23.12 (c) konvergiert die Folge jedoch nicht gegen 0: Hier ist

$$\lim_{n\to\infty} d(a_n,0) = \lim_{n\to\infty} 1 = 1 \neq 0,$$

da ja $a_n \neq 0$ für alle n gilt. In der Tat ist in der diskreten Metrik $U_1(a) = \{a\}$ für alle a – und damit konvergiert eine Folge $(a_n)_n$ in dieser Metrik genau dann gegen a, wenn fast alle Folgenglieder a_n gleich a sind.

56

Die Konvergenz einer Folge in einem metrischen Raum hängt also im Allgemeinen von der gewählten Metrik ab. Wir wollen nun aber zeigen, dass dies bei äquivalenten Normen nicht der Fall ist. So hatten wir ja z. B. in Beispiel 23.9 (b) gesehen, dass die Maximumsnorm auf $\mathbb{R}^2$ zur euklidischen Norm äquivalent ist, und dementsprechend haben wir oben in den Beispielen 23.16 (a) und (b) dasselbe Ergebnis erhalten. Diese Tatsache beruht auf der folgenden wichtigen Aussage.

Lemma 23.17. *Es seien $\|\cdot\|$ und $\|\cdot\|'$ zwei äquivalente Normen auf einem $\mathbb{K}$-Vektorraum V. Dann gilt für alle $a \in V$ und $U \subset V$:*

$$U \text{ ist eine Umgebung von } a \text{ bezüglich } \|\cdot\|' \;\Leftrightarrow\; U \text{ ist eine Umgebung von } a \text{ bezüglich } \|\cdot\|,$$

d. h. „die beiden Normen erzeugen den gleichen Umgebungsbegriff“.

Beweis. Nach Voraussetzung gibt es eine Konstante $b \in \mathbb{R}_{>0}$ mit $\|x\|' \leq b\|x\|$ für alle $x \in V$. Im Folgenden bezeichnen wir die ε-Umgebungen eines Punktes $a \in V$ bezüglich $\|\cdot\|$ und $\|\cdot\|'$ mit $U_\varepsilon(a)$ bzw. $U'_\varepsilon(a)$.

Es sei nun U eine Umgebung von a bezüglich $\|\cdot\|'$, d. h. es gilt $U'_\varepsilon(a) \subset U$ für ein $\varepsilon > 0$. Dann ist aber auch $U_{\varepsilon/b}(a) \subset U$, denn für alle $x \in U_{\varepsilon/b}(a)$ gilt $\|x-a\|' \leq b\|x-a\| < b \cdot \frac{\varepsilon}{b} = \varepsilon$, also $x \in U'_\varepsilon(a) \subset U$. Damit ist U auch eine Umgebung von a bezüglich $\|\cdot\|$.

Die andere Richtung ergibt sich analog durch Vertauschen der Rollen der beiden Normen. □

Bemerkung 23.18 (Topologie)**.** Wir hatten in Abschnitt 23.A bereits erwähnt, dass sich äquivalente Normen in vielerlei Hinsicht gleich verhalten. In Lemma 23.17 haben wir nun ein erstes und sehr wichtiges Beispiel dafür gesehen: Der von ihnen erzeugte Umgebungsbegriff ist der gleiche.

Damit stimmen bei äquivalenten Normen natürlich auch alle Eigenschaften von Objekten überein, die sich allein mit Hilfe des Umgebungsbegriffs definieren lassen. Derartige Eigenschaften, von denen wir im Folgenden noch viele kennenlernen werden, bezeichnet man als *topologische Eigenschaften*. Ein erstes Beispiel dafür ist die Folgenkonvergenz, denn nach Bemerkung 23.14 (a) konvergiert eine Folge ja genau dann gegen einen Punkt, wenn in jeder Umgebung dieses Punktes fast alle Folgenglieder liegen. So können wir nach Beispiel 23.9 (b) also z. B. für Fragen zur Folgenkonvergenz in $\mathbb{K}^n$ statt der euklidischen Norm auch die oft einfachere Maximumsnorm verwenden, da diese beiden Normen dort äquivalent sind.

In der Tat gibt es eine weitere Verallgemeinerung metrischer Räume: die sogenannten *topologischen Räume*, die ihr in der Vorlesung „Einführung in die Topologie“ des zweiten Studienjahres kennenlernen könnt. In ihnen gibt es keine Metrik mehr, aber noch einen Umgebungsbegriff. Demnach kann man dort dann zwar keine Abstände mehr messen, aber dennoch alle topologischen Eigenschaften wie z. B. die Folgenkonvergenz definieren und untersuchen.

Im konkreten Fall des normierten Raumes $\mathbb{K}^n$ ist die Untersuchung der Konvergenz von Folgen allerdings noch etwas einfacher: Wie man vielleicht schon vermutet, wollen wir jetzt kurz zeigen, dass man diese Konvergenz dort einfach koordinatenweise überprüfen kann.

Lemma 23.19. *Eine Folge $(a^{(k)})_{k\in\mathbb{N}}$ in $\mathbb{K}^n$ konvergiert genau dann gegen $a \in \mathbb{K}^n$, wenn jede Koordinatenfolge $(a_i^{(k)})_{k\in\mathbb{N}}$ mit $i = 1,\dots,n$ in $\mathbb{K}$ gegen die i-te Koordinate a_i von a konvergiert.*

Beweis. Nach Bemerkung 23.18 können wir zum Beweis die Summennorm aus Beispiel 23.3 (c) verwenden. Dann gilt

$$\begin{aligned}
\lim_{k\to\infty} a^{(k)} = a &\Leftrightarrow \lim_{k\to\infty} \|a^{(k)} - a\|_2 = 0 && \text{(Bemerkung 23.14 (b))}\\
&\Leftrightarrow \lim_{k\to\infty} \sqrt{|a_1^{(k)} - a_1|^2 + \cdots + |a_n^{(k)} - a_n|^2} = 0\\
&\overset{(*)}{\Leftrightarrow} \lim_{k\to\infty} |a_i^{(k)} - a_i| = 0 \ \text{ für alle } i = 1,\ldots,n\\
&\Leftrightarrow \lim_{k\to\infty} a_i^{(k)} = a_i \ \text{ für alle } i = 1,\ldots,n,
\end{aligned}$$

wobei in $(*)$ die Richtung „$\Leftarrow$“ die üblichen Grenzwertsätze in $\mathbb{R}$ sind, und die Richtung „$\Rightarrow$“ aus der Ungleichung $|a_i^{(k)} - a_i| \le \sqrt{|a_1^{(k)} - a_1|^2 + \cdots + |a_n^{(k)} - a_n|^2}$ folgt. □

Als Nächstes wollen wir die in $\mathbb{K}$ aus Lemma 5.8 bekannte Aussage, dass konvergente Folgen beschränkt sind, auf metrische Räume übertragen. Dazu müssen wir aber natürlich den Begriff der Beschränktheit einer Menge oder Folge in einem metrischen Raum erst einmal definieren.

Definition 23.20 (Beschränkte Mengen und Folgen)**.** Es sei M ein metrischer Raum.

(a) Eine Teilmenge $X \subset M$ heißt **beschränkt**, wenn sie in einer abgeschlossenen Kugel enthalten ist, also wenn es ein $a \in M$ und $r \in \mathbb{R}_{>0}$ gibt mit $X \subset K_r(a)$, d. h. mit $d(x,a) \le r$ für alle $x \in X$.

(b) Eine Folge in M heißt **beschränkt**, wenn die Menge ihrer Folgenglieder in M beschränkt ist.

Bemerkung 23.21.

(a) Ist $X \subset M$ beschränkt, so gilt die Bedingung aus Definition 23.20 (a) sogar für alle $a \in M$: Ist nämlich $d(x,a) \le r$ für alle $x \in X$, so gilt für einen beliebigen anderen Punkt $b \in M$ auch

$$d(x,b) \le d(x,a) + d(a,b) \le r + d(b,a) = R \qquad \text{für alle } x \in X,$$

wobei wir $R := r + d(a,b)$ gesetzt haben.

Insbesondere können wir in einem normierten Raum V also stets $a = 0$ wählen, und erhalten so die Aussage, dass eine Teilmenge $X \subset V$ genau dann beschränkt ist, wenn es ein $r \in \mathbb{R}_{>0}$ gibt mit $\|x\| \le r$ für alle $x \in X$. Für den Fall $V = \mathbb{K}$ mit der gewöhnlichen Metrik stimmt dies dann offensichtlich mit unserer alten Definition 4.22 (b) der Beschränktheit überein.

(b) Wie wir in Aufgabe 23.24 sehen werden, ist die Beschränktheit einer Menge in einem metrischen Raum keine topologische Eigenschaft im Sinne von 23.18. Dennoch stimmt sie für zwei äquivalente Normen $\|\cdot\|$ und $\|\cdot\|'$ auf einem Vektorraum V überein: Ist $\|x\|' \le b\|x\|$ für eine Konstante $b \in \mathbb{R}_{>0}$ und alle $x \in V$, und ist $X \subset V$ beschränkt bezüglich $\|\cdot\|$, d. h. gibt es ein $r \in \mathbb{R}_{>0}$ mit $\|x\| \le r$ für alle $x \in X$, so ist dann auch $\|x\|' \le b\|x\| \le br$ für alle $x \in X$, d. h. X ist auch beschränkt bezüglich $\|\cdot\|'$.

Lemma 23.22. *In einem metrischen Raum ist jede konvergente Folge beschränkt.*

Beweis. Es sei $(a_n)_n$ eine konvergente Folge mit Grenzwert a. Nach Bemerkung 23.14 (a) gilt dann $d(a_n,a) \to 0$ für $n \to \infty$. Nach Lemma 5.8 ist die konvergente reelle Folge $(d(a_n,a))_n$ also beschränkt, d. h. es gibt ein $r \in \mathbb{R}_{>0}$ mit $d(a_n,a) \le r$ für alle $n \in \mathbb{N}$. Damit ist auch $(a_n)_n$ beschränkt. □

Als einfache Folgerung können wir nun die aus Satz 5.13 bekannten Grenzwertsätze verallgemeinern – also die Aussage, dass Grenzwerte mit den üblichen Rechenoperationen vertauschen. Da es nach Definition aber in metrischen Räumen überhaupt keine Rechenoperationen und in normierten Vektorräumen nur Vektoraddition und Skalarmultiplikation gibt, müssen wir uns dabei natürlich auf diese Fälle beschränken.

Satz 23.23 (**Grenzwertsätze** in normierten Räumen)**.** *Es seien $(a_n)_n$ und $(b_n)_n$ zwei konvergente Folgen in einem normierten $\mathbb{K}$-Vektorraum V mit $a_n \to a$ und $b_n \to b$. Ferner sei $(\lambda_n)_n$ eine konvergente Folge in $\mathbb{K}$ mit $\lambda_n \to \lambda$. Dann gilt:*

(a) $a_n + b_n \to a + b$ *und* $a_n - b_n \to a - b$;

(b) $\lambda_n a_n \to \lambda a$.

Beweis.

(a) Der Beweis ist (bis auf das Ersetzen der Betragsstriche durch die Norm) wörtlich derselbe wie in Satz 5.13.

(b) Auch hier ist der Beweis analog zum eindimensionalen Fall: Es gilt

$$\begin{aligned}\|\lambda_n a_n - \lambda a\| &= \|\lambda_n a_n - \lambda a_n + \lambda a_n - \lambda a\| \\ &\le \|(\lambda_n - \lambda)a_n\| + \|\lambda(a_n - a)\| && \text{(Dreiecksungleichung)} \\ &= \underbrace{|\lambda_n - \lambda|}_{(*)} \cdot \|a_n\| + |\lambda| \cdot \underbrace{\|a_n - a\|}_{(*)}. && \text{(Definition 23.1 (a))}\end{aligned}$$

Dabei sind die mit $(*)$ bezeichneten Terme nach Voraussetzung (reelle) Nullfolgen, und wegen der Konvergenz von $(a_n)_n$ sind die Normen $\|a_n\|$ nach Lemma 23.22 beschränkt. Wie im Beweis von Satz 5.13 (b) ergibt sich nun also wieder $\|\lambda_n a_n - \lambda a\| \to 0$, und damit $\lambda_n a_n \to \lambda a$. □

Aufgabe 23.24. Für $x, y \in \mathbb{R}^2$ seien

$$d_1(x,y) := \begin{cases} \|x\|_2 + \|y\|_2 & \text{für } x \neq y, \\ 0 & \text{für } x = y \end{cases} \qquad \text{und} \qquad d_2(x,y) := \min(\|x-y\|_2, 1).$$

(a) Zeige, dass d_1 und d_2 Metriken auf $\mathbb{R}^2$ sind.

(b) Skizziere die qualitativ verschiedenen Fälle, wie abgeschlossene Kugeln bezüglich dieser beiden Metriken aussehen können.

(c) Man zeige: Eine Menge $A \subset \mathbb{R}^2$ ist bezüglich d_1 genau dann beschränkt, wenn A bezüglich der euklidischen Metrik beschränkt ist. Für d_2 gilt dies jedoch nicht.

(d) Man zeige: Eine Menge $A \subset \mathbb{R}^2$ ist bezüglich d_2 genau dann eine Umgebung eines Punktes $a \in \mathbb{R}^2$, wenn A bezüglich der euklidischen Metrik eine Umgebung von a ist. Für d_1 gilt dies jedoch nicht.

Insbesondere zeigt d_2 also, dass Beschränktheit kein topologischer Begriff ist: Diese Metrik liefert die gleichen Umgebungen wie die euklidische Metrik, aber nicht die gleichen beschränkten Mengen.

Zum Abschluss dieses Abschnitts wollen wir noch kurz das Cauchy-Kriterium aus Satz 6.25 verallgemeinern, das immer dann für den Nachweis der Konvergenz einer Folge benötigt wird, wenn ihr Grenzwert vorher noch nicht bekannt ist. Die Definition einer Cauchyfolge kann dabei unmittelbar aus Definition 6.22 übertragen werden.

Definition 23.25 (Cauchyfolgen)**.** Eine Folge $(a_n)_n$ in einem metrischen Raum M heißt **Cauchyfolge**, wenn

$$\forall \varepsilon > 0 \; \exists n_0 \in \mathbb{N} \; \forall m, n \ge n_0 : d(a_m, a_n) < \varepsilon.$$

Bemerkung 23.26.

(a) Genau wie die Beschränktheit von Mengen bzw. Folgen ist auch das Konzept von Cauchyfolgen keine topologische Eigenschaft. Mit dem gleichen Argument wie in Bemerkung 23.21 (b) stimmt es für äquivalente Normen aber dennoch überein.

(b) Wie in Bemerkung 6.23 zeigt man auch in beliebigen metrischen Räumen, dass jede konvergente Folge eine Cauchyfolge ist. Im Gegensatz zu Satz 6.25 gilt die Umkehrung in allgemeinen metrischen Räumen jedoch nicht, wie das folgende einfache Beispiel zeigt.

Beispiel 23.27 (Cauchyfolgen müssen nicht konvergieren)**.** Es seien $M = \mathbb{R}_{>0}$ mit der euklidischen Metrik und $(a_n)_{n \in \mathbb{N}_{>0}}$ die Folge in M mit $a_n = \frac{1}{n}$ für alle n. Dann ist $(a_n)_n$ nach Satz 6.25 eine Cauchyfolge, da sie in $\mathbb{R}$ gegen 0 konvergiert. Ihr Grenzwert liegt jedoch nicht in M, und damit ist die Folge in M nicht konvergent.

Da es in der Regel sehr wichtig ist zu wissen, ob Cauchyfolgen stets konvergieren und man somit die Konvergenz von Folgen mit dem Cauchy-Kriterium überprüfen kann, haben derartige metrische Räume einen besonderen Namen.

Definition 23.28 (Vollständige Räume)**.**

(a) Ein metrischer Raum heißt **vollständig**, wenn in ihm jede Cauchyfolge konvergiert (und die Cauchyfolgen nach Bemerkung 23.26 (b) damit genau die konvergenten Folgen sind).

(b) Ein vollständiger normierter Raum heißt **Banachraum**.

Satz 23.29. *$\mathbb{K}^n$ ist (bezüglich der euklidischen Norm) ein Banachraum.*

Beweis. Nach Bemerkung 23.18 und 23.26 (a) dürfen wir auf $\mathbb{K}^n$ statt der euklidischen Norm die dazu äquivalente Maximumsnorm verwenden.

Es sei nun $(a^{(k)})_{k\in\mathbb{N}}$ eine Cauchyfolge in $\mathbb{K}^n$, d. h. zu jedem $\varepsilon > 0$ gibt es ein $k_0 \in \mathbb{N}$, so dass $\|a^{(k)} - a^{(l)}\|_\infty < \varepsilon$, und damit auch $|a_i^{(k)} - a_i^{(l)}| < \varepsilon$ für alle $k, l \geq k_0$ und $i = 1, \ldots, n$ gilt. Also sind die Koordinatenfolgen $(a_i^{(k)})_{k\in\mathbb{N}}$ für alle i Cauchyfolgen in $\mathbb{K}$ und damit nach Satz 6.25 konvergent gegen gewisse $a_i \in \mathbb{K}$. Nach Lemma 23.19 konvergiert dann aber auch die ursprüngliche Folge $(a^{(k)})_{k\in\mathbb{N}}$ gegen den Vektor a mit Koordinaten $a_1, \ldots, a_n$. □

Aufgabe 23.30 (Beispiel für einen nicht-vollständigen normierten Raum)**.** Es sei $V = C^0([0,1])$ der Vektorraum der stetigen reellen Funktionen auf $[0,1]$. Man zeige:

(a) $(V, \|\cdot\|_\infty)$ ist ein Banachraum.

(b) $(V, \|\cdot\|_2)$ ist kein Banachraum.

Aufgabe 23.31. Zeige die folgenden Eigenschaften der Frobenius-Norm:

(a) Für alle $A \in \mathrm{Mat}(m \times n, \mathbb{K})$ und $B \in \mathrm{Mat}(n \times p, \mathbb{K})$ gilt $\|AB\| \leq \|A\| \cdot \|B\|$.

(b) Ist $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ mit $\|A\| < 1$, so ist $E - A$ invertierbar, und es gilt $\sum_{k=0}^{\infty} A^k = (E - A)^{-1}$.

Aufgabe 23.32. Es seien $A, B \in \mathrm{Mat}(n \times n, \mathbb{K})$.

(a) Zeige, dass der Grenzwert $\mathrm{e}^A := \sum_{k=0}^{\infty} \frac{A^k}{k!} \in \mathrm{Mat}(n \times n, \mathbb{K})$ existiert.

(b) Berechne e^A für die Matrix $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ mit $\lambda \in \mathbb{K}$.

(c) Zeige, dass im Allgemeinen *nicht* $\mathrm{e}^{A+B} = \mathrm{e}^A \cdot \mathrm{e}^B$ gilt.

23.C Offene und abgeschlossene Mengen

Im Rest dieses Kapitels wollen wir nun noch einige wichtige topologische Eigenschaften einführen. Die ersten von ihnen sind die einer offenen bzw. abgeschlossenen Menge, die eine direkte Verallgemeinerung der offenen und abgeschlossenen Intervalle bzw. Kugeln aus Notation 4.17 und Definition 23.13 (a) sind und anschaulich ausdrücken, ob eine Menge ihre Randpunkte mit enthält (siehe auch Bemerkung 23.41).

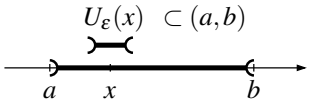
Definition 23.33 (Offene und abgeschlossene Mengen)**.** Es sei M ein metrischer Raum.

(a) Eine Teilmenge $U \subset M$ heißt **offen**, wenn sie eine Umgebung von jedem ihrer Punkte ist, also wenn es zu jedem $a \in U$ ein $\varepsilon > 0$ gibt mit $U_\varepsilon(a) \subset U$.

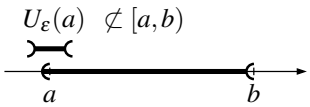
(b) Eine Teilmenge $A \subset M$ heißt **abgeschlossen**, wenn $M \backslash A$ offen ist, also wenn es zu jedem $a \in M \backslash A$ ein $\varepsilon > 0$ gibt mit $U_\varepsilon(a) \subset M \backslash A$. Man bezeichnet die Menge $M \backslash A$ oft als das **Komplement** von A in M – muss dabei aber aufpassen, dass dies natürlich nicht mit der Definition 15.7 eines Komplements von Unterräumen übereinstimmt.

Beispiel 23.34.

(a) In $\mathbb{R}$ (mit der gewöhnlichen Metrik) sind offene Intervalle (a,b) offen im Sinne von Definition 23.33: Um jeden Punkt x eines solchen Intervalls finden wir eine ε-Umgebung (nämlich für ein beliebiges $\varepsilon \leq \min(x-a, b-x)$), die ganz in (a,b) liegt. Ebenso sind die uneigentlichen Intervalle $(-\infty,b)$ und (a,∞) offen – in der Literatur werden sie daher im Gegensatz zu unserer Konvention in Notation 4.17 (a) manchmal auch zu den offenen Intervallen gezählt. Vereinigungen solcher Intervalle wie z. B. $(0,1)\cup(2,\infty)$ sind aus dem gleichen Grund ebenfalls offen. Dagegen sind die Intervalle $[a,b)$, $(a,b]$ und $[a,b]$ aufgrund der enthaltenen Randpunkte nicht offen, da es hier um die Punkte $x=a$ bzw. $x=b$, die innerhalb des Intervalls liegen, keine solche ε-Umgebung innerhalb des Intervalls mehr gibt.



(a,b) ist offen



$[a,b)$ ist nicht offen

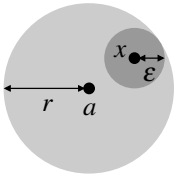
(b) Wiederum in $\mathbb{R}$ sind abgeschlossene Intervalle $[a,b]$ abgeschlossen, da ihre Komplemente $\mathbb{R}\backslash[a,b] = (-\infty,a)\cup(b,\infty)$ nach (a) offen sind. Ebenso sind die uneigentlichen Intervalle $(-\infty,b]$ und $[a,\infty)$ abgeschlossen, nicht aber die anderen Intervalltypen wie z. B. (a,b), $(a,b]$ oder (a,∞). Wie bei den offenen Intervallen werden die uneigentlichen Intervalle $(-\infty,b]$ und $[a,\infty)$ in manchen Büchern daher auch zu den abgeschlossenen Intervallen gezählt – für die beschränkten abgeschlossenen Intervalle $[a,b]$ ist dann der Name „*kompaktes Intervall*“ üblich (siehe Beispiel 23.52).

Insbesondere sehen wir an diesem Beispiel schon, dass „abgeschlossen“ nicht das Gegenteil von „offen“ ist: Das Intervall $[a,b)$ ist z. B. weder offen noch abgeschlossen, da es einen
57 Randpunkt a enthält und den anderen b nicht.

(c) In jedem metrischen Raum M sind die leere Menge $\emptyset$ und der ganze Raum M trivialerweise offen, und damit gleichzeitig auch abgeschlossen.

(d) In jedem metrischen Raum sind die offenen Kugeln $U_r(a)$ aus Definition 23.13 (a) offen: Ist $x \in U_r(a)$ beliebig, also $d(x,a) < r$, so ist wie im Bild rechts $U_\varepsilon(x) \subset U_r(a)$ mit $\varepsilon := r - d(x,a)$, denn für alle $y \in U_\varepsilon(x)$ gilt nach der Dreiecksungleichung

$$d(y,a) \leq d(y,x) + d(x,a) < \varepsilon + d(x,a) = r.$$

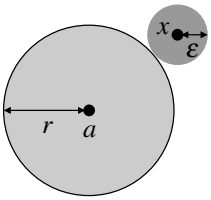


$U_r(a)$

Also ist $U_r(a)$ offen. Analog sieht man, dass jede abgeschlossene Kugel $K_r(a)$ abgeschlossen ist: Ist $x \in M\backslash K_r(a)$, also $d(x,a) > r$, so ist $U_\varepsilon(x) \subset M\backslash K_r(a)$ mit $\varepsilon := d(x,a) - r > 0$, denn für alle $y \in U_\varepsilon(x)$ gilt nun

$$d(y,a) \geq d(x,a) - d(x,y) > d(x,a) - \varepsilon = r,$$

d. h. $K_r(a)$ ist abgeschlossen. Insbesondere sind also einpunktige Mengen $\{a\} = K_0(a)$ als abgeschlossene Kugeln vom Radius 0 stets abgeschlossen.



$K_r(a)$

(e) Da in jeder ε-Umgebung eines beliebigen Punktes von $\mathbb{R}$ sowohl rationale als auch irrationale Zahlen liegen, ist weder die Menge der rationalen noch die der irrationalen Zahlen offen in $\mathbb{R}$. Mit anderen Worten ist $\mathbb{Q}$ in $\mathbb{R}$ weder offen noch abgeschlossen.

Die wichtigsten Eigenschaften offener und abgeschlossener Mengen sind die folgenden:

Lemma 23.35 (Durchschnitte und Vereinigungen offener und abgeschlossener Mengen)**.** *In jedem metrischen Raum gilt:*

(a) *Durchschnitte endlich vieler offener Mengen sind offen.*

(b) *Vereinigungen beliebig vieler (also auch unendlich vieler) offener Mengen sind offen.*

(c) *Vereinigungen endlich vieler abgeschlossener Mengen sind abgeschlossen.*

(d) *Durchschnitte beliebig vieler abgeschlossener Mengen sind abgeschlossen.*

Beweis. Es sei M ein metrischer Raum.

(a) Es seien $U_1, \dots, U_n \subset M$ offen und $a \in U_1 \cap \dots \cap U_n$. Dann ist $a \in U_i$ für alle $i = 1, \dots, n$. Da U_i offen ist, gibt es zu jedem i ein $\varepsilon_i > 0$ mit $U_{\varepsilon_i}(a) \subset U_i$. Mit $\varepsilon := \min(\varepsilon_1, \dots, \varepsilon_n)$ ist dann also $U_\varepsilon(a) \subset U_1 \cap \dots \cap U_n$ eine Umgebung von a, die ganz in $U_1 \cap \dots \cap U_n$ liegt. Also ist $U_1 \cap \dots \cap U_n$ offen.

(b) Es seien I eine beliebige Indexmenge und $U_i \subset M$ für alle $i \in I$ offen. Ist nun $a \in \bigcup_{i \in I} U_i$, so ist also $a \in U_j$ für ein $j \in I$. Da U_j offen ist, gibt es ein $\varepsilon > 0$ mit $U_\varepsilon(a) \subset U_j \subset \bigcup_{i \in I} U_i$. Also ist $\bigcup_{i \in I} U_i$ offen.

(c) Dies folgt nun durch Komplementbildung aus (a): sind $A_1, \dots, A_n \subset M$ abgeschlossen, also $M \backslash A_1, \dots, M \backslash A_n$ offen, so ist nach (a) auch

$$(M \backslash A_1) \cap \dots \cap (M \backslash A_n) = M \backslash (A_1 \cup \dots \cup A_n)$$

offen, und $A_1 \cup \dots \cup A_n$ damit abgeschlossen.

(d) ergibt sich analog aus (b): Sind $A_i \subset M$ abgeschlossen für alle i in einer Indexmenge I, also $M \backslash A_i$ offen, so ist nach (b) auch

$$\bigcup_{i \in I} (M \backslash A_i) = M \backslash \bigcap_{i \in I} A_i$$

offen, d. h. $\bigcap_{i \in I} A_i$ ist abgeschlossen. □

Bemerkung 23.36. Die Beschränkung auf endlich viele Mengen in Lemma 23.35 (a) und (c) ist wesentlich: Nach Beispiel 23.34 (d) ist in einem metrischen Raum M ja jede einpunktige Menge $\{a\}$ mit $a \in M$ abgeschlossen. Wären nun beliebige Vereinigungen abgeschlossener Mengen wieder abgeschlossen, so müsste dann ja jede Teilmenge $A \subset M$ (die ja immer Vereinigung aller ihrer Punkte ist) abgeschlossen sein – was im Allgemeinen offensichtlich falsch ist.

Aufgabe 23.37.

(a) Es sei $M = \{0\} \cup \{\frac{1}{n} : n \in \mathbb{N}_{>0}\} \subset \mathbb{R}$, aufgefasst als metrischer Raum mit der euklidischen Metrik. Gib alle offenen und alle abgeschlossenen Teilmengen dieses metrischen Raumes M an.

(b) Es sei $X \subset M$ eine Teilmenge eines metrischen Raumes M. Nach Beispiel 23.12 (a) ist dann auch X ein metrischer Raum mit der eingeschränkten Metrik.

Zeige, dass eine Menge $U \subset X$ genau dann offen in diesem metrischen Raum X ist, wenn es eine im metrischen Raum M offene Teilmenge $V \subset M$ gibt mit $U = V \cap X$.

Zum besseren Verständnis offener und abgeschlossener Mengen wollen wir jetzt noch das Konzept von Randpunkten exakt einführen, das wir oben ja schon zur Veranschaulichung dieser Begriffe verwendet haben. Auch andere verwandte Notationen haben wir bereits in Spezialfällen kennengelernt, wie z. B. dichte Teilmengen in Folgerung 4.34, den Abschluss einer Menge in Definition 8.1 und isolierte Punkte in Definition 10.1.

Definition 23.38 (Inneres, Abschluss und Rand)**.** Es seien M ein metrischer Raum, $a \in M$ ein Punkt und $X \subset M$ eine Teilmenge von M.

(a) Der Punkt a heißt **innerer Punkt** von X, wenn es eine offene Menge U gibt mit $a \in U \subset X$. In diesem Fall nennt man X auch eine **Umgebung** von a. Die Menge aller inneren Punkte von X wird mit $\mathring{X}$ bezeichnet und das **Innere** von X genannt.

(b) Der Punkt a heißt **Berührpunkt** von X, wenn jede offene Menge U mit $a \in U$ einen Punkt aus X enthält. Die Menge aller Berührpunkte von X wird mit $\overline{X}$ bezeichnet und der **Abschluss** von X genannt.

Ist $\overline{X} = M$, so heißt X **dicht** in M.

(c) Der Punkt a heißt **Randpunkt** von X, wenn jede offene Menge U mit $a \in U$ sowohl einen Punkt aus X als auch einen aus dem Komplement $M \backslash X$ enthält. Die Menge aller Randpunkte von X wird mit ∂X bezeichnet und der **Rand** von X genannt.

(d) Ein Punkt $a \in X$ heißt **Häufungspunkt** von X, wenn es in jeder offenen Menge U mit $a \in U$ einen Punkt in $X \backslash \{a\}$ gibt. Man nennt $a \in X$ einen **isolierten Punkt** von X, wenn es eine offene Menge U mit $a \in U$ gibt, die keine weiteren Punkte aus X enthält.

Beispiel 23.39. Wir betrachten den metrischen Raum $M = \mathbb{R}$ (mit der gewöhnlichen Metrik).

(a) Für ein halboffenes Intervall $X = [a,b)$ mit $a < b$ ist wie erwartet das Innere $\overset{\circ}{X} = (a,b)$, der Abschluss $\overline{X} = [a,b]$ und der Rand $\partial X = \{a,b\}$. Jeder Punkt in $[a,b]$ ist ein Häufungspunkt von X; es gibt keine isolierten Punkte in X.

(b) Für die Menge $X = \mathbb{Z}$ der ganzen Zahlen gilt $\overset{\circ}{\mathbb{Z}} = \emptyset$ und $\overline{\mathbb{Z}} = \mathbb{Z}$. Der Rand ist $\partial \mathbb{Z} = \mathbb{Z}$, und jeder dieser Randpunkte ist auch ein isolierter Punkt von $\mathbb{Z}$. Es gibt keine Häufungspunkte von $\mathbb{Z}$.

(c) Mit der gleichen Begründung wie in Beispiel 23.34 (e) ist $\overset{\circ}{\mathbb{Q}} = \emptyset$ und $\overline{\mathbb{Q}} = \mathbb{R}$, sowie $\partial \mathbb{Q} = \mathbb{R}$. Also liegt $\mathbb{Q}$ dicht in $\mathbb{R}$ (wie wir auch schon in Folgerung 4.34 gesagt hatten), und jede reelle Zahl ist ein Häufungspunkt von $\mathbb{Q}$.

Die gerade definierten Mengen $\overset{\circ}{X}$, $\overline{X}$ und ∂X erfüllen einige einfache anschauliche Eigenschaften, die mehr oder weniger direkt aus den Definitionen folgen und oft zu ihrer Berechnung benutzt werden können:

Lemma 23.40. *Für jede Teilmenge X eines metrischen Raumes M gilt:*

(a) *$\overset{\circ}{X}$ ist die Vereinigung aller offenen Mengen, die in X enthalten sind, also*

$$\overset{\circ}{X} = \bigcup_{U \subset X \text{ offen}} U.$$

Insbesondere ist $\overset{\circ}{X}$ also die größte offene Menge, die in X enthalten ist; und X ist genau dann offen, wenn $\overset{\circ}{X} = X$.

(b) *$\overline{X}$ ist der Durchschnitt aller abgeschlossenen Mengen, die X enthalten, also*

$$\overline{X} = \bigcap_{A \supset X \text{ abgeschlossen}} A.$$

Insbesondere ist $\overline{X}$ also die kleinste abgeschlossene Menge, die X enthält; und X ist genau dann abgeschlossen, wenn $\overline{X} = X$.

(c) $\overset{\circ}{X} = X \backslash \partial X$.

(d) $\overline{X} = X \cup \partial X$.

(e) $\partial X = \overline{X} \backslash \overset{\circ}{X}$. *Insbesondere ist ∂X abgeschlossen.*

Beweis.

(a) Nach Definition 23.38 (a) folgt für alle $a \in M$ sofort

$$a \in \overset{\circ}{X} \quad \Leftrightarrow \quad \text{es gibt ein } U \text{ offen mit } a \in U \subset X \quad \Leftrightarrow \quad a \in \bigcup_{U \subset X \text{ offen}} U.$$

Insbesondere ist $\overset{\circ}{X}$ als Vereinigung offener Mengen nach Lemma 23.35 (b) also offen, und damit als Vereinigung *aller* offenen Teilmengen von X auch die größte offene Teilmenge von X. Ist weiterhin X offen, so ist X selbst eine der Mengen, über die wir hier die Vereinigung bilden, und damit ist die Vereinigung $\overset{\circ}{X}$ dieser Mengen dann gleich X.

(b) zeigt man analog zu (a) durch Übergang zum Komplement: Für alle $a \in M$ gilt nach Definition 23.38 (b)

$$\begin{aligned} a \notin \overline{X} \;&\Leftrightarrow\; \text{es gibt ein } U \text{ offen mit } a \in U \text{ und } U \subset M \backslash X \\ &\Leftrightarrow\; a \in \bigcup_{U \subset M\backslash X \text{ offen}} U \\ &\Leftrightarrow\; a \notin \bigcap_{U \subset M\backslash X \text{ offen}} (M \backslash U) \\ &\Leftrightarrow\; a \notin \bigcap_{A \supset X \text{ abgeschlossen}} A \qquad\qquad (\text{mit } A = M \backslash U). \end{aligned}$$

Damit gilt die in (b) behauptete Formel, und $\overline{X}$ ist nach Lemma 23.35 (d) als Durchschnitt abgeschlossener Mengen abgeschlossen. Ist weiterhin X selbst abgeschlossen, so ist X eine der Mengen, über die hier der Durchschnitt gebildet wird, und somit ist dieser Durchschnitt (also $\overline{X}$) gleich X.

(c) Die Bedingung $a \notin \partial X$ bedeutet genau, dass es eine offene Menge U mit $a \in U$ gibt, für die entweder $U \subset X$ oder $U \subset M \backslash X$ ist. Falls $a \in X$ gilt, kann hierbei wegen $a \in U$ natürlich nur die Alternative $U \subset X$ auftreten. Also ist $a \in X \backslash \partial X$ genau dann, wenn es eine offene Menge U mit $a \in U \subset X$ gibt, also wenn $a \in \overset{\circ}{X}$.

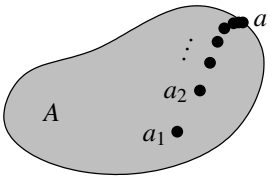
(d) „$\subset$“: Ist $a \in \overline{X}$ und $a \notin X$, so enthält jede offene Menge U mit $a \in U$ nicht nur einen Punkt aus X, sondern auch den Punkt $a \in M \backslash X$. Also ist dann $a \in \partial X$.

Die Inklusionen $X \subset \overline{X}$ und $\partial X \subset \overline{X}$ für die Rückrichtung „$\supset$“ folgen sofort aus den Definitionen.

(e) Die Gleichung folgt sofort aus (c) und (d); der Rand ∂X ist dann als Durchschnitt der nach (a) und (b) abgeschlossenen Mengen $\overline{X}$ und $M \backslash \overset{\circ}{X}$ abgeschlossen. □

Bemerkung 23.41. Insbesondere besagen die Teile (a) und (c) des Lemmas, dass eine Menge X genau dann offen ist (also $\overset{\circ}{X} = X$ gilt), wenn sie keinen ihrer Randpunkt enthält – was am Anfang dieses Abschnitts ja gerade unsere anschauliche Vorstellung von offenen Mengen war. Analog besagen (b) und (d), dass die Menge X genau dann abgeschlossen ist (also $\overline{X} = X$ gilt), wenn sie alle ihre Randpunkte enthält.

Diese Charakterisierung abgeschlossener Mengen durch ihre Randpunkte lässt sich auch gut mit Hilfe von Grenzwerten von Folgen interpretieren. Haben wir nämlich eine konvergente Folge $(a_n)_n$, deren Glieder in einer Teilmenge A eines metrischen Raumes liegen, so kann der Grenzwert dieser Folge anschaulich nur in A oder wie im Bild rechts auf dem Rand von A, also letztlich in $\overline{A}$ liegen. Damit sollte A also genau dann abgeschlossen sein, wenn dieser Grenzwert in jedem Fall wieder in A liegt. Dies besagt der folgende Satz.



Satz 23.42 (Folgenkriterium für Abgeschlossenheit)**.** *Eine Teilmenge A eines metrischen Raumes M ist genau dann abgeschlossen, wenn zu jeder konvergenten Folge $(a_n)_n$ mit $a_n \in A$ für (fast) alle n ihr Grenzwert $\lim_{n\to\infty} a_n$ ebenfalls in A liegt. (Man sagt in diesem Fall auch, dass A „abgeschlossen unter Grenzwertbildung“ ist.)*

Beweis.

„$\Rightarrow$“ Es seien $A \subset M$ abgeschlossen und $(a_n)_n$ eine konvergente Folge, deren Glieder fast alle in A liegen. Angenommen, der Grenzwert $a = \lim_{n\to\infty} a_n$ läge in der offenen Menge $M \backslash A$. Dann wäre $M \backslash A$ nach Definition 23.33 (a) eine Umgebung von a, und damit müssten nach Bemerkung 23.14 fast alle a_n in $M \backslash A$ liegen – im Widerspruch dazu, dass bereits fast alle a_n in A liegen. Also war unsere Annahme falsch, und es gilt $a \in A$.

„$\Leftarrow$“ Die Menge A sei nun abgeschlossen unter Grenzwertbildung. Angenommen, A wäre nicht abgeschlossen, also $M \backslash A$ nicht offen. Dann gäbe es einen Punkt $a \in M \backslash A$, um den keine ε-Umgebung vollständig in $M \backslash A$ liegt. Wir können also für alle $n \in \mathbb{N}_{>0}$ einen Punkt a_n in der $\frac{1}{n}$-Umgebung von a wählen, der in A liegt. Wegen

$$d(a_n, a) < \frac{1}{n} \to 0 \qquad \text{für } n \to \infty$$

ist $(a_n)_n$ dann im Widerspruch zur Annahme aber nach Bemerkung 23.14 eine konvergente Folge in A mit Grenzwert $a \notin A$. Also ist A abgeschlossen. □

Folgerung 23.43 (Vollständigkeit von Teilmengen)**.** *Eine Teilmenge eines vollständigen metrischen Raumes ist genau dann selbst wieder vollständig, wenn sie abgeschlossen ist.*

Beweis. Es sei A eine Teilmenge eines vollständigen metrischen Raumes M. Nach Definition 23.28 (a) ist A genau dann vollständig, wenn jede Cauchyfolge in A einen Grenzwert in A hat. Da M vollständig ist, konvergiert jede solche Cauchyfolge aber in jedem Fall mit einem Grenzwert in M. Also ist A genau dann vollständig, wenn der Grenzwert jeder konvergenten Folge mit Folgengliedern in A ebenfalls in A liegt, nach Satz 23.42 also genau dann, wenn A abgeschlossen ist. □

Aufgabe 23.44. Man zeige:

(a) Die Menge $U = \{x \in \mathbb{R}^2 : x_2 > x_1 + 1\}$ ist offen in $\mathbb{R}^2$.

(b) Sind $A_1, \dots, A_n \subset \mathbb{R}$ abgeschlossen, so ist $A_1 \times A_2 \times \cdots \times A_n$ abgeschlossen in $\mathbb{R}^n$.

(c) In $\mathrm{Mat}(n \times n, \mathbb{R})$ ist die Menge aller diagonalisierbaren Matrizen für $n \geq 2$ weder offen noch abgeschlossen.

Aufgabe 23.45. Es seien A und B zwei Teilmengen eines metrischen Raumes. Man zeige:

(a) Ist $A \subset B$, so auch $\mathring{A} \subset \mathring{B}$ und $\overline{A} \subset \overline{B}$.

(b) Es ist stets $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$. Gilt hier im Allgemeinen auch die Gleichheit?

Aufgabe 23.46. In einem metrischen Raum M betrachten wir zu einem Punkt $a \in M$ und einem Radius $r \in \mathbb{R}_{>0}$ die offene Kugel $U_r(a) = \{x \in M : d(x,a) < r\}$. Man zeige:

(a) Für den Rand dieser Kugel gilt $\partial U_r(a) \subset \{x \in M : d(x,a) = r\}$.

(b) In einem normierten Raum gilt in (a) wie erwartet sogar die Gleichheit, in einem beliebigen metrischen Raum jedoch im Allgemeinen nicht.

Aufgabe 23.47 (Häufungspunkte von Folgen und Mengen)**.** Es seien M ein metrischer Raum, $a \in M$ und $X \subset M$. Zeige, dass dann die folgenden Aussagen äquivalent sind:

(a) a ist ein Häufungspunkt von X.

(b) Es gibt eine Folge in X, die nirgends den Wert a annimmt, und die a als Häufungspunkt hat.

(c) Es gibt eine Folge in X, die nirgends den Wert a annimmt, und die gegen a konvergiert.

(d) $a \in \overline{X \backslash \{a\}}$.

Die Äquivalenz „(a) $\Leftrightarrow$ (b)“ stellt also insbesondere eine Beziehung zwischen den Begriffen des Häufungspunktes einer Folge (Definition 5.17 (c)) und einer Menge (Definition 23.38 (d)) her.

Aufgabe 23.48 (Ein topologisches Spiel)**.** Wir starten mit einer gegebenen Teilmenge X eines metrischen Raumes M. Ausgehend von dieser einen Teilmenge dürfen wir jetzt neue Teilmengen von M bilden, indem wir von X oder bereits vorher konstruierten Mengen entweder den Abschluss oder das Innere bilden. Starten wir z. B. im metrischen Raum $M = \mathbb{R}$ mit der Teilmenge $X = \{0\} \cup (1,2]$, so ist deren Inneres gleich $(1,2)$, und davon wieder der Abschluss gleich $[1,2]$. Außerdem ist der Abschluss von X gleich $\{0\} \cup [1,2]$. Man überprüft schnell, dass die Abschluss- oder Innerenbildung bei keiner dieser Mengen zu neuen Mengen führt. Ausgehend von der einen gewählten Menge X konnten wir hier also insgesamt vier verschiedene Mengen erzeugen.

Wie viele verschiedene Mengen kann man so bei geschickter Wahl von M und X maximal erzeugen?

Hinweis: Zeige zuerst, dass

(a) $AIAX \subset AX$ und $IAIX \supset IX$,

(b) $AIAIX = AIX$ und $IAIAX = IAX$,

wobei wir zur Übersichtlichkeit der Notation AX für den Abschluss und IX für das Innere von X geschrieben haben. Die gesuchte Maximalzahl von Mengen lässt sich bereits in $M = \mathbb{R}$ (mit der gewöhnlichen Metrik) für ein geeignetes $X \subset \mathbb{R}$ erreichen.

Falls ihr noch weiter über diese Aufgabe nachdenken wollt: Wie viele Mengen kann man maximal aus einer gegebenen Menge X durch Abschluss- und Komplementbildung erzeugen?

58

23.D Kompaktheit

In Kapitel 8 haben wir einige wichtige Eigenschaften reeller stetiger Funktionen kennengelernt, die auf einem beschränkten abgeschlossenen Intervall $[a,b] \subset \mathbb{R}$ definiert sind: Sie sind z. B. nach Satz 8.23 beschränkt, nehmen nach Satz 8.25 sogar ein Maximum und Minimum an, und sind nach Satz 8.36 auch gleichmäßig stetig. Alle diese Aussagen wären falsch, wenn wir den Definitionsbereich nur als beschränkt oder nur als abgeschlossen voraussetzen würden – wie die Beispiele der Funktionen $(0,1) \to \mathbb{R}$, $x \mapsto \frac{1}{x}$ bzw. $\mathbb{R}_{\geq 1} \to \mathbb{R}$, $x \mapsto x^2$ zeigen, die keine der drei genannten Eigenschaften erfüllen.

Für eine Teilmenge D von $\mathbb{R}$ ist diese Kombination von Beschränktheit und Abgeschlossenheit in der Praxis also sehr wichtig. Der Grund dafür ist aus den Beweisen der obigen Sätze ersichtlich: Sie alle benötigen zu einer gegebenen Folge in D die Existenz einer konvergenten Teilfolge mit Grenzwert in D. Ist nun D beschränkt, so existiert nach dem Satz 6.21 von Bolzano-Weierstraß zunächst einmal eine konvergente Teilfolge, und ist D abgeschlossen, so liegt der Grenzwert dieser Teilfolge nach Satz 23.42 dann auch in D.

Für die Verallgemeinerung dieser Aussagen auf normierte bzw. metrische Räume müssen wir daher die Existenz konvergenter Teilfolgen untersuchen. Wir übertragen dazu zunächst den Satz von Bolzano-Weierstraß ins Mehrdimensionale.

Satz 23.49 (**Satz von Bolzano-Weierstraß**)**.** *Jede beschränkte Folge in $\mathbb{K}^n$ besitzt eine konvergente Teilfolge.*

Beweis. Nach Bemerkung 23.18 können wir die Maximumsnorm verwenden. Ist dann eine Folge $(a^{(k)})_k$ beschränkt, gilt also $\|a^{(k)}\|_\infty \leq r$ für ein $r \in \mathbb{R}_{>0}$ und alle $k \in \mathbb{N}$, so ist damit auch $|a_i^{(k)}| \leq r$ für alle $i = 1, \ldots, n$ und $k \in \mathbb{N}$, d. h. es sind auch alle Koordinatenfolgen $(a_i^{(k)})_k$ beschränkt.

Nach dem Satz 6.21 von Bolzano-Weierstraß in $\mathbb{K}$ können wir also nach evtl. Auswählen einer Teilfolge annehmen, dass die erste Koordinatenfolge $(a_1^{(k)})_k$ von $(a^{(k)})_k$ konvergiert. Aus dieser Teilfolge wählen wir nun eine weitere Teilfolge aus, so dass auch die zweite Koordinatenfolge konvergiert. Setzen wir dieses Verfahren fort, so haben wir nach n Schritten eine Teilfolge von $(a^{(k)})_k$ gefunden, von der jede Koordinatenfolge konvergiert und die nach Lemma 23.19 damit in $\mathbb{K}^n$ konvergent ist. □

Mit dieser Vorarbeit können wir jetzt den zentralen Begriff dieses Abschnitts definieren und untersuchen.

Definition 23.50 (Folgenkompaktheit)**.** Eine Teilmenge A eines metrischen Raumes M heißt **kompakt** bzw. **folgenkompakt**, wenn jede Folge in A eine konvergente Teilfolge mit Grenzwert in A hat.

Satz 23.51. *Es sei A eine Teilmenge eines metrischen Raumes M.*

(a) *Ist A kompakt, so ist A beschränkt und abgeschlossen.*

(b) *Im Fall $M = \mathbb{K}^n$ gilt auch die Umkehrung, d. h. dann ist A genau dann kompakt, wenn A beschränkt und abgeschlossen ist.*

Beweis.

(a) Wäre A nicht beschränkt, dann gäbe es ein $a \in A$ und eine Folge $(a_k)_k$ in A mit $d(a_k,a) \geq k$ für alle k. Dann ist aber jede Teilfolge von $(a_k)_k$ unbeschränkt, also divergent nach Lemma 23.22. Damit kann A nicht kompakt sein.

Wäre A hingegen nicht abgeschlossen, so gäbe es nach Satz 23.42 eine konvergente Folge $(a_k)_k$ in A, deren Grenzwert a in $M \backslash A$ liegt. Dann konvergiert aber auch jede Teilfolge von $(a_k)_k$ gegen a, und damit also nicht gegen einen Punkt in A. Auch hier kann A also nicht kompakt sein.

(b) Die Menge $A \subset \mathbb{K}^n$ sei beschränkt und abgeschlossen. Jede Folge in A ist also zunächst beschränkt und besitzt damit nach dem Satz 23.49 von Bolzano-Weierstraß eine konvergente Teilfolge in $\mathbb{K}^n$. Der Grenzwert dieser Teilfolge liegt nun wegen der Abgeschlossenheit von A nach Satz 23.42 ebenfalls in A. Also ist A kompakt. □

Beispiel 23.52. Von den Intervallen in $\mathbb{R}$ wie in Notation 4.17 (a) sind nach Satz 23.51 (b) genau die Intervalle der Form $[a,b]$ kompakt.

Wie bereits angekündigt werden wir später für stetige Funktionen auf kompakten Mengen ähnliche Eigenschaften zeigen, wie wir sie in Kapitel 8 für stetige Funktionen auf beschränkten abgeschlossenen Intervallen in $\mathbb{R}$ bewiesen haben (siehe Abschnitt 24.B), z. B. dass solche Funktionen ein Maximum und Minimum annehmen. Das folgende Lemma ist eine kleine Vorbereitung dafür.

Lemma 23.53. *Jede kompakte, nicht-leere Menge in $\mathbb{R}$ besitzt ein Maximum und Minimum.*

Beweis. Aus Symmetriegründen genügt es offensichtlich, den Fall des Maximums zu betrachten. Ist $A \subset \mathbb{R}$ kompakt, so ist A nach Satz 23.51 (a) beschränkt und besitzt damit wegen $A \neq \emptyset$ nach dem Supremumsaxiom (siehe Definition 4.28) ein Supremum $s := \sup A$. Angenommen, es wäre $s \notin A$. Da A kompakt und damit nach Satz 23.51 (a) abgeschlossen ist, gäbe es dann ein $\varepsilon > 0$ mit $(s-\varepsilon, s+\varepsilon) \subset \mathbb{R} \backslash A$, d. h. $(s-\varepsilon, s+\varepsilon) \cap A = \emptyset$. Dann wäre aber nicht nur s, sondern auch $s-\varepsilon$ eine obere Schranke für A – im Widerspruch zu $s = \sup A$.

Also war unsere Annahme falsch, und es ist $s \in A$, also $s = \max A$. □

Zum Abschluss dieses Kapitels wollen wir nun noch eine oft benötigte alternative Charakterisierung der Kompaktheit untersuchen. Da sie sich sehr von unserer ursprünglichen Definition 23.50 unterscheidet, geben wir ihr zunächst einen anderen Namen – wir werden aber in Satz 23.58 sehen, dass es sich zumindest für Teilmengen von $\mathbb{K}^n$ um eine zur Folgenkompaktheit äquivalente Eigenschaft handelt.

Definition 23.54 (Überdeckungskompaktheit)**.** Eine Teilmenge A eines metrischen Raumes M heißt **überdeckungskompakt**, wenn gilt: Sind I eine beliebige Indexmenge und $U_i \subset M$ offene Teilmengen mit $A \subset \bigcup_{i \in I} U_i$, so gibt es bereits endlich viele $i_1, \dots, i_n \in I$ mit $A \subset U_{i_1} \cup \dots \cup U_{i_n}$.

Bemerkung 23.55.

(a) Man sagt auch, dass eine solche Familie offener Mengen, die in ihrer Vereinigung die Menge A enthalten, eine *offene Überdeckung* von A ist. Dementsprechend kann man die Bedingung der Überdeckungskompaktheit auch so formulieren: Jede offene Überdeckung von A besitzt eine endliche Teilüberdeckung (also eine Überdeckung aus endlich vielen der gegebenen Mengen).

(b) Sowohl die Folgenkompaktheit aus Definition 23.50 als auch die Überdeckungskompaktheit aus Definition 23.54 sind offensichtlich topologische Eigenschaften im Sinne von Bemerkung 23.18.

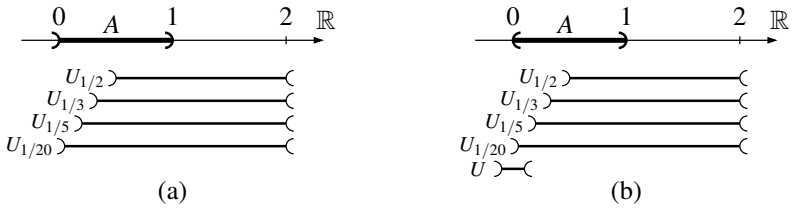
Beispiel 23.56. Es sei $M = \mathbb{R}$.

(a) Betrachten wir das halboffene Intervall $A = (0,1]$, so bilden die unendlich vielen Intervalle $U_\varepsilon = (\varepsilon, 2)$ für alle $0 < \varepsilon < 1$, von denen wir im Bild unten links exemplarisch vier eingezeichnet haben, eine offene Überdeckung von A: Alle U_ε sind offen, und ihre Vereinigung

umfasst die gesamte Menge A, da es zu jedem $x \in A = (0,1]$ noch ein ε gibt mit $0 < \varepsilon < x$, also mit $x \in U_\varepsilon$.

Treffen wir aus diesen offenen Mengen jedoch eine beliebige *endliche* Auswahl $U_{\varepsilon_1}, \dots, U_{\varepsilon_n}$, so überdecken diese endlich vielen Mengen sicher nicht mehr die ganze Menge A, denn es ist ja $U_{\varepsilon_1} \cup \dots \cup U_{\varepsilon_n} = U_\varepsilon$ mit $\varepsilon = \min(\varepsilon_1, \dots, \varepsilon_n)$, und diese Vereinigung enthält z. B. nicht mehr die Zahl $\frac{\varepsilon}{2} \in A$. (Im Bild unten links enthalten die vier dort ausgewählten Intervalle z. B. nicht mehr die Zahl $\frac{1}{40} \in A$).

Wir finden in diesem Fall also keine endliche Teilüberdeckung zu der gegebenen offenen Überdeckung von A. Damit ist A nicht überdeckungskompakt.



(b) Betrachten wir nun stattdessen das abgeschlossene Intervall $A = [0,1]$, so stellen wir zumindest fest, dass wir die Bedingung der Überdeckungskompaktheit dann nicht mehr so einfach widerlegen können: Die oben gewählten offenen Intervalle U_ε für $0 < \varepsilon < 1$ bilden nun keine Überdeckung von A mehr, da sie den Punkt $0 \in A$ nicht enthalten. Um daraus eine Überdeckung von A zu machen, müssten wir noch eine offene Menge U mit $0 \in U$ mit hinzu nehmen – z. B. $U = (-\frac{1}{10}, \frac{1}{10})$ wie im Bild oben rechts, so dass wir die offene Überdeckung $U \cup \bigcup_{0<\varepsilon<1} U_\varepsilon$ von A erhalten. In diesem Fall ist es aber einfach, daraus eine endliche Teilüberdeckung von A auszuwählen, nämlich z. B. $U \cup U_{1/20} = (-\frac{1}{10}, \frac{1}{10}) \cup (\frac{1}{20}, 2) \supset A$.

Beachte jedoch, dass wir mit diesem Argument noch nicht gezeigt haben, dass A überdeckungskompakt ist! Dazu hätten wir nämlich zu einer *beliebigen* offenen Überdeckung die Existenz einer endlichen Teilüberdeckung beweisen müssen – während wir eben ja nur eine ganz spezielle offene Überdeckung betrachtet haben. In der Tat ist A aber überdeckungskompakt, wie wir gleich in Satz 23.58 sehen werden.

Analog zu Satz 23.51 (a) zeigen wir zunächst:

Lemma 23.57. *Jede überdeckungskompakte Teilmenge eines metrischen Raumes ist beschränkt und abgeschlossen.*

Beweis. Es sei A eine überdeckungskompakte Teilmenge eines metrischen Raumes M.

A ist beschränkt: Da mit einem fest gewählten Punkt $a \in M$ die offenen Kugeln $U_r(a)$ für alle $r \in \mathbb{R}_{>0}$ offensichtlich den ganzen Raum M und damit auch A überdecken, können wir wegen der Überdeckungskompaktheit von A endlich viele $r_1, \dots, r_n \in \mathbb{R}_{>0}$ wählen mit

$$A \subset U_{r_1}(a) \cup \dots \cup U_{r_n}(a) = U_r(a),$$

wobei $r = \max(r_1, \dots, r_n)$. Nach Definition 23.20 (a) ist A also beschränkt.

A ist abgeschlossen, d. h. $M \backslash A$ ist offen: Es sei $a \in M \backslash A$. Diesmal betrachten wir die offenen Mengen

$$V_r(a) := M \backslash K_r(a) = \{x \in M : d(x,a) > r\}$$

für alle $r \in \mathbb{R}_{>0}$, die natürlich $M \backslash \{a\}$ und damit auch A überdecken. Da A überdeckungskompakt ist, gibt es also wieder endlich viele $r_1, \dots, r_n \in \mathbb{R}_{>0}$ mit

$$A \subset V_{r_1}(a) \cup \dots \cup V_{r_n}(a) = V_r(a),$$

wobei $r = \min(r_1, \dots, r_n)$. Durch Komplementbildung erhalten wir daraus $U_r(a) \subset K_r(a) \subset M \backslash A$. Also ist $M \backslash A$ offen. □

Satz 23.58 (Satz von Heine-Borel). *Eine Teilmenge von $\mathbb{K}^n$ ist genau dann überdeckungskompakt, wenn sie beschränkt und abgeschlossen ist.*

Beweis. Wir können wieder die zur euklidischen Norm äquivalente Maximumsnorm auf $\mathbb{K}^n$ verwenden. Außerdem können wir ohne Einschränkung $\mathbb{K} = \mathbb{R}$ annehmen, da wir im Fall $\mathbb{K} = \mathbb{C}$ den Raum $\mathbb{C}^n$ einfach als $\mathbb{R}^{2n}$ auffassen können.

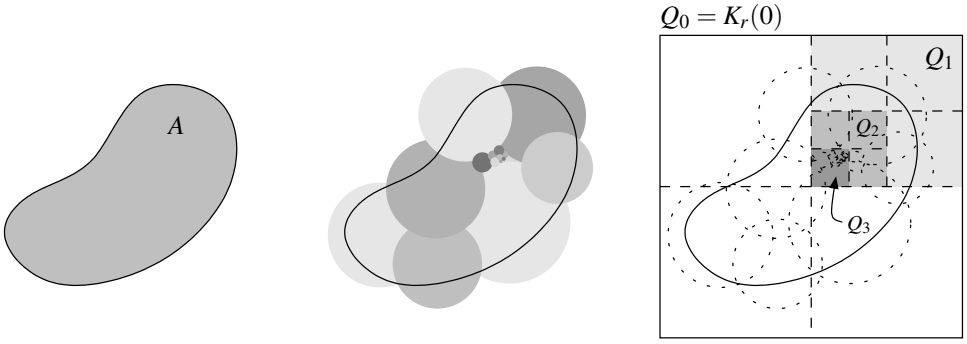
Wir haben in Lemma 23.57 bereits gesehen, dass eine überdeckungskompakte Teilmenge von $\mathbb{R}^n$ beschränkt und abgeschlossen sein muss. Es sei nun also umgekehrt $A \subset \mathbb{R}^n$ beschränkt und abgeschlossen. Wir zeigen mit einem Widerspruchsbeweis, dass A dann überdeckungskompakt ist.

Es sei also $\bigcup_{i \in I} U_i$ eine offene Überdeckung von A ohne endliche Teilüberdeckung. Im Bild unten in der Mitte ist dies dadurch angedeutet, dass die offenen Mengen der Überdeckung an einer Stelle sehr klein werden und damit sehr viele (bzw. unendlich viele) von ihnen benötigt werden, um A dort zu überdecken.

Als beschränkte Menge ist A in einer abgeschlossenen Kugel (in der Maximumsnorm), also in einem Würfel

$$K_r(0) = \{x \in \mathbb{R}^n : \|x\|_\infty \leq r\} = [-r, r]^n$$

enthalten. Wir konstruieren nun rekursiv wie im Bild unten rechts eine Folge $(Q_k)_{k \in \mathbb{N}}$ von ineinander liegenden Würfeln mit Kantenlängen $2r \cdot 2^{-k}$, so dass $A \cap Q_k$ nicht durch endlich viele U_i überdeckt werden kann:

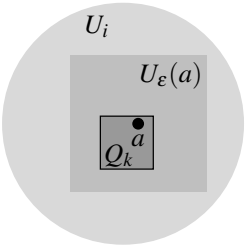


Die Menge A, ... eine offene Überdeckung von A ... und die Würfel Q_k

- Für $k = 0$ setzen wir $Q_0 := K_r(0)$; nach Voraussetzung kann $A \cap Q_0 = A$ nicht durch endlich viele U_i überdeckt werden.
- Ist Q_k für ein $k \in \mathbb{N}$ bereits konstruiert, teilen wir diesen Würfel an den Seitenmitten wie im Bild dargestellt in 2^n Teilwürfel mit jeweils halber Kantenlänge auf. Von diesen gibt es nun mindestens einen Teilwürfel, den wir dann Q_{k+1} nennen, für den der Schnitt mit A nicht durch endlich viele U_i überdeckt werden kann – denn ansonsten könnte ja auch $A \cap Q_k$ durch endlich viele U_i überdeckt werden.

Anschaulich sollten die Würfel Q_k jetzt gegen einen Punkt a konvergieren, in dessen Nähe A nicht durch endlich viele U_i überdeckt werden kann. Um diese Aussage exakt zu machen, wählen wir ein $a_k \in A \cap Q_k$ für alle $k \in \mathbb{N}$, so dass also auch $a_k \in A \cap Q_l$ für alle $k, l \in \mathbb{N}$ mit $k \geq l$ gilt. Dann ist $(a_k)_k$ eine Cauchyfolge, denn zu gegebenem $\varepsilon > 0$ können wir ein $k_0 \in \mathbb{N}$ wählen mit $2r \cdot 2^{-k_0} < \varepsilon$, und damit gilt dann $a_k, a_l \in Q_{k_0}$, also $\|a_k - a_l\|_\infty \leq 2r \cdot 2^{-k_0} < \varepsilon$ für alle $k, l \geq k_0$. Also existiert nach Satz 23.29 der Grenzwert $a := \lim_{k \to \infty} a_k$. Dieser muss nach Satz 23.42 sowohl in A als auch in allen Q_k liegen, da alle diese Mengen abgeschlossen sind und nach Konstruktion fast alle Folgenglieder enthalten.

Als Punkt in A muss a damit in einer der gegebenen Mengen U_i enthalten sein. Da U_i offen ist, enthält U_i nun aber noch eine Kugel $U_\varepsilon(a)$ (in der Maximumsnorm). Wählen wir $k \in \mathbb{N}$ mit $2r \cdot 2^{-k} < \varepsilon$, so liegt der Würfel Q_k dann wie im Bild rechts vollständig in $U_\varepsilon(a)$, und damit auch in U_i. Also wird Q_k und damit erst recht auch $A \cap Q_k$ bereits durch die eine Menge U_i überdeckt – was ein Widerspruch zur Annahme ist, dass dies nicht einmal mit endlich vielen der gegebenen offenen Mengen möglich ist.



Dieser Widerspruch zeigt, dass A überdeckungskompakt sein muss. □

59

Bemerkung 23.59 (Folgen- und Überdeckungskompaktheit)**.** Nach Satz 23.51 (b) und 23.58 stimmen die Begriffe der Folgenkompaktheit und Überdeckungskompaktheit für Teilmengen von $\mathbb{K}^n$ also überein. Man kann zeigen, dass diese Äquivalenz sogar in beliebigen metrischen Räumen gilt [M, Satz 39.30] – für alle in dieser Vorlesung betrachteten Räume sind diese beiden Begriffe also völlig gleichwertig. In der Tat wird der Begriff der Kompaktheit in den meisten Büchern über die Überdeckungskompaktheit definiert. Wir haben uns hier für die Definition über die Folgenkompaktheit entschieden, da dies wohl das deutlich anschaulichere Konzept ist.

Die Äquivalenz der Kompaktheit zur Beschränktheit und Abgeschlossenheit ist dagegen zwar gemäß Satz 23.51 (b) bzw. 23.58 für Teilmengen von $\mathbb{K}^n$ richtig, im Allgemeinen jedoch falsch. In der Tat kann man zeigen, dass diese Aussage in *jedem* unendlich-dimensionalen normierten Raum falsch ist, da in diesen Fällen bereits die (abgeschlossene) Einheitskugel $K_1(0)$ zwar beschränkt und abgeschlossen, aber nicht kompakt ist. Das Studium von solchen unendlich-dimensionalen Räumen – hauptsächlich von Räumen reellwertiger Funktionen mit bestimmten Eigenschaften – ist der Inhalt der Vorlesung „Einführung in die Funktionalanalysis“ des zweiten Studienjahres. Wir haben in den letzten Kapiteln gesehen, dass in solchen allgemeinen Räumen zwar viele Sätze noch genauso gelten wie im endlich-dimensionalen Fall, andererseits aber an entscheidenden Stellen auch große Unterschiede bestehen (siehe z. B. Aufgabe 21.18, Beispiel 21.42 oder Beispiel 23.9).

Aufgabe 23.60. Es seien A und K zwei Teilmengen eines metrischen Raumes M. Man zeige:

(a) Ist A abgeschlossen und K folgenkompakt, so ist auch $K \cap A$ folgenkompakt.

(b) Ist A abgeschlossen und K überdeckungskompakt, so ist auch $K \cap A$ überdeckungskompakt.

Die in Bemerkung 23.59 erwähnte, aber nicht bewiesene Äquivalenz zwischen Folgen- und Überdeckungskompaktheit in allgemeinen metrischen Räumen soll hierbei natürlich nicht benutzt werden.

Aufgabe 23.61. Man zeige: Sind K und L zwei kompakte Teilmengen eines normierten Raumes, so ist auch deren Summe $K + L = \{a + b : a \in K, b \in L\}$ kompakt.

24. Stetigkeit in metrischen Räumen

Wie im eindimensionalen Fall kommen wir nach unserem Studium von Grenzwerten von Folgen im letzten Kapitel jetzt zur Stetigkeit, also zu Grenzwerten von Funktionen. Auch diese können wir wieder in allgemeinen metrischen Räumen betrachten.

24.A Stetige Abbildungen

Zur Definition stetiger Abbildungen erinnern wir uns noch einmal an die entsprechende Definition 8.3 in $\mathbb{K}$: Ist $f\colon D \to \mathbb{K}$ eine Funktion auf einer Teilmenge D von $\mathbb{K}$ und a ein Punkt im Abschluss $\overline{D}$ von D, so sagen wir, dass $f(x)$ für $x \to a$ gegen ein $c \in \mathbb{K}$ konvergiert, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0}\ \forall x \in D : |x-a| < \delta \Rightarrow |f(x)-c| < \varepsilon.$$

Wie im Fall von Grenzwerten von Folgen können wir dies unmittelbar auf metrische Räume übertragen, indem wir den Abstand zweier Punkte nun mit der Metrik messen:

Definition 24.1 (Grenzwerte von Funktionen und Stetigkeit)**.** Es seien M und N metrische Räume, $D \subset M$ eine beliebige Teilmenge und $f\colon D \to N$ eine Abbildung.

(a) Ist $a \in \overline{D}$, so heißt ein Punkt $c \in N$ **Grenzwert** von f in a, wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0}\ \forall x \in D : d(x,a) < \delta \Rightarrow d(f(x),c) < \varepsilon,$$

also mit anderen Worten wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0} : f(D \cap U_\delta(a)) \subset U_\varepsilon(c)$$

(beachte dabei, dass $d(x,a)$ die Metrik in M, $d(f(x),c)$ dagegen die in N ist). Wie im eindimensionalen Fall werden wir wieder in Bemerkung 24.5 (a) sehen, dass ein solcher Grenzwert eindeutig ist, falls er existiert, so dass wir dann von *dem* Grenzwert von f in a sprechen können. Wir schreiben dies dann als

$$\lim_{x \to a} f(x) = c \qquad \text{bzw.} \qquad \lim_{\substack{x \to a \\ x \in D}} f(x) = c$$

oder auch als „$f(x) \to c$ für $x \to a$“, und sagen, dass $f(x)$ für $x \to a$ gegen c **konvergiert**. Existiert ein solcher Grenzwert nicht, so heißt f **divergent** in a.

(b) Liegt der Punkt a sogar in D, so kommt als Grenzwert von f in a mit derselben Begründung wie in Bemerkung 8.4 nur $f(a)$ in Frage. Wenn $f(x)$ dann für $x \to a$ gegen $f(a)$ konvergiert, d. h. wenn

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0}\ \forall x \in D : d(x,a) < \delta \Rightarrow d(f(x),f(a)) < \varepsilon$$

bzw.

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists \delta \in \mathbb{R}_{>0} : f(D \cap U_\delta(a)) \subset U_\varepsilon(f(a))$$

gilt, so heißt f **stetig** in a. Liegt a nicht in D, so heißt f **stetig fortsetzbar** nach a, wenn der Grenzwert $c = \lim_{x \to a} f(x)$ existiert.

(c) Die Funktion f heißt **stetig**, wenn sie in jedem Punkt $a \in D$ stetig ist.

Die anschauliche Interpretation dieser Begriffe ist in allgemeinen metrischen Räumen natürlich noch dieselbe wie in $\mathbb{K}$: So ist eine Abbildung f z. B. stetig, wenn „kleine Änderungen in der Variablen x auch nur kleine Änderungen der Funktionswerte $f(x)$ zur Folge haben“ (siehe Bemerkung 8.6). Wir werden in diesem Kapitel sehen, dass auch die Sätze über stetige Funktionen sowie ihre Beweise noch sehr ähnlich zu denen aus Kapitel 8 sind.

Bemerkung 24.2. Nach Beispiel 23.12 (a) ist jede Teilmenge D eines metrischen Raumes M mit der eingeschränkten Metrik selbst wieder ein metrischer Raum. Da wir die Metrik auf M für die Definition 24.1 (b) der Stetigkeit (also im Fall $a \in D$) auch nur für Punkte in D benötigen, können wir für Stetigkeitsbetrachtungen in metrischen Räumen in Zukunft also ohne Einschränkung $M = D$ setzen und so die Notationen etwas vereinfachen: In diesem Fall ist eine Funktion $f\colon M \to N$ nach Definition 24.1 (b) genau dann in einem Punkt $a \in M$ stetig, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt mit $f(U_\delta(a)) \subset U_\varepsilon(f(a))$.

Bei der stetigen Fortsetzbarkeit ist eine solche Vereinfachung dagegen nicht möglich.

Beispiel 24.3 (Stetigkeit der Metrik bzw. Norm)**.**

(a) In jedem metrischen Raum M ist die Abstandsfunktion

$$f\colon M \to \mathbb{R},\ x \mapsto d(x,b)$$

zu einem fest gewählten Punkt $b \in M$ stetig: Es seien $a \in M$ und $\varepsilon \in \mathbb{R}_{>0}$ gegeben; wir wählen $\delta = \varepsilon$. Dann gilt für alle $x \in M$ mit $d(x,a) < \varepsilon$ nach der Dreiecksungleichung

$$d(x,b) - d(a,b) \le d(x,a) < \varepsilon \qquad \text{sowie} \qquad d(a,b) - d(x,b) \le d(x,a) < \varepsilon,$$

und damit auch $|f(x) - f(a)| = |d(x,b) - d(a,b)| < \varepsilon$, da $|d(x,b) - d(a,b)|$ in jedem Fall eine der beiden Zahlen $d(x,b) - d(a,b)$ und $d(a,b) - d(x,b)$ ist. Also ist f stetig in jedem Punkt $a \in M$.

(b) Ist V ein normierter Raum, so ist nach (a) auch die Normfunktion

$$f\colon V \to \mathbb{R},\ x \mapsto \|x\| = d(x,0)$$

stetig.

Wie in $\mathbb{K}$ gibt es auch für metrische Räume wieder die Möglichkeit, Grenzwerte von Funktionen auf solche von Folgen zurückzuführen:

Satz 24.4 (Folgenkriterium). *Es seien M, N metrische Räume, $D \subset M$ und $f\colon D \to N$ eine Funktion.*

(a) ***(Folgenkriterium für Funktionsgrenzwerte)*** *Für $a \in \overline{D}$ und $c \in N$ gilt*

$$\lim_{x \to a} f(x) = c \quad \Leftrightarrow \quad \textit{Für jede Folge } (x_n)_n \textit{ in } D \textit{ mit } x_n \to a \textit{ gilt } f(x_n) \to c.$$

(b) ***(Folgenkriterium für Stetigkeit)*** *Für $a \in D$ gilt*

$$f \textit{ ist stetig in } a \quad \Leftrightarrow \quad \textit{Für jede Folge } (x_n)_n \textit{ in } D \textit{ mit } x_n \to a \textit{ gilt } f(x_n) \to f(a).$$

Beweis. Der Beweis ist (bis auf die Ersetzung der Betragsstriche durch die Metrik) wörtlich genauso wie der von Satz 8.11. □

Bemerkung 24.5.

(a) Da eine Folge in N nach Lemma 23.15 höchstens einen Grenzwert besitzen kann, folgt aus Satz 24.4 unmittelbar, dass auch der Grenzwert $\lim_{x \to a} f(x)$ von Funktionen im Fall der Existenz eindeutig ist.

(b) Da wir aus Bemerkung 23.18 schon wissen, dass Grenzwerte von Folgen eine topologische Eigenschaft sind, gilt dies nach Satz 24.4 nun auch für Grenzwerte bzw. die Stetigkeit von Funktionen.

Beispiel 24.6 (Stetigkeit von Koordinatenabbildungen)**.** Für $n \in \mathbb{N}_{>0}$ und $i = 1, \ldots, n$ ist die Abbildung

$$f\colon \mathbb{K}^n \to \mathbb{K},\ x \mapsto x_i,$$

die jedem Vektor seine i-te Koordinate zuordnet, stetig: Nach Bemerkung 24.5 (b) können wir dies in der zur euklidischen Norm äquivalenten Maximumsnorm überprüfen. Dann gilt für alle $\varepsilon > 0$ und $x, a \in \mathbb{K}^n$ mit $\|x - a\|_\infty < \delta := \varepsilon$, dass

$$|f(x) - f(a)| = |x_i - a_i| \le \|x - a\|_\infty < \varepsilon.$$

Wir wollen nun untersuchen, wie man die Stetigkeit konkret gegebener Abbildungen einfach überprüfen kann. Am wichtigsten ist dabei der Fall, in dem der Start- bzw. Zielraum eine Teilmenge von $\mathbb{K}^n$ ist. Als erstes Resultat hierzu besagt das folgende Lemma, dass wir die Stetigkeit in diesem Fall koordinatenweise im Zielraum überprüfen können.

Lemma 24.7 (Koordinatenweise Stetigkeit im Zielraum)**.** *Es seien D eine Teilmenge eines metrischen Raumes M und*

$$f\colon D \to \mathbb{K}^m,\ x \mapsto f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$$

eine Abbildung mit Komponentenfunktionen $f_1,\dots,f_m\colon D \to \mathbb{K}$. Ist nun $a \in \overline{D}$ und $c \in \mathbb{K}^m$, so gilt genau dann $\lim_{x\to a} f(x) = c$, wenn $\lim_{x\to a} f_i(x) = c_i$ für alle $i = 1,\dots,m$ ist, wobei c_i die Koordinaten von c bezeichnet.

Für $a \in D$ ergibt sich also insbesondere, dass f genau dann in a stetig ist, wenn alle Koordinatenfunktionen $f_i\colon D \to \mathbb{K}$ es sind.

Beweis. Es sei $(x_n)_n$ eine beliebige Folge in D, die gegen a konvergiert. Nach dem Folgenkriterium aus Satz 24.4 (angewendet sowohl auf f als auch auf $f_1,\dots,f_m$) genügt es zu zeigen, dass $f(x_n)$ genau dann gegen c konvergiert, wenn $f_i(x_n)$ für alle $i = 1,\dots,m$ gegen c_i konvergiert. Dies ergibt sich aber sofort aus Lemma 23.19. □

Wir können die Stetigkeit einer Abbildung nach $\mathbb{K}^m$ also sofort auf die Stetigkeit einer Abbildung nach $\mathbb{K}$ zurückführen. Können wir die Situation noch weiter vereinfachen und die Stetigkeit auch noch koordinatenweise im Startraum überprüfen, falls dieser eine Teilmenge von $\mathbb{K}^n$ ist? Das folgende Beispiel zeigt, dass dies leider nicht der Fall ist.

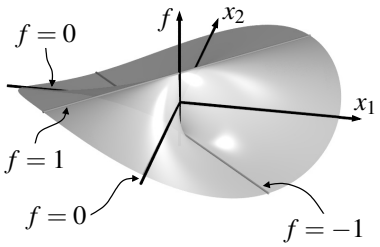
Beispiel 24.8. Wir betrachten die im Bild unten rechts dargestellte Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{cases} \frac{2x_1x_2}{x_1^2+x_2^2} & \text{für } (x_1,x_2) \neq (0,0), \\ 0 & \text{für } (x_1,x_2) = (0,0). \end{cases}$$

Man kann sie sich leicht in Polarkoordinaten $(x_1,x_2) = (r\cos\varphi, r\sin\varphi)$ vorstellen, denn der gegebene Ausdruck ist dann gleich $\frac{2x_1x_2}{x_1^2+x_2^2} = 2\cos\varphi\sin\varphi = \sin 2\varphi$. Die Funktion f ist abgesehen vom Nullpunkt also auf jeder Ursprungsgeraden konstant, z. B. auf den Koordinatenachsen gleich 0 und auf der durch $x_2 = x_1$ gegebenen Diagonalen (mit $\varphi = \frac{\pi}{4}$) gleich 1.

Daran sehen wir auch schon, dass f z. B. nach dem Folgenkriterium aus Satz 24.4 (b) im Nullpunkt unstetig ist: Die Folge $\left(\left(\frac{1}{n},\frac{1}{n}\right)^\mathsf{T}\right)_{n\in\mathbb{N}_{>0}}$ konvergiert zwar (entlang der Diagonalen) gegen den Ursprung, aber es ist natürlich

$$f\begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \end{pmatrix} = \frac{2/n^2}{2/n^2} = 1 \not\to 0 = f\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$



In der Tat ist das Folgenkriterium oft nützlich, um die *Unstetigkeit* einer Funktion zu zeigen, da es hierfür ja genügt, eine einzige Folge anzugeben, die die Bedingung des Kriteriums verletzt.

Andererseits ist aber für jedes fest gewählte $x_2 \in \mathbb{R}$ die Funktion $x_1 \mapsto f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ in *einer* Variablen x_1 stetig: Für $x_2 = 0$ ergibt sich einfach die Nullfunktion, und für alle anderen x_2 benötigen wir nur die erste Zeile in der Definition von f, die dann natürlich eine stetige Funktion in x_1 ist. Genauso gilt dies, wenn wir x_1 fest halten und x_2 als variabel betrachten.

Die Funktion f auf $\mathbb{R}^2$ ist also im Nullpunkt nicht stetig, obwohl beide Funktionen in einer Variablen, die man durch Festhalten der jeweils anderen daraus erhält, dort stetig sind. Anschaulich

liegt das daran, dass man beim Festhalten jeweils einer Variablen nur untersucht, wie sich die Funktion verhält, wenn man sich *auf einer der Koordinatenachsen* dem Nullpunkt nähert, während man sich ihm für die Überprüfung der Stetigkeit *auf einem beliebigen Weg* nähern muss. Zusammen mit Lemma 24.7 sehen wir also:

> Die Stetigkeit einer Abbildung von $\mathbb{K}^n$ nach $\mathbb{K}^m$ kann zwar koordinatenweise im Zielraum, aber nicht koordinatenweise im Startraum überprüft werden.

Dennoch wollen wir jetzt aber sehen, dass die „gewohnten Rechenregeln" aus Kapitel 8 für Summen, Differenzen, Produkte, Quotienten und Verkettungen stetiger Funktionen auch in unserem hier betrachteten allgemeineren Fall gelten.

Lemma 24.9 (**Grenzwertsätze** für Funktionen)**.** *Es seien $D \subset M$ eine Teilmenge eines metrischen Raumes M und $f,g\colon D \to \mathbb{K}$ zwei Funktionen. Weiterhin sei $a \in \overline{D}$ ein Punkt, so dass die Grenzwerte von f und g in a existieren. Dann gilt*

$$\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x),$$

und eine entsprechende Aussage auch für $f(x)-g(x)$, $f(x)\cdot g(x)$ und $\frac{f(x)}{g(x)}$ (letzteres natürlich nur falls $\lim_{x\to a} g(x) \neq 0$). Insbesondere sind für $a \in D$ also mit f und g auch $f+g$, $f-g$, $f\cdot g$ und $\frac{f}{g}$ in a stetig (letzteres wiederum nur falls $g(a) \neq 0$).

Beweis. Der Beweis läuft wörtlich genauso wie der von Satz 8.13 – nämlich indem man die Aussage mit Hilfe des Folgenkriteriums aus Satz 24.4 auf die entsprechenden Aussagen über Grenzwerte von Folgen in $\mathbb{K}$ (siehe Satz 5.13) zurückführt. □

Lemma 24.10 (Verkettung stetiger Abbildungen)**.** *Es seien $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen zwischen metrischen Räumen M,N,R. Ist dann $a \in M$, so dass f in a und g in $f(a)$ stetig sind, so ist auch $g\circ f$ stetig in a.*

Beweis. Der Beweis ist derselbe wie in Satz 8.15 bzw. Bemerkung 8.16: Ist $(x_n)_n$ eine Folge in M mit $x_n \to a$, so gilt $f(x_n) \to f(a)$ nach dem Folgenkriterium aus Satz 24.4 (b) wegen der Stetigkeit von f in a, und dann genauso $g(f(x_n)) \to g(f(a))$ wegen der Stetigkeit von g in $f(a)$. Also ist $g\circ f$ nach dem Folgenkriterium in a stetig. □

Beispiel 24.11.

(a) Die in Beispiel 24.8 betrachtete Funktion ist nach Lemma 24.9 in jedem Punkt $a \in \mathbb{R}^2\backslash\{0\}$ stetig, da sie in einer Umgebung eines jeden solchen Punktes durch Addition, Multiplikation und Division aus den nach Beispiel 24.6 stetigen Koordinatenfunktionen x_1 und x_2 zusammengesetzt ist.

(b) Nach Lemma 24.7 und 24.9 ist insbesondere jede *lineare* Abbildung $f\colon \mathbb{K}^n \to \mathbb{K}^m$, $x \mapsto Ax$ (für eine Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m\times n,\mathbb{K})$) stetig, da die zugehörigen Koordinatenfunktionen $x \mapsto a_{i,1}x_1 + \cdots + a_{i,n}x_n$ es sind. Diese Aussage erscheint zwar selbstverständlich, ist jedoch für beliebige normierte Räume falsch: Für unendlich-dimensionale Vektorräume müssen lineare Abbildungen im Sinne von Definition 13.16 nicht notwendig stetig sein, wie das folgende Beispiel zeigt.

Aufgabe 24.12 (Stetigkeit linearer Abbildungen)**.** Es sei V der Vektorraum der stetigen reellwertigen Funktionen auf $[0,1]$. Zeige, dass die lineare Abbildung $f\colon V \to \mathbb{R}$, $\varphi \mapsto \varphi(0)$ zwar bezüglich der Norm $\|\cdot\|_\infty$ auf V, aber nicht bezüglich der Norm $\|\cdot\|_2$ auf V stetig ist.

Aufgabe 24.13. Sind die folgenden Funktionen stetig in den Nullpunkt fortsetzbar?

$$f\colon \mathbb{R}^2\backslash\{0\} \to \mathbb{R},\ x \mapsto \frac{(x_1+x_2)^3}{x_1^2+x_2^2}, \qquad g\colon \mathbb{R}_{>0}\times\mathbb{R} \to \mathbb{R},\ x \mapsto x_1^{x_2}.$$

Aufgabe 24.14. Zeige, dass die Funktion

$$g\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto \begin{cases} \frac{x_1 x_2^2}{x_1^2+x_2^4} & \text{für } (x_1,x_2) \neq (0,0), \\ 0 & \text{für } (x_1,x_2) = (0,0) \end{cases}$$

zwar unstetig ist, ihre Einschränkung auf jede Gerade durch den Nullpunkt jedoch stetig ist.

Aufgabe 24.15.

(a) Es seien M ein metrischer Raum und $f,g\colon M \to \mathbb{R}$ zwei stetige Abbildungen. Zeige, dass dann auch die Abbildung $\max(f,g)\colon M \to \mathbb{R},\ x \mapsto \max(f(x),g(x))$ stetig ist.

(b) Zeige, dass die Abbildung $\mathrm{GL}(n,\mathbb{K}) \to \mathrm{GL}(n,\mathbb{K}),\ A \mapsto A^{-1}$ stetig ist.

24.B Eigenschaften stetiger Abbildungen

Nachdem wir jetzt wissen, wie wir von Abbildungen ihre Stetigkeit überprüfen können, kommen wir nun zu den Eigenschaften stetiger Funktionen. Als Erstes schauen wir uns dazu an, wie sich Umgebungen sowie offene und abgeschlossene Mengen unter stetigen Abbildungen verhalten. Unser erstes Ergebnis zeigt dabei noch einmal deutlich, dass die Stetigkeit von Funktionen eine topologische Eigenschaft ist (siehe Bemerkung 24.5 (b)).

Lemma 24.16 (Charakterisierung stetiger Funktionen durch Umgebungen)**.** *Es seien $f\colon M \to N$ eine Abbildung zwischen metrischen Räumen und $a \in M$. Dann sind äquivalent:*

(a) *f ist stetig in a.*

(b) *Für jede Umgebung $U \subset N$ von $f(a)$ ist $f^{-1}(U) \subset M$ eine Umgebung von a (man sagt auch: „Urbilder von Umgebungen sind Umgebungen“).*

Beweis. Es gilt:

$$\begin{aligned}
& f \text{ ist stetig in } a \\
\Leftrightarrow\ & \text{Für alle } \varepsilon \text{ gibt es ein } \delta \text{ mit } f(U_\delta(a)) \subset U_\varepsilon(f(a)) && \text{(Bemerkung 24.2)} \\
\Leftrightarrow\ & \text{Zu jeder Umgebung } U \text{ von } f(a) \text{ gibt es ein } \delta \text{ mit } f(U_\delta(a)) \subset U && (*) \\
\Leftrightarrow\ & \text{Zu jeder Umgebung } U \text{ von } f(a) \text{ gibt es ein } \delta \text{ mit } U_\delta(a) \subset f^{-1}(U) \\
\Leftrightarrow\ & \text{Zu jeder Umgebung } U \text{ von } f(a) \text{ ist } f^{-1}(U) \text{ eine Umgebung von } a && \text{(Definition 23.13 (b))},
\end{aligned}$$

wobei in $(*)$ die Folgerung “$\Leftarrow$” gilt, da jede ε-Umgebung eine Umgebung ist, und “$\Rightarrow$” gilt, da jede
60 Umgebung von $f(a)$ eine ε-Umgebung von $f(a)$ enthält. □

Für die Stetigkeit einer gesamten Abbildung (also nicht nur in einem einzelnen Punkt) gibt es erstaunlicherweise das folgende sehr einfache Kriterium.

Satz 24.17 (Charakterisierung stetiger Funktionen durch offene bzw. abgeschlossene Mengen)**.** *Für eine Abbildung $f\colon M \to N$ zwischen metrischen Räumen sind äquivalent:*

(a) *f ist stetig (in jedem Punkt von M).*

(b) *Für jede offene Menge $U \subset N$ ist $f^{-1}(U) \subset M$ offen (man sagt: „Urbilder offener Mengen sind offen“).*

(c) *Für jede abgeschlossene Menge $A \subset N$ ist $f^{-1}(A) \subset M$ abgeschlossen (man sagt: „Urbilder abgeschlossener Mengen sind abgeschlossen“).*

Beweis. Wir zeigen zunächst die Äquivalenz (a) $\Leftrightarrow$ (b):

$$\begin{aligned}
f \text{ ist stetig} &\overset{24.16}{\Leftrightarrow} \text{Für alle } a \in M \text{ und alle Umgebungen } U \text{ von } f(a) \text{ ist } f^{-1}(U) \text{ Umgebung von } a \\
&\overset{(*)}{\Leftrightarrow} \text{Für alle } a \in M \text{ und alle } U \text{ offen mit } f(a) \in U \text{ ist } f^{-1}(U) \text{ Umgebung von } a \\
&\Leftrightarrow \text{Für alle } U \text{ offen und alle } a \in f^{-1}(U) \text{ ist } f^{-1}(U) \text{ Umgebung von } a \\
&\overset{23.33\,(a)}{\Leftrightarrow} \text{Für alle } U \text{ offen ist } f^{-1}(U) \text{ offen},
\end{aligned}$$

wobei in $(*)$ die Folgerung „$\Rightarrow$“ gilt, weil jede offene Menge U mit $f(a) \in U$ eine Umgebung von $f(a)$ ist, und „$\Leftarrow$“ gilt, da jede Umgebung von $f(a)$ eine offene Menge U mit $f(a) \in U$ enthält.

Die Äquivalenz (b) $\Leftrightarrow$ (c) folgt direkt durch Übergang zum Komplement: Setzen wir $A = N \backslash U$, so ist A genau dann abgeschlossen, wenn U offen ist, und analog $f^{-1}(A) = f^{-1}(N \backslash U) = M \backslash f^{-1}(U)$ genau dann abgeschlossen, wenn $f^{-1}(U)$ offen ist. Dies zeigt unmittelbar die Äquivalenz der beiden Aussagen. □

Beispiel 24.18 (Offene und abgeschlossene Bedingungen)**.** Sind M ein metrischer Raum und $f\colon M \to \mathbb{R}$ eine stetige Funktion, so ist die Menge

$$\{x \in M : f(x) > 0\} = f^{-1}(\mathbb{R}_{>0})$$

als Urbild der offenen Menge $\mathbb{R}_{>0} \subset \mathbb{R}$ unter einer stetigen Abbildung nach Satz 24.17 offen in M. Ebenso gilt dies natürlich für die Mengen

$$\{x \in M : f(x) < 0\} \qquad \text{und} \qquad \{x \in M : f(x) \neq 0\}$$

als Urbilder von $\mathbb{R}_{<0}$ bzw. $\mathbb{R} \backslash \{0\}$. Dagegen ist die Menge

$$\{x \in M : f(x) \geq 0\} = f^{-1}(\mathbb{R}_{\geq 0})$$

als Urbild der abgeschlossenen Menge $\mathbb{R}_{\geq 0} \subset \mathbb{R}$ unter einer stetigen Abbildung abgeschlossen in M, genauso wie die Mengen

$$\{x \in M : f(x) \leq 0\} \qquad \text{und} \qquad \{x \in M : f(x) = 0\}.$$

Mit dieser Beobachtung kann man in vielen Fällen sehr einfach herausfinden, ob eine gegebene Menge offen bzw. abgeschlossen ist: So sieht man z. B. sofort, dass die Menge

$$\{x \in \mathbb{R}^2 : x_1^3 > \cos x_2\} \quad \subset \mathbb{R}^2$$

offen sein muss, da sie das Urbild von $\mathbb{R}_{>0}$ unter der stetigen Abbildung $\mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1^3 - \cos x_2$ ist. Man sagt in diesem Sinne oft auch, dass „$>$“, „$<$“ und „$\neq$“ (jeweils mit reellwertigen stetigen Ausdrücken auf beiden Seiten) *offene Bedingungen* sind, während „$\geq$“, „$\leq$“ und „$=$“ *abgeschlossene Bedingungen* darstellen.

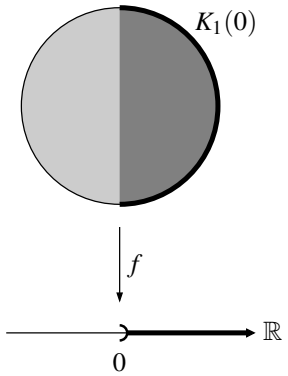
Bemerkung 24.19 (Relativ offene Mengen)**.** Es seien M und N metrische Räume, $D \subset M$ und $f\colon D \to N$ eine Abbildung. Nach Satz 24.17 ist f genau dann stetig, wenn zu jeder offenen Menge $U \subset N$ das Urbild $f^{-1}(U)$ offen im Startraum D ist.

Beachte jedoch, dass dies *nicht* heißen muss, dass $f^{-1}(U)$ auch offen in M ist! In der Tat ist $f^{-1}(U)$ nach Aufgabe 23.37 (b) genau dann offen in D, wenn es eine in M offene Teilmenge V gibt mit $f^{-1}(U) = V \cap D$ – man sagt in diesem Fall auch manchmal, dass $f^{-1}(U)$ *relativ offen* in D ist.

Als anschauliches Beispiel hierfür können wir die stetige Abbildung

$$f\colon K_1(0) \to \mathbb{R}, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1$$

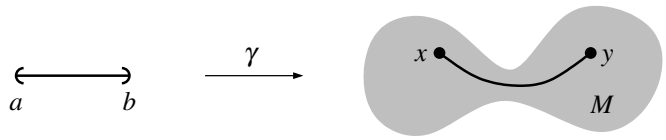
auf dem abgeschlossenen Einheitskreis $K_1(0) = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$ betrachten. Das Urbild der offenen Menge $\mathbb{R}_{>0} \subset \mathbb{R}$ unter f ist in diesem Fall (wie im Bild dunkel eingezeichnet) der rechte Halbkreis, wobei vom Rand dieses Halbkreises der Bogen mit enthalten ist, der Durchmesser jedoch nicht. Diese Menge ist wegen der enthaltenen Randpunkte auf dem Halbkreisbogen natürlich nicht offen in $\mathbb{R}^2$ – sie ist aber (relativ) offen in $K_1(0)$, denn sie ist der Durchschnitt von $K_1(0)$ mit der nach Beispiel 24.18 offenen Menge $\{x \in \mathbb{R}^2 : x_1 > 0\}$.



Als Nächstes wollen wir untersuchen, ob wir eine Verallgemeinerung des Zwischenwertsatzes 8.21 für beliebige metrische Räume finden können, d. h. ob wir für gewisse Klassen von reellwertigen Funktionen sagen können, dass sie mit zwei Werten auch jede Zahl dazwischen annehmen müssen. Die entscheidende Frage ist hierbei, ob die Definitionsmenge „nur aus einem zusammenhängenden Teil besteht“ – denn wenn sie wie z. B. die Menge $M = [0,1] \cup [2,3] \subset \mathbb{R}$ „aus zwei unverbundenen

Teilen besteht“, können wir natürlich leicht eine stetige Abbildung $f\colon M \to \mathbb{R}$ konstruieren, die die Zwischenwertbedingung nicht erfüllt (in unserem Beispiel könnten wir dafür schon die Abbildung $f\colon M \to \mathbb{R}$, $x \mapsto x$ wählen, da hierfür zwar 1 und 2, aber keine der Zahlen dazwischen im Bild von f liegen). Die mathematisch korrekte Formulierung dieser Bedingung ist die folgende:

Definition 24.20 (Wegzusammenhängende Räume)**.** Ein metrischer Raum M heißt **wegzusammenhängend**, wenn es zu je zwei Punkten $x, y \in M$ eine stetige Abbildung $\gamma\colon [a,b] \to M$ eines reellen Intervalls $[a,b]$ nach M gibt mit $\gamma(a) = x$ und $\gamma(b) = y$. Man nennt γ in diesem Fall einen **Weg** von x nach y – der Raum M heißt also wegzusammenhängend, wenn sich je zwei Punkte in M durch einen Weg verbinden lassen.



Beispiel 24.21.

(a) Jedes Intervall $M \subset \mathbb{R}$ aus Notation 4.17 (a) (offen, abgeschlossen, halboffen oder uneigentlich) ist wegzusammenhängend, denn für zwei beliebige Punkte $x, y \in M$ liegt die gerade Verbindungsstrecke $\gamma\colon [0,1] \to \mathbb{R}$, $t \mapsto x + t\,(y - x)$ von x nach y immer in M.

(b) Die Menge $M = [0,1] \cup [2,3] \subset \mathbb{R}$ von oben ist nicht wegzusammenhängend: Wäre $\gamma\colon [a,b] \to M \subset \mathbb{R}$ eine stetige Abbildung mit $\gamma(a) = 1$ und $\gamma(b) = 2$, so müsste γ nach dem Zwischenwertsatz 8.21 auch jeden Wert zwischen 1 und 2 annehmen – was aber nicht möglich ist, da diese Zahlen nicht in M liegen.

Für wegzusammenhängende Räume erhalten wir nun wie erwartet eine Verallgemeinerung des Zwischenwertsatzes 8.21 in $\mathbb{R}$:

Satz 24.22 (**Zwischenwertsatz**)**.** *Es sei $f\colon M \to N$ eine stetige Abbildung zwischen metrischen Räumen. Ferner sei M wegzusammenhängend. Dann gilt:*

(a) *Das Bild $f(M) \subset N$ ist ebenfalls wegzusammenhängend.*

(b) *Im Fall $N = \mathbb{R}$ nimmt f mit je zwei Funktionswerten auch jeden Wert dazwischen an.*

Beweis. Es seien $x, y \in f(M)$ beliebig. Wir können dann $u, v \in M$ wählen mit $x = f(u)$ und $y = f(v)$. Da M wegzusammenhängend ist, gibt es nun einen Weg $\gamma\colon [a,b] \to M$ von u nach v. Die Abbildung $f \circ \gamma\colon [a,b] \to f(M)$ ist dann ein Weg in $f(M)$ von x nach y, woraus sich bereits die Behauptung (a) ergibt. Im Fall $N = \mathbb{R}$ folgt zusätzlich mit dem eindimensionalen Zwischenwertsatz 8.21, dass $f \circ \gamma$ und damit auch f mit x und y auch jede reelle Zahl dazwischen als Wert annehmen müssen. □

Aufgabe 24.23. Es sei M ein metrischer Raum. Man zeige:

(a) Sind $a, b \in M$ und $A \subset M$ mit $a \in A$ und $b \notin A$, so enthält jeder Weg in M von a nach b einen Punkt in ∂A.

(b) Ist M wegzusammenhängend, so sind $\emptyset$ und M die einzigen Teilmengen von M, die sowohl offen als auch abgeschlossen sind.

Die Konzepte der gleichmäßigen Stetigkeit und Konvergenz aus Abschnitt 8.C übertragen sich wie erwartet auf den Fall metrischer Räume. Mit dem gleichen Argument wie z. B. bei der Beschränktheit in Bemerkung 23.21 (b) stimmen auch sie in normierten Räumen zu äquivalenten Normen überein, obwohl sie keine topologischen Eigenschaften sind.

Definition 24.24 (Gleichmäßige Stetigkeit und Konvergenz)**.** Es sei $f\colon M \to N$ eine Abbildung zwischen metrischen Räumen.

(a) Die Abbildung f heißt **gleichmäßig stetig**, wenn gilt

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, a \in M : d(x,a) < \delta \Rightarrow d(f(x), f(a)) < \varepsilon$$

(der entscheidende Punkt gegenüber der „normalen“ Stetigkeit ist also wie in Definition 8.34 (a) auch hier, dass das δ nur von ε, aber nicht vom betrachteten Punkt a abhängen darf).

(b) Für jedes $n \in \mathbb{N}$ sei eine Funktion $f_n \colon M \to N$ gegeben. Ist $\lim_{n\to\infty} f_n(x) = f(x)$ für alle $x \in M$, d. h. gilt

$$\forall x \in M \ \forall \varepsilon > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \geq n_0 : d(f_n(x), f(x)) < \varepsilon,$$

so nennt man die Funktionenfolge $(f_n)_n$ **punktweise konvergent** gegen f. Beachte, dass das n_0 dabei nicht nur von ε, sondern auch vom betrachteten Punkt x abhängen darf. Kann man n_0 jedoch auch unabhängig von x wählen, d. h. gilt sogar

$$\forall \varepsilon > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \geq n_0 \ \forall x \in M : d(f_n(x), f(x)) < \varepsilon,$$

so heißt die Funktionenfolge $(f_n)_n$ **gleichmäßig konvergent** gegen f.

Bemerkung 24.25. Da Definition 24.24 völlig analog zum eindimensionalen Fall ist, übertragen sich auch unsere damals dazu erzielten Ergebnisse und ihre Beweise unmittelbar auf die neue Situation:

(a) Ist $f \colon M \to N$ eine stetige Funktion zwischen metrischen Räumen und ist M kompakt, so ist f sogar gleichmäßig stetig (siehe Satz 8.36). Dabei bedeutet die Kompaktheit des metrischen Raumes M, dass M als Teilmenge von sich selbst im Sinne von Definition 23.50 kompakt ist.

(b) Konvergiert eine Folge stetiger Funktionen $f_n \colon M \to N$ auf einem metrischen Raum M gleichmäßig gegen $f \colon M \to N$, so ist die Grenzfunktion f ebenfalls stetig (siehe Satz 8.39).

(c) Eine Folge von Funktionen $f_n \colon M \to \mathbb{K}$ auf einem metrischen Raum M konvergiert genau dann gleichmäßig gegen $f \colon M \to \mathbb{K}$, wenn $\|f_n - f\|_{\sup} := \sup\{|f_n(x) - f(x)| : x \in M\}$ für $n \to \infty$ gegen 0 konvergiert (siehe Aufgabe 8.43).

Zum Abschluss dieses Abschnitts wollen wir schließlich noch einen oft nützlichen Satz zeigen, mit dem man die Existenz von Fixpunkten gewisser (stetiger) Abbildungen nachweisen kann. Da er im eindimensionalen Fall noch nicht besonders interessant ist, haben wir ihn damals in Kapitel 8 nicht in dieser Form bewiesen; am ehesten entspricht er vermutlich der Aussage von Aufgabe 8.30 (a).

Satz 24.26 (Banachscher Fixpunktsatz). *Es seien M ein nicht-leerer vollständiger metrischer Raum und $f \colon M \to M$ eine Abbildung mit der Eigenschaft, dass es ein $q \in [0,1)$ gibt mit*

$$d(f(x), f(y)) \leq q \cdot d(x,y)$$

für alle $x, y \in M$ (eine solche Abbildung bezeichnet man auch als Kontraktion).

Dann hat f genau einen Fixpunkt, d. h. es gibt genau ein $a \in M$ mit $f(a) = a$.

Beweis. Um die Existenz eines Fixpunktes zu zeigen, konstruieren wir zu einem beliebigen Startpunkt $x_0 \in M$ rekursiv die Folge $(x_n)_{n\in\mathbb{N}}$ in M mit $x_{n+1} = f(x_n)$ für alle $n \in \mathbb{N}$. Für die Abstände aufeinander folgender Glieder dieser Folge gilt induktiv

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq q\, d(x_n, x_{n-1}) \leq \cdots \leq q^n\, d(x_1, x_0). \qquad (*)$$

Damit ist $(x_n)_{n\in\mathbb{N}}$ eine Cauchyfolge, denn wählen wir zu gegebenem $\varepsilon \in \mathbb{R}_{>0}$ ein $n_0 \in \mathbb{N}$ mit $\frac{q^{n_0}}{1-q}\, d(x_1, x_0) < \varepsilon$, so gilt für alle $m \geq n \geq n_0$

$$\begin{aligned}
d(x_m, x_n) &\leq d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \cdots + d(x_{n+1}, x_n) && \text{(Dreiecksungleichung)}\\
&\leq (q^{m-1} + q^{m-2} + \cdots + q^n)\, d(x_1, x_0) && \text{(nach } (*))\\
&\leq q^n \cdot \frac{1}{1-q} \cdot d(x_1, x_0) && \text{(geometrische Reihe)}\\
&< \varepsilon && \text{(Wahl von } n_0).
\end{aligned}$$

Wegen der Vollständigkeit von M konvergiert $(x_n)_{n\in\mathbb{N}}$ also gegen ein $a \in M$.

Beachte, dass f in a stetig ist, denn zu gegebenem $\varepsilon \in \mathbb{R}_{>0}$ gilt für alle $x \in M$ mit $d(x,a) < \varepsilon$, dass $d(f(x), f(a)) \leq q\, d(x,a) \leq d(x,a) < \varepsilon$. Gehen wir in der Rekursionsgleichung $x_{n+1} = f(x_n)$ zum Grenzwert $n \to \infty$ über, so erhalten wir mit dem Folgenkriterium aus Satz 24.4 (b) also

$$a = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(a),$$

d. h. a ist ein Fixpunkt von f.

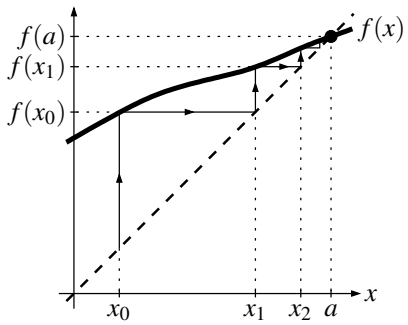
Es bleibt damit nur noch die Eindeutigkeit des Fixpunkts zu zeigen: Wäre $b \in M$ noch ein anderer Fixpunkt, so wäre $d(a,b) = d(f(a), f(b)) \leq q\, d(a,b)$, wegen $d(a,b) > 0$ erhalten wir dann also sofort den Widerspruch $1 \leq q$. □

Beispiel 24.27.

(a) Es sei $f\colon \mathbb{R} \to \mathbb{R}$ eine differenzierbare Funktion mit $|f'(x)| \leq q$ für ein $q \in [0,1)$ und alle $x \in \mathbb{R}$, so dass der Graph von f wie im Bild rechts also überall flacher als der der (gestrichelt eingezeichneten) Identität verläuft. Nach dem Mittelwertsatz 10.24 (a) gibt es dann für alle $x, y \in \mathbb{R}$ ein z zwischen x und y mit $f(x) - f(y) = f'(z) \cdot (x - y)$, und damit ist

$$|f(x) - f(y)| = |f'(z)| \cdot |x - y| \leq q\, |x - y|.$$

Also ist f eine Kontraktion und hat damit nach Satz 24.26 genau einen Fixpunkt.



61

Das Bild zeigt auch sehr anschaulich die Konstruktion des Fixpunkts wie im Beweis des Satzes: Ausgehend von einem beliebigen Punkt $x_0 \in \mathbb{R}$ bilden wir $f(x_0)$, nennen diesen Punkt x_1, bilden dann wieder $f(x_1)$ und nennen diesen Punkt x_2 usw., gehen also im Bild immer abwechselnd vertikal zum Graphen von f und horizontal zum Graphen der Identität. Auf diese Art erhalten wir eine Folge, die gegen den Fixpunkt konvergiert.

(b) Der Banachsche Fixpunktsatz kann auch oft auf Funktionenräume angewendet werden und liefert dann die Existenz von Funktionen mit bestimmten Eigenschaften. Als Beispiel dafür wollen wir untersuchen, ob es eine differenzierbare Funktion $\varphi\colon [0,1] \to \mathbb{R}$ gibt mit

$$\varphi'(x) = \frac{1}{2} \sin \varphi(x) + \sin x \quad \text{für alle } x \in [0,1].$$

Solche Gleichungen, die die Ableitung einer gesuchten Funktion durch die Funktion selbst ausdrücken, nennt man *Differentialgleichungen*, sie kommen in der Praxis an vielen Stellen vor. Sie zu lösen ist in der Regel schwierig, und in der Tat kann man für die oben angegebene Differentialgleichung mit uns bekannten Funktionen keine exakte Lösung angeben. Mit dem Banachschen Fixpunktsatz können wir jedoch die Existenz einer Lösung beweisen, indem wir die Abbildung

$$f\colon C^0([0,1]) \to C^0([0,1]), \quad f(\varphi)(x) = \int_0^x \left(\frac{1}{2} \sin \varphi(t) + \sin t \right) dt$$

betrachten: Gibt es nämlich einen Fixpunkt von f, also eine stetige Funktion

$$\varphi\colon [0,1] \to \mathbb{R} \quad \text{mit} \quad \varphi(x) = \int_0^x \left(\frac{1}{2} \sin \varphi(t) + \sin t \right) dt \quad \text{für alle } x \in [0,1],$$

so folgt daraus durch Differenzieren dieser Gleichung nach dem Hauptsatz 12.21 der Differential- und Integralrechnung sofort die gewünschte Differentialgleichung für φ. Da $C^0([0,1])$ mit der Maximumsnorm nach Aufgabe 23.30 (a) vollständig ist, müssen wir also nur zeigen, dass f bezüglich dieser Norm eine Kontraktion ist: Für alle $\varphi, \psi \in C^0([0,1])$ und

$x \in [0,1]$ gilt

$$\begin{aligned}
|f(\varphi)(x) - f(\psi)(x)| &= \left| \int_0^x \frac{1}{2}\big(\sin\varphi(t) - \sin\psi(t)\big)\,dt \right| \\
&\le \frac{1}{2}\int_0^x |\sin\varphi(t) - \sin\psi(t)|\,dt && \text{(Satz 12.13 (d))} \\
&\le \frac{1}{2}\int_0^x |\varphi(t) - \psi(t)|\,dt && \text{(Mittelwertsatz für sin wie in (a))} \\
&\le \frac{1}{2}\int_0^x \|\varphi - \psi\|_\infty\,dt \\
&\le \frac{1}{2}\|\varphi - \psi\|_\infty, && (x \le 1)
\end{aligned}$$

und damit wie benötigt $\|f(\varphi) - f(\psi)\|_\infty \le \frac{1}{2}\|\varphi - \psi\|_\infty$. Also ist f ist eine Kontraktion mit $q = \frac{1}{2}$, und damit hat die gegebene Differentialgleichung nach Satz 24.26 eine Lösung.

In der Vorlesung „Einführung in die gewöhnlichen Differentialgleichungen“ des zweiten Studienjahres wird dieses Beispiel zu einem der zentralen Sätze ausgebaut, der in nahezu allen praktisch relevanten Fällen die Existenz von Lösungen von Differentialgleichungen sichert.

Aufgabe 24.28. Es seien M ein nicht-leerer vollständiger metrischer Raum und $f\colon M \to M$ eine Abbildung. Zeige, dass f einen eindeutigen Fixpunkt besitzt, wenn eine der folgenden Bedingungen erfüllt ist:

(a) Die Abbildung f ist surjektiv, und es gibt ein $q > 1$ mit $d(f(x), f(y)) \ge q\,d(x,y)$ für alle $x, y \in M$.

(b) M ist kompakt, und es gilt $d(f(x), f(y)) < d(x,y)$ für alle $x, y \in M$ mit $x \ne y$.

Zeige ferner anhand je eines Beispiels, dass auf die Surjektivität von f bzw. die Kompaktheit von M im Allgemeinen nicht verzichtet werden kann.

24.C Stetige Bilder kompakter Mengen

Am Anfang des letzten Abschnitts hatten wir gesehen, wie sich offene und abgeschlossene Mengen unter stetigen Abbildungen verhalten. Wir wollen dies nun auch für kompakte Mengen untersuchen. Auch sie haben eine Kompatibilitätseigenschaft mit stetigen Abbildungen – überraschenderweise allerdings nicht bezüglich Urbildern, sondern bezüglich Bildern. Der folgende Satz ist dabei die wahrscheinlich wichtigste Aussage zu kompakten Mengen überhaupt. Wir werden in diesem gesamten Abschnitt nichts weiter tun als einige der wichtigsten unmittelbaren Folgerungen daraus zu untersuchen.

Satz 24.29 (Kompakte Mengen unter stetigen Abbildungen)**.** *Es sei $f\colon M \to N$ eine stetige Abbildung zwischen metrischen Räumen. Ist dann $A \subset M$ kompakt, so auch $f(A) \subset N$ (man sagt: „Bilder kompakter Mengen unter stetigen Abbildungen sind kompakt“).*

Beweis. Wir müssen das Kriterium aus Definition 23.50 für die Menge $f(A)$ nachprüfen. Es sei also $(y_n)_n$ eine Folge in $f(A)$. Für alle n gibt es dann Punkte $x_n \in A$ mit $y_n = f(x_n)$. Weil A kompakt ist, gibt es nun eine Teilfolge $(x_{n_k})_k$, die gegen einen Punkt $a \in A$ konvergiert. Nach dem Folgenkriterium für Stetigkeit aus Satz 24.4 (b) konvergiert dann aber auch $(y_{n_k})_k = (f(x_{n_k}))_k$ gegen $f(a) \in f(A)$. Also besitzt die ursprüngliche Folge $(y_n)_n$ eine konvergente Teilfolge mit Grenzwert in $f(A)$, d. h. $f(A)$ ist kompakt. □

Bemerkung 24.30.

(a) In Gegensatz zu Lemma 24.16 und Satz 24.17 gilt in Satz 24.29 nicht die Umkehrung: Auch für die offensichtlich unstetige Abbildung

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x \ge 0, \\ 0 & \text{für } x < 0 \end{cases}$$

sind sogar Bilder beliebiger Teilmengen von $\mathbb{R}$ immer kompakt (nämlich immer $\emptyset$, $\{0\}$, $\{1\}$ oder $\{0,1\}$).

(b) Ist $f\colon M \to N$ eine stetige Abbildung zwischen metrischen Räumen, so haben wir in Lemma 24.16, Satz 24.17 und Satz 24.29 jetzt also gesehen:

Urbilder von Umgebungen unter f sind Umgebungen.
Urbilder offener Mengen unter f sind offen.
Urbilder abgeschlossener Mengen unter f sind abgeschlossen.
Bilder kompakter Mengen unter f sind kompakt.

In der Tat gelten die jeweils anderen Aussagen (also mit „Bild“ und „Urbild“ vertauscht) im Allgemeinen nicht: Betrachten wir die Funktionen

$$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto 0 \qquad \text{und} \qquad g\colon \mathbb{R} \to \mathbb{R},\ x \mapsto \mathrm{e}^x,$$

so gilt in $\mathbb{R}$ offensichtlich:

- Das Intervall $U = (-1,1)$ ist eine offene Umgebung von 0, aber das Bild $f(U) = \{0\}$ ist weder offen noch eine Umgebung von 0.
- Die Menge $A = \mathbb{R}$ ist abgeschlossen in $\mathbb{R}$, aber das Bild $g(A) = \mathbb{R}_{>0}$ ist nicht abgeschlossen in $\mathbb{R}$.
- Die Menge $A = \{0\} \subset \mathbb{R}$ ist kompakt, aber das Urbild $f^{-1}(A) = \mathbb{R}$ ist nicht kompakt in $\mathbb{R}$.

Wir werden diese Unterschiede nun als Erstes geschickt ausnutzen, um einen Satz über die Existenz stetiger Umkehrabbildungen zu stetigen bijektiven Abbildungen zu beweisen.

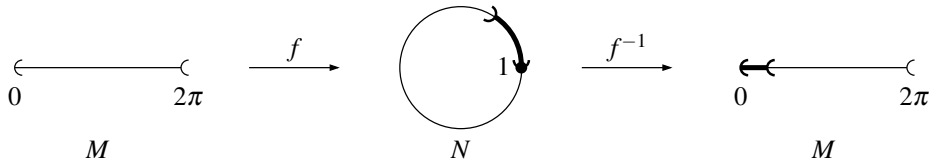
Satz 24.31 (Stetigkeit von Umkehrabbildungen)**.** *Es sei $f\colon M \to N$ eine stetige und bijektive Abbildung zwischen metrischen Räumen. Ist M kompakt, so ist auch die Umkehrabbildung $f^{-1}\colon N \to M$ stetig.*

Beweis. Nach Satz 24.17 (c) genügt es zu zeigen, dass Urbilder abgeschlossener Mengen unter f^{-1}, also Bilder abgeschlossener Mengen unter f wieder abgeschlossen sind. Es sei also $A \subset M$ abgeschlossen. Nach Aufgabe 23.60 (a) ist A als abgeschlossene Teilmenge des kompakten Raumes M dann ebenfalls kompakt. Damit ist nach Satz 24.29 aber auch $f(A)$ kompakt, insbesondere also abgeschlossen nach Satz 23.51 (a). □

Beispiel 24.32. Ohne die Zusatzvoraussetzung der Kompaktheit des Definitionsbereichs ist die Aussage von Satz 24.31 im Allgemeinen falsch: Es seien $M = [0,2\pi)$, $N = \{z \in \mathbb{C} : |z| = 1\}$ der Rand des komplexen Einheitskreises, und f die offensichtlich stetige und bijektive Abbildung

$$f\colon M \to N,\ x \mapsto \mathrm{e}^{\mathrm{i}x}$$

mit nicht kompakter Definitionsmenge M. In diesem Fall ist die Umkehrabbildung $f^{-1}\colon N \to M$ im Punkt $1 \in N$ nicht stetig: Wie im Bild unten dick eingezeichnet ist das halboffene Intervall $[0,1)$ eine Umgebung von $f^{-1}(1) = 0$ in M, aber sein Urbild unter f^{-1}, also $\{\mathrm{e}^{\mathrm{i}x} : 0 \le x < 1\}$, ist keine Umgebung von 1 in N, da hierfür Punkte in N unterhalb von 1 fehlen.



Mit kompakter Definitionsmenge $[0,2\pi]$ hätte dieses Beispiel natürlich nicht funktioniert, da f dann nicht mehr injektiv gewesen wäre.

Aufgabe 24.33. Es sei $f\colon \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1$ die Projektion auf die erste Koordinate. Man beweise oder widerlege:

(a) Bilder offener Mengen unter f sind offen.

(b) Bilder abgeschlossener Mengen unter f sind abgeschlossen.

(c) Urbilder kompakter Mengen unter f sind kompakt.

Dass stetige Abbildungen gemäß Satz 24.29 kompakte Mengen wieder auf kompakte Mengen abbilden, hat wie schon erwähnt viele weitreichende Konsequenzen. So erhalten wir z. B. die folgende wichtige Verallgemeinerung von Satz 8.25 zur Existenz von Maxima und Minima stetiger Funktionen.

Folgerung 24.34 (**Satz vom Maximum und Minimum**). *Es sei $f\colon M \to \mathbb{R}$ eine stetige Funktion auf einem nicht-leeren kompakten metrischen Raum M. Dann „nimmt f auf M ein Maximum und Minimum an", d. h. die Menge $f(M) \subset \mathbb{R}$ hat ein Maximum und Minimum. Insbesondere ist f damit also auf M beschränkt.*

Beweis. Die Menge $f(M) \subset \mathbb{R}$ ist nicht leer und nach Satz 24.29 kompakt, besitzt nach Lemma 23.53 also ein Maximum und Minimum. □

Eine weitere einfache, aber überaus wichtige Folgerung hieraus ist die bereits in Beispiel 23.9 angekündigte Äquivalenz aller Normen auf $\mathbb{K}^n$. Nach dem Beweis dieser Aussage können wir dann also auf $\mathbb{K}^n$ für alle topologischen Konzepte (und ein paar weitere wie z. B. die Beschränktheit, siehe Bemerkung 23.21 (b)) jeweils eine beliebige Norm verwenden, ohne das Ergebnis zu verändern.

Lemma 24.35 (Stetigkeit von Normen auf $\mathbb{K}^n$). *Für eine beliebige Norm $\|\cdot\|$ auf $\mathbb{K}^n$ ist die Abbildung*

$$f\colon \mathbb{K}^n \to \mathbb{R},\ x \mapsto \|x\|$$

stetig. (Beachte, dass wir $\mathbb{K}^n$ dabei wie üblich mit der euklidischen Norm als normierten Raum auffassen, und nicht mit der gegebenen Norm $\|\cdot\|$!)

Beweis. Es seien $a \in \mathbb{K}^n$ und $\varepsilon \in \mathbb{R}_{>0}$ gegeben. Nach Beispiel 24.3 (b) ist f stetig, wenn wir den Startraum als normierten Raum mit der Norm $\|\cdot\|$ versehen. Es gibt also ein $\delta \in \mathbb{R}_{>0}$, so dass $|f(x) - f(a)| < \varepsilon$ für alle $x \in \mathbb{K}^n$ mit $\|x - a\| < \delta$ gilt.

Nun gilt aber für alle $z \in \mathbb{K}^n$

$$\|z\| = \|z_1 e_1 + \cdots + z_n e_n\| \le |z_1| \cdot \|e_1\| + \cdots + |z_n| \cdot \|e_n\| \le c\,\|z\|_2$$

mit $c := \|e_1\| + \cdots + \|e_n\|$. Also folgt für alle $x \in \mathbb{K}^n$ mit $\|x - a\|_2 < \frac{\delta}{c}$ zunächst die Ungleichung $\|x - a\| \le c\,\|x - a\|_2 < c \cdot \frac{\delta}{c} = \delta$, und damit dann wieder $|f(x) - f(a)| < \varepsilon$. Dies zeigt die behauptete Stetigkeit. □

Satz 24.36 (Äquivalenz von Normen auf $\mathbb{K}^n$). *Alle Normen auf $\mathbb{K}^n$ sind zueinander äquivalent.*

Beweis. Weil die Äquivalenz von Normen eine Äquivalenzrelation ist, genügt es zu zeigen, dass jede Norm $\|\cdot\|$ auf $\mathbb{K}^n$ zur euklidischen Norm äquivalent ist, also dass es zu einer solchen Norm $a, b \in \mathbb{R}_{>0}$ gibt mit

$$a\,\|x\|_2 \le \|x\| \le b\,\|x\|_2 \quad \text{für alle } x \in \mathbb{K}^n. \tag{$*$}$$

Da man außerdem Skalare nach Definition 23.1 (a) aus beiden Normen herausziehen kann, reicht es sogar schon, diese beiden Ungleichungen für alle (euklidisch normierten) Vektoren in der Menge $A := \{x \in \mathbb{K}^n : \|x\|_2 = 1\}$ zu zeigen.

Bezüglich der euklidischen Norm ist A (also der Rand der gewöhnlichen Einheitskugel) aber offensichtlich beschränkt und abgeschlossen, also nach Satz 23.51 kompakt. Die nach Lemma 24.35 stetige Normabbildung $\|\cdot\|$ nimmt damit auf A nach Folgerung 24.34 ein Minimum a und ein Maximum b an – was sofort die behauptete Ungleichung $(*)$ zeigt. Dabei muss auch wirklich $a > 0$ sein, weil $\|x\| = 0$ nach Definition 23.1 (b) nur für $x = 0 \notin A$ gilt. □

Natürlich gilt die Äquivalenz aller Normen mit Satz 24.36 sogar auf beliebigen endlich-dimensionalen Vektorräumen, da diese stets zu einem $\mathbb{K}^n$ isomorph sind. Ein wichtiges Beispiel hierfür sind die Matrizenräume $\mathrm{Mat}(m \times n, \mathbb{K}) \cong \mathbb{K}^{mn}$. Möchte man mit Normen von Vektoren und Matrizen gleichzeitig arbeiten, ist es zur Vereinfachung der Rechnungen allerdings in der Regel geschickt, die dafür gewählten Normen aufeinander abzustimmen. In der Tat zeigen der folgende Satz und die anschließende Folgerung, dass es zu einer gewählten Vektornorm immer eine besonders natürlich definierte passende Matrixnorm gibt, die neben den reinen Normeigenschaften noch weitere nützliche Rechenregeln erfüllt.

Satz und Definition 24.37 (Matrixnormen)**.** *Für gewählte Normen $\|\cdot\|$ auf $\mathbb{K}^k$ für alle $k \in \mathbb{N}_{>0}$ und eine Matrix $A \in \mathrm{Mat}(m \times n, \mathbb{K})$ existiert das Maximum*

$$\|A\| := \max\left\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{K}^n \backslash \{0\}\right\} = \max\left\{\|Ax\| : x \in \mathbb{K}^n \text{ mit } \|x\| = 1\right\}. \qquad (*)$$

*Dies definiert eine Norm auf $\mathrm{Mat}(m \times n, \mathbb{K})$. Man nennt sie die von der gegebenen (Vektor-)Norm induzierte **Matrixnorm** und verwendet für sie die gleiche Bezeichnung, es ist also z. B.*

$$\|A\|_2 := \max\left\{\|Ax\|_2 : x \in \mathbb{K}^n \text{ mit } \|x\|_2 = 1\right\} \quad \text{und} \quad \|A\|_\infty := \max\left\{\|Ax\|_\infty : x \in \mathbb{K}^n \text{ mit } \|x\|_\infty = 1\right\}.$$

Beweis. Zunächst einmal stimmen die beiden in $(*)$ genannten Mengen überein: Da wir jeden Vektor $x \in \mathbb{K}^n \backslash \{0\}$ als $x = \lambda y$ mit $\lambda \in \mathbb{K} \backslash \{0\}$ und $y \in \mathbb{K}^n$ mit $\|y\| = 1$ schreiben können, ist

$$\begin{aligned}\left\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{K}^n \backslash \{0\}\right\} &= \left\{\frac{\|\lambda Ay\|}{\|\lambda y\|} : \lambda \in \mathbb{K} \backslash \{0\}, y \in \mathbb{K}^n \text{ mit } \|y\| = 1\right\} \\ &\overset{23.1\,(a)}{=} \left\{\|Ay\| : y \in \mathbb{K}^n \text{ mit } \|y\| = 1\right\}.\end{aligned}$$

An der hinteren Darstellung dieser Menge sehen wir auch, dass sie das Bild der (nach Satz 23.51 (b)) kompakten Menge $\{x \in \mathbb{K}^n : \|x\| = 1\}$ unter der (nach Beispiel 24.3 (b) und 24.11 (b)) stetigen Abbildung $x \mapsto \|Ax\|$ ist. Nach Folgerung 24.34 besitzt sie also ein Maximum, so dass der im Satz angegebene Ausdruck für $\|A\|$ existiert.

Die Normeigenschaften aus Definition 23.1 sind für diesen Ausdruck schnell überprüft:

(a) Für alle $\lambda \in \mathbb{K}$ ist

$$\|\lambda A\| = \max\{\|\lambda Ax\| : \|x\| = 1\} = |\lambda| \cdot \max\{\|Ax\| : \|x\| = 1\} = |\lambda| \cdot \|A\|.$$

(b) Ist $A \neq 0$, so gibt es einen Vektor $x \in \mathbb{K}^n$ mit $Ax \neq 0$ (z. B. einen Einheitsvektor zu einer Nichtnullspalte von A), also mit $\|Ax\| > 0$, und dementsprechend ist auch $\|A\| > 0$.

(c) Für eine weitere Matrix $B \in \mathrm{Mat}(m \times n, \mathbb{K})$ gilt

$$\begin{aligned}\|A + B\| &= \max\{\|(A+B)x\| : \|x\| = 1\} \\ &\leq \max\{\|Ax\| + \|Bx\| : \|x\| = 1\} \\ &\leq \max\{\|Ax\| : \|x\| = 1\} + \max\{\|Bx\| : \|x\| = 1\} \\ &= \|A\| + \|B\|.\end{aligned}$$

62 □

Folgerung 24.38 (Eigenschaften induzierter Matrixnormen)**.** *Für die von einer Vektornorm $\|\cdot\|$ induzierte Matrixnorm aus Definition 24.37 gilt für alle $A \in \mathrm{Mat}(m \times n, \mathbb{K})$:*

(a) *(Verträglichkeit mit der Vektornorm) Für alle $x \in \mathbb{K}^n$ ist $\|Ax\| \leq \|A\| \cdot \|x\|$;*

(b) *(Submultiplikativität) Für jede weitere Matrix $B \in \mathrm{Mat}(n \times p, \mathbb{K})$ ist $\|AB\| \leq \|A\| \cdot \|B\|$.*

Beweis.

(a) Für $x = 0$ ist diese Aussage trivial; für $x \neq 0$ folgt sie direkt aus der Definition der induzierten Matrixnorm, da dann ja stets $\|A\| \geq \frac{\|Ax\|}{\|x\|}$ gilt.

(b) Nach Definition der induzierten Matrixnorm $\|AB\|$ gibt es ein $x \in \mathbb{K}^p \backslash \{0\}$ mit $\|x\| = 1$ und

$$\|AB\| = \|ABx\| \overset{(a)}{\leq} \|A\| \cdot \|Bx\| \overset{(a)}{\leq} \|A\| \cdot \|B\| \cdot \|x\| = \|A\| \cdot \|B\|. \qquad \square$$

Beispiel 24.39.

(a) Nach Aufgabe 22.50 (a) ist die von der euklidischen Norm induzierte Matrixnorm $\|A\|_2$ einer Matrix $A \neq 0$ gegeben durch den größten Singulärwert von A. Im Fall einer quadratischen, symmetrischen bzw. hermiteschen und positiv semidefiniten Matrix ist dies nach Beispiel 22.46 (a) also genau der größte Eigenwert von A. Da die Menge der Eigenwerte oft als das Spektrum der Matrix bezeichnet wird, nennt man $\|A\|_2$ auch die **Spektralnorm** von A.

Beachte, dass $\|A\|_2$ (trotz dieser Notation) damit also *nicht* die Frobenius-Norm aus Beispiel 23.3 (f) ist, also nicht die euklidische Norm, wenn man $\mathrm{Mat}(m \times n, \mathbb{K})$ mit $\mathbb{K}^{mn}$ identifiziert!

(b) Wir behaupten, dass die von der Maximumsnorm induzierte Matrixnorm einer gegebenen Matrix $A = (a_{i,j})_{i,j} \in \mathrm{Mat}(m \times n, \mathbb{K})$ die sogenannte **Zeilensummennorm**

$$\|A\|_\infty = \max\left\{ \sum_{j=1}^{n} |a_{i,j}| : i = 1, \dots, m \right\} \tag{$*$}$$

ist: Für alle $x \in \mathbb{K}^n$ mit $\|x\|_\infty = 1$ gilt zunächst

$$\begin{aligned}
\|Ax\|_\infty &= \max\left\{ \left| \sum_{j=1}^{n} a_{i,j} x_j \right| : i = 1, \dots, m \right\} \\
&\overset{(1)}{\leq} \max\left\{ \sum_{j=1}^{n} |a_{i,j}| \cdot |x_j| : i = 1, \dots, m \right\} && \text{(Dreiecksungleichung)} \\
&\overset{(2)}{\leq} \max\left\{ \sum_{j=1}^{n} |a_{i,j}| : i = 1, \dots, m \right\}. && (|x_j| \leq 1 \text{ für alle } j)
\end{aligned}$$

Außerdem können wir für ein geeignetes x bei (1) und (2) die Gleichheit erhalten:

- bei (1), indem wir für das i mit maximalem Wert von $\sum_{j=1}^{n} |a_{i,j}| \cdot |x_j|$ den Winkel in Polarkoordinaten von x_j so wählen, dass alle $a_{i,j} x_j$ reell nicht-negativ sind;
- bei (2), indem wir alle x_j vom Betrag 1 wählen.

Dies zeigt den behaupteten Ausdruck $(*)$ für $\|A\|_\infty = \max\left\{ \|Ax\|_\infty : x \in \mathbb{K}^n \text{ mit } \|x\|_\infty = 1 \right\}$.

Aufgabe 24.40. Es sei $n \in \mathbb{N}_{>0}$. Man zeige:

(a) Die Menge

$$\left\{ \begin{pmatrix} x_1^2 \\ x_2 - x_1 \end{pmatrix} : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \right\}$$

ist kompakt in $\mathbb{R}^2$.

(b) Die Menge $\mathrm{O}(n)$ aller orthogonalen Matrizen ist kompakt in $\mathrm{Mat}(n \times n, \mathbb{R})$.

(c) Die Menge aller indefiniten Matrizen ist offen im Raum aller symmetrischen $n \times n$-Matrizen über $\mathbb{R}$.

Aufgabe 24.41 (Abstände von Mengen)**.** Wir definieren den *Abstand* zweier nicht-leerer abgeschlossener Teilmengen A und B eines metrischen Raumes M als

$$d(A, B) := \inf\{d(a, b) : a \in A, b \in B\} \quad \in \mathbb{R}_{\geq 0}.$$

Man zeige:

(a) Sind A und B kompakt, so ist dieses Infimum ein Minimum.

(b) Ist nur A kompakt, aber M vollständig, so ist dieses Infimum ebenfalls ein Minimum.

(c) Im Allgemeinen ist dieses Infimum kein Minimum, selbst wenn A kompakt oder M vollständig ist.

Aufgabe 24.42 (Konvergenz von Mengen)**.** Zu einem gegebenen metrischen Raum M sei $\mathscr{K}(M)$ die Menge aller nicht-leeren kompakten Teilmengen von M. Wir definieren die folgenden Abstandsfunktionen:

$$\begin{array}{ll} \text{für } a \in M \text{ und } B \in \mathscr{K}(M) \text{ sei} & d(a,B) := \min\{d(a,b) : b \in B\} \text{ wie in Aufgabe 24.41 (a);} \\ \text{für } A,B \in \mathscr{K}(M) \text{ sei} & h(A,B) := \max\left(\max\{d(a,B) : a \in A\}, \max\{d(b,A) : b \in B\}\right). \end{array}$$

Man zeige:

(a) Die Maxima in der Definition von h existieren.

(b) Die Abbildung h ist eine Metrik auf $\mathscr{K}(M)$. (Man nennt sie die *Hausdorff-Metrik*; sie misst, wie verschieden zwei Mengen voneinander sind.)

(c) Für $M = \mathbb{R}^2$ mit der euklidischen Metrik gilt $\lim\limits_{n\to\infty} K_{1+\frac{1}{n}}(0) = K_1(0)$ in $\mathscr{K}(M)$.

24.D Peano-Kurven

Zum Abschluss dieses Kapitels wollen wir eine interessante Anwendung betrachten, in der die meisten der wichtigen Sätze dieses und des vorangegangenen Kapitels geschickt miteinander kombiniert werden, um ein sehr erstaunliches Resultat zu beweisen. Es handelt sich dabei nicht so sehr um ein Resultat, das für unser weiteres Studium der Analysis von großer Bedeutung wäre, sondern einfach nur um ein sehr „schönes“ Stück Mathematik, das so verblüffend ist, dass man es fast schon als Paradoxon bezeichnen könnte: Wir werden zeigen, dass es eine stetige Abbildung vom Einheitsintervall $[0,1] \subset \mathbb{R}$ nach $\mathbb{R}^2$ (also einen Weg in der Ebene im Sinne von Definition 24.20) gibt, deren Bild das *gesamte* „Einheitsquadrat“ $[0,1] \times [0,1] \subset \mathbb{R}^2$ ist. Wir können also mit einem “eindimensionalen Objekt“ (nämlich dem Einheitsintervall) durch eine stetige Abbildung ein „zweidimensionales Objekt“ (das Einheitsquadrat) komplett ausfüllen. Spätestens hier sehen wir also, dass man mit der oft gehörten Interpretation einer stetigen Abbildung in einer Variablen als etwas, das man zeichnen kann, ohne den Stift abzusetzen, etwas vorsichtig sein muss.

In diesem Abschnitt bezeichne stets $I := [0,1] \subset \mathbb{R}$ das Einheitsintervall und $Q := I \times I = [0,1] \times [0,1]$ das Einheitsquadrat.

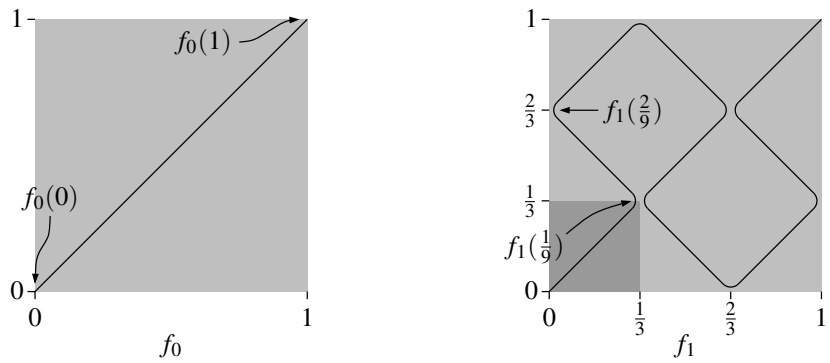
Satz 24.43 (**Peano-Kurve**)**.** *Es gibt eine stetige surjektive Abbildung $f\colon I \to Q$. (Eine solche Abbildung wird als Peano-Kurve bezeichnet.)*

Beweis. Der Beweis dieses Satzes ist konstruktiv, gibt also eine mögliche Peano-Kurve explizit an. Sie wird als Grenzwert einer rekursiv definierten Folge von Funktionen $f_n\colon I \to Q$ konstruiert.

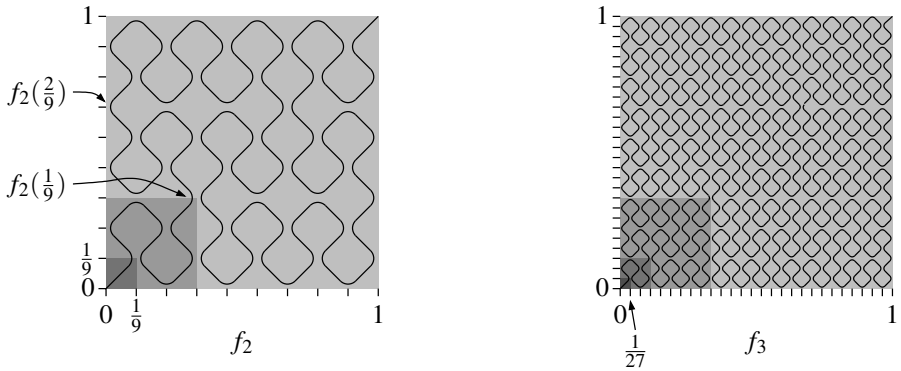
Die Funktionenfolge $(f_n)_n$ wird dabei wie folgt gebildet: Die erste Funktion f_0 ist einfach die Gerade

$$f_0\colon I \to Q,\ f_0(x) = \begin{pmatrix} x \\ x \end{pmatrix},$$

die das Einheitsquadrat auf der Diagonalen von links unten nach rechts oben durchläuft. Für die nächste Funktion f_1 teilen wir Q in 9 gleich große Teilquadrate entlang der horizontalen und vertikalen Linien bei $\frac{1}{3}$ und $\frac{2}{3}$, und durchlaufen nun diese 9 Teilquadrate der Reihe nach entlang ihrer Diagonalen wie im Bild unten dargestellt. Der Weg f_1 besteht also aus 9 Geradenstücken, die alle „mit gleicher Geschwindigkeit“ durchlaufen werden – er ist im Bild an den Ecken nur deswegen abgerundet eingezeichnet, damit man seinen Verlauf besser erkennen kann.



Der nächste Weg f_2 entsteht nun aus f_1, indem wir jedes der 9 Geradenstücke von f_1 (wie etwa das in dem oben dunkel eingezeichneten Teilquadrat) durch einen Weg ersetzen, der selbst wieder wie f_1 aussieht. Entsprechend ersetzen wir dann jedes der Geradenstücke von f_2 durch einen Weg wie f_1, um f_3 zu erhalten. Setzen wir dieses Verfahren fort, so erhalten wir eine Folge $(f_n)_n$ von Wegen $f_n\colon I \to Q$. Die Wege f_2 und f_3 sind im folgenden Bild eingezeichnet.

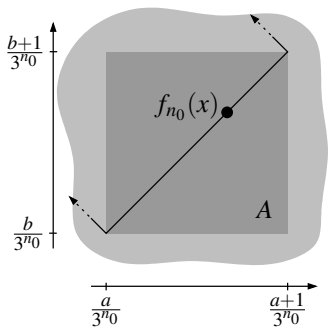


Wir wollen nun $f(x) := \lim_{n\to\infty} f_n(x)$ setzen und müssen uns dazu natürlich als Erstes davon überzeugen, dass dieser Grenzwert überhaupt existiert.

Es sei dazu $\varepsilon > 0$ beliebig. Wir wählen ein $n_0 \in \mathbb{N}$ mit $\frac{1}{3^{n_0}} < \varepsilon$. Für ein beliebiges $x \in I$ liegt $f_{n_0}(x)$ nun wie im Bild rechts auf einem der Geradenstücke, die diagonal durch eines der Teilquadrate von f_{n_0} laufen. Dieses Teilquadrat hat nach Konstruktion die Seitenlänge $\frac{1}{3^{n_0}}$ und ist von der Form

$$A = \left[\frac{a}{3^{n_0}}, \frac{a+1}{3^{n_0}}\right] \times \left[\frac{b}{3^{n_0}}, \frac{b+1}{3^{n_0}}\right] \subset \mathbb{R}^2$$

für gewisse natürliche Zahlen $0 \le a, b < 3^{n_0}$. Nach unserer rekursiven Konstruktion der Funktionen f_n ist dann klar, dass $f_n(x)$ auch für alle $n \ge n_0$ in dem Quadrat A liegt – wir ändern den Teil des Weges f_{n_0} in A für die weiteren n ja nur noch *innerhalb dieses Quadrates* A ab.



Für alle $n, m \ge n_0$ liegen $f_n(x)$ und $f_m(x)$ also innerhalb eines Quadrates mit Seitenlänge $\frac{1}{3^{n_0}}$, d. h. es ist

$$\|f_n(x) - f_m(x)\|_\infty \le \frac{1}{3^{n_0}} < \varepsilon \qquad (*)$$

für alle $n, m \ge n_0$. Dies bedeutet genau, dass die Folge $(f_n(x))_n$ für festes x eine Cauchyfolge in Q ist. Als abgeschlossene Teilmenge von $\mathbb{R}^2$ ist Q nach Satz 23.29 und Folgerung 23.43 nun aber vollständig – d. h. diese Cauchyfolge konvergiert für alle x gegen einen Punkt in Q, und wir können

unsere Funktion $f\colon I \to Q$ damit in der Tat wie gewünscht durch

$$f(x) := \lim_{n\to\infty} f_n(x)$$

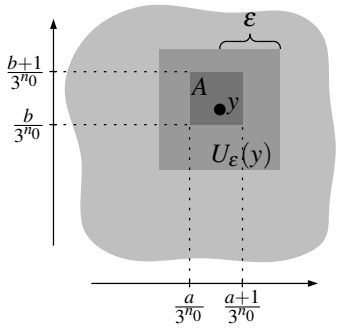
definieren. Es bleibt jetzt nur noch zu zeigen, dass f auch stetig und surjektiv ist.

Für die Stetigkeit nehmen wir den Grenzwert von $(*)$ für $m \to \infty$ und erhalten für alle $x \in I$

$$\|f_n(x) - f(x)\|_\infty \le \frac{1}{3^{n_0}} < \varepsilon$$

(beachte, dass die Normfunktion nach Beispiel 24.3 (b) stetig ist und wir die Grenzwertbildung damit nach dem Folgenkriterium für Stetigkeit aus Satz 24.4 (b) in die Normstriche hineinziehen dürfen). Dies sagt uns noch einmal, dass $f_n(x)$ für $n \to \infty$ gegen $f(x)$ konvergiert – der entscheidende Punkt ist nun aber, dass unser oben gewähltes n_0 nur von ε und nicht vom betrachteten Punkt x abhängt. Die Funktionenfolge $(f_n)_n$ ist also sogar gleichmäßig konvergent. Nach Bemerkung 24.25 (b) ist die Grenzfunktion f als gleichmäßiger Grenzwert stetiger Funktionen damit stetig.

Für die Surjektivität erinnern wir uns daran, dass nach Satz 24.29 mit I auch das Bild $f(I)$ unter der stetigen Abbildung f kompakt ist. Nach Satz 23.51 (a) ist $f(I)$ also insbesondere abgeschlossen, d. h. das Komplement $\mathbb{R}^2 \backslash f(I)$ ist offen. Angenommen, f würde nun nicht surjektiv auf das Einheitsquadrat Q abbilden, d. h. es gäbe einen Punkt $y \in Q \backslash f(I)$. Da $\mathbb{R}^2 \backslash f(I)$ offen ist, gäbe es dann eine Umgebung $U_\varepsilon(y)$ von y, die ganz in $\mathbb{R}^2 \backslash f(I)$ liegt. Wir verwenden dabei wieder die Maximumsnorm, so dass diese Umgebung wie im Bild rechts ein (offenes) Quadrat mit Seitenlänge 2ε ist.



Wählen wir nun wieder unser $n_0 \in \mathbb{N}$ mit $\frac{1}{3^{n_0}} < \varepsilon$, so muss dann mindestens ein Quadrat

$$A = \left[\frac{a}{3^{n_0}}, \frac{a+1}{3^{n_0}}\right] \times \left[\frac{b}{3^{n_0}}, \frac{b+1}{3^{n_0}}\right] \subset \mathbb{R}^2$$

der zu f_{n_0} gehörigen Unterteilung von Q ganz in $U_\varepsilon(y)$ liegen, enthält also ebenfalls keinen Punkt des Weges f. Dies ist aber ein Widerspruch, da f nach unserer Konstruktion natürlich durch jedes solche Quadrat irgendwie hindurch läuft, also insbesondere einen Punkt in A enthalten muss. Also war unsere Annahme falsch, und f ist in der Tat surjektiv. □

Nachdem wir nun gezeigt haben, dass es eine stetige surjektive Abbildung $I \to Q$ gibt, würde man vielleicht vermuten, dass es umgekehrt auch eine stetige injektive Abbildung $Q \to I$ gibt. Wie wir jetzt sehen wollen, ist dies jedoch erstaunlicherweise falsch. Das gleiche Argument zeigt dann auch, dass keine *bijektive* Peano-Kurve existiert.

Satz 24.44.

(a) *Es gibt keine stetige injektive Abbildung $Q \to I$.*

(b) *Es gibt keine stetige bijektive Abbildung $I \to Q$ (also keine bijektive Peano-Kurve).*

Beweis.

(a) Angenommen, $f\colon Q \to I$ wäre stetig und injektiv. Mit Q wäre dann nach Satz 24.29 und 24.22 (a) auch das Bild $f(Q) \subset I$ kompakt und wegzusammenhängend, also ein abgeschlossenes Intervall $[a,b]$. Damit ist die Einschränkung $f\colon Q \to [a,b]$ stetig und bijektiv.

Wir nehmen nun aus dem Intervall $[a,b]$ den Mittelpunkt $c := \frac{a+b}{2}$ heraus, und dementsprechend aus dem Quadrat Q den Urbildpunkt $f^{-1}(c)$. Die dadurch entstehende Einschränkung $f|_{Q\backslash f^{-1}(c)}\colon Q\backslash f^{-1}(c) \to [a,b]\backslash\{c\}$ ist dann natürlich immer noch stetig (und bijektiv). Dies ist aber ein Widerspruch zum Zwischenwertsatz 24.22 (a), da $Q\backslash\{f^{-1}(c)\}$ immer noch wegzusammenhängend ist, $[a,b]\backslash\{c\}$ aufgrund des fehlenden Mittelpunkts jedoch nicht.

(b) Da I kompakt ist, wäre nach Satz 24.31 mit einer stetigen bijektiven Abbildung $I \to Q$ auch ihre Umkehrung $Q \to I$ stetig und bijektiv, im Widerspruch zu (a). □

Aufgabe 24.45. Es sei $I = [0,1] \subset \mathbb{R}$. Gibt es eine stetige, surjektive Abbildung ...

(a) von I nach $I \times I \times I \subset \mathbb{R}^3$;

(b) von I nach $\mathbb{R}^2$;

(c) von $[0,1)$ nach $\mathbb{R}^2$;

(d) von I in den abgeschlossenen Einheitskreis $K_1(0) = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$?

(Hinweis: In allen Fällen oben, in denen eine solche Abbildung existiert, lässt sie sich explizit durch die Peano-Kurve $f\colon I \to I \times I$ aus Satz 24.43 ausdrücken, ohne diese aufwändige Konstruktion noch einmal zu wiederholen.)

63

25. Differenzierbarkeit im Mehrdimensionalen

Wie im eindimensionalen Fall in Kapitel 10 wollen wir uns nach der Stetigkeit von Abbildungen jetzt mit der Differenzierbarkeit beschäftigen. Wir erinnern uns dazu zunächst einmal daran, wie wir differenzierbare Funktionen damals definiert hatten: Hat D keine isolierten Punkte, ist $f\colon D \to \mathbb{K}$ eine Funktion und $a \in D$, so heißt f differenzierbar in a, wenn der Grenzwert des Differenzenquotienten

$$f'(a) := \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \quad \in \mathbb{K} \qquad (*)$$

(der dann die Ableitung von f in a genannt wird) existiert. Die anschauliche Idee hinter dieser Definition war, dass dann für x in der Nähe von a die Approximation

$$f'(a) \approx \frac{f(x) - f(a)}{x - a}, \quad \text{und damit} \quad f(x) \approx f(a) + f'(a) \cdot (x - a)$$

gilt, dass wir f also in der Nähe von a durch eine Gerade $x \mapsto f(a) + f'(a) \cdot (x - a)$ mit Steigung $f'(a)$ annähern. Wollen wir diese Idee der linearen Approximierbarkeit nun auf andere Start- und Zielbereiche verallgemeinern, stellen wir zunächst fest:

- Um lineare Approximierbarkeit formulieren zu können, müssen wir offensichtlich wissen, was lineare Abbildungen zwischen den betrachteten Räumen überhaupt sind. In metrischen Räumen wäre dies z. B. nicht der Fall, da dort im Allgemeinen ja nicht einmal eine Addition bzw. Multiplikation definiert ist. Wir werden uns daher für die Untersuchung der Differenzierbarkeit auf normierte Räume (bzw. Teilmengen davon) beschränken müssen.
- Betrachten wir als Beispiel einmal als Start- und Zielraum die normierten Räume $\mathbb{K}^n$ bzw. $\mathbb{K}^m$, so ist nun die Idee der linearen Approximierbarkeit $f(x) \approx f(a) + f'(a) \cdot (x - a)$ für $x \in \mathbb{K}^n$ in der Nähe eines gegebenen Punktes $a \in \mathbb{K}^n$ sinnvoll formulierbar, wenn wir dies als Gleichung im Zielraum $\mathbb{K}^m$ auffassen, $f'(a) \in \mathrm{Mat}(m \times n, \mathbb{K})$ eine Matrix ist und „$\cdot$“ das Matrixprodukt mit dem Vektor $x - a \in \mathbb{K}^n$ darstellt. Ableitungen werden im Mehrdimensionalen also zu Matrizen werden.
- Die Ableitungsmatrix $f'(a)$ wird uns dann von a aus für jeden kleinen Schritt $x - a$ in einer beliebigen Richtung angeben, um welchen Wert $f'(a) \cdot (x - a)$ sich f dabei ändert. Damit dies eindeutig bestimmt ist, muss man sich dem Punkt a aber offensichtlich innerhalb des Definitionsbereichs überhaupt erst einmal aus jeder Richtung nähern können. Im Gegensatz zum Eindimensionalen werden wir im Folgenden daher nur offene Definitionsbereiche betrachten, also nicht über Differenzierbarkeit in Randpunkten sprechen.

Mit diesen Vorüberlegungen werden wir im folgenden Abschnitt nun die Ableitung im Mehrdimensionalen definieren und angeben, wie man sie berechnen kann.

25.A Differenzierbare Abbildungen

Wollen wir die Definition der Ableitung ins Mehrdimensionale übertragen, haben wir zunächst ein weiteres Problem: Wir können keinen Differenzenquotienten wie in $(*)$ in der Einleitung oben bilden, da wir hierzu durch den Vektor $x - a$ teilen müssten – was natürlich nicht möglich ist. Daher werden wir als Erstes nun die eindimensionale Definition so umschreiben, dass sie sich besser auf die neue Situation übertragen lässt.

Lemma 25.1 (Differenzierbarkeit als lineare Approximierbarkeit)**.** *Es seien $D \subset \mathbb{K}$ offen, $f\colon D \to \mathbb{K}$ eine Abbildung und $a \in D$. Dann ist f genau dann in a differenzierbar mit Ableitung $c = f'(a) \in \mathbb{K}$, wenn es eine Funktion $r\colon D \to \mathbb{K}$ gibt mit*

(a) $f(x) = f(a) + c\,(x - a) + r(x)$ *für alle $x \in D$, und*

(b) $\lim\limits_{x\to a} \dfrac{r(x)}{x-a} = 0.$

Beweis. Die Bedingung (a) besagt offensichtlich, dass r die Funktion $r(x) = f(x) - f(a) - c\,(x-a)$ ist. Damit gelten die beiden Bedingungen (a) und (b) genau dann, wenn

$$\lim_{x\to a} \frac{f(x)-f(a)-c\,(x-a)}{x-a} = 0, \quad \text{also} \quad \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a} - c\right) = 0$$

ist, d. h. wenn f differenzierbar in a mit Ableitung c ist. □

Bemerkung 25.2. Man kann sich die Funktion r in Lemma 25.1 als „Restfunktion“ vorstellen, also als Differenz zwischen der eigentlichen Funktion f und ihrer Näherung $x \mapsto f(a) + c\,(x-a)$ bei a. Natürlich erhalten wir aus (a) am gegebenen Punkt $x = a$ in jedem Fall schon einmal $r(a) = 0$. Die Bedingung (b) besagt nun gerade, dass dieser Rest auch „in der Nähe des Punktes a sehr klein“, die Näherung also dort sehr gut ist: Der Quotient $\frac{r(x)}{x-a}$ ist ja bei a von der Form „$\frac{0}{0}$“, und die gegebene Bedingung besagt nun gerade, dass sich hier für $x \to a$ der Zähler $r(x)$ gegen den Nenner $x-a$ durchsetzt. Man sagt dafür auch, dass „die Restfunktion r für $x \to a$ schneller als linear gegen 0 konvergiert“.

Auf den ersten Blick sieht es nun vielleicht so aus, als ob wir mit Lemma 25.1 im Hinblick auf eine Verallgemeinerung auf das Mehrdimensionale noch nicht viel erreicht haben, weil wir in (b) ja immer noch durch $x-a$ dividieren. Dies ist aber nicht so: Die gegebene Bedingung ist ja offensichtlich äquivalent zu $\lim_{x\to a} \frac{r(x)}{|x-a|} = 0$, und dort können wir nun einfach die Betragsstriche durch die Norm ersetzen. Wir erhalten so die folgende Definition der Differenzierbarkeit, die sich von der obigen Bedingung nur dadurch unterscheidet, dass c nun eine $m \times n$-Matrix und $x - a \mapsto c\,(x-a)$ damit eine allgemeine lineare Abbildung vom Startraum $\mathbb{K}^n$ in den Zielraum $\mathbb{K}^m$ (statt wie bisher von $\mathbb{K}$ nach $\mathbb{K}$) ist.

Definition 25.3 (Differenzierbarkeit in $\mathbb{K}^n$)**.** Es seien $D \subset \mathbb{K}^n$ offen und $a \in D$. Eine Abbildung $f\colon D \to \mathbb{K}^m$ heißt **(total) differenzierbar** in a, wenn es eine Matrix $A \in \mathrm{Mat}(m \times n, \mathbb{K})$ und eine Abbildung $r\colon D \to \mathbb{K}^m$ gibt, so dass

(a) $f(x) = f(a) + A\,(x-a) + r(x)$ für alle $x \in D$, und

(b) $\lim\limits_{x\to a} \dfrac{r(x)}{\|x-a\|} = 0.$

Wir werden in Folgerung 25.12 noch sehen, dass die Matrix A (und damit auch r) in diesem Fall eindeutig bestimmt ist. Wir nennen sie dann die **(totale) Ableitung** von f in a und bezeichnen sie wie üblich mit $f'(a) \in \mathrm{Mat}(m \times n, \mathbb{K})$ (in manchen Büchern ist hierfür auch die Bezeichnung $Df(a)$ üblich, da Matrizen oftmals mit großen Buchstaben bezeichnet werden).

Die Abbildung f heißt **differenzierbar**, wenn sie in jedem Punkt $a \in D$ differenzierbar ist.

Bemerkung 25.4.

(a) Mit dem gleichen Argument wie z. B. bei der Beschränktheit in Bemerkung 23.21 (b) sieht man auch hier, dass Definition 25.3 nicht von der verwendeten Norm abhängt.

(b) Im eindimensionalen Fall $m = n = 1$ stimmt Definition 25.3 aufgrund von Lemma 25.1 offensichtlich mit der alten Definition 10.3 der Differenzierbarkeit überein, wenn wir die 1×1-Matrix $f'(a)$ als Element von $\mathbb{K}$ auffassen.

(c) Ist $f\colon D \to \mathbb{K}^m$ differenzierbar, so können wir die Ableitung ihrerseits wieder als Funktion

$$f'\colon D \to \mathrm{Mat}(m \times n, \mathbb{K}),\ x \mapsto f'(x)$$

auffassen. Beachte allerdings, dass diese Ableitungsfunktion im Gegensatz zum Eindimensionalen zwar die gleiche Startmenge D, aber nicht die gleiche Zielmenge wie die ursprüngliche Funktion $f\colon D \to \mathbb{K}^m$ hat!

Beispiel 25.5 (Differenzierbarkeit linearer Abbildungen)**.** Ist $f\colon \mathbb{R}^n \to \mathbb{R}^m$ selbst bereits eine lineare Abbildung, also $f(x) = Ax$ für eine Matrix $A \in \mathrm{Mat}(m \times n, \mathbb{R})$, so gilt $f(x) = f(a) + A(x-a)$ für alle $x \in \mathbb{R}^n$. Die Bedingungen aus Definition 25.3 sind dann also mit der Nullfunktion als Restfunktion r erfüllt. Also ist f in diesem Fall differenzierbar mit (von x unabhängiger) Ableitung $f'(x) = A$.

Bemerkung 25.6 (Differenzierbarkeit in normierten Räumen)**.** Genau wie bei unserer Untersuchung linearer Abbildungen in Kapitel 16 haben wir den Begriff der Differenzierbarkeit hier zunächst einmal nur für Abbildungen zwischen (Teilmengen von) $\mathbb{K}^n$ und $\mathbb{K}^m$ definiert. Wie ihr euch vielleicht schon denken könnt, ist eine analoge Definition aber genauso in beliebigen normierten Räumen möglich, wobei die Ableitung dann statt einer Matrix eine lineare Abbildung zwischen den gegebenen Räumen ist: Sind V und W normierte Räume und $f\colon D \to W$ eine Abbildung auf einer offenen Teilmenge $D \subset V$, so nennt man f in einem Punkt $a \in D$ *differenzierbar*, wenn es eine *stetige* lineare Abbildung $\Phi \in \mathrm{Hom}(V,W)$ und eine Abbildung $r\colon D \to W$ gibt mit

(a) $f(x) = f(a) + \Phi(x-a) + r(x)$ für alle $x \in D$, und

(b) $\displaystyle\lim_{x\to a} \frac{r(x)}{\|x-a\|} = 0.$

Auch in diesem Fall heißt Φ dann die *Ableitung* $f'(a) \in \mathrm{Hom}(V,W)$ von f in a.

Für $V = \mathbb{K}^n$ und $W = \mathbb{K}^m$ ist dies aufgrund des Isomorphismus $\mathrm{Mat}(m \times n, \mathbb{K}) \cong \mathrm{Hom}(\mathbb{K}^n, \mathbb{K}^m)$ aus Satz 16.11 offensichtlich äquivalent zu Definition 25.3. Auch für beliebige endlich erzeugte normierte Räume können wir diese neue Definition nach Wahl von Basen von V und W noch auf Definition 25.3 zurückführen, da diese Basen ja Identifikationen von V und W mit $\mathbb{K}^n$ und $\mathbb{K}^m$ erzeugen. Beachte aber, dass dieser verallgemeinerte Begriff der Differenzierbarkeit in nicht endlich erzeugten normierten Räumen in der Regel von den gewählten Normen abhängen wird, und dass die Stetigkeit der linearen Abbildung $f'(a) \in \mathrm{Hom}(V,W)$ dann auch eine echte Bedingung ist (siehe Aufgabe 24.12).

Für den Rest dieser Vorlesung werden wir der Einfachheit halber aber fast ausschließlich mit Abbildungen zwischen Teilmengen von $\mathbb{K}^n$ und $\mathbb{K}^m$ arbeiten, und können die Ableitung solcher Funktionen daher wie in Definition 25.3 als Matrizen schreiben.

Als Erstes wollen wir nun wie im Eindimensionalen (siehe Folgerung 10.7) als einfache Folgerung aus Definition 25.3 sehen, dass jede differenzierbare Funktion auch stetig ist.

Lemma 25.7. *Sind $D \subset \mathbb{K}^n$ offen und $f\colon D \to \mathbb{K}^m$ eine in einem Punkt $a \in D$ differenzierbare Funktion, so ist f auch stetig in a.*

Beweis. Erfüllt f die Differenzierbarkeitsbedingung aus Definition 25.3, so gilt

$$\lim_{\substack{x\to a\\ x\neq a}} f(x) = \lim_{\substack{x\to a\\ x\neq a}} \Big(f(a) + f'(a)\underbrace{(x-a)}_{\to 0} + \underbrace{\frac{r(x)}{\|x-a\|}}_{\to 0}\underbrace{\|x-a\|}_{\to 0} \Big) = f(a)$$

und damit auch $\lim\limits_{x\to a} f(x) = f(a)$, d. h. f ist stetig in a. □

Um Beispiele von differenzierbaren Funktionen und Ableitungen angeben zu können, müssen wir uns nun als Nächstes fragen, wie man Ableitungen überhaupt berechnen kann. Beachte, dass dies nicht offensichtlich ist, da Definition 25.3 die Ableitung im Gegensatz zum Eindimensionalen ja nicht über eine konkrete Formel, sondern als eine Matrix mit einer bestimmten Eigenschaft definiert – aber nichts darüber aussagt, wie man eine solche Matrix finden kann. Über das folgende Konzept der Richtungsableitungen können wir dieses Problem auf den uns bekannten eindimensionalen Fall zurückführen.

Definition 25.8 (Richtungsableitungen und partielle Ableitungen)**.** Es seien $D \subset \mathbb{K}^n$ offen, $a \in D$ und $f\colon D \to \mathbb{K}^m$ eine Abbildung. Ferner wählen wir einen Vektor $v \in \mathbb{K}^n \setminus \{0\}$.

(a) Existiert der Grenzwert

$$\partial_v f(a) := \lim_{t\to 0} \frac{f(a+t\,v)-f(a)}{t} \qquad \in \mathbb{K}^m,$$

so nennen wir ihn die **Richtungsableitung** von f in a in Richtung v. In der Literatur schreibt man hierfür manchmal auch $\frac{\partial f}{\partial v}(a)$.

(b) Ist speziell $v = e_i$ der i-te Einheitsvektor für ein $i \in \{1,\dots,n\}$, so heißt die entsprechende Richtungsableitung

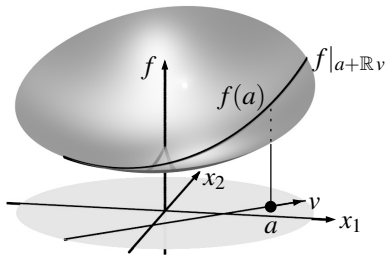
$$\partial_i f(a) := \partial_{e_i} f(a) = \lim_{t\to 0} \frac{f(a+t\,e_i)-f(a)}{t} \qquad \in \mathbb{K}^m$$

die i-te **partielle Ableitung** von f in a; hierfür ist auch die Schreibweise $\frac{\partial f}{\partial x_i}(a)$ üblich. Existieren alle diese partiellen Ableitungen (aber nicht notwendig alle Richtungsableitungen) von f in a, so nennt man f in a **partiell differenzierbar**.

Die Funktion f heißt *partiell differenzierbar*, wenn sie in jedem Punkt $a \in D$ partiell differenzierbar ist.

Bemerkung 25.9.

(a) Im Fall $m = 1$ und $\mathbb{K} = \mathbb{R}$ ist die Richtungsableitung leicht geometrisch zu interpretieren: Nach Definition 25.8 (a) ist sie einfach die gewöhnliche Ableitung der Funktion $t \mapsto f(a+tv)$ in einer Variablen t im Punkt 0, also wie im Bild rechts die Steigung der Funktion f im Punkt a, wenn man sie auf die Gerade $a+\mathbb{R}\,v$ durch a mit Richtungsvektor v einschränkt (und t als Koordinate auf dieser Geraden wählt). Die Richtungsableitung $\partial_v f(a)$ gibt also an, wie stark sich f ändert, wenn man sich von a aus ein kleines Stück in Richtung v bewegt.



(b) Im Gegensatz zur (totalen) Ableitung haben die Richtungsableitungen (und damit auch die partiellen Ableitungen) einer Funktion $f\colon D \to \mathbb{K}^m$ den Vorteil, dass sie im Fall ihrer Existenz in jedem Punkt selbst wieder Funktionen $\partial_v f\colon D \to \mathbb{K}^m$ mit der gleichen Start- und Zielmenge wie f sind (vergleiche Bemerkung 25.4 (c)).

(c) Im Fall $m = 1$ einer Abbildung mit Zielraum $\mathbb{K}$ lassen sich die partiellen Ableitungen von f leicht mit unseren Rechenregeln aus Kapitel 10 berechnen: Es ist dann nämlich

$$\partial_i f(a) = \lim_{t\to 0} \frac{1}{t} \cdot \left(f\begin{pmatrix} a_1 \\ \vdots \\ a_i+t \\ \vdots \\ a_n \end{pmatrix} - f\begin{pmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{pmatrix} \right)$$

genau die gewöhnliche Ableitung von f nach der Variablen x_i im Punkt a_i, wenn man die übrigen Variablen als konstant ansieht. Aber auch im allgemeinen Fall $m > 1$ ist die Berechnung der partiellen Ableitungen einfach: Da sich der Grenzwert aus Definition 25.8 dann nach Lemma 24.7 koordinatenweise berechnen lässt, ist $\partial_i f(a)$ einfach der Vektor mit Koordinaten $\partial_i f_j(a)$ für $j = 1,\dots,m$, wenn $f_1,\dots,f_m\colon D \to \mathbb{K}$ die Komponentenfunktionen von $f\colon D \to \mathbb{K}^m$ sind.

Beispiel 25.10. Um die partiellen Ableitungen der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 x_2 \\ x_1 \sin x_2 \end{pmatrix}$$

zu berechnen, müssen wir nach Bemerkung 25.9 (c) einfach nur die Komponentenfunktionen von f nach x_1 bzw. x_2 differenzieren und dabei jeweils die andere Variable als Konstante ansehen. Wir erhalten also mit den üblichen Rechenregeln aus Kapitel 10

$$\partial_1 f(x) = \begin{pmatrix} x_2 \\ \sin x_2 \end{pmatrix} \quad \text{und} \quad \partial_2 f(x) = \begin{pmatrix} x_1 \\ x_1 \cos x_2 \end{pmatrix}.$$

Insbesondere ist f also partiell differenzierbar.

Die partiellen Ableitungen einer Funktion können in der Regel also sehr einfach berechnet werden. Um zu sehen, wie sie uns bei der Bestimmung der totalen Differenzierbarkeit bzw. der Ableitungsmatrix helfen, benötigen wir das folgende einfache Lemma.

Lemma 25.11 (Differenzierbare Funktionen haben Richtungsableitungen)**.** *Es seien $D \subset \mathbb{K}^n$ offen, $f\colon D \to \mathbb{K}^m$ eine Funktion, und $a \in D$. Ist f in a differenzierbar, so existiert für jedes $v \in \mathbb{K}^n \backslash \{0\}$ die Richtungsableitung $\partial_v f(a)$, und es gilt*

$$\partial_v f(a) = f'(a) \cdot v$$

(wobei $f'(a) \cdot v$ das Matrixprodukt von $f'(a) \in \mathrm{Mat}(m \times n, \mathbb{K})$ mit $v \in \mathbb{K}^n$ ist).

Beweis. Setzen wir die Darstellung $f(x) = f(a) + f'(a)(x-a) + r(x)$ aus Definition 25.3 an der Stelle $x = a + t\,v$ in die Formel für die Richtungsableitung aus Definition 25.8 (a) ein, so erhalten wir

$$\partial_v f(a) = \lim_{t \to 0} \frac{f(a) + f'(a)(a + t\,v - a) + r(a + t\,v) - f(a)}{t} = f'(a) \cdot v + \underbrace{\lim_{t \to 0} \frac{r(a + t\,v)}{t}}_{(*)}.$$

Es bleibt also nur noch zu zeigen, dass der Grenzwert $(*)$ gleich 0 ist. Dies können wir z. B. mit dem Folgenkriterium tun: Es sei dazu $(t_n)_n$ eine beliebige Nullfolge in $\mathbb{K} \backslash \{0\}$, so dass $x_n := a + t_n\,v$ für alle n in D liegt. Dann gilt natürlich $x_n \to a$ für $n \to \infty$, und wegen $|t_n| = \frac{\|x_n - a\|}{\|v\|}$ haben wir

$$\lim_{n \to \infty} \frac{r(a + t_n\,v)}{|t_n|} = \|v\| \cdot \lim_{n \to \infty} \frac{r(x_n)}{\|x_n - a\|} = 0$$

nach (dem Folgenkriterium aus Satz 24.4 angewendet auf) Bedingung (b) aus Definition 25.3. Wiederum mit dem Folgenkriterium bedeutet dies nun aber $\lim_{t \to 0} \frac{r(a + t\,v)}{|t|} = 0$ und damit auch $(*) = 0$, was zu zeigen war. □

Folgerung 25.12 (Differenzierbare Funktionen sind partiell differenzierbar)**.** *Es sei $D \subset \mathbb{K}^n$ offen. Ist eine Funktion $f\colon D \to \mathbb{K}^m$ in einem Punkt $a \in D$ (total) differenzierbar, so ist f dort auch partiell differenzierbar, und es gilt*

$$f'(a) = (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,),$$

d. h. die Spalten der Ableitungsmatrix $f'(a)$ sind gerade die partiellen Ableitungen von f. Insbesondere ist die Ableitung $f'(a)$ also eindeutig bestimmt.

Beweis. Da die partiellen Ableitungen Spezialfälle der Richtungsableitungen sind, folgt die partielle Differenzierbarkeit sofort aus Lemma 25.11. Darüber hinaus ist die i-te Spalte der Matrix $f'(a)$ gerade $f'(a) \cdot e_i$ für alle $i = 1, \ldots, n$, nach Lemma 25.11 also gleich $\partial_{e_i} f(a) = \partial_i f(a)$. □

64

Definition 25.13 (Jacobi-Matrix)**.** Ist die Funktion $f\colon D \to \mathbb{K}^m$ mit Komponentenfunktionen $f_1, \ldots, f_m\colon D \to \mathbb{K}$ in einem Punkt $a \in D$ partiell differenzierbar, so nennt man die Matrix

$$Jf(a) := (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,) \;=\; (\partial_j f_i(a))_{i,j} \quad \in \mathrm{Mat}(m \times n, \mathbb{K})$$

aus Folgerung 25.12 die **Jacobi-Matrix** von f im Punkt a. Folgerung 25.12 besagt also gerade, dass die Ableitung einer differenzierbaren Funktion gleich ihrer Jacobi-Matrix sein muss.

Beispiel 25.14.

(a) Die Jacobi-Matrix der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 x_2 \\ x_1 \sin x_2 \end{pmatrix}$$

aus Beispiel 25.10 existiert in jedem Punkt und ist gleich

$$Jf(x) = (\,\partial_1 f(x) \mid \partial_2 f(x)\,) = \begin{pmatrix} x_2 & x_1 \\ \sin x_2 & x_1 \cos x_2 \end{pmatrix}.$$

Beachte jedoch, dass dies noch nicht zeigt, dass f auch differenzierbar ist, also dass $f'(a)$ existiert! In der Tat zeigt das folgende Beispiel, dass aus der partiellen Differenzierbarkeit im Allgemeinen noch nicht die totale Differenzierbarkeit folgt.

(b) (Partiell differenzierbare Funktionen müssen nicht differenzierbar sein) Wir betrachten noch einmal die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{cases} \frac{2x_1x_2}{x_1^2+x_2^2} & \text{für } (x_1,x_2) \neq (0,0), \\ 0 & \text{für } (x_1,x_2) = (0,0) \end{cases}$$

aus Beispiel 24.8. Wir haben dort gesehen, dass f im Nullpunkt nicht stetig, nach Lemma 25.7 also insbesondere auch nicht differenzierbar ist. Dennoch ist f dort partiell differenzierbar: Da f auf den Koordinatenachsen gleich 0 ist, ist natürlich

$$\partial_1 f(0) = \lim_{t\to 0} \frac{f(t\,e_1) - f(0)}{t} = 0,$$

und genauso $\partial_2 f(0) = 0$. Obwohl also die Jacobi-Matrix $Jf(0) = (\,0 \;\; 0\,) \in \mathrm{Mat}(1\times 2,\mathbb{R})$ existiert, existiert die Ableitung $f'(0)$ in diesem Fall nicht.

Algorithmus 25.15 (Differenzierbarkeit einer Abbildung)**.** Mit Folgerung 25.12 können wir nun ein explizites Verfahren angeben, um zu bestimmen, ob eine gegebene Funktion $f\colon D \to \mathbb{K}^m$ in einem Punkt $a \in D$ differenzierbar ist, und in diesem Fall die Ableitung zu berechnen:

(a) Berechne mit den eindimensionalen Methoden aus Abschnitt 10.A die partiellen Ableitungen von f in a, und damit die Jacobi-Matrix $Jf(a) = (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,)$. Existiert eine dieser partiellen Ableitungen nicht, so sind wir fertig: Dann ist f nach Folgerung 25.12 nicht differenzierbar in a.

(b) Existiert die Jacobi-Matrix $Jf(a)$, so kommt nach Folgerung 25.12 nur sie als Ableitungsmatrix $f'(a)$ in Frage. Damit muss die Restfunktion in Definition 25.3 (a) gleich $r(x) = f(x) - f(a) - Jf(a)\cdot(x-a)$ sein, und gemäß Definition 25.3 (b) ist f genau dann in a differenzierbar, wenn

$$\lim_{x\to a} \frac{f(x) - f(a) - Jf(a)\cdot(x-a)}{\|x-a\|} = 0.$$

Diesen Grenzwert können wir genau wie in Abschnitt 24.A bestimmen.

Bemerkung 25.16 (Koordinatenweise Differenzierbarkeit im Zielraum)**.** Ist $f\colon D \to \mathbb{K}^m$ eine in einem Punkt $a \in D$ partiell differenzierbare Funktion mit Jacobi-Matrix $Jf(a)$ und Komponentenfunktionen $f_1,\dots,f_m\colon D \to \mathbb{K}$, so kann die Differenzierbarkeit von f in a in diesen Komponenten separat überprüft werden: Nach Algorithmus 25.15 ist f nämlich genau dann in a differenzierbar, wenn die dort in (b) angegebene Grenzwertbedingung gilt. Dieser Grenzwert kann nach Lemma 24.7 nun aber koordinatenweise in $\mathbb{K}^m$ überprüft werden – d. h. wir erhalten die äquivalenten Bedingungen

$$\lim_{x\to a} \frac{f_i(x) - f_i(a) - Jf_i(a)\cdot(x-a)}{\|x-a\|} = 0$$

für alle $i = 1,\dots,m$, die genau besagen, dass alle Komponentenfunktionen in a differenzierbar sind. Wie im Fall der Stetigkeit kann die Differenzierbarkeit jedoch nicht koordinatenweise im Startraum überprüft werden.

Auch wenn wir mit Algorithmus 25.15 jetzt eine Möglichkeit haben, eine Abbildung in mehreren Variablen auf Differenzierbarkeit zu prüfen, können die dafür nötigen Rechnungen in der Grenzwertbetrachtung (b) schnell sehr aufwändig werden. Wünschenswert wäre es daher, wenn wir genau wie im Eindimensionalen ein allgemeines Kriterium hätten, das uns auch ohne Rechnung sagt, dass „aus schönen Funktionen zusammengesetzte Abbildungen" immer differenzierbar sind. Ein solches Kriterium wollen wir jetzt beweisen. Da wir hierfür den nur im Reellen gültigen Mittelwertsatz 10.24 brauchen, beschränken wir uns dabei auf den Fall $\mathbb{K} = \mathbb{R}$.

Satz 25.17 (Stetig partiell differenzierbare Funktionen sind differenzierbar)**.** *Es seien $D \subset \mathbb{R}^n$ offen, $f\colon D \to \mathbb{R}^m$ eine Abbildung, und $a \in D$. Wir setzen voraus, dass alle partiellen Ableitungen $\partial_1 f, \dots, \partial_n f$ in einer Umgebung von a existieren und in a stetig sind – man sagt dann auch, dass f in a* ***stetig partiell differenzierbar*** *ist.*

Dann ist f in a auch total differenzierbar (mit Ableitung $f'(a) = Jf(a)$).

Beweis. Da sowohl die Stetigkeit als auch die Differenzierbarkeit nach Lemma 24.7 bzw. Bemerkung 25.16 koordinatenweise im Zielraum überprüft werden können, genügt es, den Fall $m = 1$ zu betrachten. Nach Algorithmus 25.15 müssen wir dann die Restfunktion

$$r(x) := f(x) - f(a) - Jf(a) \cdot (x-a) = f(x) - f(a) - \sum_{i=1}^{n} \partial_i f(a) \cdot (x_i - a_i) \tag{1}$$

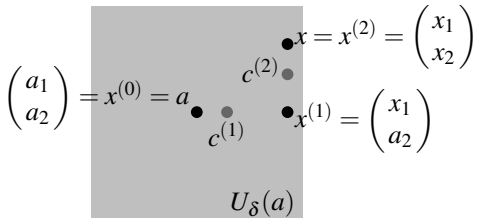
bilden und beweisen, dass $\lim\limits_{x \to a} \frac{r(x)}{\|x-a\|} = 0$ ist.

Wir zeigen dies direkt mit der Definition des Grenzwerts und verwenden dabei die Maximumsnorm auf $\mathbb{R}^n$. Es sei also $\varepsilon > 0$ gegeben. Da D offen und die partiellen Ableitungen nach Voraussetzung stetig sind, gibt es dann ein $\delta > 0$, so dass $U_\delta(a) \subset D$ gilt und

$$|\partial_i f(x) - \partial_i f(a)| < \frac{\varepsilon}{n} \quad \text{für alle } x \in U_\delta(a) \text{ und } i = 1, \dots, n. \tag{2}$$

Es sei nun $x \in U_\delta(a)$ beliebig. Wie im Bild rechts definieren wir für $i = 0, \dots, n$ dann die Punkte

$$x^{(i)} := \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} \in U_\delta(a),$$



so dass also $x^{(0)} = a$ und $x^{(n)} = x$ gilt, und jedes $x^{(i)}$ aus $x^{(i-1)}$ entsteht, indem die i-te Koordinate von a_i auf x_i abgeändert wird.

Damit können wir die Restfunktion (1) nun als

$$r(x) = \sum_{i=1}^{n} \Big(f(x^{(i)}) - f(x^{(i-1)}) - \partial_i f(a) \cdot (x_i - a_i) \Big) \tag{3}$$

umschreiben, da sich in der Summe der ersten beiden Terme alle Funktionswerte außer $f(x)$ und $f(a)$ wegheben. Wenden wir nun den Mittelwertsatz 10.24 (a) für $i = 1, \dots, n$ auf die zwischen a_i und x_i definierten reellwertigen Funktionen einer reellen Variablen

$$g_i\colon t \mapsto f \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ t \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix} \quad \text{mit Ableitung} \quad g_i'\colon t \mapsto \partial_i f \begin{pmatrix} x_1 \\ \vdots \\ x_{i-1} \\ t \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}$$

an, so erhalten wir t_i zwischen a_i und x_i und damit wie im Bild oben rechts $c^{(i)}$ auf den Verbindungsstrecken von a_i nach x_i mit

$$f(x^{(i)}) - f(x^{(i-1)}) = g(x_i) - g(a_i) = g'(t_i)\,(x_i - a_i) = \partial_i f(c^{(i)}) \cdot (x_i - a_i).$$

Setzen wir dies in (3) ein und beachten dabei, dass die Punkte $c^{(1)}, \dots, c^{(n)}$ in $U_\delta(a)$ liegen, so folgt daraus nun

$$\begin{aligned} |r(x)| &= \left| \sum_{i=1}^n (\partial_i f(c^{(i)}) - \partial_i f(a)) \cdot (x_i - a_i) \right| \le \sum_{i=1}^n |\partial_i f(c^{(i)}) - \partial_i f(a)| \cdot |x_i - a_i| \\ &\overset{(2)}{<} \sum_{i=1}^n \frac{\varepsilon}{n} \cdot \|x-a\|_\infty = \varepsilon \cdot \|x-a\|_\infty \end{aligned}$$

und damit $\frac{|r(x)|}{\|x-a\|_\infty} < \varepsilon$, woraus der behauptete Grenzwert $\lim_{x\to a} \frac{r(x)}{\|x-a\|_\infty} = 0$ folgt. □

Die Umkehrung von Satz 25.17 gilt natürlich nicht – sie ist ja schon im Eindimensionalen falsch, da wir in Aufgabe 10.36 bereits gesehen haben, dass eine differenzierbare Funktion auf $\mathbb{R}$ nicht stetig differenzierbar sein muss. Dieses Beispiel lässt sich auch problemlos auf den höherdimensionalen Fall übertragen:

Aufgabe 25.18 (Differenzierbare Funktionen müssen nicht stetig partiell differenzierbar sein)**.** Zeige, dass die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto \begin{cases} (x_1^2 + x_2^2)\sin\frac{1}{\sqrt{x_1^2+x_2^2}} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0 \end{cases}$$

(total) differenzierbar ist, die partiellen Ableitungen von f aber nicht stetig sind. Berechne auch die Ableitung f'!

Bemerkung 25.19. Insbesondere folgt aus Satz 25.17 also, dass alle Abbildungen differenzierbar sind, deren Komponentenfunktionen mit Hilfe der vier Grundrechenarten und Verkettungen aus stetig differenzierbaren Funktionen der Koordinaten des Startraums zusammengesetzt sind – denn für solche Abbildungen existieren ja nach den Rechenregeln aus Kapitel 10 dann alle partiellen Ableitungen und sind stetig. So ist z. B. die Funktion aus den Beispielen 25.10 und 25.14 (a) in jedem Punkt differenzierbar, und genauso die Abbildung aus den Beispielen 24.8 und 25.14 (b) in jedem Punkt mit Ausnahme des Nullpunkts.

Bemerkung 25.20 (Stetig differenzierbar = stetig partiell differenzierbar)**.** In Satz 25.17 haben wir eine Funktion $f\colon D \to \mathbb{R}^m$ auf einer offenen Teilmenge $D \subset \mathbb{R}^n$ aus naheliegenden Gründen *stetig partiell differenzierbar* genannt, wenn ihre partiellen Ableitungen existieren und stetig sind.

Genauso natürlich ist allerdings der schon aus Definition 11.8 (b) bekannte Begriff einer **stetig differenzierbaren** Funktion, der besagt, dass f (total) differenzierbar und die Ableitungsfunktion $f'\colon D \to \mathrm{Mat}(m \times n, \mathbb{R})$ stetig ist. Erfreulicherweise sind diese beiden Begriffe aufgrund unserer bisherigen Resultate äquivalent zueinander: Sind $f_1, \dots, f_m\colon D \to \mathbb{R}$ die Komponentenfunktionen von f, so gilt

$$\begin{aligned} f \text{ ist stetig differenzierbar} &\Leftrightarrow f \text{ ist differenzierbar und } f'\colon D \to \mathrm{Mat}(m \times n, \mathbb{R}) \text{ ist stetig} \\ &\overset{24.7}{\Leftrightarrow} f \text{ ist differenzierbar und alle } \partial_j f_i\colon D \to \mathbb{R} \text{ sind stetig} \\ &\overset{(*)}{\Leftrightarrow} f \text{ ist partiell differenzierbar und alle } \partial_j f_i\colon D \to \mathbb{R} \text{ sind stetig} \\ &\Leftrightarrow f \text{ ist stetig partiell differenzierbar,} \end{aligned}$$

wobei in $(*)$ die Richtung „$\Leftarrow$“ gerade Satz 25.17 ist. In der Regel verwendet man daher nur die einfachere der beiden Formulierungen und spricht von *stetig differenzierbaren* Funktionen.

Zusammenfassend haben wir damit nun also die folgenden Äquivalenzen bzw. Implikationen gezeigt (und gesehen, dass andere Implikationen zwischen diesen Begriffen nicht gelten):

$$\begin{tikzcd}[column sep=small]
\text{stetig differenzierbar} \arrow[d, Leftrightarrow, "25.20"'] \arrow[dr, Rightarrow, "25.17"] & & \\
\text{stetig partiell differenzierbar} \arrow[r, Rightarrow] & \text{(total) differenzierbar} \arrow[r, Rightarrow, "25.7"] \arrow[r, Rightarrow, "25.12"'] & \begin{array}{c}\text{stetig} \\ \\ \text{partiell differenzierbar}\end{array}
\end{tikzcd}$$

Aufgabe 25.21. Untersuche, welche Richtungsableitungen der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto \begin{cases} \left(1-\cos\frac{x_1^2}{x_2}\right)\cdot\sqrt{x_1^2+x_2^2} & \text{für } x_2 \neq 0, \\ 0 & \text{für } x_2 = 0. \end{cases}$$

im Nullpunkt existieren, und ob f dort differenzierbar ist.

Aufgabe 25.22.

(a) Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine quadratische Matrix. Zeige, dass die Abbildung

$$f\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto x^\mathsf{T} A x$$

differenzierbar ist, und berechne ihre Ableitung.

(b) Zeige, dass die Abbildung $f\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \|x\|$ für *keine* Norm $\|\cdot\|$ auf $\mathbb{R}^n$ mit $n > 0$ differenzierbar ist.

(c) Zeige, dass die Abbildung $f\colon \mathrm{Mat}(n \times n, \mathbb{R}) \mapsto \mathrm{Mat}(n \times n, \mathbb{R}),\ A \mapsto A^2$ differenzierbar ist, und berechne ihre Ableitung im Sinne von Bemerkung 25.6.

Aufgabe 25.23. Es sei $g\colon \mathbb{R}^2 \to \mathbb{R}$ eine stetige Abbildung. Zeige, dass $f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_1\, g(x)$ dann im Nullpunkt differenzierbar ist.

Aufgabe 25.24. Es sei $f\colon \mathbb{R}^n \to \mathbb{R}^m$ eine differenzierbare Abbildung mit $f(\lambda x) = \lambda f(x)$ für alle $\lambda \in \mathbb{R}$ und $x \in \mathbb{R}^n$. Zeige, dass f dann bereits eine lineare Abbildung ist.

Aufgabe 25.25 (*Parameterintegrale*)**.** Es seien $f\colon D \to \mathbb{R}$ eine stetige Abbildung auf einer offenen Teilmenge $D \subset \mathbb{R}^2$ sowie $[a,b] \times [c,d]$ ein in D enthaltenes Rechteck. Man zeige:

(a) Die Integralfunktion $F\colon [a,b] \to \mathbb{R},\ x_1 \mapsto \int_c^d f(x_1,x_2)\,dx_2$ ist stetig.

(b) Ist f stetig partiell nach x_1 differenzierbar, so ist F auf (a,b) differenzierbar mit Ableitung $F'(x_1) = \int_c^d \partial_1 f(x_1,x_2)\,dx_2$ (d. h. Differentiation und Integration nach verschiedenen Variablen können vertauscht werden).

(c) $\int_a^b \left(\int_c^d f(x_1,x_2)\,dx_2\right) dx_1 = \int_c^d \left(\int_a^b f(x_1,x_2)\,dx_1\right) dx_2$ (siehe auch Folgerung 28.17).

25.B Eigenschaften differenzierbarer Abbildungen

In diesem Abschnitt wollen wir differenzierbare Abbildungen im Mehrdimensionalen etwas genauer untersuchen und beginnen dazu mit einer geometrischen Deutung für den Fall, dass $\mathbb{K} = \mathbb{R}$ gilt und der Start- oder Zielraum eindimensional ist.

Bemerkung 25.26 (Geometrische Interpretation der Ableitung bei eindimensionalem Startraum)**.** Ist $f\colon D \to \mathbb{R}^m$ mit $D \subset \mathbb{R}$ offen, so können wir f als „Kurve in $\mathbb{R}^m$“ ansehen, da f dann durch nur eine reelle Variable parametrisiert wird. Ist f nun differenzierbar in einem Punkt $a \in D$, so ist

$$f(x) \approx h(x) := f(a) + f'(a)\,(x-a)$$

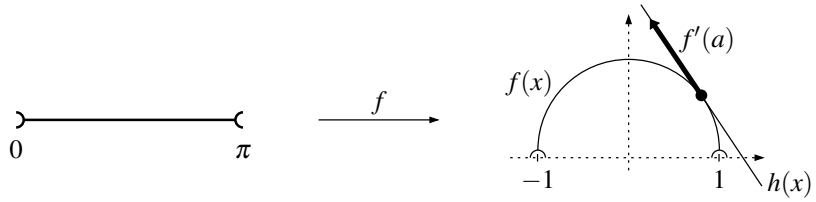
für x in der Nähe von a. Offensichtlich ist h die Parameterdarstellung einer Geraden in $\mathbb{R}^m$ mit Aufpunkt $f(a)$ und Richtungsvektor $f'(a) \in \mathrm{Mat}(m \times 1, \mathbb{R}) = \mathbb{R}^m$, und wir können uns diese Gerade als diejenige vorstellen, die die Kurve f beim Parameterwert a am besten approximiert. Mit anderen Worten ist h gerade die *Tangente* an die Kurve f im Punkt a. Ihr Richtungsvektor $f'(a)$, also die Ableitung von f in a, wird daher als **Tangentialvektor** von f in a bezeichnet.

Als konkretes Beispiel können wir wie im Bild unten den Halbkreisbogen

$$f\colon (0,\pi) \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \quad \text{mit Ableitung} \quad f'\colon (0,\pi) \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}$$

betrachten: Hier ist die Tangente an f im Punkt $a \in (0,\pi)$ wie im folgenden Bild gleich

$$h(x) = \begin{pmatrix} \cos a \\ \sin a \end{pmatrix} + \begin{pmatrix} -\sin a \\ \cos a \end{pmatrix} \cdot (x-a).$$



Bemerkung 25.27 (Geometrische Interpretation der Ableitung bei eindimensionalem Zielraum)**.** Auch für eine differenzierbare Abbildung $f\colon D \to \mathbb{R}$ mit offenem $D \subset \mathbb{R}^n$ können wir wieder mit unserer linearen Näherungsgleichung

$$f(x) \approx h(x) := f(a) + f'(a)\,(x-a)$$

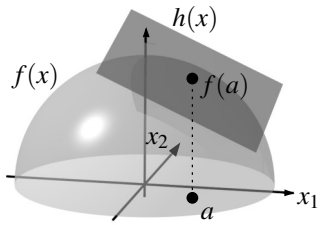
starten. In diesem Fall ist $f'(a)\,(x-a)$ jedoch ein Matrixprodukt einer $1 \times n$-Matrix mit einer $n \times 1$-Matrix, und der Graph $\{(x,h(x)) : x \in \mathbb{R}^n\}$ von h ist ein n-dimensionaler verschobener Unterraum in $\mathbb{R}^n \times \mathbb{R}$, der den Funktionsgraphen von f im Punkt a möglichst gut annähert.

Im Bild rechts ist dies für die Abbildung

$$f\colon D \to \mathbb{R},\ x \mapsto \sqrt{1-x_1^2-x_2^2}$$

mit Ableitung

$$f'(x) = \left(-\frac{x_1}{\sqrt{1-x_1^2-x_2^2}}, -\frac{x_2}{\sqrt{1-x_1^2-x_2^2}} \right)$$



auf dem offenen Einheitskreis $D = \{x \in \mathbb{R}^2 : \|x\|_2 < 1\}$ dargestellt, die offensichtlich eine Halbkugeloberfläche beschreibt. Wir können die Näherungsfunktion h hier als die *Tangentialebene* an den Graphen von f auffassen; in höheren Dimensionen $n > 2$ nennt man den durch h parametrisierten verschobenen Unterraum von $\mathbb{R}^n$ den **Tangentialraum** an den Graphen von f im Punkt a.

Beachte, dass die Ableitung in diesem Fall ein Zeilenvektor $f'(a) \in \mathrm{Mat}(1 \times n, \mathbb{R})$ ist, der damit zunächst einmal nicht wieder im Startraum $\mathbb{R}^n$ liegt. Man definiert daher im Fall eines eindimensionalen Zielraums in der Regel den **Gradienten** von f als den transponierten Vektor

$$\operatorname{grad} f(a) := f'(a)^{\mathsf{T}} \quad \in \mathbb{R}^n,$$

so dass sich die obige Näherungsgleichung mit Hilfe des Standardskalarprodukts auf $\mathbb{R}^n$ auch als

$$f(x) \approx f(a) + \langle \operatorname{grad} f(a), x-a \rangle$$

schreiben lässt.

In der Tat ermöglicht dies auch noch eine weitere geometrische Interpretation der Ableitung bzw. des Gradienten bei eindimensionalem Zielraum: Mit der Konstruktion 21.23 des Winkels zwischen zwei Vektoren können wir auch

$$f(x) \approx f(a) + \|\operatorname{grad} f(a)\|_2 \cdot \|x-a\|_2 \cdot \cos\alpha$$

schreiben, wobei α den Winkel zwischen den Vektoren $x-a$ und $\operatorname{grad} f(a)$ bezeichnet. Bewegt man sich also von a einen kleinen Schritt der Länge ε weg, d. h. geht man zu einem $x \in D$ mit $\|x-a\|_2 = \varepsilon$, so erzeugt dies in f näherungsweise eine Änderung von

$$\|\operatorname{grad} f(a)\|_2 \cdot \varepsilon \cdot \cos\alpha,$$

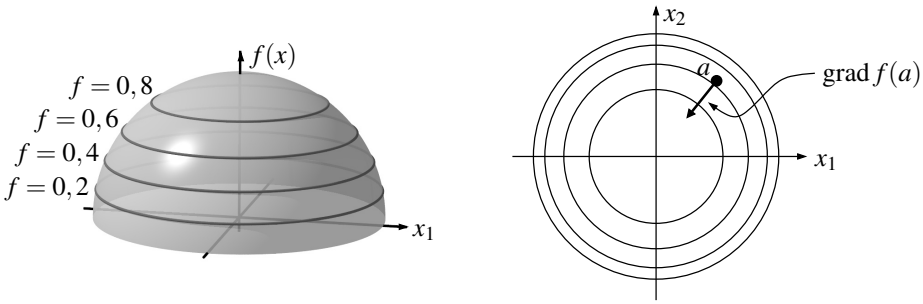
was (für festes ε) maximal wird für $\cos\alpha = 1$, d. h. für $\alpha = 0$, also wenn der gemachte Schritt $x-a$ und $\operatorname{grad} f(a)$ in dieselbe Richtung zeigen. Da die Änderung von f in diesem Fall in etwa gleich $\|\operatorname{grad} f(a)\|_2 \cdot \varepsilon$ ist, können wir den Gradienten also wie folgt geometrisch interpretieren:

Die *Richtung* von $\operatorname{grad} f(a)$ ist die Richtung des stärksten Anstiegs von f im Punkt a. Der *Betrag* von $\operatorname{grad} f(a)$ ist die Stärke des Anstiegs von f in dieser Richtung.

Das Bild unten zeigt diesen Sachverhalt für die eben als Beispiel betrachtete Funktion. Sowohl im dreidimensionalen Bild links als auch im zweidimensionalen Bild rechts sind wie auf einer Landkarte die sogenannten *Höhenlinien* von f eingezeichnet – also die Kurven, auf denen f konstant ist. Der dort eingezeichnete Gradient ist ein Vektor, der in unserem Fall mit

$$\operatorname{grad} f(x) = -\frac{1}{\sqrt{1-x_1^2-x_2^2}} \cdot x$$

direkt zum Nullpunkt hin zeigt – also in die Richtung, in der f an dieser Stelle am stärksten ansteigt.



65

Wir wollen nun noch die Rechenregeln für Ableitungen aus Kapitel 10 so weit wie möglich auf den mehrdimensionalen Fall übertragen. Auch wenn wir unsere Berechnungen durch koordinatenweise Betrachtungen mit Hilfe der partiellen Ableitungen in den meisten Fällen bereits auf den eindimensionalen Fall zurückführen können, sind in der Praxis auch Rechenregeln nützlich, die direkt mit den Ableitungsmatrizen arbeiten und die koordinatenweise Berechnung der Ableitungen damit umgehen können.

Die folgende Aussage entspricht nahezu wörtlich der von Satz 10.8. Beachte aber, dass die Ableitungen jetzt *Matrizen* sind und keine einfachen Zahlen mehr. Wegen der neuen Definition der Differenzierbarkeit können die Beweise aus dem eindimensionalen Fall auch nicht einfach übernommen werden.

Satz 25.28 (Rechenregeln für Ableitungen)**.** *Es seien $D \subset \mathbb{K}^n$ offen und $f,g\colon D \to \mathbb{K}^m$ zwei Abbildungen, die in einem Punkt $a \in D$ differenzierbar sind. Dann gilt:*

(a) *$f \pm g$ ist differenzierbar in a mit $(f \pm g)'(a) = f'(a) \pm g'(a)$.*

(b) *Für $\lambda \in \mathbb{K}$ ist λf differenzierbar in a mit $(\lambda f)'(a) = \lambda f'(a)$.*

Im Fall $m = 1$, wenn Multiplikation und Division der Funktionen ebenfalls möglich sind, gilt zusätzlich:

(c) *(**Produktregel**) fg ist differenzierbar in a mit $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.*

(d) *(**Quotientenregel**) Ist $g(a) \neq 0$, so ist $\frac{f}{g}$ differenzierbar in a mit $(\frac{f}{g})'(a) = \frac{f'(a)g(a)-f(a)g'(a)}{g(a)^2}$.*

Beweis. Nach Definition 25.3 der Differenzierbarkeit können wir $f(x) = f(a) + f'(a)(x-a) + r(x)$ und $g(x) = g(a) + g'(a)(x-a) + s(x)$ mit $\frac{r(x)}{\|x-a\|} \to 0$ und $\frac{s(x)}{\|x-a\|} \to 0$ für $x \to a$ schreiben.

(a) Addition der beiden Ausdrücke für $f(x)$ und $g(x)$ liefert

$$(f+g)(x) = (f+g)(a) + (f'(a)+g'(a))(x-a) + r(x) + s(x).$$

Wegen

$$\lim_{x\to a} \frac{r(x)+s(x)}{\|x-a\|} = \lim_{x\to a} \frac{r(x)}{\|x-a\|} + \lim_{x\to a} \frac{s(x)}{\|x-a\|} = 0$$

ergibt sich also wiederum mit Definition 25.3, dass $f'(a)+g'(a)$ die Ableitung von $f+g$ in a ist. Die Aussagen für die Differenz $f-g$ sowie für λf in (b) zeigt man natürlich genauso.

(c) Hier ergibt die Multiplikation

$$\begin{aligned}(fg)(x) &= (fg)(a) + (f'(a)g(a) + f(a)g'(a))(x-a) \\ &\quad + \underbrace{(f'(a)(x-a))(g'(a)(x-a))}_{\text{(A)}} + \underbrace{(\text{Terme mit } r(x) \text{ und / oder } s(x))}_{\text{(B)}}.\end{aligned}$$

Beachte, dass der Ausdruck $f'(a)(x-a)$ in (A) die Multiplikation einer $(1\times n)$-Matrix mit einer $(n\times 1)$-Matrix ist, die wir daher auch mit Hilfe des Gradienten und des Standardskalarprodukts als $\langle \operatorname{grad} f(a), x-a\rangle$ schreiben können (siehe Bemerkung 25.27). Analog gilt dies für $g'(a)(x-a)$, und damit folgt nach der Cauchy-Schwarz-Ungleichung aus Satz 21.20

$$\|(\mathrm{A})\|_2 \le \|\operatorname{grad} f(a)\|_2 \cdot \|\operatorname{grad} g(a)\|_2 \cdot \|x-a\|_2^2, \quad \text{d. h.} \quad \frac{(\mathrm{A})}{\|x-a\|_2} \to 0 \text{ für } x\to a.$$

Genauso gilt wegen der enthaltenen Faktoren $r(x)$ bzw. $s(x)$ natürlich $\frac{(\mathrm{B})}{\|x-a\|} \to 0$ für $x\to a$. Also können wir hier (A)+(B) als Restterm auffassen, so dass mit Definition 25.3 folgt, dass $f'(a)g(a)+f(a)g'(a)$ die Ableitung von fg in a ist.

Der Beweis von (d) verläuft ganz analog und soll daher hier nicht gegeben werden. □

Auch die Kettenregel (siehe Satz 10.10) überträgt sich wie erwartet ins Mehrdimensionale.

Satz 25.29 (Kettenregel). *Es seien $D\subset\mathbb{K}^n$ und $D'\subset\mathbb{K}^m$ offen. Ferner seien $f\colon D\to\mathbb{K}^m$ und $g\colon D'\to\mathbb{K}^p$ Abbildungen mit $f(D)\subset D'$. Ist dann f differenzierbar in $a\in D$ und g differenzierbar in $f(a)$, so ist auch die Verkettung $g\circ f$ differenzierbar in a, und es gilt*

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

d. h. „die Ableitung einer Verkettung ist das Produkt der beiden Ableitungen". (Beachte, dass es sich hierbei um ein Matrixprodukt einer $p\times m$-Matrix mit einer $m\times n$-Matrix handelt, das wie gewünscht eine $p\times n$-Matrix als Ableitung von $g\circ f$ liefert – insbesondere ist die Reihenfolge der beiden Faktoren hier also wichtig!)

Beweis. Wegen der Differenzierbarkeit von f in a und g in $f(a)$ gibt es nach Definition 25.3 Funktionen $r\colon D\to\mathbb{K}^m$ und $s\colon D'\to\mathbb{K}^p$ mit

$$f(x) = f(a) + f'(a)\,(x-a) + r(x) \tag{1}$$

$$\text{und} \qquad g(y) = g(f(a)) + g'(f(a))\,(y-f(a)) + s(y) \tag{2}$$

sowie $\lim_{x\to a}\frac{r(x)}{\|x-a\|} = 0$ und $\lim_{y\to f(a)}\frac{s(y)}{\|y-f(a)\|} = 0$. Insbesondere können wir also die Funktion

$$\tilde{s}\colon D'\backslash\{f(a)\}\to\mathbb{K}^p,\ y\mapsto \frac{s(y)}{\|y-f(a)\|}$$

durch $\tilde{s}(f(a)) := 0$ zu einer stetigen Funktion auf ganz D' fortsetzen und damit statt (2) auch

$$g(y) = g(f(a)) + g'(f(a))\,(y-f(a)) + \tilde{s}(y)\cdot\|y-f(a)\|$$

für alle $y\in D'$ schreiben. Setzen wir hier nun $y=f(x)$ ein, so ergibt sich mit (1)

$$\begin{aligned}g(f(x)) &= g(f(a)) + g'(f(a))\,(f(x)-f(a)) + \tilde{s}(f(x))\cdot\|f(x)-f(a)\| \\ &= g(f(a)) + g'(f(a))\,f'(a)\,(x-a) + \underbrace{g'(f(a))\,r(x) + \tilde{s}(f(x))\cdot\|f'(a)\,(x-a)+r(x)\|}_{=:R(x)}\end{aligned}$$

für alle $x \in D$. Verwenden wir für die auftretenden Matrizen die von der gewählten Vektornorm induzierte Matrixnorm, so können wir den Restterm R mit Folgerung 24.38 nun abschätzen durch

$$\frac{\|R(x)\|}{\|x-a\|} \le \|g'(f(a))\| \cdot \underbrace{\frac{\|r(x)\|}{\|x-a\|}}_{\to 0} + \underbrace{\|\tilde{s}(f(x))\|}_{\to 0} \cdot \|f'(a)\| \cdot \frac{\|x-a\|}{\|x-a\|} + \underbrace{\|\tilde{s}(f(x))\|}_{\to 0} \cdot \underbrace{\frac{\|r(x)\|}{\|x-a\|}}_{\to 0},$$

wobei die markierten Terme für $x \to a$ gegen 0 konvergieren (beachte dabei für den Term $\tilde{s}(f(x))$, dass f stetig in a und $\tilde{s}$ stetig in $f(a)$ ist, so dass $\tilde{s}(f(x))$ mit $x \to a$ gegen $\tilde{s}(f(a)) = 0$ konvergiert). Also gilt $\lim_{x\to a} \frac{\|R(x)\|}{\|x-a\|} = 0$ und damit auch $\lim_{x\to a} \frac{R(x)}{\|x-a\|} = 0$. Die Aussage des Satzes folgt damit aus Definition 25.3. □

Bemerkung 25.30. Auch ohne sich den genauen Beweis von Satz 25.29 anzusehen, ist einfach zu verstehen, warum bei der Formel für die Ableitung einer Verkettung Matrixprodukte auftreten: Die Ableitungen können wir ja gerade als lineare Näherungen der gegebenen Funktionen betrachten – und lineare Funktionen werden nach Lemma 16.12 verkettet, indem man die zugehörigen Matrizen miteinander multipliziert. In Formeln bedeutet dies unter Vernachlässigung der Restterme folgendes: Können wir als lineare Näherungen

$$f(x) \approx f(a) + f'(a)(x-a) \quad \text{und} \quad g(f(x)) \approx g(f(a)) + g'(f(a))(f(x) - f(a))$$

schreiben, so ergibt sich daraus durch Einsetzen

$$g(f(x)) \approx g(f(a)) + g'(f(a))f'(a)(x-a),$$

d. h. $g'(f(a))f'(a)$ ist die Ableitung von $g \circ f$ in a. (Die genaue Behandlung der Restterme, also die Präzisierung des Zeichens „≈“, ist dann natürlich die eigentliche Arbeit beim Beweis.)

Beispiel 25.31.

(a) Wir betrachten noch einmal die geometrische Deutung der Ableitung bei eindimensionalem Zielraum aus Bemerkung 25.27. Es seien dazu $D \subset \mathbb{R}^n$ offen, $f\colon D \to \mathbb{R}$ eine differenzierbare Funktion, und $\gamma\colon (a,b) \to D$ eine differenzierbare Höhenlinie von f, d. h. so dass $f \circ \gamma\colon (a,b) \to \mathbb{R}$ konstant ist. Dann ist die Ableitung dieser Verkettung natürlich in jedem Punkt $t \in (a,b)$ gleich 0, und wir erhalten somit nach Satz 25.29

$$0 = (f \circ \gamma)'(t) = f'(\gamma(t)) \cdot \gamma'(t) = \langle \operatorname{grad} f(\gamma(t)), \gamma'(t) \rangle .$$

Also ist $\operatorname{grad} f(\gamma(t)) \perp \gamma'(t)$, d. h. der Gradient von f steht in jedem Punkt senkrecht auf (dem Tangentialvektor) der Höhenlinie durch diesen Punkt. Diese Tatsache, die wir im Beispiel von Bemerkung 25.27 auch sofort im Höhenlinienbild erkennen können, ist natürlich jedem bekannt, der schon einmal eine Landkarte mit Höhenlinien benutzt hat: Um möglichst schnell nach oben zu kommen, also in Richtung des Gradienten zu laufen, geht man am besten senkrecht zu den Höhenlinien.

(b) Es sei f die Funktion

$$f\colon \mathbb{R}_{>0} \to \mathbb{R},\ x \mapsto \int_1^x \frac{\sin(xt)}{t}\, dt.$$

Offensichtlich handelt es sich hierbei um eine Funktion mit eindimensionaler Start- und Zielmenge, wie wir sie schon vor langer Zeit betrachtet haben. Man kann zeigen, dass dieses Integral nicht mit den uns bekannten speziellen Funktionen aus Kapitel 9 berechenbar ist, so dass wir f nicht auf einfachere Art hinschreiben können. Dennoch können wir aber die Ableitung von f mit der *mehrdimensionalen* Kettenregel konkret bestimmen: Dazu schreiben wir $f = g \circ h$ mit

$$g\colon \mathbb{R}^2_{>0} \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \int_1^{x_1} \frac{\sin(x_2 t)}{t}\, dt \quad \text{und} \quad h\colon \mathbb{R}_{>0} \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} x \\ x \end{pmatrix}.$$

Dann lässt sich g (und natürlich auch h) problemlos differenzieren: Nach dem Hauptsatz 12.21 der Differential- und Integralrechnung liefert die Ableitung des Integrals nach seiner

Obergrenze gerade

$$\partial_1 g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{\sin(x_2 x_1)}{x_1}.$$

Die partielle Ableitung $\partial_2 g$ dagegen lässt sich nach Aufgabe 25.25 unter dem Integral berechnen, so dass wir dafür

$$\partial_2 g = \int_1^{x_1} \partial_2\left(\frac{\sin(x_2 t)}{t}\right) dt = \int_1^{x_1} \cos(x_2 t)\, dt = \frac{1}{x_2}\left(\sin(x_2 x_1) - \sin x_2\right)$$

erhalten. Beide partiellen Ableitungen sind offensichtlich stetig, so dass g nach Satz 25.17 differenzierbar ist. Damit ergibt die Kettenregel aus Satz 25.29

$$\begin{aligned} f'(x) = g'(h(x))\, h'(x) &= \left(\frac{\sin(x^2)}{x}, \frac{\sin(x^2) - \sin x}{x}\right) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{2\sin(x^2)}{x} - \frac{\sin x}{x}. \end{aligned}$$

Aufgabe 25.32. Es sei $f\colon \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$ mit $n \geq 2$ eine differenzierbare Funktion. Man zeige:

(a) Hängt f nur von $\|x\|_2$ ab, ist also $f(x) = g(\|x\|_2)$ für eine differenzierbare Funktion $g\colon \mathbb{R}_{>0} \to \mathbb{R}$ (man sagt auch: f ist *kugelsymmetrisch*), so ist

$$\operatorname{grad} f(x) = \frac{g'(\|x\|_2)}{\|x\|_2} \cdot x \quad \text{für alle } x \in \mathbb{R}^n \backslash \{0\}.$$

(b) Gibt es umgekehrt eine Funktion $h\colon \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$ mit $\operatorname{grad} f(x) = h(x) \cdot x$ für alle x, so ist f kugelsymmetrisch. (Hinweis: Zeige, dass f entlang eines beliebigen Weges auf einer Kugeloberfläche konstant ist.)

26. Höhere Ableitungen

Im letzten Kapitel haben wir gesehen, wie man für Abbildungen zwischen mehrdimensionalen Räumen das Konzept der Differenzierbarkeit definieren und für differenzierbare Abbildungen die Ableitung berechnen kann. Analog zum eindimensionalen Fall ist eine der wichtigsten Anwendungen dieser Theorie, (lokale) Extremwerte von Funktionen zu bestimmen. Wir wollen in diesem Kapitel untersuchen, wie dies im Mehrdimensionalen funktioniert. Wie in Satz 11.19 werden wir dazu höhere Ableitungen und die Taylor-Formel benötigen, die wir daher zuerst behandeln müssen.

Um von Extremwerten überhaupt sprechen zu können, müssen wir natürlich Funktionswerte miteinander vergleichen können – was nur geht, wenn die Wertemenge der betrachteten Funktionen die Menge $\mathbb{R}$ der reellen Zahlen ist. „Mehrdimensional" bedeutet im Zusammenhang mit Extremwerten also nur, dass die *Definitionsmenge* unserer Abbildungen eine Teilmenge von $\mathbb{R}^n$ ist.

Hier ist zunächst einmal die exakte Definition von Extrema, die im Prinzip genauso ist wie im Eindimensionalen in Definition 10.19.

Definition 26.1 (Extrema)**.** Es seien $D \subset \mathbb{R}^n$, $f\colon D \to \mathbb{R}$ eine Funktion und $a \in D$. Man sagt, ...

(a) f habe in a ein **(globales) Maximum**, wenn $f(a) \geq f(x)$ für alle $x \in D$.

(b) f habe in a ein **lokales Maximum**, wenn es eine Umgebung U von a gibt mit $f(a) \geq f(x)$ für alle $x \in U$. Gilt sogar $f(a) > f(x)$ für alle $x \in U$ mit $x \neq a$, so nennt man das lokale Maximum **isoliert**.

Analog definiert man globale und lokale (isolierte) **Minima**. Hat f in a ein (globales, lokales, isoliertes) Maximum oder Minimum, so sagt man auch, dass f dort ein (globales, lokales, isoliertes) **Extremum** hat.

Einfach zu sehen ist genau wie im Eindimensionalen (siehe Lemma 10.21), dass die (erste) Ableitung einer Funktion ein einfaches *notwendiges* Kriterium für ein lokales Extremum liefert, das nicht auf dem Rand der Definitionsmenge liegt:

Lemma 26.2 (Notwendige Bedingung für Extrema)**.** *Es seien $D \subset \mathbb{R}^n$ und $f\colon D \to \mathbb{R}$ eine Abbildung. Hat f dann in einem Punkt $a \in \mathring{D}$ im Inneren der Definitionsmenge ein lokales Extremum und ist f dort differenzierbar, so gilt $f'(a) = 0$.*

*Wie im Eindimensionalen in Bemerkung 10.22 (a) werden Punkte $a \in \mathring{D}$ mit $f'(a) = 0$ als **kritische Punkte** bezeichnet.*

Beweis. Es sei $v \in \mathbb{R}^n \backslash \{0\}$. Wegen $a \in \mathring{D}$ ist die reellwertige Funktion $g\colon t \mapsto f(a+tv)$ einer reellen Variablen t in einer Umgebung von 0 definiert, und hat dort natürlich ebenfalls ein lokales Extremum. Nach Lemma 10.21 ist also

$$0 = g'(0) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t} = \partial_v f(a),$$

d. h. alle Richtungsableitungen von f sind in a gleich 0. Insbesondere verschwinden damit auch alle partiellen Ableitungen an diesem Punkt, und es folgt $f'(a) = (\,\partial_1 f(a) \mid \cdots \mid \partial_n f(a)\,) = 0$. □

66

Beispiel 26.3. Wir wollen die (lokalen) Extrema der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ f(x) = (x_1 + x_2)^2 + \mathrm{e}^{x_1^2}$$

finden. Nach Bemerkung 25.19 ist f differenzierbar mit

$$f'(x) = (2(x_1 + x_2) + 2x_1 \mathrm{e}^{x_1^2}, 2(x_1 + x_2)).$$

Ein notwendiges Kriterium für ein lokales Extremum am Punkt $x \in \mathbb{R}^2$ ist also

$$2(x_1+x_2)+2x_1\mathrm{e}^{x_1^2}=0 \quad \text{und} \quad 2(x_1+x_2)=0,$$

was offensichtlich nur für $x_1=x_2=0$, also im Nullpunkt erfüllt ist. In der Tat kann man in diesem Fall einfach sehen, dass im Nullpunkt ein (sogar globales) Minimum vorliegt, denn für alle $x \in \mathbb{R}^2$ ist ja

$$f(x)=(x_1+x_2)^2+\mathrm{e}^{x_1^2} \geq 0+1=1=f(0).$$

Bemerkung 26.4. Bis hierher funktioniert die Extremwertberechnung also genau wie im Eindimensionalen. Es stellen sich nun natürlich sofort die folgenden Fragen:

(a) Wie kann man von einem Punkt $a \in \overset{\circ}{D}$ mit $f'(a)=0$ möglichst einfach überprüfen, ob dort wirklich ein lokales Extremum vorliegt, d. h. welche *hinreichenden* Kriterien für lokale Extrema gibt es? Wie bereits angekündigt werden wir hierfür höhere Ableitungen und eine mehrdimensionale Taylor-Entwicklung benötigen, die wir in diesem Kapitel untersuchen werden.

(b) Wie kann man Extrema am *Rand* der Definitionsmenge D finden, wenn D nicht offen ist? Betrachten wir z. B. die Funktion f aus Beispiel 26.3 auf dem abgeschlossenen Einheitskreis $K_1=\{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$, so muss f dort nach Folgerung 24.34 irgendwo ein Maximum annehmen – und da wir außer dem Minimum bei 0 keine weiteren Stellen gefunden haben, an denen die Ableitung von f verschwindet, muss dieses Maximum auf dem Rand des Einheitskreises liegen. Im Gegensatz zum Eindimensionalen besteht dieser Rand jetzt aber nicht mehr nur aus zwei Punkten, sondern aus einem in diesem Fall selbst eindimensionalen Objekt, nämlich der Einheitskreislinie. Wir können also nicht mehr einfach alle Randpunkte in f einsetzen, um das Maximum zu finden, sondern brauchen ein geschickteres Verfahren, um Randextrema zu finden. Dies wird der Inhalt von Abschnitt 27.C sein.

26.A Die mehrdimensionale Taylor-Entwicklung

Zur Einführung höherer Ableitungen, die wir für die Taylor-Entwicklung benötigen werden, gibt es zwei ganz unterschiedliche Herangehensweisen:

Bemerkung 26.5 (Höhere Ableitungen)**.** Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^m$ eine Funktion.

(a) Ist f differenzierbar, so ist die Ableitung von f nach den Bemerkungen 25.4 (c) und 25.6 eine Funktion $f'\colon D \to \mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m) \cong \mathrm{Mat}(m \times n,\mathbb{R})$, also selbst wieder eine Funktion zwischen (offenen Teilmengen von) normierten Räumen – für die wir ja definiert haben, was Differenzierbarkeit bedeutet. Es liegt also eigentlich nichts näher als zu sagen, dass f zweimal differenzierbar heißt, wenn f' selbst wieder differenzierbar ist, und f'' dann als die Ableitung von f' zu definieren. Da f' eine Abbildung von $D \subset \mathbb{R}^n$ nach $\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)$ ist, ist f'' nach Bemerkung 25.6 dann eine Abbildung

$$f''\colon D \to \mathrm{Hom}(\mathbb{R}^n,\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)).$$

Dies kann man natürlich fortsetzen, und erhält (im Fall der Existenz) z. B. als dritte Ableitung von f eine Abbildung

$$f'''\colon D \to \mathrm{Hom}(\mathbb{R}^n,\mathrm{Hom}(\mathbb{R}^n,\mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m))).$$

Auch wenn dies eigentlich die natürlichste Art der Definition höherer Ableitungen im mehrdimensionalen Fall ist, sollte hieraus schon ersichtlich sein, dass es sich dabei allein schon von der Notation her nicht unbedingt um die bequemste Art handelt. Wir werden im Folgenden daher einen anderen Ansatz verfolgen:

(b) Da die Komponentenfunktionen der Ableitungsmatrix von f gerade die partiellen Ableitungen sind, können wir zur Betrachtung k-facher Ableitungen auch einfach mehrfache partielle Ableitungen der Form $\partial_{i_1}\cdots\partial_{i_k}f$ betrachten. Diese haben im Gegensatz zum Ansatz aus (a) den großen Vorteil, im Fall ihrer Existenz selbst wieder Funktionen von D nach $\mathbb{R}^m$ zu sein –

ihr Nachteil besteht lediglich darin, dass es nicht nur eine, sondern n^k solcher k-fachen partiellen Ableitungen gibt, da wir bei jeder der k Differentiationen wieder eine der n Variablen auswählen müssen.

Man kann allerdings zumindest hoffen, dass es bei derartigen mehrfachen partiellen Ableitungen nicht auf die Reihenfolge der Ableitungen ankommt. Der folgende Satz besagt, dass dies in den meisten Fällen auch wirklich so ist.

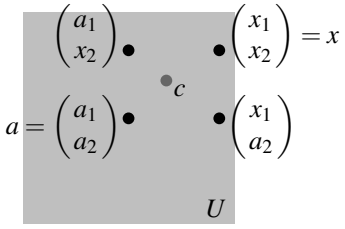
Satz 26.6 (**Satz von Schwarz**: Vertauschbarkeit partieller Ableitungen)**.** *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^m$ eine Abbildung. Wir nehmen an, dass alle zweifachen partiellen Ableitungen $\partial_i \partial_j f$ für $i, j = 1, \dots, n$ auf D existieren und in einem Punkt $a \in D$ stetig sind. Dann gilt $\partial_i \partial_j f(a) = \partial_j \partial_i f(a)$ für alle $i, j = 1, \dots, n$.*

Beweis. Wir können ohne Einschränkung $m = 1$ annehmen, da man sonst jede Komponentenfunktion von f einzeln betrachten kann. Weiterhin können wir natürlich $i \neq j$ voraussetzen und alle Variablen außer x_i und x_j als konstante Parameter betrachten. Nach Umbenennung der Variablen können wir uns daher auf den Fall $n = 2$, $i = 1$ und $j = 2$ beschränken.

Da D offen ist, können wir eine offene Kugel U um a in der Maximumsnorm wählen, die noch ganz in D liegt. Für einen Punkt $x \in U$ mit $x_1 \neq a_1$ und $x_2 \neq a_2$ betrachten wir nun den „doppelten Differenzenquotienten“

$$Q(x) := \frac{f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - f\begin{pmatrix} a_1 \\ x_2 \end{pmatrix} - f\begin{pmatrix} x_1 \\ a_2 \end{pmatrix} + f\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}}{(x_1 - a_1)(x_2 - a_2)},$$

der sich aus den Funktionswerten von f an den vier Eckpunkten des Rechtecks wie im Bild rechts ergibt. Wir wollen zeigen, dass $Q(x)$ für $x \to a$ sowohl gegen $\partial_1 \partial_2 f(a)$ als auch gegen $\partial_2 \partial_1 f(a)$ konvergiert, woraus dann die Behauptung folgt.



Dazu wenden wir zunächst den Mittelwertsatz 10.24 (a) auf die zwischen a_2 und x_2 definierte reellwertige Funktion

$$g\colon t \mapsto f\begin{pmatrix} x_1 \\ t \end{pmatrix} - f\begin{pmatrix} a_1 \\ t \end{pmatrix}$$

an und erhalten ein c_2 zwischen a_2 und x_2 mit

$$g(x_2) - g(a_2) = g'(c_2)\,(x_2 - a_2) = \left(\partial_2 f\begin{pmatrix} x_1 \\ c_2 \end{pmatrix} - \partial_2 f\begin{pmatrix} a_1 \\ c_2 \end{pmatrix}\right)(x_2 - a_2),$$

also mit

$$Q(x) = \frac{g(x_2) - g(a_2)}{(x_1 - a_1)(x_2 - a_2)} = \frac{\partial_2 f\begin{pmatrix} x_1 \\ c_2 \end{pmatrix} - \partial_2 f\begin{pmatrix} a_1 \\ c_2 \end{pmatrix}}{(x_1 - a_1)}.$$

Nun wenden wir den Mittelwertsatz erneut auf die zwischen a_1 und x_1 definierte Funktion

$$h\colon t \mapsto \partial_2 f\begin{pmatrix} t \\ c_2 \end{pmatrix}$$

an und erhalten so ein c_1 zwischen a_1 und x_1 mit

$$h(x_1) - h(a_1) = h'(c_1)\,(x_1 - a_1) = \partial_1 \partial_2 f\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}(x_1 - a_1),$$

d. h. mit

$$Q(x) = \frac{h(x_1) - h(a_1)}{x_1 - a_1} = \partial_1 \partial_2 f(c)$$

für ein $c \in U$ wie im Bild oben rechts, dessen Koordinaten zwischen a_1 und x_1 bzw. a_2 und x_2 liegen.

Wähle nun eine beliebige Folge $(x^{(k)})_{k \in \mathbb{N}}$, die gegen a konvergiert und deren Glieder nicht auf der horizontalen oder vertikalen Geraden durch a liegen. Mit $x^{(k)} \to a$ konvergiert dann natürlich auch

die Folge der oben gefundenen zugehörigen Punkte $c^{(k)}$ gegen a, und so erhalten wir wegen der vorausgesetzten Stetigkeit von $\partial_1\partial_2 f$ in a nach dem Folgenkriterium aus Satz 24.4

$$\lim_{k\to\infty} Q(x^{(k)}) = \lim_{k\to\infty} \partial_1\partial_2 f(c^{(k)}) = \partial_1\partial_2 f(a).$$

Nun ist der Ausdruck $Q(x)$ aber symmetrisch in den beiden Koordinaten des Startraums, und daher ergibt sich mit genau dem gleichen Argument (nur indem man den Mittelwertsatz zuerst auf die erste und dann auf die zweite Koordinate anwendet) auch $\lim_{k\to\infty} Q(x^{(k)}) = \partial_2\partial_1 f(a)$. Die Behauptung folgt damit aus der Eindeutigkeit des Grenzwerts. □

Die Vertauschbarkeit partieller Ableitungen ist also gewährleistet, wenn die partiellen Ableitungen existieren *und stetig sind*. Dass man auf diese Zusatzforderung der Stetigkeit leider nicht verzichten kann, zeigt die folgende Aufgabe:

Aufgabe 26.7 (Partielle Ableitungen müssen nicht miteinander vertauschen)**.** Zeige, dass die partiellen Ableitungen $\partial_1\partial_2 f$ und $\partial_2\partial_1 f$ der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad x \mapsto \begin{cases} x_1 x_2 \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{für } (x_1, x_2) \neq (0,0), \\ 0 & \text{für } (x_1, x_2) = (0,0) \end{cases}$$

zwar existieren, aber nicht übereinstimmen.

Um die Probleme solcher nicht miteinander vertauschbaren Ableitungen zu umgehen, wollen wir uns daher ab jetzt auf Abbildungen beschränken, deren partielle Ableitungen nicht nur existieren, sondern auch stetig sind.

Definition 26.8 (Mehrfach stetig differenzierbare Funktionen)**.** Es sei $f\colon D \to \mathbb{R}^m$ eine Abbildung auf einer offenen Teilmenge $D \subset \mathbb{R}^n$. Für $r \geq 0$ sagt man, f sei **r-mal stetig differenzierbar** oder **r-mal stetig partiell differenzierbar**, wenn alle r-fachen partiellen Ableitungen $\partial_{i_1} \cdots \partial_{i_r} f$ für $i_1, \ldots, i_r \in \{1, \ldots, n\}$ auf D existieren und stetig sind. Man schreibt diese Ableitungen auch als

$$\frac{\partial^r f}{\partial x_{i_1} \cdots \partial x_{i_r}}$$

bzw. bei mehrfachen Ableitungen nach derselben Variablen in „Potenzschreibweise“, d. h. z. B. als

$$\partial_1^2 f \quad \text{oder} \quad \frac{\partial^2 f}{\partial x_1^2}$$

für die zweifache partielle Ableitung nach x_1. Im Fall $m = 1$ eines eindimensionalen Zielraums schreibt man die Menge aller r-mal stetig differenzierbaren Funktionen auf D wie in Definition 11.8 als $C^r(D)$.

Bemerkung 26.9.

(a) In Bemerkung 25.20 hatten wir bereits gesehen, dass im Fall $r = 1$ die Begriffe „(einmal) stetig differenzierbar“ und „(einmal) stetig partiell differenzierbar“ zusammenfallen, wenn man sie auf die natürliche Art definiert. Man kann zeigen, dass dieselbe Aussage auch für höhere Ableitungen gilt, wenn man „r-mal stetig differenzierbar“ als Stetigkeit der r-ten Ableitung wie in Bemerkung 26.5 (a) und „r-mal stetig partiell differenzierbar“ wie in Definition 26.8 interpretiert. Da wir mit den höheren Ableitungen wie in Bemerkung 26.5 (a) hier nicht arbeiten werden, wollen wir dies hier allerdings nicht beweisen, und erwähnen diese Tatsache nur als Motivation dafür, dass wir die Begriffe „r-mal stetig differenzierbar“ und „r-mal stetig partiell differenzierbar“ oben als gleichwertig definiert haben.

(b) Ist f eine r-mal stetig differenzierbare Funktion wie in Definition 26.8, so sind nicht nur die partiellen Ableitungen der Stufe r, sondern auch die aller Stufen $k \leq r$ stetig: Da die partiellen Ableitungen $\partial_{i_1} \cdots \partial_{i_r} f$ nach Voraussetzung stetig sind, sind die $(r-1)$-fachen partiellen Ableitungen $\partial_{i_2} \cdots \partial_{i_r} f$ stetig partiell differenzierbar, nach Satz 25.17 und Lemma 25.7 also auch total differenzierbar und damit stetig. Mit Induktion sind demnach dann auch alle niedrigeren partiellen Ableitungen stetig.

(c) Nach Satz 26.6 kommt es bei bis zu r-fachen partiellen Ableitungen einer r-fach stetig differenzierbaren Funktion nicht auf die Reihenfolge dieser Ableitungen an, da man in einem Ausdruck der Form $\partial_{i_1} \cdots \partial_{i_k} f$ (mit $k \leq r$) durch fortgesetztes Vertauschen zweier benachbarter partieller Ableitungen jede andere Reihenfolge erzeugen kann.

Aufgabe 26.10. Es sei f die in einer Umgebung des Ursprungs von $\mathbb{R}^2$ definierte reellwertige Funktion mit

$$f(x) = \frac{\cos(x_1^5 + x_2^5)}{1 - x_1^3 - x_2^3}.$$

Berechne mit einer geeigneten Potenzreihenentwicklung (und ohne Computer) die partielle Ableitung $\partial_1^{11} \partial_2^8 f(0)$.

Mit Hilfe dieser höheren Ableitungen können wir nun zur Verallgemeinerung der Taylor-Formel aus Satz 11.14 auf den mehrdimensionalen Fall kommen. Damit das darin auftretende Taylor-Polynom formal analog zum eindimensionalen Fall aussieht, führen wir dazu zunächst einige nützliche Schreibweisen ein.

Notation 26.11 (Multi-Indizes)**.** Ein Element $I = (i_1, \dots, i_n)$ in $\mathbb{N}^n$ nennt man einen **Multi-Index**. Für derartige Multi-Indizes definieren wir die suggestiven Notationen

$$\begin{aligned} |I| &:= i_1 + \cdots + i_n, \\ I! &:= i_1! \cdots i_n!, \\ \partial^I &:= \partial_1^{i_1} \cdots \partial_n^{i_n}, \\ x^I &:= x_1^{i_1} \cdots x_n^{i_n} \quad \text{für } x \in \mathbb{R}^n. \end{aligned}$$

Zusätzlich wird im Folgenden die Konstruktion $I + e_j = (i_1, \dots, i_{j-1}, i_j + 1, i_{j+1}, \dots, i_n)$ auftreten, wobei e_j mit $j \in \{1, \dots, n\}$ wie üblich für den j-ten Einheitsvektor steht.

Definition 26.12 (Taylor-Polynom)**.** Es seien $D \subset \mathbb{R}^n$ offen, $r \in \mathbb{N}$, $f \colon D \to \mathbb{R}^m$ eine r-mal stetig differenzierbare Funktion, und $a \in D$. Dann heißt

$$T_{f,a}^r \colon D \to \mathbb{R}^m,\ x \mapsto \sum_{I : |I| \leq r} \frac{\partial^I f(a)}{I!} (x-a)^I$$

das **r-te Taylor-Polynom** von f mit Entwicklungspunkt a.

Wir werden dieses Taylor-Polynom nahezu ausschließlich im Fall $m = 1$ eines eindimensionalen Zielraums betrachten – dies ist natürlich auch der für die Untersuchung von Extremwerten relevante Fall. Zunächst einmal können wir dann feststellen, dass es mit unserer alten Definition des eindimensionalen Taylor-Polynoms aus Definition 11.11 (a) übereinstimmt, wenn $n = 1$ ist (und I damit aus nur einem Index besteht). Auch für allgemeine n gibt die Zahl $|I|$ in den Summanden offensichtlich an, wie oft f differenziert wird. Die Terme mit kleinem $|I|$ lassen sich auf die folgende Art mit Hilfe von Matrixprodukten auch anders schreiben:

Definition 26.13 (Hesse-Matrix)**.** Es seien $D \subset \mathbb{R}^n$ offen und $f \colon D \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Dann heißt für $a \in D$ die (nach Satz 26.6 symmetrische) Matrix der zweiten partiellen Ableitungen

$$Hf(a) := (\partial_i \partial_j f(a))_{i,j} \in \mathrm{Mat}(n \times n, \mathbb{R})$$

die **Hesse-Matrix** von f in a.

Bemerkung 26.14 (Alternative Schreibweise der Taylor-Polynome vom Grad höchstens 2)**.** Für die ersten Terme in den Taylor-Polynomen einer Funktion $f \colon D \to \mathbb{R}$ mit $D \subset \mathbb{R}^n$ können die zugehörigen Multi-Indizes explizit angegeben werden:

(a) Für $|I| = 0$ ist nur $I = (0, \dots, 0)$ möglich; der zugehörige Term im Taylor-Polynom ist einfach der Funktionswert $f(a)$ am Entwicklungspunkt.

(b) Für $|I| = 1$ ist notwendigerweise $I = e_j = (0,\dots,0,1,0,\dots,0)$ mit der Eins an einer beliebigen Position $j \in \{1,\dots,n\}$. Die entsprechenden Summanden im Taylor-Polynom lassen sich also schreiben als

$$\sum_{j=1}^{n} \frac{\partial_j f(a)}{1!}(x_j - a_j) = f'(a)\cdot(x-a).$$

(c) Für $|I| = 2$ haben die Multi-Indizes die Form $I = (0,\dots,0,1,0,\dots,0,1,0,\dots,0)$ mit Einsen an Positionen $i < j$ sowie $I = (0,\dots,0,2,0,\dots,0)$ mit einer Zwei an Position i. Im Taylor-Polynom führt dies zu den Termen

$$\begin{aligned}\sum_{i<j} \frac{\partial_i\partial_j f(a)}{1!\cdot 1!}(x_i-a_i)(x_j-a_j) + \sum_{i=1}^{n} \frac{\partial_i^2 f(a)}{2!}(x_i-a_i)^2 &= \frac{1}{2}\cdot\sum_{i,j=1}^{n} \partial_i\partial_j f(a)\cdot(x_i-a_i)\cdot(x_j-a_j)\\ &= \frac{1}{2}(x-a)^{\mathsf{T}}\cdot Hf(a)\cdot(x-a).\end{aligned}$$

Das zweite Taylor-Polynom einer zweimal stetig differenzierbaren Funktion f an einem Entwicklungspunkt a ist also

$$T^2_{f,a}(x) = f(a) + f'(a)\cdot(x-a) + \frac{1}{2}(x-a)^{\mathsf{T}}\cdot Hf(a)\cdot(x-a).$$

Beispiel 26.15. Wir betrachten die Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_1 \sin x_2$$

und wollen dazu das zweite Taylor-Polynom um $a = 0$ berechnen. Dazu benötigen wir die ersten und zweiten partiellen Ableitungen

$$\begin{aligned}&\partial_1 f(x) = \sin x_2,\ \ \partial_2 f(x) = x_1\cos x_2\\ \text{und}\quad &\partial_1^2 f(x) = 0,\ \ \partial_1\partial_2 f(x) = \cos x_2,\ \ \partial_2^2 f(x) = -x_1\sin x_2,\end{aligned}$$

im Nullpunkt also

$$f(0) = \partial_1 f(0) = \partial_2 f(0) = \partial_1^2 f(0) = \partial_2^2 f(0) = 0 \quad\text{und}\quad \partial_1\partial_2 f(0) = 1.$$

Das gesuchte Taylor-Polynom ist damit

$$\begin{aligned}T^2_{f,0}(x) &= \sum_{i_1+i_2\le 2} \frac{\partial_1^{i_1}\partial_2^{i_2} f(0)}{i_1!\,i_2!} x_1^{i_1} x_2^{i_2}\\ &= \underbrace{f(0)}_{=0} + \underbrace{\frac{\partial_1^1 f(0)}{1!}}_{=0} x_1^1 + \underbrace{\frac{\partial_2^1 f(0)}{1!}}_{=0} x_2^1 + \underbrace{\frac{\partial_1^2 f(0)}{2!}}_{=0} x_1^2 + \underbrace{\frac{\partial_1^1\partial_2^1 f(0)}{1!\,1!}}_{=1} x_1^1 x_2^1 + \underbrace{\frac{\partial_2^2 f(0)}{2!}}_{=0} x_2^2\\ &= x_1 x_2.\end{aligned}$$

Alternativ ist wie in Bemerkung 26.14

$$f'(0) = (0\ \ 0) \quad\text{und}\quad Hf(0) = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},$$

und damit wieder

$$T^2_{f,0}(x) = 0 + (0\ \ 0)\cdot\begin{pmatrix}x_1\\ x_2\end{pmatrix} + \frac{1}{2}(x_1\ \ x_2)\cdot\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\cdot\begin{pmatrix}x_1\\ x_2\end{pmatrix} = x_1 x_2.$$

67

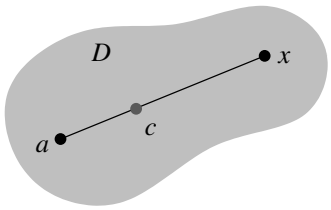
Das entscheidende Resultat über Taylor-Polynome ist nun, dass $T^r_{f,a}$ wie im Eindimensionalen als „beste Näherung“ der Funktion f im Punkt a durch ein Polynom vom Grad r angesehen werden kann. Präzise ausgedrückt ist die Differenz $f(x) - T^r_{f,a}(x)$ wie in Satz 11.14 ein (in der Regel kleines) Restglied, das genauso aussieht wie der nächste Term der Ordnung $r+1$ der Taylor-Entwicklung, allerdings mit der Ableitung an einer Zwischenstelle berechnet statt am Entwicklungspunkt.

Satz 26.16 (Taylor-Formel). *Es seien $D \subset \mathbb{R}^n$ offen, $r \in \mathbb{N}$ und $f\colon D \to \mathbb{R}$ eine $(r+1)$-mal stetig differenzierbare Funktion. Ferner seien $a, x \in D$, so dass die gesamte Verbindungsstrecke*

$$\overline{ax} := \{a + t\,(x-a) : t \in [0,1]\}$$

von a nach x in D liegt. Dann gibt es einen Punkt $c \in \overline{ax}$ auf dieser Strecke, so dass

$$f(x) - T^r_{f,a}(x) = \sum_{I:|I|=r+1} \frac{\partial^I f(c)}{I!}(x-a)^I.$$



*Wie in Satz 11.14 bezeichnet man diesen Ausdruck auch als das **Restglied** der Taylor-Entwicklung.*

Beweis. Wir betrachten die Funktion $g\colon [0,1] \to \mathbb{R}$, $t \mapsto f(a + t\,(x-a))$, die die Einschränkung von f auf $\overline{ax}$ beschreibt, und wollen zeigen, dass die eindimensionale Taylor-Formel aus Satz 11.14 angewendet auf g exakt die Behauptung unseres Satzes ist.

Dazu müssen wir die höheren Ableitungen von g berechnen: Wir zeigen für alle $t \in [0,1]$ mit Induktion über $k = 0, \dots, r+1$, dass

$$g^{(k)}(t) = k! \cdot \sum_{I:|I|=k} \frac{\partial^I f(a + t\,(x-a))}{I!}(x-a)^I. \qquad (*)$$

Der Induktionsanfang für $k = 0$, also $g^{(0)}(t) = g(t) = f(a + t\,(x-a))$, ist dabei trivial. Wir nehmen nun an, dass diese Formel für ein $k \in \{0, \dots, r\}$ gilt, und müssen diesen Ausdruck erneut nach t differenzieren, um $g^{(k+1)}$ zu berechnen. Dazu bemerken wir zunächst, dass die Ableitung der Funktion $t \mapsto \partial^I f(a + t\,(x-a))$ nach der Kettenregel aus Satz 25.29 das Matrixprodukt der Ableitungen von $\partial^I f$ und $t \mapsto a + t\,(x-a)$ ist, also gleich

$$\sum_{j=1}^n \partial_j \partial^I f(a + t\,(x-a)) \cdot (x_j - a_j) = \sum_{j=1}^n \partial^{I+e_j} f(a + t\,(x-a)) \cdot (x_j - a_j).$$

Damit erhalten wir als Ableitung von $(*)$, wobei wir wieder $I = (i_1, \dots, i_n)$ schreiben,

$$\begin{aligned}
g^{(k+1)}(t) &= k! \cdot \sum_{j=1}^n \sum_{I:|I|=k} \frac{\partial^{I+e_j} f(a + t\,(x-a))}{I!}(x-a)^{I+e_j} \\
&= k! \cdot \sum_{j=1}^n \sum_{\substack{I:|I|=k+1 \\ i_j>0}} \frac{\partial^I f(a + t\,(x-a))}{(I-e_j)!}(x-a)^I && \text{(Indexverschiebung } i_j \to i_j - 1) \\
&= k! \cdot \sum_{I:|I|=k+1} \underbrace{\Bigg(\sum_{j:i_j>0} i_j\Bigg)}_{=k+1} \cdot \frac{\partial^I f(a + t\,(x-a))}{I!}(x-a)^I && \text{(Erweitern mit } i_j) \\
&= (k+1)! \cdot \sum_{I:|I|=k+1} \frac{\partial^I f(a + t\,(x-a))}{I!}(x-a)^I,
\end{aligned}$$

was die behauptete Formel $(*)$ mit Induktion zeigt. Damit folgt nun mit der eindimensionalen Taylor-Formel für g

$$f(x) - T^r_{f,a}(x) = g(1) - \sum_{k=0}^r \frac{g^{(k)}(0)}{k!} \cdot 1^k = g(1) - T^r_{g,0}(1) \overset{11.14}{=} \frac{g^{(r+1)}(t)}{(r+1)!} \cdot 1^{r+1} \quad \text{für ein } t \in (0,1),$$

was wegen $(*)$ für $c := a + t\,(x-a)$ genau die Aussage des Satzes ist. □

Die wichtigsten Fälle in Satz 26.16 sind sicher $r = 0$ und $r = 1$. Während $r = 1$ zur im nächsten Abschnitt betrachteten Extremwertberechnung führt, ergibt sich für $r = 0$ einfach die mehrdimensionale Entsprechung des Mittelwertsatzes 10.24 (a):

Folgerung 26.17 (Mittelwertsatz). *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}$ eine stetig differenzierbare Funktion. Ferner seien $a, x \in D$ zwei Punkte mit $\overline{ax} \subset D$. Dann gibt es eine Zwischenstelle $c \in \overline{ax}$ mit*

$$f(x) - f(a) = f'(c) \cdot (x - a).$$

Beweis. Mit der Schreibweise aus Bemerkung 26.14 (a) und (b) ist die behauptete Gleichung genau die Taylor-Formel aus Satz 26.16 für $r = 0$. □

Bemerkung 26.18. Bereits am Mittelwertsatz sieht man gut, dass die Voraussetzung einer Funktion mit nur eindimensionaler Wertemenge in Satz 26.16 wesentlich ist: Ist $f\colon D \to \mathbb{R}^m$ mit $D \subset \mathbb{R}^n$ eine stetig differenzierbare Abbildung mit Komponentenfunktionen $f_1, \dots, f_m$ und sind $a, x \in D$ wie oben, so können wir zwar durch Anwendung des gerade gezeigten Mittelwertsatzes auf $f_1, \dots, f_m$ Punkte $c_1, \dots, c_m \in \overline{ax}$ mit

$$f_i(x) - f_i(a) = f_i'(c_i) \cdot (x - a)$$

finden – da die Komponentenfunktionen $f_1, \dots, f_m$ nichts miteinander zu tun haben müssen, werden diese Punkte aber im Allgemeinen verschieden sein, so dass wir nicht erwarten können, ein gemeinsames $c \in \overline{ax}$ mit

$$f(x) - f(a) = f'(c) \cdot (x - a)$$

zu finden. Ganz explizit kann man dies z. B. an der Abbildung

$$f\colon \mathbb{R} \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} \cos x \\ \sin x \end{pmatrix} \quad \text{mit Ableitung} \quad f'(x) = \begin{pmatrix} -\sin x \\ \cos x \end{pmatrix}$$

sehen: Für $a = 0$ und $x = 2\pi$ ist $f(x) - f(a) = 0$, aber es gibt offensichtlich kein $c \in [0, 2\pi]$ mit $0 = f'(c) \cdot (2\pi - 0)$.

Der Mittelwertsatz (und allgemeiner die Taylor-Formel aus Satz 26.16) gilt also nur für Funktionen nach $\mathbb{R}$ – bzw. bei Funktionen nach $\mathbb{R}^m$ nur für die Komponentenfunktionen einzeln. Für mehrdimensionale Wertebereiche können wir lediglich wie folgt eine dem Mittelwertsatz ähnliche Abschätzung angeben, die wir später noch mehrfach benötigen werden.

Folgerung 26.19. *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^m$ eine stetig differenzierbare Funktion mit Komponentenfunktionen $f_1, \dots, f_m$. Sind $a, x \in D$ zwei Punkte mit $\overline{ax} \subset D$, so gilt*

$$\|f(x) - f(a)\|_\infty \le \|f'|_{\overline{ax}}\|_\infty \cdot \|x - a\|_\infty,$$

wobei mit $\|f'|_{\overline{ax}}\|_\infty$ analog zu Beispiel 24.39 (b) „das Maximum der Zeilensummennorm von f' auf der Strecke $\overline{ax}$“ gemeint ist, also

$$\|f'|_{\overline{ax}}\|_\infty := \max\left\{ \sum_{j=1}^n |\partial_j f_i(c)| : i \in \{1, \dots, m\}, c \in \overline{ax} \right\}.$$

Beweis. Beachte zunächst, dass das angegebene Maximum existiert, da die nach Voraussetzung stetigen Funktionen $\sum_{j=1}^n |\partial_j f_i|$ auf der kompakten Menge $\overline{ax}$ gemäß Folgerung 24.34 ein Maximum annehmen.

Nach dem Mittelwertsatz aus Folgerung 26.17 gibt es nun für alle $i = 1, \dots, m$ einen Punkt $c_i \in \overline{ax}$ mit $f_i(x) - f_i(a) = f_i'(c_i) \cdot (x - a)$, also insbesondere mit

$$\begin{aligned} |f_i(x) - f_i(a)| &= \left| \sum_{j=1}^n \partial_j f_i(c_i) \cdot (x_j - a_j) \right| \le \sum_{j=1}^n |\partial_j f_i(c_i)| \cdot |x_j - a_j| \le \sum_{j=1}^n |\partial_j f_i(c_i)| \cdot \|x - a\|_\infty \\ &\le \|f'|_{\overline{ax}}\|_\infty \cdot \|x - a\|_\infty. \end{aligned}$$

Mit der Definition der Maximumsnorm auf $\mathbb{R}^m$ folgt daraus die behauptete Aussage. □

26.B Extremwerte

Als Nächstes betrachten wir nun den für Extremwertberechnungen wichtigen Fall $r = 1$ in der Taylor-Formel aus Satz 26.16, in dem das Restglied also aus den zweiten Ableitungen der betrachteten Funktion besteht. Wir hatten in Lemma 26.2 ja schon gesehen, dass an einem lokalen Extremum die Ableitung der betrachteten Funktion notwendigerweise gleich Null sein muss. Wie im eindimensionalen Fall in Satz 11.19 geben die zweiten Ableitungen auch hier nun an einer solchen Stelle oft Auskunft darüber, ob wirklich ein Extremum vorliegt und ob es sich dabei um ein Maximum oder Minimum handelt – und zwar abhängig davon, ob die Hesse-Matrix gemäß Definition 21.9 (b) und Bemerkung 21.10 (b) positiv definit, negativ definit oder indefinit ist.

Satz 26.20 (Extremwertkriterium). *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion. Ferner sei $a \in D$ ein kritischer Punkt von f, also $f'(a) = 0$. Ist dann die Hesse-Matrix $Hf(a)$ ...*

(a) *positiv definit, so hat f in a ein isoliertes lokales Minimum.*

(b) *negativ definit, so hat f in a ein isoliertes lokales Maximum.*

(c) *indefinit, so hat f in a kein lokales Extremum.*

Beweis. Nach eventuellem Verkleinern von D auf eine Kugel mit Mittelpunkt a können wir annehmen, dass Verbindungsstrecken von a zu einem beliebigen Punkt in D wieder in D liegen. Die Taylor-Formel aus Satz 26.16 für $r = 1$ in der Notation von Bemerkung 26.14 liefert dann zu jedem $x \in D$ ein $c \in \overline{ax} \subset D$ mit

$$f(x) = f(a) + f'(a) \cdot (x-a) + \frac{1}{2}(x-a)^\mathsf{T} \cdot Hf(c) \cdot (x-a).$$

Da nach Voraussetzung $f'(a) = 0$ gilt, müssen wir also untersuchen, ob der Ausdruck

$$f(x) - f(a) = \frac{1}{2}(x-a)^\mathsf{T} \cdot Hf(c) \cdot (x-a) \qquad (*)$$

für x in einer Umgebung von a immer positiv ist, immer negativ ist, oder wechselndes Vorzeichen hat.

(a) Nach dem Hurwitz-Kriterium aus Satz 21.43 ist die Hesse-Matrix $Hf(x)$ für ein $x \in D$ genau dann positiv definit, wenn

$$\det((\partial_i \partial_j f(x))_{i,j=1,\dots,k}) > 0 \quad \text{für alle } k = 1, \dots, n.$$

Da f nach Voraussetzung zweimal stetig differenzierbar ist, sind diese Determinanten nun stetige Funktionen in x, und damit ist die Menge

$$U := \{x \in D : Hf(x) \text{ ist positiv definit}\} = \bigcap_{k=1}^{n} \left\{x \in D : \det((\partial_i \partial_j f(x))_{i,j=1,\dots,k}) > 0\right\}$$

nach Lemma 23.35 (a) und Beispiel 24.18 als endlicher Durchschnitt offener Mengen offen und enthält nach Voraussetzung den Punkt a. Nach evtl. erneutem Verkleinern von D können wir daher $D \subset U$ annehmen.

Dann ist $Hf(c)$ in $(*)$ aber stets positiv definit, und damit $f(x) - f(a) > 0$, d. h. $f(x) > f(a)$ für alle $x \neq a$. Also hat f in a ein isoliertes lokales Minimum.

(b) folgt genauso wie (a), indem wir die Bedingungen an U gemäß Satz 22.33 (b) so abändern, dass sie der negativen Definitheit entsprechen.

(c) Wir zeigen zunächst, dass f in a kein lokales Maximum hat.

Da $Hf(a)$ indefinit ist, gibt es einen Vektor $v \in \mathbb{R}^n$ mit $v^\mathsf{T} \cdot Hf(a) \cdot v > 0$. Wie in (a) sind die Einträge der Hesse-Matrix $Hf(x)$ stetige Funktionen in x, und damit ist die Menge

$$U := \{x \in D : v^\mathsf{T} \cdot Hf(x) \cdot v > 0\}$$

nach Beispiel 24.18 eine offene Umgebung von a. Wie oben können wir nach evtl. Verkleinern von D also wieder $D \subset U$ annehmen. Dann ist in $(*)$ aber für alle $t > 0$, so dass $x := a + tv$ (und damit auch c) noch in D liegt,

$$f(x) - f(a) = \frac{1}{2} t^2 v^\mathsf{T} \cdot Hf(c) \cdot v > 0$$

und damit $f(x) > f(a)$. Also hat f in a kein lokales Maximum.

Natürlich zeigt man mit Hilfe eines Vektors $w \in \mathbb{R}^n$ mit $w^\mathsf{T} \cdot Hf(a) \cdot w < 0$ analog, dass f in a auch kein lokales Minimum hat. □

Bemerkung 26.21. Beachte, dass das Kriterium aus Satz 26.20 nicht in jedem Fall entscheiden kann, ob an einem kritischen Punkt wirklich ein Extremum vorliegt: Hat die Hesse-Matrix den Eigenwert 0, und sonst nur positive oder nur negative Eigenwerte, so trifft nach Satz 22.33 (a) keiner der drei Fälle von Satz 26.20 zu. In diesem Fall könnte man nun (analog zum Eindimensionalen in Satz 11.19) höhere als zweite Ableitungen von f betrachten und dafür ähnliche Kriterien beweisen. Da diese höheren Ableitungen im Mehrdimensionalen aufgrund der Vielzahl der partiellen Ableitungen recht kompliziert zu untersuchen sind, werden wir dies hier aber nicht weiter ausführen.

Bemerkung 26.22 (Geometrische Interpretation des Extremwertkriteriums)**.** Die Aussage von Satz 26.20 ist auch leicht anschaulich zu verstehen: Nach Bemerkung 26.14 erhalten wir aus der Taylor-Formel für $r = 2$ um einen Punkt a mit $f'(a) = 0$ ja die Näherung

$$f(x) \approx f(a) + \frac{1}{2}(x-a)^\mathsf{T} \cdot Hf(a) \cdot (x-a).$$

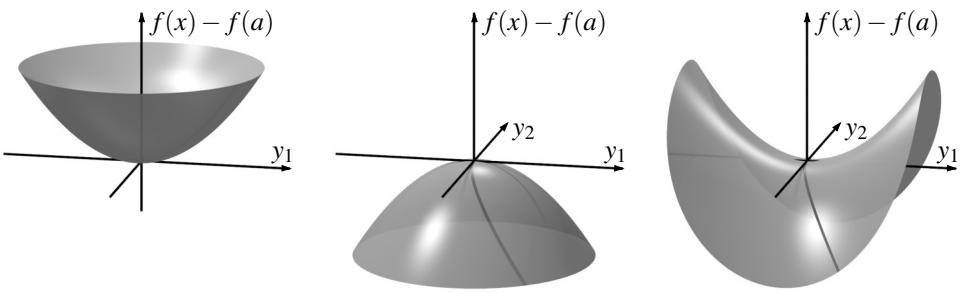
Zu der symmetrischen Hesse-Matrix $Hf(a)$ können wir nun nach dem Trägheitssatz von Sylvester aus Satz 22.36 bzw. Bemerkung 22.37 eine invertierbare Matrix $T \in \mathrm{GL}(n, \mathbb{R})$ finden, so dass

$$T^\mathsf{T} \cdot Hf(a) \cdot T = \mathrm{diag}(\underbrace{1, \ldots, 1}_{k}, \underbrace{-1, \ldots, -1}_{l}, 0, \ldots, 0) =: D$$

gilt, wobei k und l die Anzahl der positiven bzw. negativen Eigenwerte von $Hf(a)$ ist. Mit der linearen Koordinatentransformation $x = a + Ty$ wird die obige Näherungsformel dann zu

$$f(x) \approx f(a) + \frac{1}{2} y^\mathsf{T} \cdot D \cdot y = f(a) + \frac{1}{2}(y_1^2 + \cdots + y_k^2 - y_{k+1}^2 - \cdots - y_{k+l}^2).$$

Wie in den Bildern unten im zweidimensionalen Fall dargestellt, kann man hieraus leicht das lokale Verhalten von f um a ablesen: Im positiv definiten Fall $k = n$ ist die Differenz $f(x) - f(a)$ als Summe von Quadraten für $x \neq a$ positiv, und damit liegt dort dann ein lokales Minimum vor. Der Schnitt des Graphen von f mit der y_1- oder y_2-Koordinatenachse ist in diesem Fall näherungsweise eine nach oben geöffnete quadratische Parabel. Für eine negativ definite Hesse-Matrix mit $l = n$ haben wir analog ein Maximum.



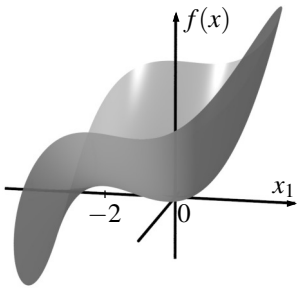
$Hf(a)$ positiv definit	$Hf(a)$ negativ definit	$Hf(a)$ indefinit
$f(x) - f(a) \approx \frac{1}{2}(y_1^2 + y_2^2)$	$f(x) - f(a) \approx \frac{1}{2}(-y_1^2 - y_2^2)$	$f(x) - f(a) \approx \frac{1}{2}(y_1^2 - y_2^2)$
Minimum	Maximum	Sattelpunkt

Im indefiniten Fall hat f auf den verschiedenen Geraden durch a manchmal ein Maximum und manchmal ein Minimum – man sagt auch, dass in diesem Fall ein **Sattelpunkt** vorliegt. Ist die Hesse-Matrix schließlich weder positiv definit, negativ definit noch indefinit, ist sie also z. B. einfach die Nullmatrix, so sagt die obige Näherung einfach $f(x) \approx f(a)$ für manche oder sogar alle x, und in diesem Fall müsste man sich diese „näherungsweise Gleichheit" noch genauer anschauen, um entscheiden zu können, ob $f(x)$ nun etwas größer oder etwas kleiner ist als $f(a)$.

Beispiel 26.23. Wir wollen die (lokalen) Extrema der Funktion

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto 3x_1^2 + x_1^3 + x_2^2$$

(siehe Bild rechts) berechnen. Dazu bestimmen wir zunächst gemäß Lemma 26.2 die kritischen Punkte als mögliche Kandidaten für Extrema, also die Stellen mit Ableitung Null: Es ist



$$f'(x) = (6x_1 + 3x_1^2, 2x_2)$$

und damit $f'(x) = 0$ genau dann, wenn

$$3x_1(2+x_1) = 0 \quad \text{und} \quad 2x_2 = 0,$$

woraus sich die möglichen Extremstellen $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$ ergeben.

Für unser Extremwertkriterium müssen wir jetzt die Hesse-Matrix berechnen: Es ist

$$Hf(x) = \begin{pmatrix} \partial_1\partial_1 f(x) & \partial_1\partial_2 f(x) \\ \partial_2\partial_1 f(x) & \partial_2\partial_2 f(x) \end{pmatrix} = \begin{pmatrix} 6+6x_1 & 0 \\ 0 & 2 \end{pmatrix}.$$

An der Stelle $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ist also

$$Hf\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{mit} \quad \det(6) = 6 > 0 \quad \text{und} \quad \det\begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix} = 12 > 0,$$

womit Hf dort nach Satz 21.43 positiv definit ist. Also hat f an diesem Punkt nach Satz 26.20 (a) ein isoliertes lokales Minimum (wie auch im Bild ersichtlich ist). An der Stelle $\begin{pmatrix} -2 \\ 0 \end{pmatrix}$ hingegen ist

$$Hf\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} -6 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{mit} \quad \det(-6) = -6 < 0 \quad \text{und} \quad \det\begin{pmatrix} -6 & 0 \\ 0 & 2 \end{pmatrix} = -12 < 0,$$

und damit ist Hf dort nach Satz 22.33 (b) indefinit. Es liegt an dieser Stelle nach Satz 26.20 (c) also kein lokales Extremum, sondern ein Sattelpunkt vor.

Aufgabe 26.24. Bestimme alle lokalen Minima und Maxima der Funktionen

(a) $f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_1^3 + x_2^3 + 3x_1x_2$;

(b) $g\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_2(x_2 - \cos x_1 - 1) + \cos x_1$.

Gib zusätzlich von der Funktion g das zweite Taylor-Polynom mit Entwicklungspunkt $\begin{pmatrix} \pi/2 \\ 0 \end{pmatrix}$ an.

Aufgabe 26.25. Zeige, dass die Funktion $f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto (x_2 - x_1^2)(x_2 - 2x_1^2)$ keine lokalen Extrema hat, dass die Einschränkung von f auf jede Gerade durch den Ursprung aber ein lokales Minimum in 0 besitzt.

Aufgabe 26.26. Für gegebene Punkte $a_1, \dots, a_k \in \mathbb{R}^n$ betrachten wir die Summe der Abstandsquadrate

$$f\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \sum_{i=1}^{k} \|x - a_i\|_2^2.$$

Bestimme alle lokalen und globalen Extrema von f.

Aufgabe 26.27. Es sei $f\colon \mathbb{R}^n \to \mathbb{R}$ eine zweimal stetig differenzierbare Funktion, für die die Hesse-Matrix an jedem Punkt positiv definit ist. Zeige, dass f höchstens ein lokales Extremum besitzt.

68

27. Implizite Funktionen

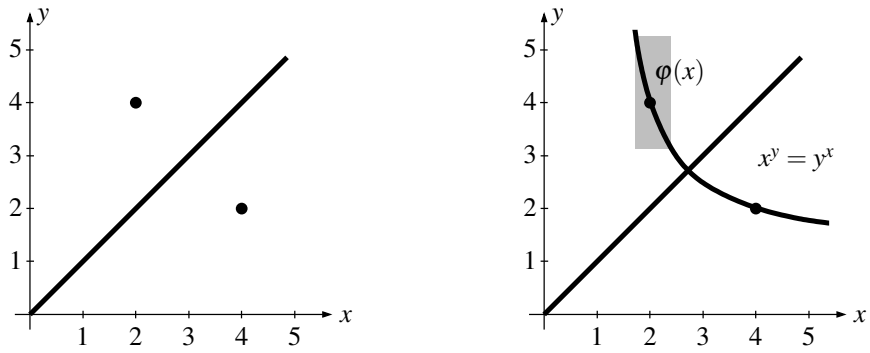
Bevor wir uns ab dem nächsten Kapitel der Integration widmen, wollen wir zum Abschluss unseres Studiums differenzierbarer Abbildungen noch auf das in der Praxis sehr wichtige Thema der sogenannten impliziten Funktionen eingehen, bei dem es um die Auflösbarkeit von Gleichungen nach bestimmten Variablen geht. Die Idee dieser Situation lässt sich am besten an einem Beispiel erklären.

Beispiel 27.1. Für $x, y \in \mathbb{R}_{>0}$ wollen wir die Lösungsmenge der Gleichung $x^y = y^x$ untersuchen.

Diese Gleichung lässt sich mit den uns bekannten speziellen Funktionen aus Kapitel 9 nicht nach x oder y auflösen, da beide Variablen sowohl in der Basis als auch im Exponenten auftreten. Wir sehen allerdings schon einmal die Lösungen

- $y = x$ für beliebige $x, y > 0$, sowie
- $(x, y) = (2, 4)$ und $(x, y) = (4, 2)$,

also die im folgenden Bild links eingezeichnete Punktmenge.



Dieses Bild sieht natürlich sehr merkwürdig aus: Ist z. B. der Punkt $(2, 4)$ wirklich ein isolierter Punkt der Lösungsmenge oder gibt es in einer Umgebung davon noch weitere Lösungen? Um dies zunächst einmal numerisch herauszufinden, könnten wir die Lösungsmenge der betrachteten Gleichung von einem Computer berechnen lassen, der einfach alle Punkte der Ebene abtastet und diejenigen Paare (x, y) zeichnet, bei denen x^y gleich bzw. sehr nahe bei y^x ist. Das Ergebnis, das wir so erhalten würden, ist im Bild oben rechts dargestellt.

Danach sieht es so aus, als ob die Lösungsmenge der gegebenen Gleichung aus zwei Zweigen besteht: den Punkten mit $y = x$, und einer weiteren Kurve, die sich als Graph einer (stetig differenzierbaren) Funktion schreiben lässt. In einer (im Bild oben grau eingezeichneten) Umgebung des Punktes $(2, 4)$ kann man die gegebene Gleichung $x^y = y^x$ also z. B. anscheinend nach y auflösen und als Funktionsgleichung $y = \varphi(x)$ schreiben – auch wenn wir diese Funktion nicht explizit angeben können. Man sagt, dass diese Funktion φ *implizit* durch die Gleichung $x^y = y^x$ definiert ist. Am Kreuzungspunkt der beiden Zweige oben lässt sich die gegebene Gleichung jedoch nicht nach einer der beiden Variablen auflösen und z. B. y als Funktion von x schreiben, weil in einer Umgebung dieses Punktes ja für jeden Wert von x *zwei* mögliche Werte y mit $x^y = y^x$ existieren.

Ziel dieses Kapitels ist es, derartige Aussagen exakt zu beweisen. Dabei werden wir auch sehen, dass man mit solchen nicht nach einer Variablen aufgelösten Funktionsdefinitionen durchaus arbeiten kann. So werden wir z. B. im obigen Bild den Schnittpunkt der beiden Zweige und die Ableitung der Funktion φ im Punkt $(2, 4)$ bestimmen können, auch ohne diese Funktion explizit zu kennen (siehe Beispiel 27.12 (b)).

27.A Umkehrfunktionen

Wir beginnen unser Studium impliziter Funktionen mit einem wichtigen Spezialfall: Sind $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^n$ eine gegebene Funktion, so wollen wir untersuchen, ob wir die Gleichung $y = f(x)$ (mit $x \in D$ und $y \in \mathbb{R}^n$) nach x auflösen, also eine Umkehrfunktion $x = f^{-1}(y)$ finden können. Im eindimensionalen Fall wissen wir dies bereits:

Beispiel 27.2 (Umkehrbarkeit im Eindimensionalen)**.** Sind $D = (a,b)$ ein offenes Intervall und $f\colon D \to \mathbb{R}$ eine stetig differenzierbare Funktion mit $f'(x) \neq 0$ für alle $x \in D$, so ist f' zunächst einmal entweder überall positiv oder überall negativ, da f' sonst nach dem Zwischenwertsatz 8.21 auch irgendwo den Wert 0 annehmen müsste. Also ist f nach Folgerung 10.25 dann streng monoton und damit injektiv. Auf dem Bildbereich $f(D)$, der nach dem Zwischenwertsatz ebenfalls ein Intervall ist, existiert also eine Umkehrfunktion f^{-1} von f, und diese ist nach Satz 10.11 ebenfalls differenzierbar mit Ableitung $(f^{-1})'(f(x)) = \frac{1}{f'(x)}$. Mit anderen Worten ist die Gleichung $y = f(x)$ somit auf dem betrachteten Intervall nach x auflösbar, nämlich durch die Umkehrfunktion $x = f^{-1}(y)$.

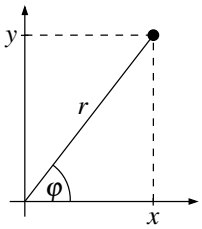
Bemerkung 27.3 (Ableitung einer Umkehrfunktion)**.** Die Formel für die Ableitung der Umkehrfunktion lässt sich sofort auf den mehrdimensionalen Fall verallgemeinern: Ist $f\colon D \to D'$ eine differenzierbare Funktion zwischen offenen Teilmengen D und D' von $\mathbb{R}^n$, und wissen wir bereits, dass f bijektiv ist und eine ebenfalls differenzierbare Umkehrfunktion $f^{-1}\colon D' \to D$ besitzt, so erhält man durch Differenzieren der Gleichung $f^{-1}(f(x)) = x$ für alle $x \in D$ mit der Kettenregel aus Satz 25.29 sofort $(f^{-1})'(f(x)) \cdot f'(x) = E$, da die Ableitung von $x \mapsto x = Ex$ nach Beispiel 25.5 die Einheitsmatrix ist. Also muss die Matrix $f'(x) \in \mathrm{Mat}(n \times n, \mathbb{R})$ invertierbar sein, d. h. es muss $\det f'(x) \neq 0$ gelten, und für die Ableitung von f^{-1} ist analog zum Eindimensionalen die inverse Matrix

$$(f^{-1})'(f(x)) = (f'(x))^{-1}.$$

Wo wir im Eindimensionalen $f'(x) \neq 0$ für alle $x \in D$ vorausgesetzt haben, müssen wir nun also verlangen, dass die Matrix $f'(x)$ überall invertierbar ist. Überraschend ist dabei allerdings, dass die Invertierbarkeit von $f'(x)$ für $n > 1$ im Gegensatz zum eindimensionalen Fall in Beispiel 27.2 nicht mehr hinreichend für die Existenz einer Umkehrfunktion ist, wie das folgende einfache Beispiel zeigt.

Beispiel 27.4 (Umkehrbarkeit der Polarkoordinaten)**.** Wir betrachten noch einmal die Polarkoordinatenabbildung in $\mathbb{C} = \mathbb{R}^2$

$$f\colon \mathbb{R}_{>0} \times \mathbb{R} \to \mathbb{R}^2,\ \begin{pmatrix} r \\ \varphi \end{pmatrix} \mapsto \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix} =: \begin{pmatrix} x \\ y \end{pmatrix}$$



(siehe Satz 9.26 und Definition 9.12), die dem Betrag und Winkel einer komplexen Zahl ihren Real- und Imaginärteil zuordnet. Dann ist die Ableitungsmatrix von f

$$f' = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}, \tag{1}$$

und damit nach Satz 9.14 (b)

$$\det f' = r(\cos^2\varphi + \sin^2\varphi) = r > 0.$$

Die Ableitungsmatrix f' ist also überall invertierbar. Trotzdem ist f aber nicht injektiv, da die Addition von Vielfachen von 2π zum Winkel φ nichts am Funktionswert ändert. Um eine bijektive Abbildung zu erhalten, müssen wir f einschränken: Betrachten wir z. B. wie im Bild unten nur die Werte von r und φ mit $1 < r < 2$ und $0 < \varphi < \frac{\pi}{2}$, so ist f auf dieser offenen Teilmenge U des Definitionsbereichs injektiv. Das Bild dieser Teilmenge unter f ist der unten rechts im Bild dargestellte Viertelkreisring V, so dass die Einschränkung $f|_U\colon U \to V$ nun bijektiv ist und damit eine Umkehrabbildung $f^{-1}\colon V \to U$ besitzt. In der Tat können wir diese Umkehrabbildung auch sofort aus dem

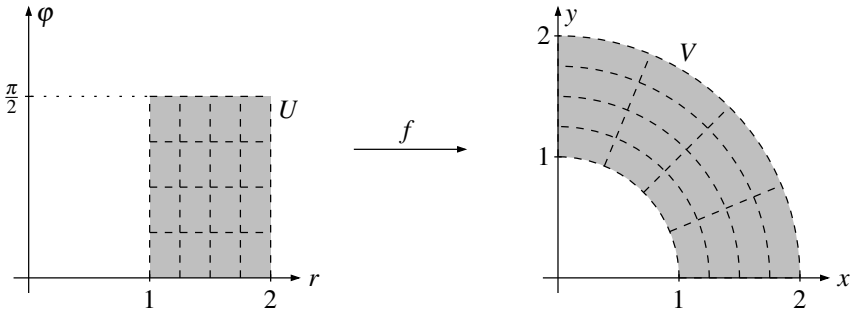
geometrischen Bild oben rechts ablesen: Man kann r und φ in diesem Winkelbereich offensichtlich mit den (stetig differenzierbaren) Formeln

$$r = \sqrt{x^2+y^2} \quad \text{und} \quad \varphi = \arctan\frac{y}{x}$$

aus x und y zurückgewinnen, so dass also

$$f^{-1}\colon V \to U,\ \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{x^2+y^2} \\ \arctan\frac{y}{x} \end{pmatrix} \tag{2}$$

dort die Umkehrfunktion ist.



Gemäß Bemerkung 27.3 können wir die Ableitung dieser Umkehrfunktion g nun als die zu f' inverse Matrix berechnen: Mit (1) und Beispiel 18.21 erhalten wir

$$(f^{-1})' = \frac{1}{r}\cdot\begin{pmatrix} r\cos\varphi & r\sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix},$$

wie man durch Differenzieren der expliziten Formel (2) für f^{-1} natürlich auch direkt bestätigen könnte.

Auch bei überall invertierbarer Ableitungsmatrix können wir also nur lokal (d. h. nach geeigneter Einschränkung des Definitionsbereichs auf eine offene Umgebung eines Punktes) erwarten, eine bijektive Abbildung zu erhalten. Dass dieses Phänomen erst bei einer Raumdimension $n > 1$ sichtbar wird, hat letztlich topologische Gründe: Wählen wir in unserem Beispiel einen „zu großen" Definitionsbereich wie z. B. $U = (1,2)\times(0,3\pi)$, so würde das Bild von $f|_U$ den Kreisring mit innerem Radius 1 und äußerem Radius 2 eineinhalbmal durchlaufen, so dass f dort dann nicht mehr injektiv wäre – für solche „Schleifen" um den Nullpunkt herum ist auf der eindimensionalen reellen Zahlengeraden aber „nicht genug Platz".

Nach diesen Vorbemerkungen wollen wir nun den technisch etwas aufwändigen, aber sicher nicht mehr unerwarteten Satz beweisen, dass eine stetig differenzierbare Abbildung mit invertierbarer Ableitungsmatrix stets lokal umkehrbar ist. Wir beginnen dazu mit dem folgenden Lemma, in dem wir einige spezielle Koordinatenwahlen getroffen haben, um den Beweis einfacher zu halten. Es enthält die eigentliche technische Arbeit, die für die Sätze in diesem Kapitel erforderlich ist – die weiteren Aussagen werden sich daraus dann durch geeignete Koordinatentransformationen als Anwendungen ergeben.

Lemma 27.5 (Lemma über lokale Umkehrfunktionen)**.** *Es seien $D \subset \mathbb{R}^n$ offen mit $0 \in D$, und $f\colon D \to \mathbb{R}^n$ eine stetig differenzierbare Abbildung mit $f(0) = 0$ und $f'(0) = E_n$.*

Dann ist die Funktion f lokal um den Ursprung umkehrbar, d. h. es gibt offene Umgebungen U und V von $0 \in \mathbb{R}^n$ mit $U \subset D$, so dass die Einschränkung $f|_U\colon U \to V$ bijektiv ist. Darüber hinaus ist die dann existierende Umkehrfunktion $f^{-1}\colon V \to U$ ebenfalls differenzierbar in 0 mit Ableitung E_n.

Beweis. Die Beweisidee dieses Lemmas besteht darin, Urbilder unter f als Fixpunkte einer geeigneten Hilfsfunktion umzuschreiben, und diese Fixpunkte dann mit Hilfe des Banachschen Fixpunktsatzes 24.26 zu untersuchen. Wir verwenden im Beweis durchgehend die Maximumsnorm auf $\mathbb{R}^n$ und die davon induzierte Zeilensummennorm auf $\mathrm{Mat}(n \times n, \mathbb{R})$ (siehe Beispiel 24.39 (b)).

Da D nach Voraussetzung eine offene Umgebung von 0 ist und f' stetig mit $f'(0) = E$, gibt es ein $\varepsilon > 0$, so dass $x \in D$ und

$$\|E - f'(x)\| < \frac{1}{2} \tag{$*$}$$

für alle $x \in \mathbb{R}^n$ mit $\|x\| < \varepsilon$ gilt. Es sei nun $y \in \mathbb{R}^n$ mit $\|y\| < \frac{\varepsilon}{4}$. Um Urbilder von y unter f zu suchen, betrachten wir die (von y abhängige) Hilfsfunktion

$$\varphi\colon D \to \mathbb{R}^n,\ x \mapsto x - f(x) + y,$$

so dass $f(x) = y$ genau dann gilt wenn $\varphi(x) = x$, die Urbilder von y unter f also genau die Fixpunkte von φ sind. Um den Banachschen Fixpunktsatz anwenden zu können, zeigen wir nun, dass φ auf der abgeschlossenen Kugel $K_{\frac{\varepsilon}{2}}(0)$ eine Kontraktion ist:

(a) Für alle $x, \tilde{x} \in K_{\frac{\varepsilon}{2}}(0)$ gilt

$$\|\varphi(x) - \varphi(\tilde{x})\| \overset{26.19}{\leq} \|\varphi'|_{\overline{x\tilde{x}}}\| \cdot \|x - \tilde{x}\| = \|(E - f')|_{\overline{x\tilde{x}}}\| \cdot \|x - \tilde{x}\| \overset{(*)}{\leq} \frac{1}{2}\|x - \tilde{x}\|.$$

(b) Für alle $x \in K_{\frac{\varepsilon}{2}}(0)$ gilt

$$\|\varphi(x)\| \leq \|\varphi(x) - \varphi(0)\| + \|\varphi(0)\| \overset{(a)}{\leq} \frac{1}{2}\|x\| + \|y\| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}$$

und damit auch $\varphi(x) \in K_{\frac{\varepsilon}{2}}(0)$.

Da $K_{\frac{\varepsilon}{2}}(0)$ als abgeschlossene Teilmenge von $\mathbb{R}^n$ nach Satz 23.29 und Folgerung 23.43 vollständig ist, hat φ nach Satz 24.26 dort also genau einen Fixpunkt, d. h. y besitzt in $K_{\frac{\varepsilon}{2}}(0)$ genau ein Urbild unter f. Da für $\varphi(x) = x$ aus (b) außerdem $\|x\| < \frac{\varepsilon}{2}$ folgt, liegt dieses Urbild sogar in der offenen Kugel $U_{\frac{\varepsilon}{2}}(0)$. Insgesamt ist damit die eingeschränkte Abbildung

$$f\colon \underbrace{U_{\frac{\varepsilon}{2}}(0) \cap f^{-1}(U_{\frac{\varepsilon}{4}}(0))}_{=:U} \to \underbrace{U_{\frac{\varepsilon}{4}}(0)}_{=:V}$$

bijektiv. Da U und V nach Lemma 23.35 (a) und Satz 24.17 (b) offen sind, zeigt dies die behauptete Existenz einer lokalen Umkehrabbildung.

Es bleibt also nur noch zu zeigen, dass f^{-1} im Nullpunkt mit Ableitung E differenzierbar ist. Wir prüfen dies wie in Algorithmus 25.15 nach: Ist $y = f(x)$, also $\varphi(x) = x$, so folgt $\|x\| \leq \frac{1}{2}\|x\| + \|y\|$ und damit $\|x\| \leq 2\|y\|$ aus (b), und damit wie gewünscht

$$\frac{f^{-1}(y) - f^{-1}(0) - Ey}{\|y\|} = \frac{x - f(x)}{\|y\|} = -\underbrace{\frac{f(x) - f(0) - Ex}{\|x\|}}_{\to 0} \cdot \underbrace{\frac{\|x\|}{\|y\|}}_{\leq 2} \to 0$$

für $y \to 0$, da mit $y \to 0$ wegen $\|x\| \leq 2\|y\|$ auch $x \to 0$ folgt und f in 0 differenzierbar mit Ableitung E ist. □

Satz 27.6 (Satz über lokale Umkehrfunktionen). *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^n$ eine stetig differenzierbare Abbildung. Weiterhin sei $a \in D$ ein Punkt, an dem* $\det f'(a) \neq 0$ *ist.*

Dann ist f lokal um den Punkt $(a, f(a))$ umkehrbar, d. h. es gibt offene Umgebungen $U \subset D$ von a und $V \subset \mathbb{R}^n$ von $f(a)$, so dass die Einschränkung $f|_U\colon U \to V$ bijektiv ist. Darüber hinaus ist die dann existierende Umkehrfunktion $f^{-1}\colon V \to U$ ebenfalls stetig differenzierbar mit Ableitung

$$(f^{-1})'(f(x)) = (f'(x))^{-1} \quad \textit{für alle } x \in U.$$

Beweis. Wir führen den Beweis in mehreren Reduktionsschritten auf Lemma 27.5 zurück.

(a) Indem wir D evtl. auf die nach Beispiel 24.18 offene Umgebung $\{x \in D : \det f'(x) \neq 0\}$ von a verkleinern, können wir annehmen, dass die Voraussetzung $\det f' \neq 0$ des Satzes nicht nur bei a, sondern sogar an jedem Punkt von D gilt.

(b) Es genügt zu zeigen, dass eine lokale Umkehrfunktion $f^{-1}\colon V \to U$ existiert und im Punkt $f(a)$ differenzierbar ist: Nach (a) können wir dieses Ergebnis dann auf jeden Punkt $x \in D$ anwenden und erhalten, dass f^{-1} sogar auf ganz V differenzierbar ist. Die Formel $(f^{-1})'(f(x)) = (f'(x))^{-1}$ für die Ableitung der Umkehrfunktion hatten wir dann bereits in Bemerkung 27.3 hergeleitet. Mit $y = f(x)$ ist also $(f^{-1})'(y) = (f'(f^{-1}(y)))^{-1}$, und da dieser Ausdruck aus stetigen Funktionen zusammengesetzt ist, ist f^{-1} wie behauptet auch stetig differenzierbar.

(c) Nach geeigneter Koordinatentransformation im Start- und Zielraum können wir $a = f(a) = 0$ annehmen: Führen wir die neuen Koordinaten $\tilde{x} = x - a$ und $\tilde{y} = y - f(a)$ ein und finden wir von der entsprechend transformierten Funktion $\tilde{f}\colon \tilde{x} \mapsto \tilde{y} = f(x) - f(a) = f(\tilde{x} + a) - f(a)$ eine lokale Umkehrfunktion $\tilde{g}$ um $\tilde{x} = 0$, so dass also $\tilde{g}(\tilde{y}) = \tilde{x}$ gilt, so ist natürlich durch $g(y) := \tilde{g}(y - f(a)) + a = x$ eine lokale Umkehrfunktion von f um $x = a$ gegeben.

(d) Durch eine weitere Koordinatentransformation im Zielraum können wir zusätzlich $f'(0) = E$ annehmen: Mit $A := f'(0)$ betrachten wir die Funktion $\tilde{f}\colon x \mapsto \tilde{y} := A^{-1} \cdot y = A^{-1} \cdot f(x)$. Diese hat im Nullpunkt nach der Kettelregel die Ableitung $\tilde{f}'(0) = A^{-1} \cdot f'(0) = E$ – und wenn sie um 0 eine lokale Umkehrfunktion $\tilde{g}$ besitzt, so dass also $\tilde{g}(\tilde{y}) = \tilde{g}(A^{-1} \cdot f(x)) = x$ gilt, so ist damit $g(y) := \tilde{g}(A^{-1} \cdot y) = \tilde{g}(\tilde{y}) = x$ eine lokale Umkehrfunktion um 0 von f.

Die Behauptung des Satzes folgt damit aus Lemma 27.5. □

Aufgabe 27.7. Überprüfe, ob die Abbildung $f\colon \operatorname{Mat}(2 \times 2, \mathbb{R}) \to \operatorname{Mat}(2 \times 2, \mathbb{R})$, $A \mapsto A^2$ lokal um $(A, f(A))$ mit

$$\text{(a) } A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad \text{(b) } A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

umkehrbar ist. Ist das Bild von f eine Umgebung der Einheitsmatrix $E \in \operatorname{Mat}(2 \times 2, \mathbb{R})$?

Aufgabe 27.8. Es seien $U \subset \mathbb{R}^n$ offen und beschränkt sowie $f\colon \overline{U} \to \mathbb{R}^n$ eine stetige Funktion, die auf U stetig differenzierbar ist. Ferner sei $y \in \mathbb{R}^n$ ein Punkt, so dass jeder Urbildpunkt $x \in f^{-1}(y)$ in U liegt und $\det f'(x) \neq 0$ erfüllt. Zeige, dass es dann nur endlich viele solcher Urbilder gibt.

69

27.B Der Satz über implizite Funktionen

Nach dem Studium von Umkehrfunktionen kommen wir nun wieder auf die impliziten Funktionen wie in Beispiel 27.1 zurück. Im Sinne der mehrdimensionalen Analysis wollen wir dabei aber nicht nur *eine* Gleichung nach *einer* der darin vorkommenden Variablen auflösen können, sondern n gegebene Gleichungen nach n darin enthaltenen Variablen. Für den folgenden Satz fixieren wir dazu die folgenden Notationen: Auf einer offenen Teilmenge $D \subset \mathbb{R}^r \times \mathbb{R}^n$ mit Koordinaten $u = (u_1, \ldots, u_r)$ (von $\mathbb{R}^r$) und $x = (x_1, \ldots, x_n)$ (von $\mathbb{R}^n$) betrachten wir eine stetig differenzierbare Abbildung $f\colon D \to \mathbb{R}^n$, deren Ableitungsmatrix also die Form

$$f' = \left(\frac{\partial f}{\partial u_1} \,\middle|\, \cdots \,\middle|\, \frac{\partial f}{\partial u_r} \,\middle|\, \frac{\partial f}{\partial x_1} \,\middle|\, \cdots \,\middle|\, \frac{\partial f}{\partial x_n} \right) =: \left(\frac{\partial f}{\partial u} \,\middle|\, \frac{\partial f}{\partial x} \right)$$

hat, wobei die $n \times r$-Matrix $\frac{\partial f}{\partial u}$ die ersten r Spalten und die $n \times n$-Matrix $\frac{\partial f}{\partial x}$ die letzten n Spalten von f' bezeichnet. Wir wollen untersuchen, ob wir die n Gleichungen $f(u,x) = 0 \in \mathbb{R}^n$ in den $r + n$ Variablen u und x in einer Umgebung eines Punktes $(a,b) \in D$ mit $f(a,b) = 0$ nach den n Variablen x auflösen können. In Beispiel 27.1 wäre also $n = r = 1$ und $f\colon \mathbb{R}_{>0} \times \mathbb{R}_{>0}$, $(u,x) \mapsto u^x - x^u$, und z. B. $(a,b) = (2,4)$.

Der Satz über implizite Funktionen, den wir jetzt aus dem Satz über lokale Umkehrfunktionen herleiten wollen, besagt in dieser Situation gerade, dass eine solche lokale Auflösung nach x immer möglich ist, wenn die Ableitungsmatrix $\frac{\partial f}{\partial x}$ an der betrachteten Stelle invertierbar ist. Als zusätzliche

Aussage können wir in diesem Fall auch eine Formel für die Ableitung dieser Auflösungsfunktion angeben.

Satz 27.9 (**Satz über implizite Funktionen**). *Es seien $D \subset \mathbb{R}^r \times \mathbb{R}^n$ eine offene Teilmenge und $f\colon D \to \mathbb{R}^n$, $(u,x) \mapsto f(u,x)$ eine stetig differenzierbare Abbildung. Ferner sei $(a,b) \in D$ ein Punkt mit $f(a,b) = 0$ und $\det \frac{\partial f}{\partial x}(a,b) \neq 0$.*

Dann lässt sich die Gleichung $f(u,x) = 0$ lokal um (a,b) nach x auflösen, d. h. es gibt offene Umgebungen U von $a \in \mathbb{R}^r$ und V von $b \in \mathbb{R}^n$ mit $U \times V \subset D$ sowie eine stetig differenzierbare Funktion $\varphi\colon U \to V$ mit $\varphi(a) = b$, so dass für alle $u \in U$ und $x \in V$ gilt, dass

$$f(u,x) = 0 \quad \textit{genau dann, wenn} \quad x = \varphi(u).$$

Beweis. Wir betrachten die Hilfsfunktion $F\colon D \to \mathbb{R}^r \times \mathbb{R}^n$ mit $F(u,x) = (u, f(u,x))$. Wegen

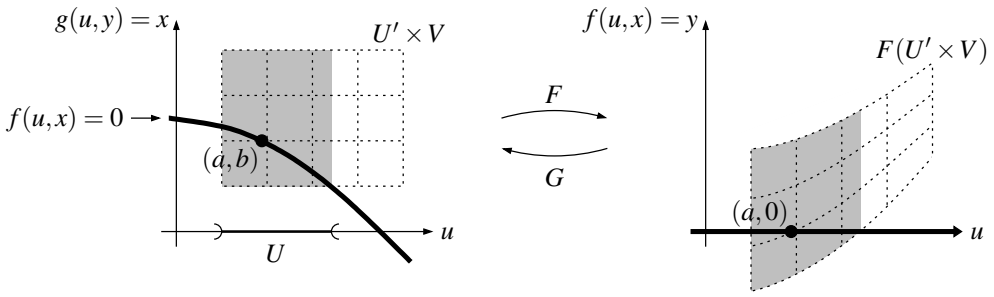
$$F' = \begin{pmatrix} E_r & 0 \\ \frac{\partial f}{\partial u} & \frac{\partial f}{\partial x} \end{pmatrix}, \quad \text{also} \quad \det F'(a,b) = \det E_r \cdot \det \frac{\partial f}{\partial x}(a,b) \neq 0$$

können wir den Satz 27.6 über lokale Umkehrfunktionen auf F an der Stelle (a,b) anwenden und erhalten offene Umgebungen von (a,b) und $F(a,b) = (a,0)$, zwischen denen F bijektiv ist und eine stetig differenzierbare Umkehrfunktion G besitzt. Weil die Umkehrung der ersten Komponente von F, also der Identität, natürlich trivial ist, hat G notwendigerweise die Form $G(u,y) = (u, g(u,y))$ mit

$$g(u, f(u,x)) = x \quad \text{und} \quad f(u, g(u,y)) = y \tag{$*$}$$

für alle u,x,y in den gefundenen Umgebungen, wobei $G(a,0) = (a,b)$ und damit $g(a,0) = b$ gilt.

Durch evtl. Verkleinern können wir erreichen, dass die offene Umgebung von (a,b) von der Form $U' \times V$ mit offenen Umgebungen U' von a und V von b ist (wähle z. B. eine offene Kugel in der Maximumsnorm). Die entsprechende Situation ist dann im folgenden Bild dargestellt, wobei F und G die beiden gestrichelt umrandeten Bereiche bijektiv aufeinander abbilden und die dick eingezeichneten Kurven die gesuchten Stellen mit $f = 0$ markieren. Beachte, dass die erste Komponente von F und G die Identität ist und die vertikalen gestrichelten Linien damit wieder auf vertikale Linien an der gleichen Stelle u abgebildet werden.



Wie im Bild grau markiert verkleinern wir U' nun auf die Menge $U = \{u \in U' : (u,0) \in F(U' \times V)\}$, die wegen $(a,0) = F(a,b)$ immer noch den Punkt a enthält. Als Urbild der offenen Menge $F(U' \times V)$ unter der stetigen Abbildung $u \mapsto (u,0)$ ist U nach Satz 24.17 (b) außerdem offen, also eine offene Umgebung von a. Nach Konstruktion sind dann G und damit auch g auf den Punkten $(u,0)$ mit $u \in U$ definiert, und die Funktion

$$\varphi\colon U \to V,\ u \mapsto g(u,0)$$

leistet das Gewünschte, denn es ist $\varphi(a) = g(a,0) = b$, und nach $(*)$ außerdem $f(u,x) = 0$ genau dann, wenn $x = g(u,0)$. □

Folgerung 27.10 (Ableitung der Auflösungsfunktion). *In der Situation und mit den Notationen von Satz 27.9 ist die Ableitung der Auflösungsfunktion φ für alle $u \in U$ gegeben durch*

$$\varphi'(u) = -\left(\frac{\partial f}{\partial x}(u,x)\right)^{-1} \cdot \frac{\partial f}{\partial u}(u,x)$$

mit $x = \varphi(u)$.

Beweis. Nach Konstruktion der Auflösungsfunktion φ gilt $f(u,\varphi(u)) = 0$ für alle $u \in U$. Differenzieren dieser Gleichung nach u ergibt mit der Kettenregel aus Satz 25.29

$$0 = \left(\frac{\partial f}{\partial u} \,\middle|\, \frac{\partial f}{\partial x}\right) \cdot \begin{pmatrix} E_r \\ \varphi' \end{pmatrix} = \frac{\partial f}{\partial u} + \frac{\partial f}{\partial x} \cdot \varphi',$$

woraus durch Auflösen nach φ' die behauptete Formel folgt. □

Bemerkung 27.11. In Satz 27.9 haben wir die Variablen $x_1,\dots,x_n$, nach denen die gegebene Gleichung $f = 0$ aufgelöst werden soll, mit anderen Buchstaben bezeichnet als die übrigen Variablen $u_1,\dots,u_r$. Für die Formulierung und den Beweis des Satzes war dies sehr praktisch – in konkreten Anwendungen werden wir dies aber in der Regel nicht tun. Die Aussage von Satz 27.9 ist dann anschaulich formuliert: Haben wir n Gleichungen $f(x_1,\dots,x_m) = 0$ in $m \geq n$ Variablen gegeben und können wir n Spalten aus der Ableitungsmatrix $f' \in \mathrm{Mat}(n \times m, \mathbb{R})$ so auswählen, dass die resultierende quadratische Matrix an einem gegebenen Punkt der Lösungsmenge invertierbar ist, so lässt sich die Gleichung $f = 0$ lokal um diesen Punkt nach den entsprechenden Variablen auflösen.

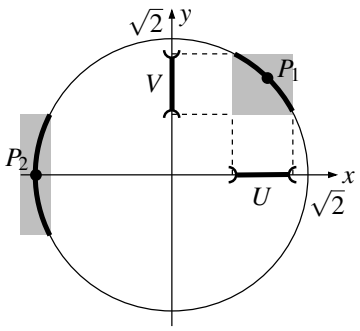
Beispiel 27.12.

(a) Zum besseren Verständnis beginnen wir mit einem Beispiel, bei dem wir die Auflösungsfunktion bereits kennen und uns Satz 27.9 daher nichts Neues sagt: Wir betrachten eine einfache Kreisgleichung $x^2 + y^2 = 2$, d. h. die Gleichung $f(x,y) = 0$ mit

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ f(x,y) = x^2 + y^2 - 2$$

und damit $f' = (2x, 2y)$.

Der Punkt $P_1 = (1,1)$ erfüllt offensichtlich die Gleichung $f(x,y) = 0$, und es ist $\frac{\partial f}{\partial y}(P_1) = 2 \cdot 1 \neq 0$. Nach Satz 27.9 bzw. Bemerkung 27.11 (hier haben wir lediglich *eine* Gleichung, daher ist die Determinante der 1×1-Matrix $\frac{\partial f}{\partial y}(P_1)$ einfach gleich dieser Zahl selbst) kann man die gegebene Gleichung also lokal um P_1 stetig differenzierbar nach y auflösen – was wir natürlich schon vorher wussten, denn in einer Umgebung $U \times V$ von P_1 ist $f(x,y) = 0$ ja einfach äquivalent zu $y = \sqrt{2 - x^2}$.



Am Punkt $P_2 = (-\sqrt{2}, 0)$ hingegen ist $\frac{\partial f}{\partial y}(P_2) = 0$, und in der Tat können wir hier in keiner Umgebung von P_2 die Variable y als Funktion von x ausdrücken: In jeder Umgebung gibt es für x in der Nähe von $-\sqrt{2}$ ja immer zwei oder gar keinen möglichen Wert von y.

Es ist jedoch $\frac{\partial f}{\partial x}(P_2) = -2\sqrt{2} \neq 0$, und daher können wir analog zum Fall von P_1 oben die Gleichung lokal um P_2 zumindest nach x auflösen, also x als Funktion von y schreiben – auch das wussten wir bereits vorher, die auflösende Gleichung ist hier $x = -\sqrt{2 - y^2}$.

(b) Wie in Beispiel 27.1 betrachten wir noch einmal die Gleichung $f(x,y) := x^y - y^x = 0$ für $x, y \in \mathbb{R}_{>0}$. Wir hatten dort bereits vermutet, dass sich diese Gleichung lokal um den Punkt $(2,4)$ nach y auflösen lässt, also dass sich die Lösungsmenge als Graph einer (stetig differenzierbaren) Funktion $\varphi\colon x \mapsto y = \varphi(x)$ schreiben lässt. Dies können wir nun mit Satz 27.9 bestätigen, denn es ist

$$\frac{\partial f}{\partial y} = x^y \log x - x y^{x-1} \quad \text{und damit} \quad \frac{\partial f}{\partial y}(2,4) = 16 \log 2 - 8 \neq 0.$$

Auch wenn wir die Funktion φ nicht explizit hinschreiben können, können wir mit Folgerung 27.10 noch ihre Ableitung an dieser Stelle berechnen: Wegen

$$\frac{\partial f}{\partial x} = y x^{y-1} - y^x \log y \quad \text{und damit} \quad \frac{\partial f}{\partial x}(2,4) = 32 - 16 \log 4$$

ist die Steigung des Graphen am Punkt $(2,4)$

$$\varphi'(2) = -\left(\frac{\partial f}{\partial y}(2,4)\right)^{-1} \cdot \frac{\partial f}{\partial x}(2,4) = -\frac{32-16\log 4}{16\log 2 - 8} = -3,177\ldots\ .$$

Darüber hinaus haben wir am Bild in Beispiel 27.1 gesehen, dass die durch die Gleichung $f(x,y)=0$ gegebene Punktmenge aus zwei Zweigen besteht, die sich in einem Punkt schneiden. Wiederum mit Hilfe von Satz 27.9 können wir nun auch die exakte Position des Kreuzungspunktes dieser beiden Zweige berechnen: Dort kann die Gleichung $f(x,y)=0$ ja nicht lokal nach y auflösbar sein, und daher muss an diesem Punkt $\frac{\partial f}{\partial y}=0$ sein. Da für diesen Punkt außerdem offensichtlich $y=x$ gilt, ergibt sich mit der obigen Formel für die Ableitung von f am Kreuzungspunkt also

$$x x^{x-1} - x^x \log x = 0, \quad \text{d. h.} \quad x^x(1-\log x) = 0$$

und damit $x = \mathrm{e}$. Der Kreuzungspunkt liegt also bei (e,e).

(c) (Nullstellen von Polynomen) Für ein gegebenes $n \in \mathbb{N}$ sei

$$F\colon \mathbb{R}^{n+1} \times \mathbb{R} \to \mathbb{R},\ (u,x) \mapsto f_u(x) := u_n x^n + \cdots + u_1 x + u_0,$$

wobei $u_0,\ldots,u_n$ die Koordinaten von $u \in \mathbb{R}^{n+1}$ sind. Wir können F also als Polynomfunktion in x auffassen, wobei wir aber auch die Koeffizienten u des Polynoms als variabel ansehen.

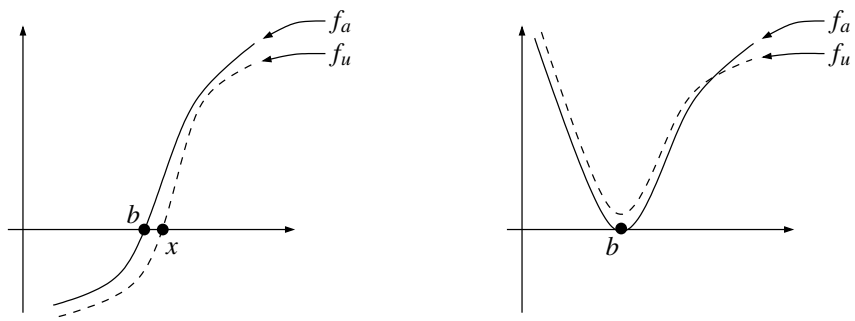
Natürlich gilt dann

$$\frac{\partial F}{\partial x} = n u_n x^{n-1} + \cdots + u_1 = f_u'(x).$$

Ist also $(a,b) \in \mathbb{R}^{n+1} \times \mathbb{R}$ mit

$$F(a,b) = f_a(b) = 0 \quad \text{und} \quad \frac{\partial F}{\partial x}(a,b) = f_a'(b) \neq 0$$

gegeben, d. h. ist b eine *einfache* Nullstelle von f_a, so besagt Satz 27.9 gerade, dass es in einer Umgebung von (a,b) eine stetig differenzierbare Funktion $\varphi\colon u \mapsto x = \varphi(u)$ gibt, so dass x dort genau dann eine Nullstelle des Polynoms f_u ist, also $F(u,x) = f_u(x) = 0$ gilt, wenn $x = \varphi(u)$ gilt. Wir können φ für die betrachtete Polynomgleichung also lokal als eine Lösungsformel ansehen, die den Koeffizienten u eine Nullstelle von f_u zuordnet. Da diese Funktion stetig differenzierbar ist, sagt man auch: *Die Nullstellen eines Polynoms hängen stetig differenzierbar von seinen Koeffizienten ab.* Dies ist im Bild unten links dargestellt: Eine kleine Änderung im Polynom von f_a nach f_u hat auch nur eine kleine Änderung der Nullstelle von b nach x zur Folge.



b ist einfache Nullstelle von f_a
Nullstelle x variiert stetig differenzierbar mit u

b ist Nullstelle höherer Ordnung von f_a
keine Nullstelle von f_u in der Nähe von b

Ist hingegen $\frac{\partial F}{\partial x}(a,b) = f_a'(b) = 0$, also b eine Nullstelle höherer Ordnung von f_a, so ist der Satz über implizite Funktionen nicht wie oben anwendbar. In der Tat zeigt hier auch das Bild oben rechts, dass es in einer kleinen Umgebung von a, also für ein Polynom f_u mit u in der Nähe von a, in der Regel keine (eindeutige) Nullstelle von f_u in der Nähe von b gibt.

Aufgabe 27.13. Zeige, dass das Gleichungssystem

$$\mathrm{e}^{xz} - x^2 + y^2 = xy^3 + x^2z + yz^2 = 1$$

in einer Umgebung des Punktes $x_0 = 1$, $y_0 = 1$, $z_0 = 0$ nach y und z aufgelöst, also als $\begin{pmatrix} y \\ z \end{pmatrix} = \varphi(x)$ für eine stetig differenzierbare Funktion φ geschrieben werden kann. Berechne auch die Ableitung $\varphi'(1)$ an diesem Punkt!

Aufgabe 27.14. Die *Lemniskate* L ist definiert als die Menge aller Punkte in $\mathbb{R}^2$, für die das Produkt der (euklidischen) Abstände zu den beiden Punkten $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ gleich 1 ist. Zeige, dass L in einer Umgebung des Punktes $\begin{pmatrix} \frac{2}{3} \\ \frac{1}{3}\sqrt{2} \end{pmatrix} \in L$ als Graph einer differenzierbaren Funktion $x_2 = \varphi(x_1)$ geschrieben werden kann.

Aufgabe 27.15. Berechne alle isolierten Punkte der Menge $M = \{(x,y) \in \mathbb{R}^2 : x^2 + x^3 + y^2 + y^3 = 0\}$.

Aufgabe 27.16. Für $a \in \mathbb{R}$ betrachten wir die Funktion

$$f_a \colon \mathbb{R}^2 \to \mathbb{R},\ f_a(x,y) = x^2 + y^2 + a\,\mathrm{e}^{x+2y}.$$

Für $a = 0$ hat diese Funktion offensichtlich ein (lokales und globales) Minimum im Nullpunkt.

(a) Zeige, dass es eine auf einer Umgebung $U \subset \mathbb{R}$ von 0 definierte und stetig differenzierbare Funktion $\varphi \colon U \to \mathbb{R}^2$ gibt, so dass für alle $a \in U$ die Funktion f_a ein lokales Minimum in $\varphi(a)$ hat.

(b) Berechne die Ableitung $\varphi'(0)$.

70

Wir haben nun also gesehen, dass man ein gegebenes Gleichungssystem $f(x) = 0$ mit einer stetig differenzierbaren Funktion $f \colon D \to \mathbb{R}^n$ auf $D \subset \mathbb{R}^m$ mit $m \geq n$ unter Umständen lokal nach n der Variablen $x_1, \ldots, x_m$ auflösen kann. Ob und nach welchen Variablen dies möglich ist, hängt dabei natürlich in der Regel wie in Beispiel 27.12 (a) nicht nur von f, sondern auch vom betrachteten Punkt ab. Ist eine solche Auflösung in *jedem* Punkt möglich – wenn auch je nach Punkt nach unterschiedlichen Variablen – so hat dies für die Lösungsmenge der Gleichung $f(x) = 0$ eine besondere geometrische Bedeutung, die wir jetzt kurz untersuchen wollen.

Definition 27.17 (Untermannigfaltigkeiten)**.** Es seien $r, n \in \mathbb{N}$ und $m := r + n$. Eine Teilmenge M von $\mathbb{R}^m$ heißt **r-dimensionale Untermannigfaltigkeit** von $\mathbb{R}^m$, wenn es zu jedem $c \in M$ eine offene Umgebung U und eine stetig differenzierbare Funktion $f \colon U \to \mathbb{R}^n$ gibt mit $\operatorname{rk} f'(c) = n$, so dass $M \cap U = \{x \in U : f(x) = 0\}$.

Beispiel 27.18. In Definition 27.17 darf die Funktion f von dem betrachteten Punkt $c \in M$ abhängen. Oft ist dies in der Praxis jedoch gar nicht der Fall, und es gibt eine (globale) stetig differenzierbare Funktion $f \colon \mathbb{R}^m \to \mathbb{R}^n$, so dass $M = \{x \in \mathbb{R}^m : f(x) = 0\}$ deren Nullstellenmenge ist und $\operatorname{rk} f'(c) = n$ an jedem Punkt $c \in M$ gilt.

So ist z. B. die Kreislinie $M = \{(x,y) : f(x,y) = 0\} \subset \mathbb{R}^2$ mit $f(x,y) = x^2 + y^2 - 2$ aus Beispiel 27.12 (a) eine 1-dimensionale Untermannigfaltigkeit von $\mathbb{R}^2$, denn die Ableitungsmatrix $f' = (2x \;\; 2y)$ hat an jedem Punkt $(x,y) \in M$ Rang 1 (beachte, dass der einzige Punkt $(x,y) = (0,0)$, an dem $\operatorname{rk} f' = 0$ gilt, nicht zu M gehört).

Lemma 27.19 (Untermannigfaltigkeiten als Graphen)**.** *Es seien $r, n \in \mathbb{N}$ und $m := r + n$. Für eine Teilmenge $M \subset \mathbb{R}^m$ sind äquivalent:*

(a) *M ist eine r-dimensionale Untermannigfaltigkeit.*

(b) *Zu jedem $c \in M$ gibt es*

- *eine Aufteilung der m Koordinaten von $\mathbb{R}^m$ in r Koordinaten $u_1, \ldots, u_r$ und n Koordinaten $x_1, \ldots, x_n$, mit entsprechender Aufteilung $c = (a,b)$,*
- *offene Umgebungen U von a in $\mathbb{R}^r$ und V von b in $\mathbb{R}^n$,*

- *eine stetig differenzierbare Funktion* $\varphi\colon U \to V$ *mit* $M \cap (U \times V) = \{(u, \varphi(u)) : u \in U\}$ *und* $\varphi(a) = b$.

(Mit anderen Worten lässt sich M also lokal als Graph einer stetig differenzierbaren Funktion in r Variablen schreiben, wobei um jeden Punkt von M unterschiedlich sein darf, welche der m Variablen als Variablen des Startraums und welche als Variable des Zielraums aufgefasst werden.)

Beweis.

(a) $\Rightarrow$ (b): Ist c ein Punkt einer r-dimensionalen Untermannigfaltigkeit M von $\mathbb{R}^m$, so ist M in einer offenen Umgebung W von c die Menge aller Nullstellen einer stetig differenzierbaren Funktion $f\colon W \to \mathbb{R}^n$ mit $\operatorname{rk} f'(c) = n$. Nach Aufgabe 16.46 gibt es in der Ableitungsmatrix $f'(c) \in \operatorname{Mat}(n \times m, \mathbb{R})$ also eine invertierbare $n \times n$-Teilmatrix. Bezeichnen wir die Koordinaten, die diesen Spalten entsprechen, mit $x_1, \dots, x_n$, und die restlichen mit $u_1, \dots, u_r$, so ergibt sich die Behauptung (b) direkt aus dem Satz 27.9 über implizite Funktionen.

(b) $\Rightarrow$ (a): Es sei $c \in M$ beliebig. Nach Voraussetzung gibt es dann bei geeigneter Aufteilung $u_1, \dots, u_r, x_1, \dots, x_n$ der Koordinaten von $\mathbb{R}^m$ eine offene Umgebung $U \times V$ von c, auf der M gegeben ist durch $x = \varphi(u)$ für eine stetig differenzierbare Funktion $\varphi\colon U \to V$, also durch die Gleichung $f(u,x) = 0$ mit

$$f\colon U \times V \to \mathbb{R}^n,\ (u,x) \mapsto x - \varphi(u).$$

Wegen $\frac{\partial f}{\partial x}(c) = E_n$ ist $\operatorname{rk} f'(c) = n$. Damit ist M nach Definition 27.17 eine r-dimensionale Untermannigfaltigkeit von $\mathbb{R}^m$. □

Bemerkung 27.20 (Eine Untermannigfaltigkeit ist lokal ein gekrümmter $\mathbb{R}^r$)**.** Nach Lemma 27.19 sind die r-dimensionalen Untermannigfaltigkeiten genau die Teilmengen von $\mathbb{R}^m$, die sich lokal als Graph einer stetig differenzierbaren Funktion in r der Variablen von $\mathbb{R}^m$ schreiben lassen. Wie im Eindimensionalen bedeutet dies anschaulich, dass M als Graph solcher stetig differenzierbaren Funktionen „keine Knicke haben kann". Anschaulich können wir also sagen: *Eine r-dimensionale Untermannigfaltigkeit ist ein Raum, der lokal wie ein gekrümmter (aber nicht geknickter)* $\mathbb{R}^r$ *aussieht* – so wie z. B. die Kreislinie in Beispiel 27.12 (a) bzw. 27.18. Dementsprechend zeigen die ersten beiden Bilder unten 1-dimensionale Untermannigfaltigkeiten von $\mathbb{R}^2$, die anderen beiden wegen der grau markierten Kreuzungs- bzw. Knickpunkte jedoch nicht.

1-dimensionale Untermannigfaltigkeiten keine Untermannigfaltigkeiten

Aufgabe 27.21. Es sei $M = \{x \in \mathbb{R}^2 : x_1^3 = x_2^2\}$. Zeige, dass $M \backslash \{0\}$ eine 1-dimensionale Untermannigfaltigkeit von $\mathbb{R}^2$ ist, M jedoch nicht.

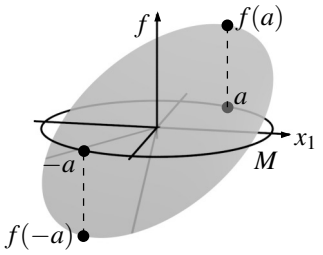
Skizziere auch die Menge M. Könnt ihr das Ergebnis der Aufgabe auch in dieser Skizze erkennen?

27.C Extrema unter Nebenbedingungen

Eine Anwendung von Untermannigfaltigkeiten, die wir zum Abschluss dieses Kapitels noch betrachten wollen, ist das bereits in Bemerkung 26.4 (b) beschriebene Problem der *Randextrema* bzw. der Extrema unter Nebenbedingungen. Haben wir eine differenzierbare Funktion $f\colon D \to \mathbb{R}$ auf einer offenen Teilmenge $D \subset \mathbb{R}^n$ und eine Untermannigfaltigkeit $M \subset D$, so geht es dabei wie im folgenden Beispiel darum, die (lokalen) Extrema der eingeschränkten Funktion $f|_M$ zu bestimmen.

Beispiel 27.22 (Randextrema durch Einsetzen der Nebenbedingungen)**.** Wir wollen die Extrema der unten rechts dargestellten Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1 + 2x_2$ auf der abgeschlossenen Kreisscheibe $K_1(0) = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1\}$ finden.

Da $K_1(0)$ als beschränkte und abgeschlossene Teilmenge von $\mathbb{R}^2$ nach Satz 23.51 (b) kompakt ist, nimmt f dort nach Folgerung 24.34 tatsächlich ein Maximum und Minimum an. Allerdings besitzt f als lineare Funktion keine kritischen Punkte, denn die Ableitung $f' = (1 \;\; 2)$ ist nirgends gleich Null – und damit kann f nach Lemma 26.2 keine Extrema im Inneren von $K_1(0)$ haben. Maximum und Minimum von f auf $K_1(0)$ werden also notwendigerweise auf dem Rand $M := \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$ angenommen.



Im Eindimensionalen, also wenn die betrachtete Definitionsmenge ein Intervall ist, besteht der Rand dieses Intervalls ja höchstens aus zwei Punkten, und daher konnten wir derartige Randextrema dann einfach durch Einsetzen aller Randpunkte bestimmen. In unserem momentanen Fall ist dies jedoch nicht mehr möglich, da der Rand M von $K_1(0)$ unendlich viele Punkte hat. Stattdessen müssen wir nun die Extrema der eingeschränkten Funktion $f|_M$, also die Extrema von f unter der Nebenbedingung $x_1^2 + x_2^2 = 1$ suchen.

Die einfachste Möglichkeit dafür ist hier vermutlich, die Nebenbedingung nach einer der Variablen aufzulösen und in die Zielfunktion f einzusetzen: Nehmen wir zunächst $x_2 > 0$ an, so ist M dort gegeben durch die aufgelöste Gleichung $x_2 = \sqrt{1 - x_1^2}$. Einsetzen in f zeigt also, dass wir – nun ohne Nebenbedingungen – eine Stelle $x_1 \in (-1, 1)$ suchen, an der die Funktion

$$h(x_1) := f\left(x_1, \sqrt{1 - x_1^2}\right) = x_1 + 2\sqrt{1 - x_1^2}$$

ein Extremum besitzt. Dieses können wir nun leicht mit Hilfe unserer eindimensionalen Techniken finden: Es ist

$$h'(x_1) = 1 - \frac{2x_1}{\sqrt{1 - x_1^2}},$$

und die (einzige) Nullstelle dieser Funktion berechnet man schnell zu $a_1 = \frac{1}{\sqrt{5}}$, woraus durch Einsetzen in die Nebenbedingung $a_2 = \frac{2}{\sqrt{5}}$ folgt. Der entsprechende Punkt a ist im Bild rechts oben eingezeichnet, und man sieht leicht (z. B. mit Satz 11.19 durch Untersuchung der zweiten Ableitung von h in diesem Punkt), dass hier ein lokales Maximum vorliegt. Untersuchen wir zusätzlich auch noch die Randpunkte von h (also $x_1 = \pm 1$) sowie die Punkte mit $x_2 < 0$, so sehen wir, dass dies sogar das globale Maximum von f auf M ist. Entsprechend findet man das Minimum von f am gegenüberliegenden Punkt $-a$.

Diese Strategie, die Nebenbedingung nach einer der Variablen aufzulösen und in die Zielfunktion einzusetzen, führt aber zum einen wie oben schnell zu Fallunterscheidungen, und funktioniert zum anderen natürlich auch nur, wenn man diese Auflösungsfunktion explizit berechnen kann und das Einsetzen dieser Funktion nicht zu einer zu komplizierten Zielfunktion führt. Andernfalls muss man stattdessen mit dem Satz über implizite Funktionen arbeiten und erhält so das folgende Resultat.

Satz 27.23 (Extrema unter Nebenbedingungen). *Es seien $r, n \in \mathbb{N}$, $m := r + n$ und $f : D \to \mathbb{R}$ eine differenzierbare Funktion auf einer offenen Teilmenge $D \subset \mathbb{R}^m$. Ferner sei $g : D \to \mathbb{R}^n$ eine stetig differenzierbare Funktion mit $\operatorname{rk} g'(c) = n$ für alle $c \in M := \{x \in D : g(x) = 0\}$, so dass M also eine r-dimensionale Untermannigfaltigkeit von $\mathbb{R}^m$ ist.*

Hat dann die Einschränkung $f|_M$ ein lokales Extremum in einem Punkt $c \in M$, so gibt es ein $\lambda \in \mathbb{R}^n$ mit

$$f'(c) = \lambda^\mathsf{T} \cdot g'(c),$$

also mit $f'(c) = \lambda_1 g_1'(c) + \cdots + \lambda_n g_n'(c)$, wobei $\lambda_1, \ldots, \lambda_n$ und $g_1, \ldots, g_n$ die Komponenten von λ bzw. g sind.

*Man bezeichnet die Koeffizienten $\lambda^\mathsf{T} = (\lambda_1, \ldots, \lambda_n)$ als **Lagrange-Multiplikatoren** und spricht bei der Anwendung dieses Satzes daher auch vom Verfahren der Lagrange-Multiplikatoren.*

Beweis. Da M nach Voraussetzung lokal um c eine n-dimensionale Untermannigfaltigkeit ist, können wir die Koordinaten von $\mathbb{R}^m$ nach Lemma 27.19 so als $u_1,\dots,u_r$ und $x_1,\dots,x_n$ wählen, dass M dort durch eine aufgelöste Gleichung der Form $x=\varphi(u)$ gegeben ist. Für diese Auflösungsfunktion φ gilt ferner nach Folgerung 27.10

$$\varphi' = -\left(\frac{\partial g}{\partial x}\right)^{-1}\cdot\frac{\partial g}{\partial u}. \tag{1}$$

Wie in Beispiel 27.22 müssen wir nun also nach Extrema ohne Nebenbedingungen der Funktion $h(u):=f(u,\varphi(u))$ suchen. An einem solchen Extremum gilt nach Lemma 26.2 mit der Kettenregel aus Satz 25.29

$$0=h'=\left(\frac{\partial f}{\partial u}\,\middle|\,\frac{\partial f}{\partial x}\right)\cdot\begin{pmatrix}E\\ \varphi'\end{pmatrix}=\frac{\partial f}{\partial u}+\frac{\partial f}{\partial x}\cdot\varphi' \overset{(1)}{=} \frac{\partial f}{\partial u}-\frac{\partial f}{\partial x}\cdot\left(\frac{\partial g}{\partial x}\right)^{-1}\cdot\frac{\partial g}{\partial u}. \tag{2}$$

Setzen wir nun

$$\lambda^\mathsf{T} := \frac{\partial f}{\partial x}(c)\cdot\left(\frac{\partial g}{\partial x}(c)\right)^{-1} \qquad \in \operatorname{Mat}(1\times n,\mathbb{R}), \tag{3}$$

so ergibt sich am Punkt c wie behauptet

$$\begin{aligned}
f' &= \left(\frac{\partial f}{\partial u}\,\middle|\,\frac{\partial f}{\partial x}\right)\\
&\overset{(2)}{=} \left(\frac{\partial f}{\partial x}\cdot\left(\frac{\partial g}{\partial x}\right)^{-1}\cdot\frac{\partial g}{\partial u}\,\middle|\,\frac{\partial f}{\partial x}\cdot\left(\frac{\partial g}{\partial x}\right)^{-1}\cdot\frac{\partial g}{\partial x}\right)\\
&\overset{(3)}{=} \lambda^\mathsf{T}\cdot\left(\frac{\partial g}{\partial u}\,\middle|\,\frac{\partial g}{\partial x}\right)\\
&= \lambda^\mathsf{T}\cdot g'.
\end{aligned}$$

□

Bemerkung 27.24. Die Lagrange-Multiplikatoren $\lambda^\mathsf{T}=(\lambda_1,\dots,\lambda_n)$ in Satz 27.23 sind in der Regel zunächst unbekannt und somit n weitere Variablen, die zusätzlich zu den Koordinaten in $\mathbb{R}^m$ an einem Extremum berechnet werden müssen. Insgesamt sind also $n+m$ Variablen zu bestimmen. Da wir zusätzlich zu den m Gleichungen für $f'(c)$ aber auch noch die n Nebenbedingungsgleichungen $g=0$ haben, kann man dieses (in der Regel nichtlineare) Gleichungssystem oftmals lösen. Das folgende Beispiel verdeutlicht dies.

Beispiel 27.25 (Randextrema durch Lagrange-Multiplikatoren)**.** Wir wollen noch einmal die Extremwerte der Funktion $f(x)=x_1+2x_2$ aus Beispiel 27.22 auf der abgeschlossenen Kreisscheibe $K_1(0)=\{x\in\mathbb{R}^2 : x_1^2+x_2^2\le 1\}$ finden – diesmal aber mit dem Verfahren der Lagrange-Multiplikatoren aus Satz 27.23.

Wir hatten schon gesehen, dass wir dazu nur noch die Extrema von f auf dem Rand der Kreisscheibe

$$M=\{x\in\mathbb{R}^2 : g(x)=0\} \quad \text{mit} \quad g(x):=x_1^2+x_2^2-1$$

finden müssen. Wegen $g'(x)=(2x_1\ \ 2x_2)\neq 0$, also $\operatorname{rk} g'(x)=1$ für alle $x\in M$ ist Satz 27.23 anwendbar und liefert uns, dass an jedem solchen Extremum die Gleichung $f'=\lambda\, g'$ für ein $\lambda\in\mathbb{R}$ gelten muss. Mit $f'=(1\ \ 2)$ bedeutet dies

$$1=\lambda\cdot 2x_1 \quad \text{und} \quad 2=\lambda\cdot 2x_2 \qquad \text{zusammen mit der Nebenbedingung} \quad x_1^2+x_2^2-1=0.$$

Dieses Gleichungssystem mit 3 Variablen und Gleichungen können wir nun leicht lösen: Lösen wir die ersten beiden Gleichungen nach x_1 bzw. x_2 auf und setzen dies in die Nebenbedingung ein, so erhalten wir

$$\frac{1}{4\lambda^2}+\frac{1}{\lambda^2}=1 \quad \text{und damit} \quad \lambda=\pm\frac{\sqrt{5}}{2},$$

woraus sich mit den obigen Gleichungen sofort $x_1=\pm\frac{1}{\sqrt{5}}$ und $x_2=\pm\frac{2}{\sqrt{5}}$ ergibt – also das gleiche Ergebnis, das wir auch in Beispiel 27.22 herausbekommen hatten.

Beispiel 27.26 (Hauptachsentransformation in der Analysis)**.** Es sei $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ eine symmetrische, positiv definite Matrix. Wir betrachten die Menge

$$M = \{x \in \mathbb{R}^n : g(x) = 0\} \quad \text{mit} \quad g(x) = x^\mathsf{T} A x - 1,$$

die nach der Hauptachsentransformation aus Beispiel 22.38 der Rand eines (in der Regel gedrehten) Ellipsoids ist. Wir wollen auf M die Punkte mit maximalem und minimalem euklidischen Abstand zum Nullpunkt bestimmen – also die Punkte, an denen die Funktion $f(x) = \|x\|_2^2 = x^\mathsf{T} x$ wie im Bild unten links ihre Extrema annimmt.

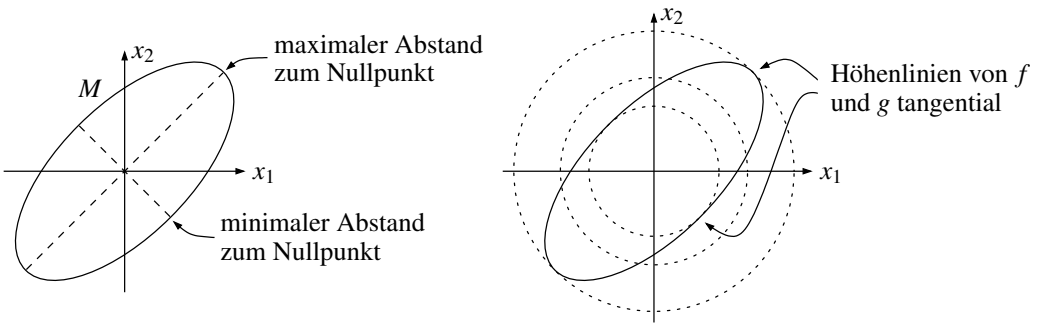
Dazu berechnen wir zunächst wieder die Ableitungen: Nach Aufgabe 25.22 (a) ist

$$g'(x) = 2x^\mathsf{T} A \quad \text{und} \quad f'(x) = 2x^\mathsf{T}.$$

Beachte, dass $g'(x) \neq 0$ und damit auch $\operatorname{rk} g'(x) = 1$ für alle $x \in M$ gilt: Andernfalls wäre nämlich $0 = g'(x) \cdot x = 2x^\mathsf{T} A x$, woraus aus der positiven Definitheit von A sofort $x = 0$ und damit $x \notin M$ folgt. Also ist die Rangbedingung an g' aus Satz 27.23 erfüllt. Damit können Extremstellen von f auf M nach diesem Satz nur dort vorliegen, wo es ein $\lambda \in \mathbb{R}$ gibt mit $f'(x) = \lambda\, g'(x)$. Durch Einsetzen der Ableitungen und Transponieren folgt daraus die Gleichung

$$2x = \lambda \cdot 2Ax \quad \text{und damit} \quad Ax = \frac{1}{\lambda} x.$$

Der maximale und minimale Abstand von M zum Nullpunkt tritt damit nur an Eigenvektoren auf – was mit dem Ergebnis von Beispiel 22.38 übereinstimmt. Die Lagrange-Multiplikatoren entsprechen hier also (bis auf Kehrwertbildung) genau dem größten und kleinsten Eigenwert von A.



Zusätzlich können wir an diesem Beispiel auch gut eine geometrische Interpretation der Extremwertbedingung $f' = \lambda\, g'$ aus Satz 27.23 angeben: Nach Transponieren besagt diese Gleichung ja gerade $\operatorname{grad} f = \lambda \operatorname{grad} g$, d. h. die Gradienten sind parallel zueinander. Da die Gradienten nach Beispiel 25.31 (a) aber senkrecht auf den Höhenlinien stehen, ist dies äquivalent dazu, dass die Höhenlinien von f und g in diesen Punkten tangential zueinander sind – wie man auch gut im Bild oben rechts erkennen kann, in dem die Höhenlinien von f bzw. g als gestrichelte bzw. durchgezogene Linien dargestellt sind.

Aufgabe 27.27. Es sei noch einmal $M = \{x \in \mathbb{R}^2 : x_1^3 = x_2^2\}$ wie in Aufgabe 27.21. Berechne alle lokalen Extrema der Funktion $f\colon M \to \mathbb{R}$, $x \mapsto x_1 + x_2$.

Aufgabe 27.28. Finde mit Hilfe von Lagrange-Multiplikatoren eine Matrix $A \in \mathrm{Mat}(2 \times 2, \mathbb{R})$ mit $\det A = 0$, deren Abstand zur Matrix $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ in der euklidischen Norm minimal ist.

71

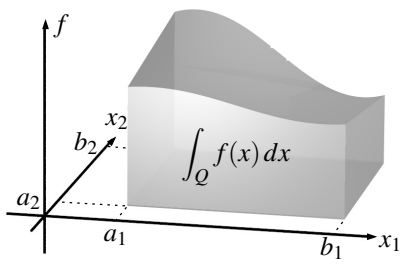
28. Integralrechnung im Mehrdimensionalen

Wir haben nun das Studium der Differenzierbarkeit in mehreren Veränderlichen beendet und werden uns als Nächstes wie im Eindimensionalen der Integration zuwenden.

Zum Einstieg wollen wir uns dabei anschauen, was wir von einer solchen Verallgemeinerung des Integralbegriffs ins Mehrdimensionale überhaupt erwarten würden. Dazu erinnern wir uns daran, dass wir im eindimensionalen Fall in Kapitel 12 zwei Interpretationsmöglichkeiten für das Integral hatten:

- (Flächenberechnung) Wir konnten mit Hilfe des Integrals die Fläche unter dem Graphen einer Funktion in einer reellen Variablen berechnen.

  Die direkte Verallgemeinerung ins Mehrdimensionale besteht hier offensichtlich darin, eine reellwertige Funktion in n Variablen zu betrachten und wie im Bild rechts nach dem $(n+1)$-dimensionalen Volumen des Gebiets unter dem Graphen von f zu fragen, z. B. über einem Quader $Q = [a_1,b_1] \times \cdots \times [a_n,b_n]$. In der Tat ist die exakte Formulierung und Lösung dieses Problems der Inhalt dieses Kapitels. Wir werden das Ergebnis als das n-dimensionale Integral von f über Q bezeichnen und als $\int_Q f(x)\,dx$ schreiben.



- (Umkehrung der Differentiation) Im Eindimensionalen ließ sich das Problem der Flächenberechnung unter dem Graphen einer Funktion f nach dem Hauptsatz der Differential- und Integralrechnung 12.21 auf die Suche nach einer Stammfunktion zurückführen, also einer Funktion F mit $F' = f$. Im Mehrdimensionalen besteht dieser direkte Zusammenhang jedoch nicht mehr: Auch hier kann man sich zwar fragen, ob man z. B. zu einer gegebenen Funktion $f\colon \mathbb{R}^n \to \mathrm{Mat}(1\times n,\mathbb{R})$ eine differenzierbare Funktion $F\colon \mathbb{R}^n \to \mathbb{R}$ mit $F' = f$ finden kann – aber allein schon daran, dass wir hierfür eine Funktion f mit Zielraum $\mathrm{Mat}(1\times n,\mathbb{R})$ benötigen, da die Ableitung von F ja eine derartige Matrix ist, kann man erkennen, dass dies nicht dieselbe Fragestellung wie die der Volumenberechnung unter dem Graphen von f sein kann.

  Dieses Problem der Umkehrung der Differentiation im Mehrdimensionalen erfordert vielmehr andere Methoden, die ihr in der Vorlesung „Vektoranalysis“ des zweiten Studienjahres kennenlernen könnt. Die folgende Aufgabe soll dabei nur einen kleinen Vorgeschmack auf die dort zu erwartenden Resultate geben: Sie zeigt, dass die Existenz einer Funktion F mit $F' = f$ im Gegensatz zum eindimensionalen Fall auch für stetiges f keinesfalls automatisch ist und darüber hinaus auch noch von der Form der betrachteten Definitionsmenge abhängt.

Aufgabe 28.1 (Umkehrung der Differentiation in $\mathbb{R}^2$)**.** Man zeige:

(a) Sind $f_1, f_2\colon \mathbb{R}^2 \to \mathbb{R}$ zwei stetig differenzierbare Funktionen, so gibt es genau dann eine zweimal stetig differenzierbare Funktion $F\colon \mathbb{R}^2 \to \mathbb{R}$ mit $F' = (\,f_1 \mid f_2\,)$, wenn $\partial_1 f_2 = \partial_2 f_1$.

(b) Im Fall der Definitionsmenge $D = \mathbb{R}^2 \backslash \{0\}$ hingegen ist diese Aussage falsch: Die Funktionen

$$f_1\colon D \to \mathbb{R},\ x \mapsto \frac{-x_2}{x_1^2+x_2^2} \quad \text{und} \quad f_2\colon D \to \mathbb{R},\ x \mapsto \frac{x_1}{x_1^2+x_2^2}$$

erfüllen zwar $\partial_1 f_2 = \partial_2 f_1$, aber es gibt keine zweimal stetig differenzierbare Funktion $F\colon D \to \mathbb{R}$ mit $F' = (\,f_1 \mid f_2\,)$.

(Hinweis: Eine Möglichkeit besteht darin, die Formel $F(x) - F(a) = \int_0^1 F'(\gamma(t)) \cdot \gamma'(t)\,dt$ für einen stetig differenzierbaren Weg $\gamma\colon [0,1] \to \mathbb{R}^2$ von a nach x zu zeigen und zu benutzen.)

Bemerkung 28.2 (Riemann- und Lebesgue-Integral)**.** Um die obige Frage nach dem Volumen unter dem Graphen einer Funktion in mehreren Variablen formal exakt zu untersuchen, gibt es verschiedene Herangehensweisen. Wir werden in dieser Vorlesung die sogenannte *Riemannsche Integrationstheorie* behandeln, die das Integral völlig analog zum eindimensionalen Fall in Kapitel 12 über Unter- und Obersummen zu immer feiner werdenden Zerlegungen des Integrationsbereichs definiert. Der Hauptvorteil dieser Theorie ist, dass sie relativ einfach ist und für viele Funktionen, z. B. für stetige Funktionen auf $\mathbb{R}^n$, schnell zum gewünschten Ergebnis führt. Für Anwendungen z. B. in der Stochastik ist es allerdings unabdingbar, den Integralbegriff deutlich zu erweitern. Analog zur Differentiation, die man nach Bemerkung 25.6 nicht nur in $\mathbb{R}^n$, sondern sogar in allgemeinen normierten Räumen einführen kann, kann man auch Integrale nicht nur von Funktionen auf $\mathbb{R}^n$, sondern auch von solchen auf allgemeineren Räumen, den sogenannten *Maßräumen*, definieren. Diese allgemeinere Integrationstheorie, das sogenannte *Lebesgue-Integral*, ist vom theoretischen Standpunkt her zwar deutlich schöner, allerdings auch deutlich komplizierter als die Riemannsche Theorie, und wird daher erst in der für das spätere Studium angebotenen Vorlesung „Maß- und Integrationstheorie“ ausführlich behandelt.

28.A Das mehrdimensionale Riemann-Integral

Wir beginnen damit, analog zu Kapitel 12 Unter- und Obersummen beschränkter Funktionen und damit schließlich das Integral solcher Funktionen in mehreren Variablen einzuführen. Viele Definitionen und Resultate sind dabei bis auf kleine Modifikationen dieselben wie im eindimensionalen Fall. Beweise, die wörtlich genauso funktionieren wie in Kapitel 12, werden wir dabei natürlich nicht noch einmal hinschreiben.

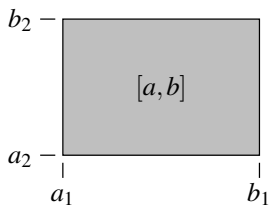
Der Einfachheit halber beschäftigen wir uns zunächst einmal nur mit Integralen von Funktionen, die auf einem Quader (also einem Produkt von Intervallen) definiert sind. Den Fall allgemeinerer Integrationsbereiche werden wir dann im nächsten Kapitel darauf zurückführen.

Notation 28.3 (Quader)**.** Es seien $n \in \mathbb{N}_{>0}$ und $a, b \in \mathbb{R}^n$ mit Koordinaten $a_1, \dots, a_n$ bzw. $b_1, \dots, b_n$, so dass $a_i \le b_i$ für alle $i = 1, \dots, n$. Dann heißt die Menge

$$[a,b] := [a_1, b_1] \times \cdots \times [a_n, b_n] \quad \subset \mathbb{R}^n$$

ein **Quader** in $\mathbb{R}^n$. Wir definieren das **Volumen** eines solchen Quaders als

$$\mathrm{vol}([a,b]) := (b_1 - a_1) \cdot \cdots \cdot (b_n - a_n).$$



Die folgenden beiden Definitionen sind völlig analog, und im eindimensionalen Fall $n = 1$ auch äquivalent zu Definition 12.1.

Definition 28.4 (Zerlegungen)**.** Es sei $[a,b] = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ ein Quader.

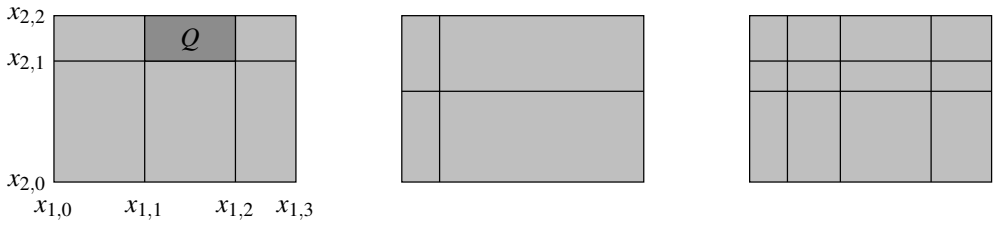
(a) Eine **Zerlegung** $Z = (Z_1, \dots, Z_n)$ des Quaders $[a,b]$ besteht aus Zerlegungen Z_i der Intervalle $[a_i, b_i]$ im Sinne von Definition 12.1 (a) für alle $i = 1, \dots, n$. Es ist also jedes $Z_i = \{x_{i,0}, \dots, x_{i,k_i}\}$ eine endliche Teilmenge von $[a_i, b_i]$ mit $a_i, b_i \in Z_i$, von der wir vereinbaren, dass in dieser Schreibweise immer $a_i = x_{i,0} < x_{i,1} < \cdots < x_{i,k_i} = b_i$ gilt. Ein **Teilquader** von Z ist wie im Bild unten links ein Quader Q der Form

$$[x_{1,j_1-1}, x_{1,j_1}] \times \cdots \times [x_{n,j_n-1}, x_{n,j_n}] \quad \subset [a,b]$$

für gewisse $j_1, \dots, j_n$ mit $1 \le j_i \le k_i$ für alle $i = 1, \dots, n$. Wir bezeichnen die Menge aller solcher Teilquader von Z mit $\mathrm{TQ}(Z)$.

(b) Sind $Z = (Z_1, \dots, Z_n)$ und $Z' = (Z'_1, \dots, Z'_n)$ zwei Zerlegungen von $[a,b]$, so nennen wir Z' eine **Verfeinerung** von Z und schreiben dies als $Z \subset Z'$, wenn $Z_i \subset Z'_i$ für alle $i = 1, \dots, n$ gilt, also jedes Z'_i im Sinne von Definition 12.1 (a) eine Verfeinerung von Z_i ist.

(c) Für zwei beliebige Zerlegungen $Z = (Z_1, \dots, Z_n)$ und $Z' = (Z'_1, \dots, Z'_n)$ von $[a,b]$ definieren wir wie im folgenden Bild dargestellt $Z \cup Z' := (Z_1 \cup Z'_1, \dots, Z_n \cup Z'_n)$; offensichtlich ist dies eine gemeinsame Verfeinerung von Z und Z'.



Eine Zerlegung Z mit einem Teilquader Q, ... eine weitere Zerlegung Z'... und die gemeinsame Verfeinerung $Z \cup Z'$

Bemerkung 28.5. Ist Z eine Zerlegung eines Quaders $[a,b]$, so sollte es anschaulich klar sein, dass die Summe der Volumina aller Teilquader in Z das Gesamtvolumen $\mathrm{vol}([a,b])$ ergibt. Wir können dies auch leicht explizit nachrechnen: Mit den Notationen aus Definition 28.4 ist

$$\begin{aligned}\sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) &= \sum_{j_1=1}^{k_1} \cdots \sum_{j_n=1}^{k_n} (x_{1,j_1} - x_{1,j_1-1}) \cdot \cdots \cdot (x_{n,j_n} - x_{n,j_n-1}) \\ &= \Bigg(\underbrace{\sum_{j_1=1}^{k_1} (x_{1,j_1} - x_{1,j_1-1})}_{b_1 - a_1} \Bigg) \cdot \cdots \cdot \Bigg(\underbrace{\sum_{j_n=1}^{k_n} (x_{n,j_n} - x_{n,j_n-1})}_{b_n - a_n} \Bigg) \\ &= \mathrm{vol}([a,b]).\end{aligned}$$

Definition 28.6 (Unter- und Obersummen)**.** Es sei $f\colon [a,b] \to \mathbb{R}$ eine beschränkte Funktion auf einem Quader $[a,b] \subset \mathbb{R}^n$. Für eine Zerlegung Z von $[a,b]$ nennen wir

$$\mathrm{US}(f,Z) := \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \inf f(Q) \qquad \text{die } \textbf{Untersumme} \text{ und}$$

$$\mathrm{OS}(f,Z) := \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \sup f(Q) \qquad \text{die } \textbf{Obersumme}$$

von f bezüglich Z.

Der Weg von den Unter- und Obersummen zum Integral ist derselbe wie im Eindimensionalen in Lemma 12.3, Definition 12.5, Lemma 12.6 und Definition 12.7: Wir wollen, dass sich Unter- und Obersummen dem gesuchten Integral annähern, also dass das Supremum aller Untersummen gleich dem Infimum aller Obersummen ist.

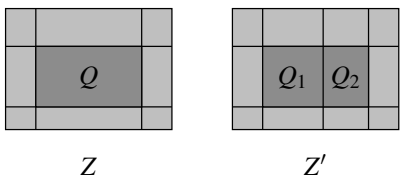
Lemma 28.7 (Eigenschaften von Unter- und Obersummen)**.** *Es seien $f\colon [a,b] \to \mathbb{R}$ eine beschränkte Funktion auf einem Quader $[a,b]$ und Z, Z' zwei Zerlegungen von $[a,b]$. Dann gilt:*

(a) *Ist Z' eine Verfeinerung von Z, so gilt $\mathrm{US}(f,Z') \geq \mathrm{US}(f,Z)$ und $\mathrm{OS}(f,Z') \leq \mathrm{OS}(f,Z)$.*

(b) $\mathrm{US}(f,Z) \leq \mathrm{OS}(f,Z')$.

Beweis. Der Beweis dieses Lemmas ist nahezu der gleiche wie der von Lemma 12.3.

(a) Da jede Verfeinerung von Z durch endliches Hinzufügen von weiteren Unterteilungspunkten in den Seitenintervallen von $[a,b]$ entsteht, genügt es, den Fall zu betrachten, dass Z' wie im Bild rechts aus Z durch Hinzufügen eines weiteren Zwischenpunktes in einer der Koordinaten entsteht.



Ist Q dann ein Teilquader von Z, der in Z' in zwei Teilquader Q_1 und Q_2 aufgeteilt wird, so gilt

$$\begin{aligned}\operatorname{vol}(Q_1)\cdot\inf f(Q_1)+\operatorname{vol}(Q_2)\cdot\inf f(Q_2) &\geq \operatorname{vol}(Q_1)\cdot\inf f(Q)+\operatorname{vol}(Q_2)\cdot\inf f(Q)\\ &= \operatorname{vol}(Q)\cdot\inf f(Q).\end{aligned}$$

Aufsummieren dieser Ungleichungen über alle Teilquader liefert damit die behauptete Ungleichung $\mathrm{US}(f,Z')\geq\mathrm{US}(f,Z)$ für die Untersummen. Die Aussage über die Obersummen beweist man natürlich genauso.

(b) Da $Z\cup Z'$ eine gemeinsame Verfeinerung von Z und Z' ist, folgt aus (a)

$$\mathrm{US}(f,Z)\leq\mathrm{US}(f,Z\cup Z')\leq\mathrm{OS}(f,Z\cup Z')\leq\mathrm{OS}(f,Z'),$$

wobei die mittlere Ungleichung gilt, weil das Infimum einer Menge immer kleiner oder gleich dem Supremum ist. □

Definition 28.8 (Integrierbarkeit)**.** Es sei $f\colon[a,b]\to\mathbb{R}$ eine beschränkte Funktion auf einem Quader $[a,b]\subset\mathbb{R}^n$. Dann heißt

$$\begin{aligned}\mathrm{UI}(f) &:= \sup\{\mathrm{US}(f,Z) : Z \text{ Zerlegung von } [a,b]\} && \text{das } \textbf{Unterintegral}\text{, und analog}\\ \mathrm{OI}(f) &:= \inf\{\mathrm{OS}(f,Z) : Z \text{ Zerlegung von } [a,b]\} && \text{das } \textbf{Oberintegral}\end{aligned}$$

von f (beachte, dass dieses Supremum bzw. Infimum existiert, da die Menge aller Untersummen nach Lemma 28.7 (b) durch jede Obersumme nach oben beschränkt, und die Menge aller Obersummen durch jede Untersumme nach unten beschränkt ist). Wie in Lemma 12.6 zeigt man, dass stets $\mathrm{UI}(f)\leq\mathrm{OI}(f)$ gilt. Ist sogar $\mathrm{UI}(f)=\mathrm{OI}(f)$, so nennen wir f **(Riemann-)integrierbar**, und definieren das **Integral** von f als diesen Wert

$$\int_{[a,b]} f(x)\,dx := \mathrm{UI}(f)=\mathrm{OI}(f).$$

Möchte man explizit schreiben, dass hier über mehrere Variablen integriert wird, sind für dieses Integral auch die Schreibweisen

$$\int_{[a,b]} f(x)\,d(x_1,\ldots,x_n) \qquad \text{und} \qquad \int_{[a,b]} f(x)\,d^n x$$

üblich. Im Gegensatz zum Eindimensionalen wird im mehrdimensionalen Fall aber immer der Integrationsbereich als Menge unten ans Integralzeichen geschrieben – eine Schreibweise $\int_a^b$ wie im Eindimensionalen werden wir nicht verwenden, zumal sie für allgemeinere Integrationsbereiche als Quader, wie wir sie im nächsten Kapitel betrachten werden, auch gar nicht möglich wäre.

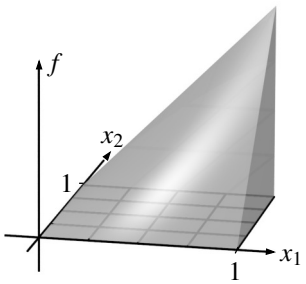
Beispiel 28.9. Im eindimensionalen Fall sind alle obigen Definitionen offensichtlich exakt dieselben wie in Kapitel 12, so dass sich in diesem Fall auch der gleiche Integralbegriff ergibt. Hier sind daher zwei Beispiele für die explizite Berechnung eines mehrdimensionalen Integrals mit Hilfe von Unter- und Obersummen.

(a) Es sei $f\colon[a,b]\to\mathbb{R},\ f(x)=c$ eine konstante Funktion auf einem n-dimensionalen Quader $[a,b]$. Dann sind alle Suprema und Infima auf von f gleich c, und demzufolge gilt sowohl für die Untersumme als auch für die Obersumme von f zu einer beliebigen Zerlegung Z:

$$\mathrm{US}(f,Z)=\mathrm{OS}(f,Z)=\sum_{Q\in\mathrm{TQ}(Z)}\operatorname{vol}(Q)\cdot c \overset{28.5}{=} c\cdot\operatorname{vol}([a,b]).$$

Damit sind also auch das Unter- und Oberintegral von f gleich $c\cdot\operatorname{vol}([a,b])$, d. h. f ist wie erwartet integrierbar mit $\int_{[a,b]} f(x)\,dx = c\cdot\operatorname{vol}([a,b])$.

(b) Es sei
$$f\colon [0,1]\times[0,1]\to\mathbb{R},\ x\mapsto x_1\cdot x_2$$
die rechts dargestellte Funktion. Für ein $n\in\mathbb{N}_{>0}$ wollen wir die (im Bild für $n=4$ eingezeichnete) Zerlegung
$$Z=\left(\left\{0,\frac{1}{n},\frac{2}{n},\dots,1\right\},\left\{0,\frac{1}{n},\frac{2}{n},\dots,1\right\}\right)$$
von $[0,1]\times[0,1]$ in n^2 Teilquadrate der Seitenlänge $\frac{1}{n}$ betrachten und analog zu Beispiel 12.2 die zugehörige Unter- und Obersumme von f berechnen.



Da f auf jedem solchen Teilquadrat $\left[\frac{i-1}{n},\frac{i}{n}\right]\times\left[\frac{j-1}{n},\frac{j}{n}\right]$ das Minimum am linken unteren Punkt $\left(\frac{i-1}{n},\frac{j-1}{n}\right)$ annimmt, berechnet sich die Untersumme zu
$$\begin{aligned}\mathrm{US}(f,Z)&=\sum_{i=1}^n\sum_{j=1}^n\frac{1}{n^2}\cdot f\left(\frac{i-1}{n},\frac{j-1}{n}\right)\\&=\frac{1}{n^4}\cdot\left(\sum_{i=1}^n(i-1)\right)\cdot\left(\sum_{j=1}^n(j-1)\right)\\&=\frac{1}{n^4}\cdot\frac{n(n-1)}{2}\cdot\frac{n(n-1)}{2}\\&=\frac{(n-1)^2}{4n^2}.\end{aligned}$$
Genauso nimmt f das Maximum auf jedem Teilquadrat am rechten oberen Punkt an, und man erhält mit der entsprechenden Rechnung
$$\mathrm{OS}(f,Z)=\frac{(n+1)^2}{4n^2}.$$
Damit haben wir die Abschätzungen
$$\frac{(n-1)^2}{4n^2}=\mathrm{US}(f,Z)\le\mathrm{UI}(f)\le\mathrm{OI}(f)\le\mathrm{OS}(f,Z)=\frac{(n+1)^2}{4n^2},$$
da das Unter- bzw. Oberintegral nach Definition eine obere bzw. untere Schranke für die Unter- bzw. Obersummen ist. Bilden wir von dieser Ungleichungskette nun den Grenzwert für $n\to\infty$, so sehen wir, dass
$$\frac{1}{4}\le\mathrm{UI}(f)\le\mathrm{OI}(f)\le\frac{1}{4},$$
d. h. f ist integrierbar mit $\int_{[0,1]\times[0,1]}f(x)\,dx=\mathrm{UI}(f)=\mathrm{OI}(f)=\frac{1}{4}$.

Um die Integrierbarkeit von Funktionen in Zukunft einfacher beweisen zu können, halten wir als Nächstes das folgende sehr wichtige Kriterium fest, das wir auch schon aus dem Eindimensionalen kennen (siehe Lemma 12.11).

Lemma 28.10 (Riemannsches Integrabilitätskriterium). *Es sei $f\colon[a,b]\to\mathbb{R}$ eine beschränkte Funktion auf einem Quader $[a,b]$.*

(a) *f ist genau dann integrierbar, wenn es zu jedem $\varepsilon>0$ eine Zerlegung Z von $[a,b]$ gibt mit $\mathrm{OS}(f,Z)-\mathrm{US}(f,Z)<\varepsilon$.*

(b) *f ist genau dann integrierbar mit Integral $\int_{[a,b]}f(x)\,dx=c$, wenn es zu jedem $\varepsilon>0$ Zerlegungen Z und Z' von $[a,b]$ gibt mit $\mathrm{OS}(f,Z)<c+\varepsilon$ und $\mathrm{US}(f,Z')>c-\varepsilon$.*

Beweis. Der Beweis ist wörtlich derselbe wie der von Lemma 12.11. □

Ebenfalls wie im Eindimensionalen können wir mit Hilfe dieses Kriteriums nun zeigen, dass jede stetige Funktion integrierbar ist.

Satz 28.11. *Ist $f\colon [a,b] \to \mathbb{R}$ eine stetige Funktion auf einem Quader $[a,b]$, so ist f auch integrierbar auf $[a,b]$.*

Beweis. Der Beweis verläuft ähnlich zu dem von Satz 12.12. Da der Quader $[a,b]$ beschränkt und nach Aufgabe 23.44 (b) auch abgeschlossen ist, ist er nach Satz 23.51 (b) kompakt. Also ist f nach Folgerung 24.34 beschränkt, so dass wir die Begriffe dieses Kapitels anwenden können. Wir zeigen nun die Integrierbarkeit von f mit dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (a).

Es sei also $\varepsilon > 0$ gegeben. Wegen der Kompaktheit von $[a,b]$ ist f nach Bemerkung 24.25 (a) sogar gleichmäßig stetig. Es gibt also ein $\delta > 0$, so dass $|f(y) - f(z)| < \frac{\varepsilon}{\mathrm{vol}([a,b])}$ für alle $y,z \in [a,b]$ mit $\|y-z\|_\infty < \delta$ gilt.

Wir wählen nun eine Zerlegung Z von $[a,b]$, so dass alle Kantenlängen der Teilquader von Z kleiner als δ sind. Dann gilt zunächst

$$\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) = \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \big(\sup f(Q) - \inf f(Q)\big).$$

Da alle hier auftretenden Teilquader Q kompakt sind, nimmt f dort als stetige Funktion nach Folgerung 24.34 ein Maximum und Minimum an. Es gibt also Punkte $y_Q, z_Q \in Q$ mit $f(y_Q) = \sup f(Q)$ und $f(z_Q) = \inf f(Q)$. Weil y_Q und z_Q beide in dem Quader Q liegen, dessen Kantenlängen ja kleiner als δ sind, ist natürlich auch $\|y_Q - z_Q\|_\infty < \delta$ und damit $|f(y_Q) - f(z_Q)| < \frac{\varepsilon}{\mathrm{vol}([a,b])}$ nach Wahl von δ. Wir können oben also weiterrechnen und erhalten

$$\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) = \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot (f(y_Q) - f(z_Q)) < \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \frac{\varepsilon}{\mathrm{vol}([a,b])} = \varepsilon,$$

da die Summe aller Volumina der Teilquader nach Beispiel 28.9 (a) gleich $\mathrm{vol}([a,b])$ ist. Die Behauptung des Satzes ergibt sich nun also aus Lemma 28.10 (a). □

Eine wesentliche Verallgemeinerung dieser Aussage werden wir in Satz 29.25 kennenlernen. Zunächst einmal wollen wir aber die wichtigsten elementaren Eigenschaften des Riemann-Integrals zusammenfassen, die sich genauso wie im Eindimensionalen in Satz 12.13 ergeben.

Satz 28.12 (Eigenschaften des Integrals)**.** *Es seien $f,g\colon [a,b] \to \mathbb{R}$ integrierbare Funktionen auf einem Quader $[a,b] \subset \mathbb{R}^n$ und $c \in \mathbb{R}$.*

(a) *$f+g$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt*

$$\int_{[a,b]} (f(x)+g(x))\,dx = \int_{[a,b]} f(x)\,dx + \int_{[a,b]} g(x)\,dx.$$

(b) *$c\,f$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt*

$$\int_{[a,b]} c\,f(x)\,dx = c \cdot \int_{[a,b]} f(x)\,dx.$$

(c) *Ist $f \le g$, d. h. $f(x) \le g(x)$ für alle $x \in [a,b]$, so ist $\int_{[a,b]} f(x)\,dx \le \int_{[a,b]} g(x)\,dx$.*

(d) *$|f|$ ist ebenfalls integrierbar auf $[a,b]$, und es gilt die **Dreiecksungleichung***

$$\left| \int_{[a,b]} f(x)\,dx \right| \le \int_{[a,b]} |f(x)|\,dx.$$

Beweis. Der Beweis verläuft genauso wie in Satz 12.13, wobei sich auch der Beweis der dort zentral verwendeten Aussage aus Aufgabe 12.4 wörtlich auf den mehrdimensionalen Fall überträgt. □

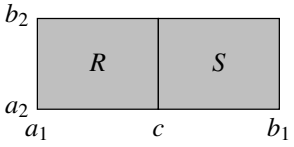
Aufgabe 28.13 (Integrale über Randfunktionen, siehe Beispiel 12.9 (c))**.** Es sei $f\colon Q \to \mathbb{R}$ eine beschränkte Funktion auf einem Quader $[a,b] \subset \mathbb{R}^n$. Man zeige: Ist $f(x) = 0$ für alle $x \in \mathring{Q}$, so ist f integrierbar mit $\int_{[a,b]} f(x)\,dx = 0$.

Aufgabe 28.14 (Additivität des Integrals, siehe Satz 12.14)**.** Es sei $f\colon Q \to \mathbb{R}$ eine beschränkte Funktion auf einem Quader $Q = [a_1,b_1] \times \cdots \times [a_n,b_n] \subset \mathbb{R}^n$. Man zeige: Ist $c \in (a_1,b_1)$, so ist f genau dann auf Q integrierbar, wenn f auf

$$\begin{aligned} R &= [a_1,c] \times [a_2,b_2] \times \cdots \times [a_n,b_n] \\ \text{und} \quad S &= [c,b_1] \times [a_2,b_2] \times \cdots \times [a_n,b_n] \end{aligned}$$

integrierbar ist, und in diesem Fall gilt

$$\int_Q f(x)\,dx = \int_R f(x)\,dx + \int_S f(x)\,dx.$$



Aufgabe 28.15. Es seien $f,g\colon [a,b] \to \mathbb{R}$ integrierbare Funktionen auf einem Quader $[a,b] \subset \mathbb{R}^n$. Man zeige:

(a) Die Funktion $f \cdot g$ ist integrierbar auf $[a,b]$.

(Hinweis: Man kann sich etwas Arbeit sparen, indem man die Aussage zunächst nur im Fall $g = f$ zeigt und sich dann überlegt, warum daraus bereits der allgemeine Fall folgt.)

(b) (Mittelwertsatz der Integralrechnung) Ist f sogar stetig und $g \geq 0$, so gibt es ein $c \in [a,b]$ mit

$$\int_{[a,b]} f(x)\,g(x)\,dx = f(c) \cdot \int_{[a,b]} g(x)\,dx.$$

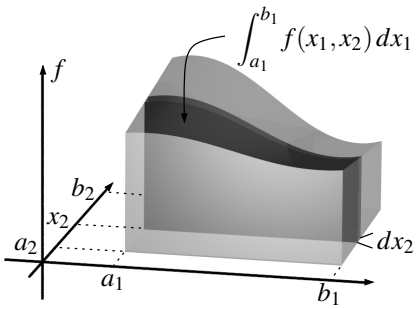
28.B Der Satz von Fubini

Wir haben die eindimensionale Theorie des Riemann-Integrals jetzt so weit ins Mehrdimensionale übertragen, wie dies ohne nennenswerte Änderungen möglich ist. Natürlich müssen wir nun aber auch noch untersuchen, wie man mehrdimensionale Integrale überhaupt in der Praxis berechnen kann, ohne jedesmal Grenzwerte von Unter- und Obersummen bestimmen zu müssen.

Dazu betrachten wir einmal die Situation wie im Bild rechts, bei der wir das Integral einer Funktion $f\colon [a_1,b_1] \times [a_2,b_2] \to \mathbb{R}$, also das Volumen unter dem Graphen dieser Funktion berechnen wollen. Halten wir dann zunächst einmal einen Wert $x_2 \in [a_2,b_2]$ fest, so ist das eindimensionale Integral

$$F(x_2) := \int_{a_1}^{b_1} f(x_1,x_2)\,dx_1$$

natürlich gerade die rechts dunkel eingezeichnete Querschnittsfläche des betrachteten Volumens mit einer senkrechten Ebene beim festen Wert x_2.



Anschaulich können wir uns den Ausdruck $F(x_2)\,dx_2$ also als eine dünne Scheibe des betrachteten Volumens vorstellen, die die Querschnittsfläche $F(x_2)$ und die Dicke dx_2 hat. Summieren wir nun die Volumina dieser dünnen Scheiben von a_2 bis b_2 auf, bilden wir also das Integral

$$\int_{a_2}^{b_2} F(x_2)\,dx_2 = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1,x_2)\,dx_1 \right) dx_2,$$

so würden wir demnach erwarten, dass wir damit das gesuchte Gesamtvolumen unter dem Graphen von f erhalten: Das ursprüngliche zweidimensionale Integral sollte sich gemäß

$$\int_{[a_1,b_1]\times[a_2,b_2]} f(x_1,x_2)\,d(x_1,x_2) = \int_{a_2}^{b_2} \left(\int_{a_1}^{b_1} f(x_1,x_2)\,dx_1 \right) dx_2$$

als doppeltes eindimensionales Integral schreiben lassen, das wir dann natürlich mit unseren bekannten Methoden aus Kapitel 12 berechnen können.

72

Wir werden nun beweisen, dass dies (unter recht milden Voraussetzungen an f) in der Tat richtig ist. Da wir später natürlich nicht nur zweidimensionale Integrale untersuchen wollen, betrachten wir

dabei gleich den allgemeinen Fall, in dem die Variablen x_1 und x_2 oben selbst wieder aus mehreren Komponenten bestehen. Zum besseren anschaulichen Verständnis des folgenden Satzes, der im Wesentlichen die obige Idee zu einem exakten Beweis macht, ist es aber vermutlich von Vorteil, wenn ihr euch die beiden Einzelintegrale erst einmal wie oben als jeweils eindimensional vorstellt, in der Notation des Satzes also an den Fall $n = n' = 1$ denkt.

Satz 28.16 (**Satz von Fubini** für Quader)**.** *Es seien $[a,b]$ und $[a',b']$ zwei Quader in $\mathbb{R}^n$ bzw. $\mathbb{R}^{n'}$ mit Koordinaten x bzw. x'. Ferner sei $f\colon [a,b]\times[a',b'] \to \mathbb{R}$ eine integrierbare Funktion, für die auch die Integrale*

$$F(x') := \int_{[a,b]} f(x,x')\,dx$$

über die erste Komponente für alle $x' \in [a',b']$ existieren.

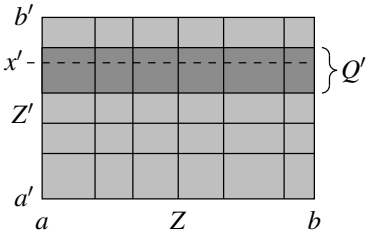
Dann ist auch $F\colon [a',b'] \to \mathbb{R}$ integrierbar, und es gilt

$$\int_{[a',b']} F(x')\,dx' = \int_{[a',b']} \left(\int_{[a,b]} f(x,x')\,dx \right) dx' = \int_{[a,b]\times[a',b']} f(x,x')\,d(x,x').$$

Beweis. Die erste Gleichung in der Behauptung ist natürlich einfach nur die Definition von F. Wir zeigen nun die Integrierbarkeit von F zusammen mit der zweiten Gleichung über das Riemannsche Integrabilitätskriterium aus Lemma 28.10 (b).

Es sei also $\varepsilon > 0$ gegeben. Da f nach Voraussetzung integrierbar ist, gibt es nach Lemma 28.10 (b) eine Zerlegung (Z,Z') von $[a,b]\times[a',b']$ mit

$$\mathrm{US}(f,(Z,Z')) > \int_{[a,b]\times[a',b']} f(x,x')\,d(x,x') - \varepsilon. \tag{1}$$



Betrachten wir wie im Bild rechts einen Teilquader Q' der so gewonnenen Zerlegung Z' von $[a',b']$, so gilt für alle $x' \in Q'$

$$\begin{aligned}
F(x') &= \int_{[a,b]} f(x,x')\,dx && \text{(Definition von } F\text{)} \qquad (2)\\
&\geq \mathrm{US}(f|_{[a,b]\times\{x'\}}, Z) && \text{(Integral obere Schranke für Untersummen)}\\
&= \sum_{Q\in\mathrm{TQ}(Z)} \mathrm{vol}(Q)\cdot \inf f(Q\times\{x'\}) && \text{(Definition der Untersumme)}\\
&\geq \sum_{Q\in\mathrm{TQ}(Z)} \mathrm{vol}(Q)\cdot \inf f(Q\times Q'), && \text{(größere Menge} \Rightarrow \text{kleineres Infimum)}
\end{aligned}$$

und damit auch für das Infimum dieser Zahlen

$$\inf F(Q') \geq \sum_{Q\in\mathrm{TQ}(Z)} \mathrm{vol}(Q)\cdot \inf f(Q\times Q'). \tag{3}$$

Wir erhalten so für die Untersumme von F bezüglich Z' die Abschätzung

$$\begin{aligned}
\mathrm{US}(F,Z') &= \sum_{Q'\in\mathrm{TQ}(Z')} \mathrm{vol}(Q')\cdot \inf F(Q')\\
&\overset{(3)}{\geq} \sum_{Q'\in\mathrm{TQ}(Z')} \mathrm{vol}(Q')\cdot \sum_{Q\in\mathrm{TQ}(Z)} \mathrm{vol}(Q)\cdot \inf f(Q\times Q')\\
&= \sum_{Q\times Q'\in\mathrm{TQ}(Z,Z')} \mathrm{vol}(Q\times Q')\cdot \inf f(Q\times Q')\\
&= \mathrm{US}(f,(Z,Z'))\\
&\overset{(1)}{>} \int_{[a,b]\times[a',b']} f(x,x')\,d(x,x') - \varepsilon.
\end{aligned}$$

Genauso zeigt man für die Obersumme von F

$$\mathrm{OS}(F,Z') < \int_{[a,b]\times[a',b']} f(x,x')\,d(x,x') + \varepsilon.$$

Damit folgt aus dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (b), dass F' integrierbar ist mit

$$\int_{[a',b']} F(x')\,dx' = \int_{[a,b]\times[a',b']} f(x,x')\,d(x,x'). \qquad \square$$

Folgerung 28.17. *Es seien $[a,b] = [a_1,b_1]\times\cdots\times[a_n,b_n] \subset \mathbb{R}^n$ ein Quader und $f\colon [a,b]\to\mathbb{R}$ eine stetige Funktion. Dann lässt sich das Integral über f als n-faches eindimensionales Integral*

$$\int_{[a,b]} f(x)\,dx = \int_{a_n}^{b_n}\left(\cdots\int_{a_2}^{b_2}\left(\int_{a_1}^{b_1} f(x)\,dx_1\right)dx_2\cdots\right)dx_n$$

berechnen. Außerdem darf die Reihenfolge dieser eindimensionalen Integrale dabei beliebig vertauscht werden (was wir auch bereits in Aufgabe 25.25 (c) gesehen hatten).

Beweis. Dies folgt sofort mit Induktion aus Satz 28.16: Mit $Q = [a_1,b_1]\times\cdots\times[a_{n-1},b_{n-1}]$ ergibt dieser Satz zunächst

$$\int_{[a,b]} f(x)\,dx = \int_{a_n}^{b_n}\left(\int_Q f(x)\,d(x_1,\dots,x_{n-1})\right)dx_n,$$

da sowohl f als auch die eingeschränkte Funktion $(x_1,\dots,x_{n-1})\mapsto f(x_1,\dots,x_{n-1},x_n)$ für festes x_n stetig und damit nach Satz 28.11 integrierbar sind. Anwenden der Induktionsvoraussetzung auf das innere Integral liefert dann sofort die Behauptung. Da die Nummerierung der Variablen beliebig ist, erhält man die gleiche Aussage natürlich auch mit einer anderen Reihenfolge der Integrale. $\square$

Notation 28.18. In der Aussage von Folgerung 28.17 lässt man die Klammern oft weg und schreibt das Integral damit als

$$\int_{[a,b]} f(x)\,dx = \int_{a_n}^{b_n}\cdots\int_{a_2}^{b_2}\int_{a_1}^{b_1} f(x)\,dx_1\,dx_2\cdots dx_n,$$

wobei dann darauf zu achten ist, dass die Integrationsgrenzen an den Integralzeichen in umgekehrter Reihenfolge wie die Differentiale $dx_1,\dots,dx_n$ geschrieben werden.

Ein einfacher Spezialfall des Satzes von Fubini ergibt sich in der oft auftretenden Situation, dass der Integrand ein Produkt von zwei Funktionen ist, die nur von jeweils einer der Variablen abhängen:

Folgerung 28.19. *Es seien $[a,b]$ und $[a',b']$ zwei Quader in $\mathbb{R}^n$ bzw. $\mathbb{R}^{n'}$ sowie $f\colon [a,b]\to\mathbb{R}$ und $g\colon [a',b']\to\mathbb{R}$ zwei stetige Funktionen. Dann gilt*

$$\int_{[a,b]\times[a',b']} f(x)g(x')\,d(x,x') = \left(\int_{[a,b]} f(x)\,dx\right)\cdot\left(\int_{[a',b']} g(x')\,dx'\right).$$

Beweis. Als stetige Funktion ist das Produkt $(x,x')\mapsto f(x)g(x')$ nach Satz 28.11 integrierbar. Der Satz 28.16 von Fubini ergibt daher unmittelbar

$$\begin{aligned}\int_{[a,b]\times[a',b']} f(x)g(x')\,d(x,x') &= \int_{[a',b']}\left(\int_{[a,b]} f(x)g(x')\,dx\right)dx' \\ &\overset{(1)}{=} \int_{[a',b']}\left(\int_{[a,b]} f(x)\,dx\right)\cdot g(x')\,dx' \\ &\overset{(2)}{=} \left(\int_{[a,b]} f(x)\,dx\right)\cdot\left(\int_{[a',b']} g(x')\,dx'\right),\end{aligned}$$

da man von der Integrationsvariablen unabhängige Faktoren ($g(x')$ in (1) bzw. das Integral über x in (2)) nach Satz 28.12 (b) aus dem Integral herausziehen kann. $\square$

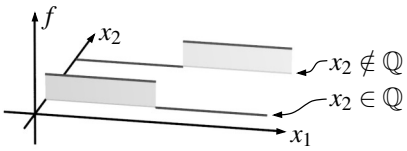
Beispiel 28.20. Mit Hilfe von Folgerung 28.19 können wir das Integral $\int_{[0,1]\times[0,1]} x_1x_2\,dx$ aus Beispiel 28.9 (b) nun auch viel einfacher berechnen, indem wir es auf zwei eindimensionale Integrale zurückführen und die Methoden aus Kapitel 12 benutzen: Es ist

$$\int_{[0,1]\times[0,1]} x_1x_2\,dx = \left(\int_0^1 x_1\,dx_1\right)\cdot\left(\int_0^1 x_2\,dx_2\right) = \left(\left[\frac{x^2}{2}\right]_0^1\right)^2 = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Bemerkung 28.21 (Notwendigkeit der Integrierbarkeitsvoraussetzungen im Satz von Fubini)**.**

(a) Die Voraussetzung der Integrierbarkeit der Gesamtfunktion f in Satz 28.16 ist wirklich notwendig: Dass es nicht genügt, die Existenz des Doppelintegrals vorauszusetzen, zeigt das Beispiel der Funktion

$$f\colon [0,1]\times[0,1]\to\mathbb{R},\ x\mapsto \begin{cases} 1 & \text{für } x_1\le\frac{1}{2}, x_2\in\mathbb{Q},\\ 0 & \text{für } x_1>\frac{1}{2}, x_2\in\mathbb{Q},\\ 0 & \text{für } x_1\le\frac{1}{2}, x_2\notin\mathbb{Q},\\ 1 & \text{für } x_1>\frac{1}{2}, x_2\notin\mathbb{Q}.\end{cases}$$



Für diese Funktion ist offensichtlich

$$\int_0^1 f(x_1,x_2)\,dx_1=\frac{1}{2}$$

für alle $x_2\in[0,1]$, und damit existiert das Doppelintegral

$$\int_0^1\left(\int_0^1 f(x_1,x_2)\,dx_1\right)dx_2=\int_0^1\frac{1}{2}\,dx_2=\frac{1}{2}.$$

Die Funktion f ist jedoch auf dem Quader $[0,1]\times[0,1]$ nicht integrierbar, da jeder Teilquader einer Zerlegung Z von $[0,1]\times[0,1]$ Punkte mit Funktionswert 0 und solche mit Funktionswert 1 enthält, so dass also stets $\mathrm{US}(f,Z)=0$ und $\mathrm{OS}(f,Z)=1$ und damit $\mathrm{UI}(f)=0$ und $\mathrm{OI}(f)=1$ gelten.

Beachte, dass in diesem Beispiel das andere Doppelintegral

$$\int_0^1\left(\int_0^1 f(x_1,x_2)\,dx_2\right)dx_1$$

nicht existiert, da f bei festem x_1 nicht über $x_2\in[0,1]$ integrierbar ist. Wir werden allerdings in Beispiel 28.24 sehen, dass es auch möglich ist, dass beide Doppelintegrale existieren, aber f trotzdem nicht integrierbar ist.

(b) Die in Satz 28.16 vorausgesetzte Existenz des inneren Integrals $F(x')$ ist hingegen *nicht* wirklich essentiell: Existiert dieses Integral nicht für alle x', so können wir $F(x')$ stattdessen als das (immer existierende) Unter- oder Oberintegral der Funktion $x\mapsto f(x,x')$ auf $[a,b]$ definieren. Tun wir dies, ändert sich im Beweis des Satzes lediglich der Ausdruck für $F(x')$ in (2) – aber die darauf folgende Abschätzung und damit auch der Rest des Beweises bleiben weiterhin gültig, da sowohl das Unter- als auch das Oberintegral eine obere Schranke für jede Untersumme ist.

In der Literatur wird der Satz von Fubini daher auch oft ohne die Voraussetzung der Existenz des inneren Integrals angegeben. Die Aussage des Satzes ist dann also so zu verstehen, dass man das innere Integral wahlweise als Unter- oder als Oberintegral interpretieren kann, falls diese nicht übereinstimmen.

Aufgabe 28.22. Berechne das Integral $\displaystyle\int_{[1,2]\times[1,2]}\frac{1}{x+y}\,d(x,y)$.

Aufgabe 28.23. Man zeige: Ist $f\colon[a,b]\to\mathbb{R}_{>0}$ eine stetige Funktion auf einem Quader $Q\subset\mathbb{R}^n$, so gilt

$$\int_Q f(x)\,dx\cdot\int_Q\frac{1}{f(x)}\,dx\ge(\mathrm{vol}(Q))^2.$$

(Hinweis: Beweise und verwende die Ungleichung $\frac{f(x)}{f(y)}+\frac{f(y)}{f(x)}\ge 2$ für alle $x,y\in Q$.)

Aufgabe 28.24. Es sei $Q=[0,1]\times[0,1]$. Man zeige:

(a) Es gibt eine Menge $D\subset Q$, für die gilt:

- Für alle $x,y\in D$ mit $x\neq y$ gilt $x_1\neq y_1$ und $x_2\neq y_2$.
- Für alle $a\in Q$ und $\varepsilon\in\mathbb{R}_{>0}$ enthält $U_\varepsilon(a)$ einen Punkt aus D.

(b) Für eine Menge D wie in (a) sei

$$f\colon Q \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{falls } x \in D, \\ 0 & \text{falls } x \notin D. \end{cases}$$

Dann existieren die Doppelintegrale

$$\int_0^1 \left(\int_0^1 f(x)\, dx_1 \right) dx_2 \quad \text{und} \quad \int_0^1 \left(\int_0^1 f(x)\, dx_2 \right) dx_1$$

und haben den gleichen Wert, aber f ist trotzdem nicht integrierbar auf Q.

29. Messbare Mengen

Im letzten Kapitel haben wir in direkter Analogie zu Kapitel 12 das mehrdimensionale Riemann-Integral über Quader in $\mathbb{R}^n$ eingeführt. Im Gegensatz zum eindimensionalen Fall gibt es in $\mathbb{R}^n$ aber natürlich noch sehr viel mehr anders geformte Mengen – z. B. Kreise, Ellipsen, oder generell durch irgendwelche Ungleichungen in den Koordinaten gegebene Teilmengen – über die man in der Praxis auch integrieren möchte. Wir wollen daher in diesem Kapitel sehen, wie man unsere Konstruktionen und Ergebnisse auf solche allgemeinen Integrationsbereiche ausdehnen kann.

29.A Integrale über beschränkte Mengen

Für unsere Verallgemeinerung des Riemann-Integrals wollen wir zunächst nur Integrationsbereiche D betrachten, die zwar nicht notwendig Quader, aber zumindest noch in einem Quader enthalten (also beschränkt) sind. Auch die zu integrierende Funktion $f\colon D \to \mathbb{R}$ werden wir erst einmal noch als beschränkt voraussetzen. In diesem Fall ist die Grundidee für die Konstruktion des Integrals sehr einfach: Wir setzen f auf einen beliebigen Quader $[a,b] \supset D$ fort, indem wir die Funktionswerte außerhalb von D gleich 0 setzen. Integrieren wir die so fortgesetzte Funktion dann wie in Kapitel 28 über $[a,b]$, so sollte dieses Integral keine Beiträge von Punkten außerhalb von D enthalten, und wir können es als das Integral von f über D auffassen.

Wir führen daher als Erstes eine Notation für derartige durch 0 fortgesetzte Funktionen ein.

Definition 29.1 (Indikatorfunktionen und Fortsetzung durch 0)**.** Es sei $f\colon M \to \mathbb{R}$ eine Funktion auf einer Menge $M \subset \mathbb{R}^n$. Ferner sei D eine Teilmenge von M.

(a) Wir bezeichnen die durch 0 außerhalb von D fortgesetzte Funktion f mit

$$f_D\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \begin{cases} f(x) & \text{für } x \in D, \\ 0 & \text{sonst.} \end{cases}$$

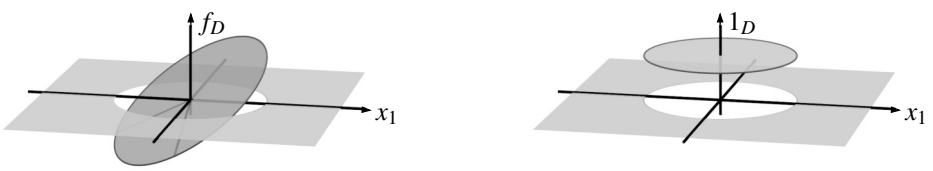
In der Regel werden wir hierbei nur den Fall $D = M$ benötigen.

(b) Speziell für die konstante Funktion 1 heißt

$$1_D\colon \mathbb{R}^n \to \mathbb{R},\ x \mapsto \begin{cases} 1 & \text{für } x \in D, \\ 0 & \text{sonst} \end{cases}$$

die **Indikatorfunktion** oder **charakteristische Funktion** von D.

Das folgende Bild zeigt diese beiden Konstruktionen im Fall des Kreises $D = \{x \in \mathbb{R}^2 : \|x\|_2 \le 1\}$ und der Funktion $f\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto x_1 + x_2$.



Bemerkung 29.2.

(a) Für die Indikatorfunktion 1_D einer beschränkten Menge $D \subset \mathbb{R}^n$ erwarten wir offensichtlich, dass man ihr Integral als das (n-dimensionale) Volumen von D interpretieren kann. Wir werden dies in Definition 29.5 (b) zur Definition des Volumens benutzen.

(b) Beachte, dass die Funktion f_D aus Definition 29.1 auch bei stetigem f in den Punkten von ∂D in der Regel unstetig ist. Hier liegt auch schon das Hauptproblem der obigen Idee, das Integral von f über D als das von f_D über einem Quader $[a,b] \supset D$ zu definieren: Wegen dieser Unstetigkeitsstellen von f_D ist nicht klar, ob f_D überhaupt integrierbar ist – Satz 28.11 ist hier ja nicht anwendbar. In der Tat werden wir sehen, dass es auch entscheidend von der Form von D abhängt, ob dieses Integral existiert oder nicht.

Wenn wir das Integral einer Funktion $f\colon D \to \mathbb{R}$ als das Integral von f_D über einem Quader $[a,b] \supset D$ definieren wollen, müssen wir aber natürlich als Erstes zeigen, dass das Ergebnis nicht von der Wahl dieses Quaders abhängt. Dies wollen wir jetzt tun, und benötigen dafür zuvor noch eine allgemeine Konstruktion von Zerlegungen eines Quaders, die in diesem Kapitel noch mehrmals vorkommen wird.

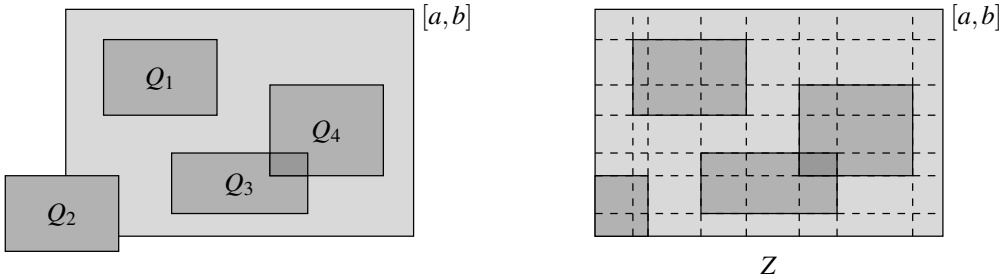
Konstruktion 29.3 (Induzierte Zerlegungen)**.** Es seien $[a,b]$ und $Q_1,\dots,Q_k$ für $k \in \mathbb{N}$ Quader in $\mathbb{R}^n$ mit

$$[a,b] = [a_1,b_1] \times \dots \times [a_n,b_n] \quad \text{und} \quad Q_i = [a_{i,1},b_{i,1}] \times \dots \times [a_{i,n},b_{i,n}]$$

für alle $i = 1,\dots,k$ (wobei $Q_1,\dots,Q_k$ in $[a,b]$ enthalten sein können, aber nicht müssen). Dann ist

$$Z = (Z_1,\dots,Z_n) \quad \text{mit} \quad Z_j = [a_j,b_j] \cap \Big(\{a_j,b_j\} \cup \{a_{i,j} : i = 1,\dots,k\} \cup \{b_{i,j} : i = 1,\dots,k\}\Big)$$

eine Zerlegung von $[a,b]$, die wie im Bild unten anschaulich dadurch entsteht, dass wir alle Seiten der Quader $Q_1,\dots,Q_k$ über diese Quader hinaus fortsetzen. Wir nennen sie die von $Q_1,\dots,Q_k$ *induzierte Zerlegung*. Sie hat offensichtlich die Eigenschaft, dass für jeden Quader Q_i der Schnitt $Q_i \cap [a,b]$ gleich der Vereinigung aller $Q \in \mathrm{TQ}(Z)$ mit $Q \subset Q_i$ ist.



73

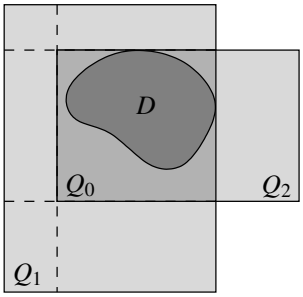
Lemma 29.4. *Es sei $f\colon D \to \mathbb{R}$ eine beschränkte Funktion auf einer beschränkten Teilmenge D von $\mathbb{R}^n$. Ferner seien Q_1 und Q_2 Quader, die D enthalten.*

Dann ist die durch 0 fortgesetzte Funktion f_D genau dann über Q_1 integrierbar, wenn sie über Q_2 integrierbar ist, und im Fall der Integrierbarkeit gilt

$$\int_{Q_1} f_D(x)\,dx = \int_{Q_2} f_D(x)\,dx.$$

Beweis. Es sei $Q_0 = Q_1 \cap Q_2$; auch dies ist offensichtlich ein Quader, der D enthält.

Wir betrachten nun wie im Bild gestrichelt eingezeichnet die von Q_0 induzierte Zerlegung Z von Q_1. Der Quader Q_0 ist dann einer der Teilquader dieser Zerlegung. Auf dem Inneren aller anderen Teilquader Q ist f_D gleich 0, und damit ist f_D nach Aufgabe 28.13 auf diesen Q integrierbar mit $\int_Q f_D(x)\,dx = 0$. Nach der Additivität des Integrals (siehe Aufgabe 28.14) angewendet auf die Zerlegung Z ist f_D also genau dann auf Q_1 integrierbar, wenn f_D auf Q_0 integrierbar ist, und in diesem Fall gilt



$$\int_{Q_1} f_D(x)\,dx = \int_{Q_0} f_D(x)\,dx + \sum_{\substack{Q \in \mathrm{TQ}(Z) \\ Q \neq Q_0}} \underbrace{\int f_D(x)\,dx}_{=0} = \int_{Q_0} f_D(x)\,dx.$$

Genauso folgt diese Aussage nun für Q_2 und Q_0, und damit die Behauptung. □

Mit diesem Lemma können wir nun die am Anfang dieses Kapitels motivierte Definition exakt hinschreiben.

Definition 29.5 (Integrierbarkeit und messbare Mengen)**.** Es sei $f\colon D \to \mathbb{R}$ eine beschränkte Funktion auf einer beschränkten Menge D. Ferner wählen wir einen Quader $Q \supset D$.

(a) Die Funktion f heißt **(Riemann-)integrierbar** (auf D), wenn f_D im Sinne von Definition 28.8 auf Q integrierbar ist. In diesem Fall setzen wir

$$\int_D f(x)\,dx := \int_Q f_D(x)\,dx.$$

(b) Die Menge D heißt **(Jordan-)messbar**, wenn ihre Indikatorfunktion 1_D integrierbar ist. In diesem Fall nennen wir

$$\operatorname{vol}(D) := \int_D 1\,dx = \int_Q 1_D(x)\,dx \quad \in \mathbb{R}_{\geq 0}$$

das **Volumen** von D.

Aufgrund von Lemma 29.4 sind diese Definitionen unabhängig von der Wahl von Q.

Beispiel 29.6.

(a) Ist D in Definition 29.5 selbst ein Quader, so können wir dort offensichtlich $Q = D$ wählen. Die Begriffe des Integrals über Q und des Volumens von Q stimmen dann also mit den alten aus Definition 28.8 bzw. Notation 28.3 überein (siehe Beispiel 28.9 (a)). Insbesondere sind Quader also stets messbar.

(b) Die Menge $D = \mathbb{Q} \cap [0,1] \subset \mathbb{R}$ ist nicht messbar, da ihre Indikatorfunktion nach Beispiel 12.9 (d) nicht integrierbar ist. Anschaulich heißt das, dass wir dieser Menge in unserer betrachteten Riemannschen Theorie nicht sinnvoll ein (1-dimensionales) Volumen zuordnen können. Dies sollte nicht weiter erstaunlich sein, wenn man bedenkt, dass ja in jeder Umgebung von jedem Punkt von $[0,1]$ sowohl Punkte in D als auch Punkte nicht in D liegen.

(c) Betten wir diese Menge jedoch auf der x_1-Achse in den $\mathbb{R}^2$ ein, betrachten wir also

$$D := \{(x,0) : x \in \mathbb{Q} \cap [0,1]\} \quad \subset \mathbb{R}^2,$$

so ist D nach Aufgabe 28.13 nun messbar mit Volumen 0: Sie liegt komplett im Rand des Quaders $[0,1] \times [0,1]$, und damit ist 1_D auch nur dort ungleich 0. Anschaulich gesehen hat die Menge D zwar immer noch dieselbe seltsame Form wie vorher in (b), aber sie ist trotzdem „klein genug“, um ihr das 2-dimensionale Volumen, also den Flächeninhalt 0 zuzuordnen.

Natürlich müssen wir nun Methoden finden, mit denen wir die Messbarkeit von Mengen überprüfen bzw. ihr Volumen berechnen können, und mit denen wir Integrale über allgemeine beschränkte Definitionsmengen bestimmen können. Als Erstes übertragen wir dazu die uns für Quader bereits bekannten Eigenschaften aus Satz 28.12 und erweitern sie noch um ein paar weitere.

Folgerung 29.7 (Eigenschaften des Integrals)**.** *Es seien $f, g\colon D \to \mathbb{R}$ integrierbare Funktionen auf einer beschränkten Menge $D \subset \mathbb{R}^n$ und $c \in \mathbb{R}$.*

(a) *$f + g$ ist ebenfalls integrierbar auf D, und es gilt*

$$\int_D (f(x) + g(x))\,dx = \int_D f(x)\,dx + \int_D g(x)\,dx.$$

(b) *$c\,f$ ist ebenfalls integrierbar auf D, und es gilt*

$$\int_D c\,f(x)\,dx = c \cdot \int_D f(x)\,dx.$$

(c) *Ist $f \leq g$, d. h. $f(x) \leq g(x)$ für alle $x \in D$, so ist $\int_D f(x)\,dx \leq \int_D g(x)\,dx$.*

(d) *$|f|$ ist ebenfalls integrierbar auf D, und es gilt die* ***Dreiecksungleichung***

$$\left|\int_D f(x)\,dx\right| \le \int_D |f(x)|\,dx.$$

Beweis. Alle diese Aussagen folgen mit Definition 29.5 (a) sofort aus Satz 28.12 angewendet auf f_D und g_D. □

Folgerung 29.8. *Es seien $f,g\colon D \to \mathbb{R}$ beschränkte Funktionen auf einer beschränkten Teilmenge $D \subset \mathbb{R}^n$. Dann gilt:*

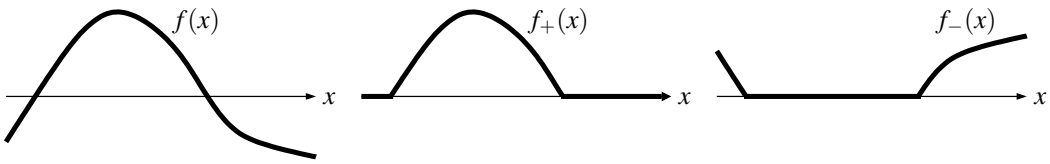
(a) *Sind f und g integrierbar auf D, so sind auch die Funktionen $\max(f,g)$ und $\min(f,g)$ auf D integrierbar.*

(b) *Die Funktion f ist genau dann auf D integrierbar, wenn die beiden unten im Bild dargestellten Funktionen*

$$f_+ := \max(f,0) \quad \textit{und} \quad f_- := \max(-f,0)$$

integrierbar sind, und in diesem Fall gilt

$$\int_D f(x)\,dx = \int_D f_+(x)\,dx - \int_D f_-(x)\,dx.$$



Beweis.

(a) Wegen

$$\max(f,g) = \frac{f+g}{2} + \left|\frac{f-g}{2}\right| \quad \text{und} \quad \min(f,g) = \frac{f+g}{2} - \left|\frac{f-g}{2}\right|$$

ergibt sich dies sofort aus Folgerung 29.7.

(b) Ist f integrierbar, so sind nach (a) auch f_+ und f_- integrierbar. Sind umgekehrt f_+ und f_- integrierbar, so ist nach Folgerung 29.7 wegen $f = f_+ - f_-$ auch f integrierbar, und es gilt die behauptete Formel. □

Mit diesen Ergebnissen können wir nun bereits ein Resultat analog zu Satz 12.14 und Aufgabe 28.14 über die Additivität von Integrationsbereichen bzw. Volumina zeigen.

Folgerung 29.9 (Additivität des Integrals und des Volumens)**.** *Es seien $A,B \subset \mathbb{R}^n$ zwei beschränkte Mengen und $f\colon A \cup B \to \mathbb{R}$ eine beschränkte Funktion.*

(a) *Ist f auf A und B integrierbar, so auch auf $A \cup B$, $A \cap B$ und $A \backslash B$, und es gilt*

$$\int_{A\cup B} f(x)\,dx + \int_{A\cap B} f(x)\,dx = \int_A f(x)\,dx + \int_B f(x)\,dx$$

$$\textit{und} \qquad \int_{A\backslash B} f(x)\,dx = \int_A f(x)\,dx - \int_{A\cap B} f(x)\,dx.$$

(b) *Sind A und B messbar, so auch $A \cup B$, $A \cap B$ und $A \backslash B$, und es gilt*

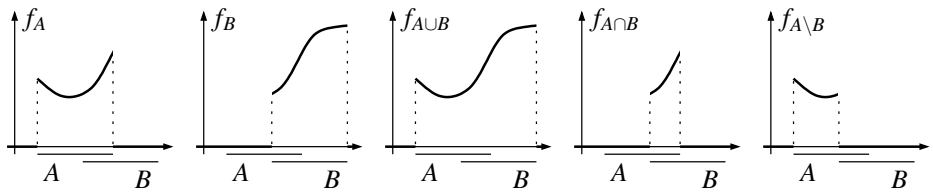
$$\mathrm{vol}(A\cup B) + \mathrm{vol}(A\cap B) = \mathrm{vol}(A) + \mathrm{vol}(B) \quad \textit{und} \quad \mathrm{vol}(A\backslash B) = \mathrm{vol}(A) - \mathrm{vol}(A\cap B).$$

Beweis.

(a) Nach Folgerung 29.8 (b) genügt es, die Aussage für die dort eingeführten Funktionen f_+ und f_- zu zeigen, deren Funktionswerte ja überall nicht-negativ sind. Mit anderen Worten können wir uns auf den Fall beschränken, dass $f(x) \geq 0$ für alle $x \in A \cup B$ gilt. Dann gilt aber wie im Bild unten

$$f_{A\cup B} = \max(f_A, f_B), \quad f_{A\cap B} = \min(f_A, f_B) \quad \text{und} \quad f_{A\setminus B} = f_A - f_{A\cap B},$$

so dass sich die Integrierbarkeit dieser drei Funktionen, also die Integrierbarkeit von f auf $A \cup B$, $A \cap B$ und $A \setminus B$ aus Folgerung 29.7 und 29.8 ergibt. Da außerdem offensichtlich $f_{A\cup B} + f_{A\cap B} = f_A + f_B$ und $f_{A\setminus B} = f_A - f_{A\cap B}$ gelten, zeigt Folgerung 29.7 auch die beiden behaupteten Gleichungen für diese Integrale.



(b) folgt sofort aus (a) angewendet auf die konstante Funktion 1. □

Bemerkung 29.10 (Eigenschaften des Volumens)**.** Insbesondere ergibt sich aus Folgerung 29.9 (b) direkt:

(a) Endliche Vereinigungen, Durchschnitte und Differenzmengen messbarer Mengen sind messbar.

(b) Sind A und B messbar und $B \subset A$, so ist

$$\operatorname{vol}(B) = \operatorname{vol}(A \cap B) = \operatorname{vol}(A) - \operatorname{vol}(A \setminus B) \leq \operatorname{vol}(A)$$

(man sagt auch, das Volumen ist *monoton*).

(c) Sind A und B messbar, so gilt für die damit ebenfalls messbare Menge $A \cup B$

$$\operatorname{vol}(A \cup B) = \operatorname{vol}(A) + \operatorname{vol}(B) - \operatorname{vol}(A \cap B) \leq \operatorname{vol}(A) + \operatorname{vol}(B)$$

(man sagt auch, das Volumen ist *subadditiv*). Im Fall $A \cap B$ gilt offensichtlich sogar

$$\operatorname{vol}(A \cup B) = \operatorname{vol}(A) + \operatorname{vol}(B)$$

(man sagt dafür auch, dass das Volumen *additiv* ist).

Mit vollständiger Induktion bedeutet dies natürlich auch

$$\operatorname{vol}(A_1 \cup \cdots \cup A_k) \leq \operatorname{vol}(A_1) + \cdots + \operatorname{vol}(A_k)$$

für messbare Mengen $A_1, \ldots, A_k$, wobei im Fall paarweise disjunkter Mengen wieder die Gleichheit gilt.

74

Beispiel 29.11 (Volumina offener Quader)**.** Es sei $Q = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ ein Quader.

(a) Der Rand ∂Q des Quaders ist nach Aufgabe 28.13 messbar mit Volumen 0 (da die Funktion $1_{\partial Q}$ auf $\overset{\circ}{Q}$ gleich 0 ist). Aus der Additivität des Volumens folgt also, dass auch $\overset{\circ}{Q}$ messbar ist mit

$$\operatorname{vol}\left(\overset{\circ}{Q}\right) = \operatorname{vol}(Q) - \operatorname{vol}(\partial Q) = \operatorname{vol}(Q),$$

d. h. das Herausnehmen des Randes ändert nichts an der Messbarkeit oder am Volumen des Quaders.

(b) Vergrößern wir Q etwas zu einem offenen Quader über seine Kanten hinaus, so vergrößert sich dadurch sein Volumen nur wenig: Setzen wir

$$R_\delta = [a_1-\delta, b_1+\delta] \times \cdots \times [a_n-\delta, b_n+\delta] \quad \text{für } \delta \in \mathbb{R}_{>0},$$

so gilt $Q \subset \overset{\circ}{R}_\delta$ für alle δ, und wegen der Stetigkeit der Funktion $\delta \mapsto \operatorname{vol}(R_\delta)$ gibt es zu jedem $\varepsilon \in \mathbb{R}_{>0}$ ein $\delta \in \mathbb{R}_{>0}$ mit $\operatorname{vol}(R_\delta) < \operatorname{vol}(Q) + \varepsilon$.

Aufgabe 29.12. Zeige, dass das Riemann-Integral *translationsinvariant* ist: Sind f eine integrierbare Funktion auf einer beschränkten Menge $D \subset \mathbb{R}^n$ und $a \in \mathbb{R}^n$, so ist auch die Funktion $x \mapsto f(x-a)$ auf der Menge $D+a := \{x+a : x \in D\}$ integrierbar, und es gilt

$$\int_{D+a} f(x-a)\,dx = \int_D f(x)\,dx.$$

Aufgabe 29.13 (Offene, abgeschlossene bzw. kompakte Mengen müssen nicht messbar sein)**.** Wir wissen nach Beispiel 5.58 (a), dass die Menge $\mathbb{Q} \cap [0,1]$ abzählbar ist. Es sei nun also $(q_n)_{n\in\mathbb{N}}$ eine solche Abzählung.

(a) Zeige, dass die Menge

$$D = \bigcup_{n\in\mathbb{N}} \left(q_n - \frac{1}{2^{n+3}}, q_n + \frac{1}{2^{n+3}} \right) \quad \subset \mathbb{R}$$

offen und nicht messbar ist.

(b) Finde eine kompakte (und damit auch abgeschlossene) Teilmenge von $\mathbb{R}$, die nicht messbar ist.

(Hinweis: Da die obigen Intervalle nur eine Gesamtlänge von $\frac{1}{2}$ haben, kann D ja wohl nicht das ganze Einheitsintervall $[0,1]$ überdecken. Andererseits enthält D aber um jede rationale Zahl noch ein offenes Intervall und muss damit doch eigentlich auch alle irrationalen Zahlen in $[0,1]$ enthalten, also doch das ganze Intervall $[0,1]$ überdecken? Wo ist der Denkfehler?)

29.B Nullmengen

Von besonderer Bedeutung sind messbare Mengen, deren Volumen gleich 0 ist: Sie spielen nicht nur bei der Volumenberechnung, sondern auch – wie wir gleich sehen werden – bei Integralen keine Rolle, so dass sie dabei also beliebig hinzugefügt oder weggelassen werden können, bzw. die zu integrierende Funktion dort beliebig abgeändert werden kann, ohne etwas an der Integrierbarkeit oder dem Wert des Integrals zu ändern. Wir wollen derartige Mengen jetzt genauer untersuchen.

Definition 29.14 (Nullmengen)**.** Eine beschränkte Menge $N \subset \mathbb{R}^n$ heißt eine **(Jordan-)Nullmenge**, wenn sie messbar ist mit $\operatorname{vol}(N) = 0$.

Lemma 29.15 (Äquivalente Kriterien für Nullmengen)**.** *Für eine beschränkte Teilmenge $N \subset \mathbb{R}^n$ sind äquivalent:*

(a) *N ist eine Nullmenge.*

(b) *Für alle $\varepsilon > 0$ gibt es eine Zerlegung Z eines Quaders $[a,b] \supset N$ mit*

$$\sum_{Q \in \mathrm{TQ}(Z) : Q \cap N \neq \emptyset} \operatorname{vol}(Q) < \varepsilon.$$

(c) *Für alle $\varepsilon > 0$ gibt es endlich viele Quader $Q_1, \dots, Q_k$ mit*

$$N \subset Q_1 \cup \cdots \cup Q_k \quad \textit{und} \quad \sum_{i=1}^k \operatorname{vol}(Q_i) < \varepsilon.$$

(d) *Für alle $\varepsilon > 0$ gibt es endlich viele Quader $Q_1, \dots, Q_k$ mit*

$$N \subset \overset{\circ}{Q}_1 \cup \cdots \cup \overset{\circ}{Q}_k \quad \textit{und} \quad \sum_{i=1}^k \operatorname{vol}(Q_i) < \varepsilon.$$

Beweis.

(a) $\Rightarrow$ (b): Nach Voraussetzung ist die Indikatorfunktion 1_N über einen Quader $[a,b] \supset N$ integrierbar mit Integral 0. Dies bedeutet nach dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (b), dass es zu jedem $\varepsilon > 0$ eine Zerlegung Z von $[a,b]$ gibt mit

$$\mathrm{OS}(1_N, Z) = \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \sup 1_N(Q) < \varepsilon.$$

Da dieses Supremum aber gerade 1 ist falls $Q \cap N \neq \emptyset$, und 0 sonst, folgt damit schon die Behauptung (b).

(b) $\Rightarrow$ (c): Dies ist klar, da wir für $Q_1, \ldots, Q_k$ einfach die Teilquader Q mit $Q \cap N \neq \emptyset$ der aus (b) vorausgesetzten Zerlegung nehmen können.

(c) $\Rightarrow$ (d): Nach Voraussetzung gibt es zu jedem $\varepsilon > 0$ Quader $Q_1, \ldots, Q_k$ mit $N \subset Q_1 \cup \cdots \cup Q_k$ und $\sum_{i=1}^k \mathrm{vol}(Q_i) < \frac{\varepsilon}{2}$. Wie in Beispiel 29.11 (b) vergrößern wir nun jeden dieser Quader Q_i zu einem Quader R_i mit $Q_i \subset \mathring{R}_i$ und $\mathrm{vol}(R_i) < \mathrm{vol}(Q_i) + \frac{\varepsilon}{2k}$. Dann überdecken die offenen Quader $\mathring{R}_1, \ldots, \mathring{R}_k$ die Menge N, und es gilt

$$\sum_{i=1}^k \mathrm{vol}(R_i) < \sum_{i=1}^k \left(\mathrm{vol}(Q_i) + \frac{\varepsilon}{2k} \right) < \frac{\varepsilon}{2} + k \cdot \frac{\varepsilon}{2k} = \varepsilon.$$

(d) $\Rightarrow$ (a): Wir zeigen mit dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (b), dass das Integral über die Indikatorfunktion 1_N auf einem Quader $[a,b] \supset N$ existiert und gleich 0 ist.

Es sei dazu $\varepsilon > 0$ gegeben. Wir wählen Quader $Q_1, \ldots, Q_k$ wie in (d), und betrachten die von ihnen induzierte Zerlegung Z von $[a,b]$ wie in Konstruktion 29.3. Da N vollständig in $\mathring{Q}_1 \cup \cdots \cup \mathring{Q}_k$ enthalten ist, ist die Indikatorfunktion 1_N höchstens auf den Teilquadern $Q \subset Q_1 \cup \cdots \cup Q_k$ nicht identisch 0, und dort ist ihr Supremum dann höchstens gleich 1. Damit folgt

$$\begin{aligned} \mathrm{OS}(1_N, Z) &= \sum_{Q \in \mathrm{TQ}(Z)} \mathrm{vol}(Q) \cdot \sup 1_N(Q) \leq \sum_{\substack{Q \in \mathrm{TQ}(Z): \\ Q \subset Q_1 \cup \cdots \cup Q_k}} \mathrm{vol}(Q) \\ &= \mathrm{vol}(Q_1 \cup \cdots \cup Q_k) \overset{29.10\,(c)}{\leq} \mathrm{vol}(Q_1) + \cdots + \mathrm{vol}(Q_k) < \varepsilon. \end{aligned}$$

Da die Ungleichung $\mathrm{US}(1_N, Z) \geq 0$ wegen $1_N \geq 0$ offensichtlich ist, ist damit die Integrierbarkeit von 1_N gezeigt. □

Bemerkung 29.16 (Teilmengen von Nullmengen sind Nullmengen)**.** Insbesondere folgt aus der Charakterisierung (c) von Nullmengen aus Lemma 29.15, dass jede Teilmenge A einer Nullmenge N wieder eine Nullmenge ist – denn Quader, die N überdecken, überdecken dann natürlich auch A. Beachte, dass die entsprechende Aussage für allgemeine messbare Mengen falsch ist: Eine Teilmenge einer messbaren Menge muss nicht wieder messbar sein, wie das Beispiel $\mathbb{Q} \cap [0,1] \subset [0,1]$ in $\mathbb{R}$ zeigt.

Ein weiteres Nullmengenkriterium, das aus Lemma 29.15 folgt und sehr viele Nullmengen als solche erkennen kann, ist das folgende.

Folgerung 29.17 (Graphen integrierbarer Funktionen sind Nullmengen)**.** *Es sei $f\colon D \to \mathbb{R}$ eine integrierbare Funktion auf einer beschränkten Menge $D \subset \mathbb{R}^n$. Dann ist der Graph*

$$N := \{(x, f(x)) : x \in D\} \quad \subset \mathbb{R}^n \times \mathbb{R}$$

von f eine Nullmenge in $\mathbb{R}^{n+1}$.

Beweis. Es sei $\varepsilon > 0$ gegeben. Da f_D nach Voraussetzung auf einem Quader $[a,b] \supset D$ integrierbar ist, gibt es nach dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (a) eine Zerlegung Z

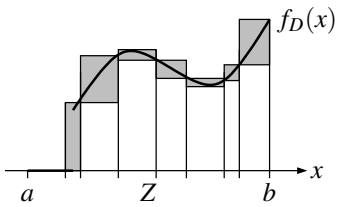
von $[a,b]$, so dass

$$\mathrm{OS}(f_D,Z)-\mathrm{US}(f_D,Z)=\sum_{Q\in\mathrm{TQ}(Z)}\mathrm{vol}(Q)\cdot\big(\sup f_D(Q)-\inf f_D(Q)\big)<\varepsilon.$$

Dies ist aber gerade die Summe der Volumina der rechts im Bild grau eingezeichneten Quader

$$Q\times\big[\inf f_D(Q),\sup f_D(Q)\big]$$

für $Q\in\mathrm{TQ}(Z)$, die offensichtlich den Graphen von f_D und damit auch den von f überdecken. Also ist N nach dem Kriterium aus Lemma 29.15 (c) eine Nullmenge. □



Beispiel 29.18. Die Kreislinie $\{x\in\mathbb{R}^2:\|x\|_2=1\}$ ist nach Bemerkung 29.10 (c) und Folgerung 29.17 eine Nullmenge in $\mathbb{R}^2$, da sie die Vereinigung der Graphen der beiden stetigen und damit integrierbaren Funktionen $[-1,1]\to\mathbb{R},\ x\mapsto\pm\sqrt{1-x^2}$ ist.

Allgemein kann man sich die Graphen von (integrierbaren) Funktionen $f\colon D\to\mathbb{R}$ auf $D\subset\mathbb{R}^n$ anschaulich als „n-dimensionale Objekte“ in $\mathbb{R}^{n+1}$ vorstellen, denen man daher das $(n+1)$-dimensionale Volumen 0 zuordnen kann.

Aufgabe 29.19. Zeige, dass jede beschränkte Teilmenge von $\mathbb{R}^n$, die nur endlich viele Häufungspunkte besitzt, eine Nullmenge ist.

Aufgabe 29.20. Es seien $A\subset B\subset C$ beschränkte Teilmengen von $\mathbb{R}^n$. Man zeige: Sind A und C messbar mit $\mathrm{vol}(C)=\mathrm{vol}(A)$, so ist auch B messbar mit $\mathrm{vol}(B)=\mathrm{vol}(A)$.

Aufgabe 29.21. Es sei $(r_n)_{n\in\mathbb{N}}$ eine Nullfolge positiver reeller Zahlen. Zeige, dass dann die Menge $M=\{x\in\mathbb{R}^2:\text{es gibt ein } n\in\mathbb{N}\text{ mit }\|x\|_2=r_n\}$ eine Nullmenge ist.

Nachdem wir von einer gegebenen Menge nun gut nachprüfen können, ob sie eine Nullmenge ist, wollen wir jetzt wie bereits angekündigt sehen, dass Nullmengen nicht nur beim Volumen, sondern auch bei der Integration beliebiger (beschränkter) Funktionen keine Rolle spielen.

Satz 29.22 (Integrale über Nullmengen sind 0)**.** *Jede beschränkte Funktion $f\colon N\to\mathbb{R}$ auf einer Nullmenge $N\subset\mathbb{R}^n$ ist integrierbar mit $\int_N f(x)\,dx=0$.*

Beweis. Wir zeigen die Integrierbarkeit und den Wert des Integrals mit dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (b).

Da f beschränkt ist, gibt es ein $c\in\mathbb{R}_{>0}$ mit $|f(x)|\le c$ für alle $x\in N$. Weiterhin ist N eine Nullmenge, und damit gibt es zu gegebenem $\varepsilon\in\mathbb{R}_{>0}$ nach Lemma 29.15 (b) eine Zerlegung Z eines Quaders $[a,b]\supset N$, so dass die N schneidenden Teilquader von Z ein Gesamtvolumen kleiner als $\frac{\varepsilon}{c}$ haben. Dann gilt aber

$$\mathrm{OS}(f_N,Z)=\sum_{Q\in\mathrm{TQ}(Z)}\mathrm{vol}(Q)\cdot\sup f_N(Q)\le\sum_{Q\in\mathrm{TQ}(Z):Q\cap N\neq\emptyset}\mathrm{vol}(Q)\cdot c<\frac{\varepsilon}{c}\cdot c=\varepsilon,$$

da das Supremum auf den Teilquadern Q mit $Q\cap N=\emptyset$ gleich 0 und auf allen anderen höchstens gleich c ist. Analog zeigt man $\mathrm{US}(f_N,Z)>-\varepsilon$, so dass die Behauptung des Satzes aus Lemma 28.10 (b) folgt. □

Notation 29.23 („fast überall“)**.** Man sagt, eine von einem Punkt $a\in\mathbb{R}^n$ abhängige Eigenschaft gilt **fast überall**, wenn es eine Nullmenge N gibt, so dass die Eigenschaft für alle $a\notin N$ gilt. So sagt man z. B. für zwei Funktionen $f,g\colon D\to\mathbb{R}$ auf einer beschränkten Menge $D\subset\mathbb{R}^n$:

- $f=g$ fast überall, wenn es eine Nullmenge $N\subset D$ gibt mit $f|_{D\setminus N}=g|_{D\setminus N}$;
- f ist fast überall stetig, wenn es eine Nullmenge $N\subset D$ gibt, so dass f in jedem Punkt von $D\setminus N$ stetig ist.

Wir können Satz 29.22 damit also so umformulieren: Ist f eine beschränkte Funktion, die fast überall gleich 0 ist, so ist f integrierbar mit Integral 0.

Bemerkung 29.24. Für viele Aussagen über Integrale reicht es aufgrund von Satz 29.22 aus, wenn die von den beteiligten Funktionen geforderten Voraussetzungen nur fast überall gelten. Hier sind zwei Beispiele dafür:

(a) Es seien $f,g\colon D\to\mathbb{R}$ beschränkt auf einer beschränkten Menge $D\subset\mathbb{R}^n$. Ist dann f integrierbar und gilt fast überall $f=g$, so ist auch g integrierbar mit $\int_D g(x)\,dx=\int_D f(x)\,dx$: Dann ist nämlich $g-f$ beschränkt und nur auf einer Nullmenge ungleich 0, nach Satz 29.22 also integrierbar mit Integral 0 – und damit ergibt sich die Behauptung mit der Formel

$$\int_D g(x)\,dx=\int_D f(x)\,dx+\int_D (g(x)-f(x))\,dx$$

aus Folgerung 29.7 (a).

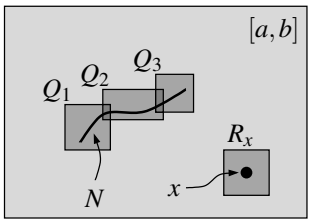
(b) Sind $f,g\colon D\to\mathbb{R}$ zwei integrierbare Funktionen auf einer beschränkten Menge $D\subset\mathbb{R}^n$ und gilt fast überall $f\le g$, so ist $\int_D f(x)\,dx\le\int_D g(x)\,dx$: Dann können wir f und g nämlich nach (a) ohne Änderung ihrer Integrale auf der Nullmenge, auf der nicht $f\le g$ gilt, gleich 0 setzen. Damit gilt aber überall $f\le g$, und die Behauptung ergibt sich aus Folgerung 29.7 (c).

In dieses Schema passt auch gut der folgende zentrale Satz, dass eine fast überall stetige Funktion auf einem Quader integrierbar ist – da wir ja in Satz 12.12 gesehen haben, dass dies für überall stetige Funktionen gilt. Beachte jedoch, dass diese Aussage nicht wie die in Bemerkung 29.24 direkt aus Satz 29.22 folgt, da es in der Regel nicht möglich ist, eine fast überall stetige Funktion so auf einer Nullmenge abzuändern, dass sie überall stetig wird.

Satz 29.25 (**Lebesguesches Integrabilitätskriterium** für Quader)**.** *Jede beschränkte, fast überall stetige Funktion $f\colon[a,b]\to\mathbb{R}$ auf einem Quader $[a,b]$ ist integrierbar.*

Beweis. Wir zeigen den Satz mit dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (a). Es sei also $\varepsilon>0$ gegeben. Nach den Voraussetzungen an f gibt es ferner ein $c\in\mathbb{R}_{>0}$ mit $|f(x)|\le c$ für alle $x\in[a,b]$ sowie eine Nullmenge $N\subset[a,b]$, so dass f in jedem Punkt von $[a,b]\backslash N$ stetig ist. Wir konstruieren nun eine offene Überdeckung des Quaders $[a,b]$ und setzen dazu $\varepsilon':=\frac{\varepsilon}{2c+\mathrm{vol}([a,b])}$.

- Weil N eine Nullmenge ist, gibt es nach Lemma 29.15 (d) wie im Bild rechts Quader $Q_1,\dots,Q_k$ mit
$$\sum_{i=1}^{k}\mathrm{vol}(Q_i)<\varepsilon',\tag{1}$$
deren Inneres die Menge N überdeckt.
- Für jeden Punkt $x\in[a,b]\backslash N$ hingegen ist f stetig in x, und damit gibt es einen Quader R_x mit Mittelpunkt x, so dass



$$\sup f(R_x\cap[a,b])-\inf f(R_x\cap[a,b])<\varepsilon'.\tag{2}$$

75

Die offenen Quader $\mathring{Q}_1,\dots,\mathring{Q}_k$ sowie $\mathring{R}_x$ für alle $x\in[a,b]\backslash N$ überdecken nun nach Konstruktion die kompakte Menge $[a,b]$. Nach dem Satz 23.58 von Heine-Borel können wir hieraus also eine endliche Teilüberdeckung wählen, d. h. es ist $[a,b]\subset\mathring{Q}_1\cup\dots\cup\mathring{Q}_k\cup\mathring{R}_{x_1}\cup\dots\cup\mathring{R}_{x_l}$ für gewisse $x_1,\dots,x_l\in[a,b]\backslash N$. Es sei Z die von diesen Quadern $Q_1,\dots,Q_k,R_{x_1},\dots,R_{x_l}$ induzierte Zerlegung wie in Konstruktion 29.3. Dann liegt jeder Teilquader $Q\in\mathrm{TQ}(Z)$ in einem Q_i für $i=1,\dots,k$ oder

R_{x_j} für $j = 1,\dots,l$ (oder beides), und damit folgt

$$\begin{aligned}
\mathrm{OS}(f,Z) - \mathrm{US}(f,Z) &= \sum_{Q\in \mathrm{TQ}(Z)} \mathrm{vol}(Q)\cdot(\sup f(Q) - \inf f(Q)) \\
&\le \underbrace{\sum_{\substack{Q\in \mathrm{TQ}(Z):\\ Q\subset Q_i \text{ für ein } i}} \mathrm{vol}(Q)}_{\le \mathrm{vol}(Q_1\cup\dots\cup Q_k) < \varepsilon' \text{ nach (1)}} \cdot \underbrace{(\sup f(Q) - \inf f(Q))}_{\le 2c} \\
&\quad + \underbrace{\sum_{\substack{Q\in \mathrm{TQ}(Z):\\ Q\subset R_{x_j} \text{ für ein } j}} \mathrm{vol}(Q)}_{\le \mathrm{vol}([a,b])} \cdot \underbrace{(\sup f(Q) - \inf f(Q))}_{< \varepsilon' \text{ nach (2)}} \\
&< (2c + \mathrm{vol}([a,b]))\cdot \varepsilon' = \varepsilon.
\end{aligned}$$

□

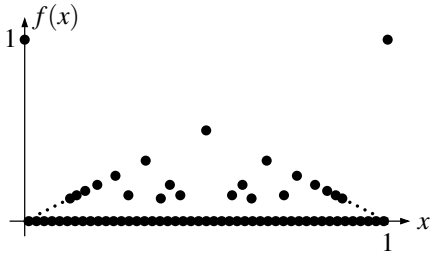
Beispiel 29.26 (Stückweise stetige Funktionen)**.** Wir hatten in Beispiel 12.16 bereits gesehen, dass jede stückweise stetige Funktion auf einem Intervall $[a,b] \subset \mathbb{R}$ integrierbar ist. Diese Aussage ist offensichtlich ein Spezialfall von Satz 29.25, da die Menge der Punkte, an denen eine solche Funktion unstetig ist, endlich und damit natürlich eine Nullmenge ist. Aber auch für z. B. die Funktion

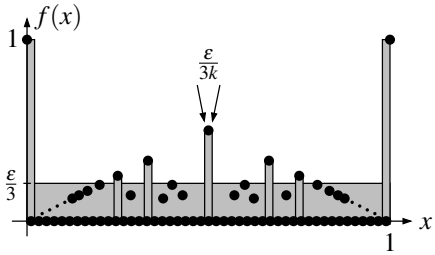
$$f\colon [-1,1] \to \mathbb{R},\ x \mapsto \begin{cases} \sin\frac{1}{x} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0, \end{cases}$$

die zwar nicht stückweise stetig, aber beschränkt und stetig außerhalb von 0 ist, garantiert Satz 29.25 die Integrierbarkeit (siehe auch Aufgabe 12.17 (b)).

Beispiel 29.27 (Die Umkehrung des Lebesgueschen Integrabilitätskriteriums gilt nicht)**.** Das folgende Beispiel zeigt, dass die Umkehrung von Satz 29.25 falsch ist: Wie im Bild unten links schematisch dargestellt sei

$$f\colon [0,1] \to \mathbb{R}, x \mapsto \begin{cases} \frac{1}{q} & \text{für } x \in \mathbb{Q} \text{ mit gekürzter Darstellung } x = \frac{p}{q},\ p \in \mathbb{N},\ q \in \mathbb{N}_{>0}, \\ 0 & \text{für } x \notin \mathbb{Q}. \end{cases}$$





Dann ist f nach dem Riemannschen Integrabilitätskriterium aus Lemma 12.11 (b) integrierbar mit Integral 0: Zu gegebenem $\varepsilon > 0$ gibt es wie im Bild rechts nur endlich viele $x_1,\dots,x_k \in [0,1]$ mit Funktionswert größer als $\frac{\varepsilon}{3}$. Wählen wir eine Zerlegung Z von $[0,1]$, die um jeden dieser Punkte ein Intervall der Länge höchstens $\frac{\varepsilon}{3k}$ enthält, so erhält die Obersumme $\mathrm{OS}(f,Z)$ von diesen Intervallen einen Beitrag von höchstens $k\cdot\frac{\varepsilon}{3k} = \frac{\varepsilon}{3}$, und der Beitrag der übrigen Intervalle ist ebenfalls höchstens $\frac{\varepsilon}{3}$. Insgesamt ist dann also $\mathrm{OS}(f,Z) \le \frac{2\varepsilon}{3} < \varepsilon$, so dass die Integrierbarkeit von f mit Integral 0 aus Lemma 12.11 (b) folgt.

Die Funktion f ist jedoch an allen rationalen Punkten unstetig, weil in jeder Umgebung eines solchen Punktes auch irrationale Zahlen und damit Punkte mit Funktionswert 0 liegen. Da $\mathbb{Q}\cap[0,1]$ nach Beispiel 29.6 (b) aber keine Nullmenge ist, ist f damit nicht fast überall stetig. Dieses Beispiel zeigt also, dass die Umkehrung von Satz 29.25 nicht gilt.

Man kann jedoch zeigen, dass die Umkehrung von Satz 29.25 richtig wird, wenn man im Gegensatz zu Lemma 29.15 auch solche Mengen als Nullmengen bezeichnet, für die es zu jedem $\varepsilon > 0$ eine Überdeckung durch *abzählbar viele* Quader mit Volumensumme kleiner als ε gibt [M, Satz

45.13]. Man spricht in diesem Fall auch von *Lebesgue-Nullmengen* im Gegensatz zu unseren Jordan-Nullmengen, da dies genau die Nullmengen in der Lebesgueschen Integrationstheorie (siehe Bemerkung 28.2) sind. So ist dann z. B. jede abzählbare Menge, also z. B. $\mathbb{Q}$ oder $\mathbb{Q}\cap[0,1]$, trivialerweise eine Lebesgue-Nullmenge, da sie die abzählbare Vereinigung aller ihrer Punkte, also von Quadern mit Volumen 0 ist.

Man sagt dann auch, dass eine Eigenschaft *Lebesgue-fast überall* gilt, wenn sie an jedem Punkt mit Ausnahme einer Lebesgue-Nullmenge gilt. Die Form, in der man das Lebesguesche Integrabilitätskriterium in der Literatur oft findet, ist also: Eine beschränkte Funktion auf einem Quader ist genau dann Riemann-integrierbar, wenn sie Lebesgue-fast überall stetig ist. Wir werden diese allgemeinere Aussage im Folgenden aber nicht benötigen.

Als erste wichtige Folgerung von Satz 29.25 wollen wir nun ein Kriterium beweisen, mit dem man in der Regel einfach überprüfen kann, ob eine gegebene Menge messbar ist.

Folgerung 29.28 (messbar $\Leftrightarrow$ Rand ist Nullmenge)**.** *Eine beschränkte Menge $A\subset\mathbb{R}^n$ ist genau dann messbar, wenn ihr Rand ∂A eine Nullmenge ist.*

Beweis.

„$\Rightarrow$“ Wir wählen einen Quader $[a,b]$, der $\overline{A}$ im Inneren enthält. Da A als messbar vorausgesetzt ist, die Indikatorfunktion 1_A also über $[a,b]$ integrierbar ist, gibt es nach dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (a) zu jedem $\varepsilon>0$ eine Zerlegung Z von $[a,b]$ mit

$$\mathrm{OS}(1_A,Z)-\mathrm{US}(1_A,Z)=\sum_{Q\in\mathrm{TQ}(Z)}\mathrm{vol}(Q)\cdot\underbrace{(\sup 1_A(Q)-\inf 1_A(Q))}_{(*)}<\varepsilon.$$

Offensichtlich gilt für dieses Supremum und Infimum

$$\sup 1_A(Q)=\begin{cases}1 & \text{falls } Q\cap A\neq\emptyset,\\ 0 & \text{falls } Q\cap A=\emptyset\end{cases}$$

$$\text{und}\quad \inf 1_A(Q)=\begin{cases}0 & \text{falls } Q\cap(\mathbb{R}^n\backslash A)\neq\emptyset,\\ 1 & \text{falls } Q\cap(\mathbb{R}^n\backslash A)=\emptyset.\end{cases}$$

Also ist die Differenz $(*)$ nur dann ungleich 0, und zwar dann gleich 1, wenn $Q\cap A\neq\emptyset$ und $Q\cap(\mathbb{R}^n\backslash A)\neq\emptyset$ gilt, also wenn Q sowohl einen Punkt in A als auch einen Punkt nicht in A enthält. Sind $Q_1,\dots,Q_k$ diese Teilquader von Z, so ist deren Volumensumme also kleiner als ε.

Gleichzeitig überdecken diese Teilquader aber auch den Rand ∂A: Ist $x\in\partial A$, so wählen wir eine offene Umgebung von x, die keine weiteren Teilquader außer denen trifft, in denen x liegt: Diese Umgebung enthält dann nach Definition 23.38 (c) einen Punkt y, so dass von den Punkten x und y genau einer in A liegt. Nach Wahl der Umgebung liegt y nun in einem Teilquader von Z, in dem auch x liegt, und der demzufolge in den $Q_1,\dots,Q_k$ enthalten ist.

Also ist ∂A nach Lemma 29.15 (c) eine Nullmenge.

„$\Leftarrow$“ Die Indikatorfunktion 1_A ist genau am Rand ∂A unstetig. Ist dieser also eine Nullmenge, so ist 1_A nach Satz 29.25 integrierbar, d. h. A ist messbar. □

Beispiel 29.29.

(a) Die Kreisscheibe $A=\{x\in\mathbb{R}^2:\|x\|_2\le 1\}$ ist nach Folgerung 29.28 messbar, da ihr Rand $\partial A=\{x\in\mathbb{R}^2:\|x\|_2=1\}$ nach Beispiel 29.18 eine Nullmenge ist.

(b) Wir wussten bereits aus Beispiel 29.6 (b), dass die Menge $A=\mathbb{Q}\cap[0,1]$ nicht messbar ist. Alternativ sieht man dies nun leicht mit Folgerung 29.28, da der Rand $\partial A=[0,1]$ Volumen 1 hat und damit keine Nullmenge ist.

(c) Ist $A\subset\mathbb{R}^n$ messbar, so kann man beliebig zu A Randpunkte hinzunehmen oder von A wegnehmen, ohne etwas an der Messbarkeit oder dem Volumen zu ändern: Nach Bemerkung 29.16 und Folgerung 29.28 ist jede Teilmenge B von ∂A eine Nullmenge, und damit sind nach Folgerung 29.9 auch $A\cup B$ und $A\backslash B$ messbar mit $\mathrm{vol}(A\cup B)=\mathrm{vol}(A\backslash B)=\mathrm{vol}(A)$.

Verbindet man die Ergebnisse von Satz 29.25 und Folgerung 29.28 miteinander, so erhält man die folgende Aussage, mit der man für die meisten in der Praxis auftretenden Funktionen sofort sehen kann, dass sie integrierbar sind.

Folgerung 29.30 (**Lebesguesches Integrabilitätskriterium** für messbare Mengen)**.** *Jede beschränkte, fast überall stetige Funktion $f\colon D \to \mathbb{R}$ auf einer messbaren Menge $D \subset \mathbb{R}^n$ ist integrierbar.*

Beweis. Es sei $N \subset D$ eine Nullmenge, so dass f in jedem Punkt von $D \backslash N$ stetig ist. Dann ist die Funktion f_D auf einem Quader $[a,b] \supset D$ höchstens in den Punkten von $N \cup \partial D$ unstetig. Da ∂D aber nach Folgerung 29.28 eine Nullmenge ist, gilt dies nach Bemerkung 29.10 (c) auch für $N \cup \partial D$. Nach Satz 29.25 ist also f_D auf $[a,b]$, und damit nach Definition auch f auf D integrierbar. □

Aufgabe 29.31. Untersuche, ob die folgenden Mengen messbar sind, und bestimme im Fall der Messbarkeit ihr Volumen:

(a) $\{(x,y) \in \mathbb{Q} \times \mathbb{R} : x^2 + y^2 = 1\} \subset \mathbb{R}^2$;

(b) $\{(x,y) \in \mathbb{Q} \times \mathbb{R} : x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$;

(c) $A \times N \subset \mathbb{R}^{m+n}$ für eine beschränkte Menge $A \subset \mathbb{R}^m$ und eine Nullmenge $N \subset \mathbb{R}^n$.

Aufgabe 29.32. Es seien $f\colon D \to \mathbb{R}$ eine integrierbare Funktionen auf einer beschränkten Menge $D \subset \mathbb{R}^n$ und $A \subset D$ eine messbare Teilmenge von D. Zeige, dass f dann auch auf A integrierbar ist.

Aufgabe 29.33 (Integrale als Volumen zwischen Graphen)**.** Es seien $D \subset \mathbb{R}^n$ eine messbare Menge und $f,g\colon D \to \mathbb{R}$ zwei beschränkte stetige Funktionen mit $f(x) \leq g(x)$ für alle $x \in D$. Zeige, dass die Menge

$$M = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in D,\ f(x) \leq y \leq g(x)\} \quad \subset \mathbb{R}^{n+1}$$

zwischen den Graphen von f und g messbar ist mit $\operatorname{vol}(M) = \displaystyle\int_D (g(x) - f(x))\,dx$.

29.C Normalbereiche

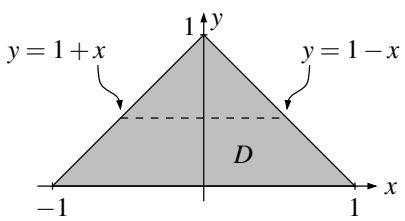
Wir haben uns inzwischen gute Kriterien erarbeitet, mit denen wir überprüfen können, ob eine gegebene Menge messbar bzw. eine Nullmenge, und ob eine gegebene Funktion integrierbar ist. Natürlich müssen wir nun aber auch noch untersuchen, wie man diese Volumina bzw. Integrale im Fall der Existenz dann auch berechnen kann.

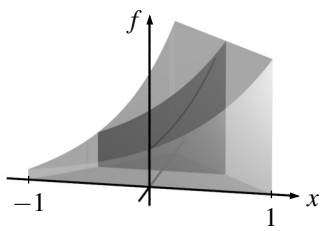
Die erfreuliche Nachricht hierzu ist, dass wir dies bereits können – der Satz 28.16 von Fubini liefert uns schon alles, was wir dazu benötigen. Um dies zu sehen, beginnen wir am besten mit einem Beispiel.

Beispiel 29.34. Für die Funktion $f\colon \mathbb{R}^2 \to \mathbb{R},\ (x,y) \mapsto \mathrm{e}^{x+y}$ wollen wir das Integral

$$\int_D f(x,y)\,d(x,y)$$

über das links unten eingezeichnete Dreieck D berechnen, also das Volumen unter dem Graphen von f wie im Bild rechts.





Beachte zunächst, dass dieses Integral existiert: Der Rand ∂D ist eine Vereinigung aus den beiden Funktionsgraphen zu $[-1,1] \to \mathbb{R},\ x \mapsto 0$ und $[-1,1] \to \mathbb{R},\ x \mapsto 1-|x|$ und damit nach Bemerkung 29.10 (c) und Folgerung 29.17 eine Nullmenge. Also ist D nach Folgerung 29.28 messbar, und die stetige Funktion f darauf nach dem Lebesgueschen Integrabilitätskriterium aus Folgerung 29.30 integrierbar.

Nach Definition 29.5 (a) ist das gesuchte Integral nun gleich $\int_Q f_D(x,y)\,d(x,y)$ für einen beliebigen Quader $Q \supset D$ – wir können hier $Q = [-1,1] \times [0,1]$ wählen. Da dieses Integral existiert, sagt uns der Satz 28.16 von Fubini also, dass

$$\int_D \mathrm{e}^{x+y}\,d(x,y) = \int_0^1 \left(\int_{-1}^1 f_D(x,y)\,dx \right) dy,$$

zumindest sofern auch das innere Integral existiert (was wir gleich durch eine explizite Berechnung dieses Integrals sehen werden).

Um dieses innere Integral auszuwerten, betrachten wir nun ein festes $y \in [0,1]$ – so wie auf der gestrichelten Linie im Bild oben links bzw. auf der dunklen Fläche im Bild oben rechts. Die Funktion $x \mapsto f_D(x,y)$ stimmt dort nach Definition mit f überein, wenn $(x,y) \in D$ liegt, und ist sonst gleich 0. Da die diagonalen Seitenlinien von D durch die Gleichungen $y = 1+x$ bzw. $y = 1-x$ gegeben sind, stimmt f_D für unser festes y also genau zwischen $x = y-1$ und $x = 1-y$ mit f überein. Das innere Integral ist demnach ein Integral über f zwischen $y-1$ und $1-y$, und wir erhalten so

$$\int_D \mathrm{e}^{x+y}\,d(x,y) = \int_0^1 \left(\int_{y-1}^{1-y} \mathrm{e}^{x+y}\,dx \right) dy.$$

Wir haben die spezielle Form von D also letztlich dadurch berücksichtigt, dass wir die Integrationsgrenzen angepasst haben – und zwar so, dass die Grenzen des inneren Integrals über x von der äußeren Integrationsvariable y abhängen und angeben, für welche x bei festem y der betrachtete Punkt in D liegt.

Der Rest der Rechnung lässt sich nun problemlos mit unseren eindimensionalen Methoden aus Kapitel 12 durchführen: Wir erhalten als Endergebnis

$$\int_0^1 \left(\int_{y-1}^{1-y} \mathrm{e}^{x+y}\,dx \right) dy = \int_0^1 \left[\mathrm{e}^{x+y}\right]_{x=y-1}^{x=1-y} dy = \int_0^1 (\mathrm{e} - \mathrm{e}^{2y-1})\,dy = \left[\mathrm{e}y - \frac{1}{2}\mathrm{e}^{2y-1}\right]_0^1 = \frac{1}{2}(\mathrm{e} + \mathrm{e}^{-1}).$$

Wir wollen diese Idee zur Berechnung mehrdimensionaler Integrale nun noch zu einem allgemeinen Kriterium ausbauen. Dies hat den Vorteil, dass wir im Fall der Anwendbarkeit dieses Kriteriums im Gegensatz zum obigen Beispiel dann keinerlei Integrierbarkeitsvoraussetzungen mehr überprüfen, sondern nur noch die eigentliche Rechnung durchführen müssen. Die Gebiete, für die dies am einfachsten möglich ist, sind die sogenannten Normalbereiche. Diese Normalbereiche sind also lediglich als Hilfsmittel für Berechnungen zu verstehen; sie haben in der Mathematik keine theoretische Bedeutung.

Definition 29.35 (Normalbereiche)**.** Ein (n-dimensionaler) **Normalbereich** in $\mathbb{R}^n$ ist rekursiv über n wie folgt definiert:

(a) Für $n = 1$ ist ein Normalbereich in $\mathbb{R}^1 = \mathbb{R}$ ein abgeschlossenes Intervall $[a_1, b_1] \subset \mathbb{R}$ mit $a_1 \le b_1$.

(b) Für $n > 1$ ist ein Normalbereich in $\mathbb{R}^n$ eine Menge der Form

$$\{(x,y) \in \mathbb{R}^{n-1} \times \mathbb{R} : x \in M \text{ und } a_n(x) \le y \le b_n(x)\} \quad \subset \mathbb{R}^{n-1} \times \mathbb{R} = \mathbb{R}^n,$$

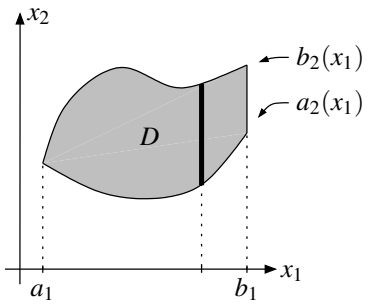
wobei $M \subset \mathbb{R}^{n-1}$ ein $(n-1)$-dimensionaler Normalbereich und $a_n, b_n \colon M \to \mathbb{R}$ stetige Funktionen mit $a_n(x) \le b_n(x)$ für alle $x \in M$ sind.

Beispiel 29.36.

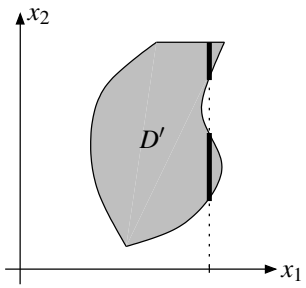
(a) Für $n = 2$ ist ein Normalbereich in $\mathbb{R}^2$ also von der Form

$$D = \{(x_1, x_2) \in \mathbb{R}^2 : a_1 \le x_1 \le b_1 \text{ und } a_2(x_1) \le x_2 \le b_2(x_1)\} \quad \subset \mathbb{R}^2$$

für gewisse $a_1, b_1 \in \mathbb{R}$ mit $a_1 \le b_1$ und stetige Funktionen $a_2, b_2 \colon [a_1, b_1] \to \mathbb{R}$ mit $a_2 \le b_2$. Anschaulich wird D also wie im Bild unten links dadurch beschrieben, dass x_1 zwischen den Grenzen a_1 und b_1 läuft, während x_2 dann für ein solches festes $x_1 \in [a_1, b_1]$ zwischen den (von x_1 abhängenden) Grenzen a_2 und b_2, also wie im Bild durch die dicke Linie angedeutet im Intervall $[a_2(x_1), b_2(x_1)]$ variiert.



Normalbereich bezüglich (x_1, x_2)



kein Normalbereich bezüglich (x_1, x_2), aber Normalbereich bezüglich (x_2, x_1)

Spiegelt man diese Menge D an der Diagonalen, vertauscht man also wie im Bild oben rechts die beiden Koordinaten, so ist die daraus resultierende Menge D' in diesem Fall jedoch nach unserer Definition kein Normalbereich mehr, denn für den eingezeichneten Wert von x_1 variiert x_2 in D' (wie durch die dicken Linien dargestellt) nicht in einem Intervall, sondern in zwei disjunkten Intervallen. Dagegen könnte man diese gespiegelte Menge aber natürlich durch Vertauschen der Variablen wieder wie oben in der Form

$$D' = \{(x_1, x_2) \in \mathbb{R}^2 : a_2 \le x_2 \le b_2 \text{ und } a_1(x_2) \le x_1 \le b_1(x_2)\} \quad \subset \mathbb{R}^2$$

schreiben, in der nun zunächst x_2 zwischen zwei Grenzen a_2 und b_2, und x_1 dann für ein festes x_2 in einem Intervall $[a_1(x_2), b_1(x_2)]$ variiert. Will man dies ausdrücken, so sagt man, dass z. B. die Mengen in $\mathbb{R}^2$ mit der Eigenschaft aus Definition 22.22 wie im Bild oben links *Normalbereiche bezüglich der Variablenordnung* (x_1, x_2) sind, während die Menge D' wie im Bild oben rechts ein Normalbereich bezüglich der Variablenordnung (x_2, x_1) ist. Um die Notationen einfach zu halten, werden wir Normalbereiche in $\mathbb{R}^n$ im Folgenden immer als Normalbereiche bezüglich $(x_1, \dots, x_n)$ annehmen; analoge Aussagen gelten aber natürlich auch für jede andere Variablenordnung.

(b) Für allgemeines n hat ein Normalbereich in $\mathbb{R}^n$ analog die Form

$$\begin{aligned}\{(x_1, \dots, x_n) \in \mathbb{R}^n : a_1 \le x_1 \le b_1,\\ a_2(x_1) \le x_2 \le b_2(x_1), \dots,\\ a_n(x_1, \dots, x_{n-1}) \le x_n \le b_n(x_1, \dots, x_{n-1})\}\end{aligned}$$

für stetige Funktionen $a_1, \dots, a_n, b_1, \dots, b_n$ in den entsprechenden Variablen mit $a_i \le b_i$ für alle $i = 1, \dots, n$. Das Gebiet muss sich also so beschreiben lassen, dass der Laufbereich jeder Variablen ein Intervall ist, dessen Grenzen nur von den vorherigen Variablen abhängt.

Wir wollen nun zeigen, dass stetige Funktionen auf Normalbereichen immer integrierbar sind, und dass sich ihre Integrale wie in Beispiel 29.34 als mehrfache eindimensionale Integrale berechnen lassen. Dazu benötigen wir zunächst das folgende Lemma.

Lemma 29.37. *Jeder Normalbereich ist kompakt und messbar.*

Beweis. Es sei D ein n-dimensionaler Normalbereich in $\mathbb{R}^n$. Wir zeigen die Aussage mit Induktion über n. Der Induktionsanfang ist dabei klar, denn für $n = 1$ ist D ein abgeschlossenes Intervall, also kompakt nach Beispiel 23.52 und messbar nach Beispiel 29.6 (a).

Für den Induktionsschritt sei $D = \{(x,y) \in \mathbb{R}^{n-1} \times \mathbb{R} : x \in M \text{ und } a_n(x) \le y \le b_n(x)\}$ wie in Definition 29.35 (b), wobei M ein Normalbereich in $\mathbb{R}^{n-1}$ ist. Nach Satz 23.51 (b) genügt es zu zeigen, dass D beschränkt, abgeschlossen und messbar ist.

- D ist beschränkt: Der Normalbereich M ist nach Induktionsvoraussetzung kompakt. Die stetigen Funktionen a_n und b_n nehmen auf dieser kompakten Menge nach Folgerung 24.34 also ein Minimum c bzw. Maximum d an. Da M als kompakte Menge nach Satz 23.51 (a) beschränkt ist, ist damit auch $D \subset M \times [c,d]$ beschränkt.
- D ist abgeschlossen: Wir zeigen dies mit dem Folgenkriterium für Abgeschlossenheit aus Satz 23.42. Es sei also $(x_k, y_k)_k \in \mathbb{R}^{n-1} \times \mathbb{R}$ eine konvergente Folge mit Grenzwert (x,y), deren Folgenglieder in D liegen. Da M nach Induktionsvoraussetzung kompakt und damit abgeschlossen ist, ist mit $x_k \in M$ für alle k dann auch $x \in M$. Bilden wir weiterhin in der Ungleichung $a_n(x_k) \le y_k \le b_n(x_k)$ den Grenzwert für $k \to \infty$, so erhalten wir wegen der Stetigkeit von a_n und b_n mit Satz 24.4 (b) auch die Ungleichung $a_n(x) \le y \le b_n(x)$, also insgesamt $(x,y) \in D$. Damit ist D abgeschlossen.
- D ist messbar: Dies folgt unmittelbar aus Aufgabe 29.33. □

76

Folgerung 29.38 (**Satz von Fubini** für Normalbereiche)**.** *Wie in Beispiel 29.36 (b) sei*

$$\begin{aligned} D = \{(x_1,\dots,x_n) \in \mathbb{R}^n : a_1 \le x_1 \le b_1,\\ a_2(x_1) \le x_2 \le b_2(x_1),\ \dots,\\ a_n(x_1,\dots,x_{n-1}) \le x_n \le b_n(x_1,\dots,x_{n-1})\} \end{aligned}$$

ein Normalbereich in $\mathbb{R}^n$. Dann ist jede auf D stetige Funktion f integrierbar, und ihr Integral lässt sich als n-faches eindimensionales Integral

$$\int_D f(x)\,dx = \int_{a_1}^{b_1} \left(\int_{a_2(x_1)}^{b_2(x_1)} \left(\cdots \int_{a_n(x_1,\dots,x_{n-1})}^{b_n(x_1,\dots,x_{n-1})} f(x_1,\dots,x_n)\,dx_n \cdots \right) dx_2 \right) dx_1$$

berechnen.

Beweis. Auch diese Aussage beweisen wir wieder mit Induktion über n; für $n = 1$ ist nichts zu zeigen.

Da D nach Lemma 29.37 kompakt und damit beschränkt ist, können wir für den Induktionsschritt im Fall $n > 1$ einen Quader der Form $[a_1,b_1] \times Q$ mit $Q \subset \mathbb{R}^{n-1}$ wählen, der D enthält. Wir wollen nun den Satz 28.16 von Fubini für Quader auf die Funktion f_D auf $[a_1,b_1] \times Q$ wie in Definition 29.1 (a) anwenden und müssen daher die Voraussetzungen dieses Satzes überprüfen:

- Die Funktion f ist nach Voraussetzung auf der gemäß Lemma 29.37 kompakten Menge D stetig und nach Folgerung 24.34 damit beschränkt. Da D nach Lemma 29.37 außerdem messbar ist, ist f nach dem Lebesgueschen Integrabilitätskriterium aus Folgerung 29.30 auf D integrierbar.
- Betrachten wir für ein festes $x_1 \in [a_1,b_1]$ die Menge

  $$D_{x_1} := \{(x_2,\dots,x_n) \in Q : (x_1,\dots,x_n) \in D\}$$

  aller Punkte von D mit diesem festen x_1-Wert, so ist dies ein $(n-1)$-dimensionaler Normalbereich. Nach Induktionsvoraussetzung ist die Funktion $(x_2,\dots,x_n) \mapsto f(x_1,x_2,\dots,x_n)$ also auf D_{x_1} integrierbar mit

  $$\begin{aligned} \int_Q f_D(x_1,\dots,x_n)\,d(x_2,\dots,x_n) &= \int_{D_{x_1}} f(x_1,\dots,x_n)\,d(x_2,\dots,x_n) \\ &= \int_{a_2(x_1)}^{b_2(x_1)} \left(\cdots \int_{a_n(x_1,\dots,x_{n-1})}^{b_n(x_1,\dots,x_{n-1})} f(x_1,\dots,x_n)\,dx_n \cdots \right) dx_2. \end{aligned}$$

Damit erhalten wir nun wie behauptet

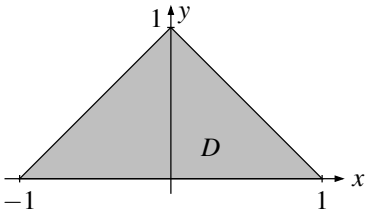
$$\begin{aligned}\int_D f(x)\,dx &= \int_{[a_1,b_1]\times Q} f_D(x)\,dx \\ &= \int_{a_1}^{b_1}\left(\int_Q f_D(x_1,\dots,x_n)\,d(x_2,\dots,x_n)\right)dx_1 && \text{(Satz 28.16)} \\ &= \int_{a_1}^{b_1}\left(\int_{a_2(x_1)}^{b_2(x_1)}\left(\cdots\int_{a_n(x_1,\dots,x_{n-1})}^{b_n(x_1,\dots,x_{n-1})} f(x_1,\dots,x_n)\,dx_n\cdots\right)dx_2\right)dx_1. \qquad \square\end{aligned}$$

Beispiel 29.39. Das rechts abgebildete Dreieck können wir wie in Beispiel 29.34 als Normalbereich

$$D = \{(x,y) : 0 \le y \le 1 \text{ und } y-1 \le x \le 1-y\}$$

bezüglich (y,x) schreiben. Dementsprechend ist jede auf D stetige Funktion f nach Folgerung 29.38 integrierbar, und es gilt

$$\int_D f(x,y)\,d(x,y) = \int_0^1\left(\int_{y-1}^{1-y} f(x,y)\,dx\right)dy,$$



wie wir auch schon im obigen Beispiel gesehen haben. Genauso können wir D auch als Normalbereich

$$D = \{(x,y) : -1 \le x \le 1 \text{ und } 0 \le y \le 1-|x|\}$$

bezüglich (x,y) schreiben, und erhalten für jede stetige Funktion f auf D die analoge Formel

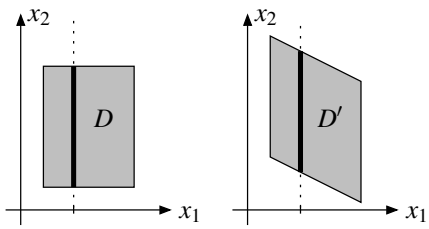
$$\int_D f(x,y)\,d(x,y) = \int_{-1}^1\left(\int_0^{1-|x|} f(x,y)\,dy\right)dx$$

zur Berechnung des Integrals über f als doppeltes eindimensionales Integral.

Beispiel 29.40 (Prinzip von Cavalieri). Wie im Bild rechts seien D und D' zwei messbare Mengen in $\mathbb{R}^n$, so dass für jedes feste $x_1 \in \mathbb{R}$ die Mengen

$$\begin{aligned} D_{x_1} &= \{(x_2,\dots,x_n) : (x_1,\dots,x_n) \in D\} \\ \text{und}\quad D'_{x_1} &= \{(x_2,\dots,x_n) : (x_1,\dots,x_n) \in D'\}\end{aligned}$$

der Punkte von D bzw. D' mit diesem festen x_1-Wert messbar sind und das gleiche $(n-1)$-dimensionale Volumen haben.



Dann haben auch D und D' dasselbe Volumen: Nach dem Satz 28.16 von Fubini gilt zunächst

$$\mathrm{vol}(D) = \int 1_D(x)\,dx = \int\left(\int 1_{D_{x_1}}(x)\,d(x_2,\dots,x_n)\right)dx_1 = \int \mathrm{vol}(D_{x_1})\,dx_1$$

(mit Integralen über genügend große Quader), und analog dann natürlich auch

$$\mathrm{vol}(D') = \int_{a_1}^{b_1} \mathrm{vol}(D'_{x_1})\,dx_1.$$

Wegen $\mathrm{vol}(D_{x_1}) = \mathrm{vol}(D'_{x_1})$ bedeutet dies also wie behauptet $\mathrm{vol}(D) = \mathrm{vol}(D')$. In der Literatur wird diese einfache Folgerung aus dem Satz von Fubini oft als das *Prinzip von Cavalieri* bezeichnet.

Aufgabe 29.41.

(a) Berechne $\mathrm{vol}(D)$ für das Tetraeder D, das von den drei Koordinatenebenen und der durch die Gleichung $x_3 = 2 - 2x_1 - 2x_2$ beschriebenen Ebene in $\mathbb{R}^3$ begrenzt wird.

(b) Berechne das Integral $\int_D |y| \cdot \cos x\, d(x,y)$, wobei D die durch die Gleichung $4x^2 + y^2 \le 4$ gegebene Ellipse in $\mathbb{R}^2$ ist.

29.D Uneigentliche Integrale

Zum Abschluss dieses Kapitels wollen wir nun noch uneigentliche Integrale analog zu Definition 12.27 konstruieren, also die bisher immer gemachte Voraussetzung fallen lassen, dass die Definitionsmenge und die zu integrierende Funktion beschränkt sind. Wie auch schon im Eindimensionalen werden wir dies natürlich tun, indem wir z. B. ein Integral über einen unbeschränkten Bereich als Grenzwert von Integralen über immer größer werdende beschränkte Bereiche definieren. Im Gegensatz zu Kapitel 12 ist die exakte Formulierung dieser Idee im Mehrdimensionalen aber etwas komplizierter, da es in $\mathbb{R}^n$ sehr viele ganz unterschiedlich geformte unbeschränkte Gebiete gibt und ein solches uneigentliches Integral daher nicht einfach als Integral mit „Obergrenze ∞" aufgefasst werden kann. Wir müssen daher zunächst einmal exakt definieren, was wir darunter verstehen wollen, dass „immer größer werdende beschränkte Bereiche eine gegebene Menge annähern".

Definition 29.42 (Ausschöpfende Folgen)**.** Es sei $D \subset \mathbb{R}^n$ eine beliebige Teilmenge. Eine **ausschöpfende Folge** für D ist eine Folge $D_0 \subset D_1 \subset D_2 \subset \cdots$ messbarer (also insbesondere beschränkter) Teilmengen von D, so dass für alle $r \in \mathbb{R}_{>0}$ gilt:

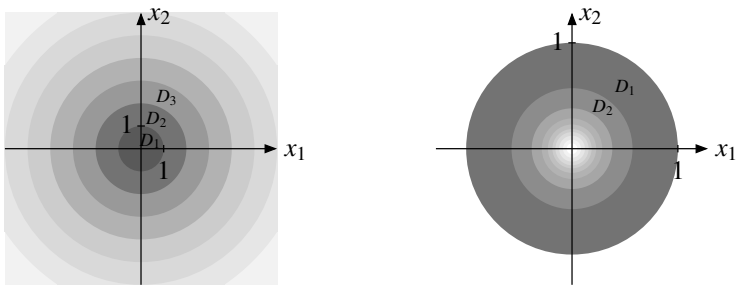
(a) Für alle $k \in \mathbb{N}$ ist die Menge $(D \backslash D_k) \cap K_r(0)$ messbar, wobei $K_r(0)$ wie in Definition 23.13 (a) die euklidische Kugel um 0 mit Radius r bezeichnet.

(b) $\lim\limits_{k\to\infty} \operatorname{vol}\big((D\backslash D_k) \cap K_r(0)\big) = 0$.

Bemerkung 29.43. Die Bedingung (a) in Definition 29.42 ist offensichtlich notwendig, um Bedingung (b) formulieren zu können, also damit das Volumen der Menge $(D\backslash D_k) \cap K_r(0)$ überhaupt definiert ist. Diese zweite Bedingung besagt dann anschaulich, dass in jedem beschränkten Bereich (also in jeder Kugel $K_r(0)$) die Teilmengen D_k der gegebenen Folge die Menge D vom Volumen her beliebig gut annähern. Wie erwartet werden wir dabei in Bemerkung 29.47 (b) noch sehen, dass die Wahl der euklidischen Norm hierbei letztlich keine Rolle spielt.

Beispiel 29.44.

(a) Die Menge $D = \mathbb{R}^n$ wird offensichtlich wie im Bild unten links dargestellt durch die euklidischen Kugeln $D_k = K_k(0)$ ausgeschöpft, denn in diesem Fall gilt $(D\backslash D_k) \cap K_r(0) = \emptyset$ für alle $k \geq r$ und damit natürlich $\lim_{k\to\infty} \operatorname{vol}\big((D\backslash D_k) \cap K_r(0)\big) = 0$ für alle fest gewählten r. Genauso schöpfen auch die Würfel $D_k = [-k,k]^n$ den ganzen $\mathbb{R}^n$ aus, denn auch hier gilt $(D\backslash D_k) \cap K_r(0) = \emptyset$ für alle $k \geq r$.



(a) (b)

(b) Die Menge $D = \{x \in \mathbb{R}^n : 0 < \|x\|_2 \leq 1\}$ wird von den im Bild oben rechts eingezeichneten Teilmengen $D_k = \{x \in \mathbb{R}^n : \frac{1}{k} \leq \|x\|_2 \leq 1\}$ mit $k \geq 1$ ausgeschöpft, denn die Mengen $(D\backslash D_k) \cap K_r(0)$ sind nach Folgerung 29.9 (b) offensichtlich messbar, und ihr Volumen konvergiert wegen

$$(D\backslash D_k) \cap K_r(0) \subset D\backslash D_k = \left\{x \in \mathbb{R}^n : 0 < \|x\|_2 < \frac{1}{k}\right\} \subset \left[-\frac{1}{k}, \frac{1}{k}\right]^n$$

und $\operatorname{vol}([-\frac{1}{k}, \frac{1}{k}])^n = (\frac{2}{k})^n$ mit $k \to \infty$ gegen 0.

(c) Ist D selbst messbar, so ist die konstante Folge mit $D_k = D$ für alle k offensichtlich eine ausschöpfende Folge für D. Insbesondere wäre diese Wahl einer konstanten Folge also in (b) auch möglich gewesen.

Ist andererseits D beschränkt und $D_0 \subset D_1 \subset D_2 \subset \cdots$ eine ausschöpfende Folge für D, so muss D bereits messbar sein: Ist dann nämlich $r \in \mathbb{R}_{>0}$ mit $D \subset K_r(0)$, so ist die Menge $(D \backslash D_k) \cap K_r(0) = D \backslash D_k$ nach Definition 29.42 (a) für alle k messbar, und mit D_k nach Folgerung 29.9 (b) dann auch $(D \backslash D_k) \cup D_k = D$.

Insbesondere sehen wir also, dass nicht jede Teilmenge von $\mathbb{R}^n$ eine ausschöpfende Folge besitzt: Die Menge $\mathbb{Q} \cap [0,1]$ ist nach Beispiel 29.6 (b) beschränkt und nicht messbar, und kann damit auch keine ausschöpfende Folge besitzen.

Definition 29.45 (Uneigentliche Integrale)**.** Es sei $f\colon D \to \mathbb{R}$ eine beliebige Funktion auf einer Teilmenge $D \subset \mathbb{R}^n$.

(a) Es sei zunächst $f \geq 0$. Gibt es dann eine ausschöpfende Folge $D_0 \subset D_1 \subset D_2 \subset \cdots$ für D, so dass f auf jeder Menge D_k (beschränkt und) integrierbar ist, so nennen wir

$$\int_D f(x)\,dx := \lim_{k\to\infty} \int_{D_k} f(x)\,dx \quad \in \mathbb{R}_{\geq 0} \cup \{\infty\}$$

das **uneigentliche Integral** von f auf D – beachte, dass dieser Grenzwert nach Bemerkung 5.43 existiert, da die Folge der Integrale wegen $f \geq 0$ und $D_k \subset D_{k+1}$ für alle k monoton wachsend ist. Wir werden außerdem gleich in Lemma 29.46 sehen, dass diese Definition nicht von der Wahl der ausschöpfenden Folge abhängt. Liegt dieser Grenzwert zusätzlich in $\mathbb{R}$, ist er also nicht gleich ∞, so heißt das uneigentliche Integral $\int_D f(x)\,dx$ **konvergent**, andernfalls **divergent**.

(b) Für beliebiges f betrachten wir den positiven und negativen Anteil f_+ bzw. f_- wie in Folgerung 29.8 (b) und definieren das uneigentliche Integral von f auf D als

$$\int_D f(x)\,dx := \int_D f_+(x)\,dx - \int_D f_-(x)\,dx \quad \in \mathbb{R} \cup \{\pm\infty\},$$

sofern beide uneigentlichen Integrale auf der rechten Seite gemäß (a) existieren und die Differenz nicht von der Form „$\infty - \infty$“ ist. Wie in (a) spricht man auch hier wieder von der Konvergenz dieses Integrals, wenn sein Wert in $\mathbb{R}$ liegt.

Lemma 29.46. *Die Definition 29.45 (a) des uneigentlichen Integrals nicht-negativer Funktionen ist unabhängig von der Wahl der ausschöpfenden Folge.*

Beweis. Es sei $f\colon D \to \mathbb{R}_{\geq 0}$ eine nicht-negative Funktion auf einer Teilmenge $D \subset \mathbb{R}^n$. Weiterhin sei $D_0 \subset D_1 \subset D_2 \subset \cdots$ eine ausschöpfende Folge für D, so dass f auf jeder Menge D_k integrierbar ist. Wir betrachten nun die beiden Werte

$$I := \lim_{k\to\infty} \int_{D_k} f(x)\,dx \qquad \text{und} \qquad S := \sup\left\{ \int_C f(x)\,dx : C \subset D \text{ messbar und } f \text{ integrierbar auf } C \right\},$$

die beide in $\mathbb{R}_{\geq 0} \cup \{\infty\}$ existieren. Offensichtlich genügt es, die Gleichheit $I = S$ zu zeigen, denn dann ist das uneigentliche Integral I ja gleich dem von der ausschöpfenden Folge unabhängigen Supremum S.

$I \leq S$: Für alle k gilt $\int_{D_k} f(x)\,dx \leq S$, da dieses Integral in der Menge enthalten ist, von der für S das Supremum gebildet wird. Grenzwertbildung liefert mit Satz 5.23 (a) also auch $I \leq S$.

$S \leq I$: Es genügt zu zeigen, dass I eine obere Schranke für alle Integrale $\int_C f(x)\,dx$ ist, wobei $C \subset D$ messbar und f auf C (beschränkt und) integrierbar ist – die kleinste obere Schranke S dieser Integrale kann dann nämlich höchstens so groß wie I sein.

Da eine solche Menge C insbesondere beschränkt ist, können wir ein $r \in \mathbb{R}_{>0}$ wählen mit $C \subset K_r(0)$. Dann ist $C \backslash D_k$ messbar mit

$$\operatorname{vol}(C \backslash D_k) = \operatorname{vol}\big((C \backslash D_k) \cap K_r(0)\big) \leq \operatorname{vol}\big((D \backslash D_k) \cap K_r(0)\big) \quad \to 0 \quad \text{für } k \to \infty \qquad (*)$$

nach Definition 29.42 (b). Ist nun c eine obere Schranke für f auf C, so gilt für alle k

$$\int_C f(x)\,dx \le \int_{C\cup D_k} f(x)\,dx = \int_{C\setminus D_k} f(x)\,dx + \int_{D_k} f(x)\,dx \le c\cdot \operatorname{vol}(C\setminus D_k) + \int_{D_k} f(x)\,dx,$$

und daher mit $k\to\infty$ wegen $(*)$ auch wie behauptet

$$\int_C f(x)\,dx \le c\cdot 0 + I = I. \qquad \square$$

Bemerkung 29.47.

(a) Sind $D\subset\mathbb{R}^n$ eine (beschränkte und) messbare Teilmenge und f eine (beschränkte und) integrierbare Funktion auf D, so haben wir den Ausdruck $\int_D f(x)\,dx$ nun zweimal definiert: einmal als gewöhnliches Integral wie in Definition 29.5 (a) und einmal als uneigentliches Integral in Definition 29.45. Die beiden Definitionen stimmen jedoch überein, da man nach Beispiel 29.44 (c) in der Konstruktion des uneigentlichen Integrals dann die konstante ausschöpfende Folge $D\subset D\subset D\subset\cdots$ wählen kann.

(b) Die alternative Darstellung des uneigentlichen Integrals als Supremum S im Beweis von Lemma 29.46 zeigt auch, dass das vielleicht etwas künstlich erscheinende Auftreten der euklidischen Einheitskugeln in der Definition 29.42 der ausschöpfenden Folgen letztlich keine Rolle spielt.

77

Bemerkung 29.48 (Uneigentliche Integrale ohne Aufteilung in f_+ und f_-)**.** Die Formel aus Definition 29.45 (b), mit der wir das uneigentliche Integral beliebiger Funktionen definiert haben, ist in der Praxis etwas unbequem, weil die Integrale über die „künstlich abgeschnittenen“ Funktionen f_+ und f_- in der Regel aufwändiger zu berechnen sind als die über f. Glücklicherweise ist eine solche Aufteilung aber oft nicht nötig: Nehmen wir einmal an, wir haben eine Funktion $f\colon D\to\mathbb{R}$ auf einer Teilmenge $D\subset\mathbb{R}^n$, von der wir bereits wissen, dass das uneigentliche Integral $\int_D f(x)\,dx$ wie in Definition 29.45 existiert. Ferner sei $D_0\subset D_1\subset D_2\subset\cdots$ eine ausschöpfende Folge für D, so dass f auf jedem D_k beschränkt und integrierbar ist.

Dann sind nach Folgerung 29.8 (b) auch f_+ und f_- auf jedem D_k integrierbar, und nach Definition 29.45 (a) gilt

$$\int_D f_+(x)\,dx = \lim_{k\to\infty}\int_{D_k} f_+(x)\,dx \qquad \text{und} \qquad \int_D f_-(x)\,dx = \lim_{k\to\infty}\int_{D_k} f_-(x)\,dx.$$

Damit gilt nach Definition 29.45 (b) also

$$\int_D f(x)\,dx = \lim_{k\to\infty}\int_{D_k} f_+(x)\,dx - \lim_{k\to\infty}\int_{D_k} f_-(x)\,dx = \lim_{k\to\infty}\int_{D_k}(f_+(x)-f_-(x))\,dx = \lim_{k\to\infty}\int_{D_k} f(x)\,dx.$$

Wenn wir also bereits wissen, dass das Integral $\int_D f(x)\,dx$ existiert, können wir seinen Wert auch ohne Aufteilung in f_+ und f_- mit der gleichen Formel wie in Definition 29.45 (a) berechnen. In Beispiel 29.49 (c) werden wir jedoch sehen, dass diese Aussage ohne die Voraussetzung der Existenz des Integrals in der Regel falsch ist.

Beispiel 29.49.

(a) Um das uneigentliche Integral $\int_{\mathbb{R}^2_{\ge 0}} \mathrm{e}^{-x-y}\,d(x,y)$ zu berechnen, können wir z. B. die ausschöpfende Folge $(D_k)_k$ mit $D_k=[0,k]^2$ wählen und erhalten gemäß Definition 29.45 (a)

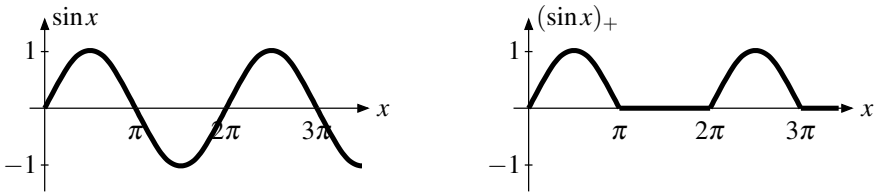
$$\begin{aligned}
\int_{\mathbb{R}^2_{\ge 0}} \mathrm{e}^{-x-y}\,d(x,y) &= \lim_{k\to\infty}\int_{[0,k]^2} \mathrm{e}^{-x-y}\,d(x,y)\\
&\overset{28.19}{=} \lim_{k\to\infty}\left(\int_0^k \mathrm{e}^{-x}\,dx\cdot\int_0^k \mathrm{e}^{-y}\,dy\right)\\
&= \lim_{k\to\infty}\left[-\mathrm{e}^{-x}\right]_0^k\cdot\left[-\mathrm{e}^{-y}\right]_0^k\\
&= \lim_{k\to\infty}(1-\mathrm{e}^{-k})^2\\
&= 1.
\end{aligned}$$

(b) Das Integral $\int_{(0,1)^2} \frac{x}{y}\, d(x,y)$ ist ebenfalls nur als uneigentliches Integral definiert, da die Funktion $\frac{x}{y}$ auf dem Integrationsbereich $(0,1)^2$ unbeschränkt ist. In diesem Fall können wir als ausschöpfende Folge $(D_k)_k$ mit $D_k = (0,1) \times (\frac{1}{k}, 1)$ wählen, da f auf jeder dieser Mengen beschränkt ist. Da der Rand von D_k eine Nullmenge und damit für das Integral unerheblich ist, erhalten wir

$$\begin{aligned}
\int_{(0,1)^2} \frac{x}{y}\, d(x,y) &= \lim_{k\to\infty} \int_{(0,1)\times(\frac{1}{k},1)} \frac{x}{y}\, d(x,y) \\
&= \lim_{k\to\infty} \int_{[0,1]\times[\frac{1}{k},1]} \frac{x}{y}\, d(x,y) \\
&\overset{28.19}{=} \lim_{k\to\infty} \left(\int_0^1 x\,dx \cdot \int_{\frac{1}{k}}^1 \frac{1}{y}\, dy \right) \\
&= \lim_{k\to\infty} \frac{1}{2} \log k \\
&= \infty.
\end{aligned}$$

(c) Wir betrachten das (eindimensionale) Integral $\int_{\mathbb{R}_{\geq 0}} \sin x\, dx$. Wählen wir als ausschöpfende Folge hier $(D_k)_k$ mit $D_k = [0, 2\pi k]$, so ist (siehe Bild unten)

$$\lim_{k\to\infty} \int_{D_k} (\sin x)_+\, dx = \lim_{k\to\infty} k \cdot \int_0^\pi \sin x\, dx = \lim_{k\to\infty} 2k = \infty.$$



Da analog natürlich auch $\int_{\mathbb{R}_{\geq 0}} (\sin x)_-\, dx = \infty$ gilt, existiert das uneigentliche Integral $\int_{\mathbb{R}_{\geq 0}} \sin x\, dx$ gemäß Definition 29.45 (b) also nicht. Allerdings ist

$$\lim_{k\to\infty} \int_{D_k} \sin x\, dx = \lim_{k\to\infty} \int_0^{2\pi k} \sin x\, dx = \lim_{k\to\infty} 0 = 0.$$

Wie schon in Bemerkung 29.48 erwähnt, sehen wir hier also noch einmal explizit, dass die Grenzwertformel wie in Definition 29.45 (a) für Funktionen mit wechselndem Vorzeichen nur angewendet werden darf, wenn die Existenz des Integrals bereits bekannt ist.

Aufgabe 29.50. Zeige, dass das uneigentliche Integral $\int_{\mathbb{R}^n} \mathrm{e}^{-\|x\|_2}\, dx$ für alle $n \in \mathbb{N}_{>0}$ konvergiert.

30. Der Transformationssatz für mehrdimensionale Integrale

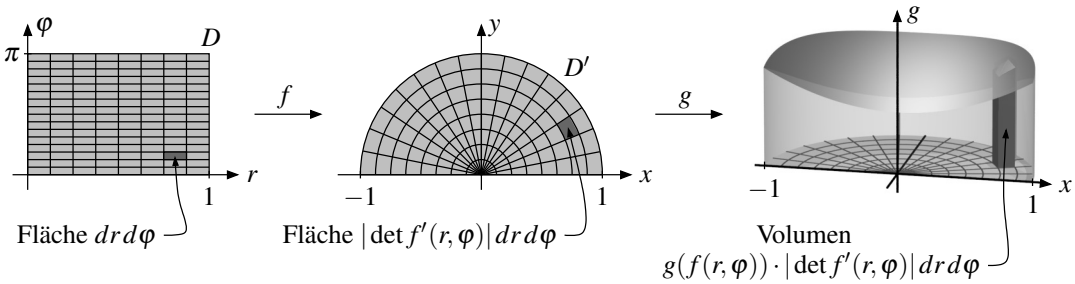
In den letzten beiden Kapiteln haben wir gesehen, wie man mehrdimensionale Integrale über gewisse Funktionen auf einer Teilmenge von $\mathbb{R}^n$ definieren und mit Hilfe des Satzes von Fubini auf n eindimensionale Integrale zurückführen kann. Ein wesentliches Phänomen dabei ist, dass die Integrationsgrenzen der zuerst ausgeführten Integrale dabei wie in Beispiel 29.34 oder Folgerung 29.38 von den späteren Integrationsvariablen abhängen, wenn der Integrationsbereich nicht gerade ein Quader ist. So würde z. B. wie im Bild unten rechts das Integral über eine Funktion $g\colon D' \to \mathbb{R}$ auf der oberen Hälfte des Einheitskreises

$$D' = \left\{(x,y) \in \mathbb{R}^2 : -1 \le x \le 1 \text{ und } 0 \le y \le \sqrt{1-x^2}\right\}$$

zu einem Doppelintegral der Form

$$\int_{-1}^{1} \left(\int_0^{\sqrt{1-x^2}} g(x,y)\,dy \right) dx$$

mit variabler Grenze für y führen. Dieses Einsetzen der variablen Grenzen im y-Integral kann natürlich schnell zur Folge haben, dass der Integrand für das äußere x-Integral zu kompliziert wird, um seine Stammfunktion explizit zu bestimmen.



Es wäre daher schön, wenn wir für die Berechnung des Integrals andere Koordinaten verwenden könnten. So würden sich z. B. für die gerade betrachtete obere Hälfte des Einheitskreises Polarkoordinaten r, φ wie in Satz 9.26 und Beispiel 27.4 anbieten, denn in diesen Koordinaten lässt sich diese Menge D' gerade durch den Quader D mit den Grenzen $0 \le r \le 1$ und $0 \le \varphi \le \pi$ wie im Bild oben links beschreiben – formal ist also $D' = f(D)$ mit der Polarkoordinatenabbildung $f\colon (r,\varphi) \mapsto (x,y) = (r\cos\varphi, r\sin\varphi)$. Könnten wir das Integral auch in diesen Koordinaten berechnen, so ergäbe sich ein Doppelintegral

$$\int_0^1 \left(\int_0^{\pi} \cdots\, d\varphi \right) dr,$$

in dem die Integrationsgrenzen nun konstant sind und die zu integrierende Funktion durch das Einsetzen der inneren Grenzen damit nicht unnötig verkompliziert wird. Wir wollen daher jetzt untersuchen, wie sich mehrdimensionale Integrale unter derartigen Variablentransformationen verhalten, also wie die Integrale über g und $g \circ f$ miteinander zusammenhängen, wenn f eine Koordinatentransformation wie oben ist. Mit anderen Worten wollen wir eine mehrdimensionale Verallgemeinerung der Substitutionsregel aus Satz 12.31 finden.

Aus dem Bild oben lässt sich bereits ablesen, wie eine solche Transformationsformel für Integrale aussehen sollte: Wollen wir das Integral $\int_{D'} g(x,y)\,d(x,y)$, also das Volumen unter dem Graphen von g über D' berechnen, so können wir uns dieses Volumen wie gewohnt als Summe der Volumina von kleinen Säulen wie im Bild oben rechts vorstellen, deren Grundflächen nun aber aufgrund der Variablentransformation keine Quader mehr sind, sondern aus kleinen Quadern mit Seitenlängen dr

und $d\varphi$ wie im linken Bild durch Abbilden mit f entstehen. Wir müssen also herausfinden, wie sich Flächeninhalte (oder in höheren Dimensionen Volumina) unter solchen Koordinatentransformationen ändern – in der Tat ist dies der wesentliche Teil in der Herleitung unserer gewünschten Formel. Dazu werden wir zunächst in Satz 30.7 zeigen, dass sich Volumina von Quadern unter einer *linearen* Abbildung mit dem Betrag der Determinante der zugehörigen Abbildungsmatrix multiplizieren. Ist f keine lineare Abbildung, so müssen wir für einen kleinen Quader stattdessen die entsprechende lineare Approximation nehmen, also die Ableitung $f'(r,\varphi)$ an diesem Punkt. Im Bild oben rechts hat die dunkel eingezeichnete Säule also die Höhe $g(f(r,\varphi))$ (nämlich den Funktionswert von g am betrachteten Punkt) und die Grundfläche $|\det f'(r,\varphi)| \cdot dr\,d\varphi$. Summiert man dies nun auf, so würde man also erwarten, dass sich das Integral über g als

$$\int_{D'} g(x,y)\,d(x,y) = \int_D g(f(r,\varphi)) \cdot |\det f'(r,\varphi)|\,d(r,\varphi)$$

mit $D = [0,1] \times [0,\pi]$ in den neuen Koordinaten berechnen lässt. In der Tat ist dies bereits die gesuchte Transformationsformel (im Fall der obigen Koordinaten (x,y) bzw (r,φ)).

Das Ziel dieses Kapitels ist es, diese anschaulichen Argumente zu einem exakten Beweis zu machen. Der daraus resultierende wichtige Satz 30.12 ist ohne Zweifel die am aufwändigsten zu zeigende Aussage in diesem Skript, und daher ist es beim ersten Durchlesen dieses Kapitels vermutlich sinnvoll, seinen Beweis und seine ganzen Vorbereitungen im folgenden Abschnitt 30.A zunächst einmal zu überspringen und direkt bei der Aussage bzw. den anschließenden Beispielen von Satz 30.12 in Abschnitt 30.B weiterzulesen.

30.A Transformation von Volumina

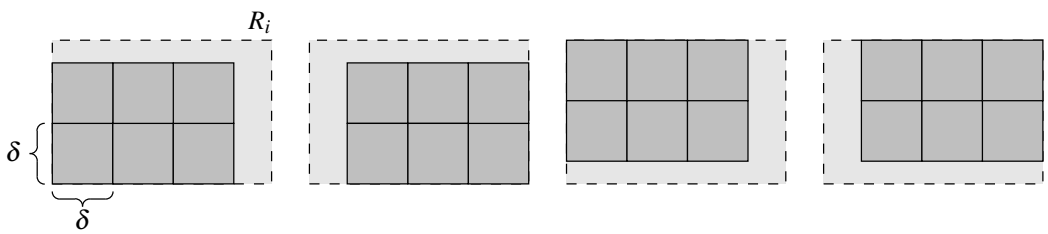
Auf dem Weg zum Beweis des Transformationssatzes beginnen wir mit der ersten oben erwähnten Frage, wie sich Volumina unter Koordinatentransformationen verändern. Leider haben wir dabei gleich noch ein weiteres technisches Problem, da wir ja in Kapitel 29 gesehen haben, dass wir gar nicht jeder beliebigen Menge sinnvoll ein Volumen zuordnen können. Wir müssen daher zunächst erst einmal überprüfen, ob messbare Mengen durch Koordinatentransformationen überhaupt wieder in ebenfalls messbare Mengen umgewandelt werden. Aufgrund unseres Messbarkeitskriteriums aus Folgerung 29.28 können wir dies durch die Untersuchung des Randes auf die Frage zurückführen, ob Nullmengen durch geeignete Koordinatentransformationen wieder in Nullmengen überführt werden. Mit diesem Problem werden wir daher nun unsere Arbeit an dem letztlich gesuchten Transformationssatz für Integrale beginnen. Als Erstes benötigen wir dazu ein weiteres Nullmengenkriterium analog zu Lemma 29.15.

Lemma 30.1. *Eine beschränkte Menge $N \subset \mathbb{R}^n$ ist genau dann eine Nullmenge, wenn es für alle $\varepsilon > 0$ endlich viele Würfel $Q_1,\dots,Q_k$ (also Quader, deren Seitenlängen alle gleich sind) gibt mit*

$$N \subset Q_1 \cup \dots \cup Q_k \quad \textit{und} \quad \sum_{i=1}^k \mathrm{vol}(Q_i) < \varepsilon.$$

Beweis. Gibt es stets solche Würfel, so ist N nach Lemma 29.15 natürlich eine Nullmenge.

Für die umgekehrte Richtung sei nun N eine Nullmenge. Nach Lemma 29.15 gibt es dann endlich viele Quader $R_1,\dots,R_l$ mit $\sum_{i=1}^l \mathrm{vol}(R_i) < \frac{\varepsilon}{2^n}$, die N überdecken. Wir wählen ein δ kleiner als die minimale Seitenlänge aller dieser Quader, und füllen jeden Quader R_i wie im Bild unten von jeder seiner 2^n Ecken kommend mit Würfeln der Kantenlänge δ auf.



Da die in R_i von einer Ecke kommenden Würfel ein Gesamtvolumen von höchstens $\operatorname{vol}(R_i)$ haben, haben alle diese Würfel zusammen nun ein Gesamtvolumen von höchstens $\sum_{i=1}^{l} 2^n \operatorname{vol}(R_i) < \varepsilon$. Außerdem überdecken die von einer Ecke kommenden Würfel den Quader R_i mindestens bis zu seinem Mittelpunkt. Daher überdecken alle Würfel zusammen genommen alle Quader R_i, und damit auch N. □ 78

Satz 30.2 (Bilder kompakter Nullmengen)**.** *Es seien $D \subset \mathbb{R}^n$ offen und $f\colon D \to \mathbb{R}^n$ eine stetig differenzierbare Funktion. Ist $N \subset D$ dann eine kompakte Nullmenge, so auch $f(N)$.*

Beweis. Nach Satz 24.29 ist $f(N)$ natürlich kompakt. Für die Nullmengeneigenschaft unterscheiden wir zwei Fälle:

1. Fall: N ist in einem Quader $Q \subset D$ enthalten. Da f stetig differenzierbar ist, existiert dann auf diesem kompakten Quader nach Folgerung 24.34 das Maximum der Zeilensummennorm

$$c := \|f'|_Q\|_\infty := \max\left\{\sum_{j=1}^{n} |\partial_j f_i(x)| : i \in \{1,\dots,n\}, x \in Q\right\}$$

von f'. Ist nun $\varepsilon > 0$ gegeben, so gibt es nach Lemma 30.1 Würfel $Q_1,\dots,Q_k$ mit $\sum_{i=1}^{k} \operatorname{vol}(Q_i) < \frac{\varepsilon}{c^n}$, die N überdecken, und von denen wir annehmen können, dass sie in Q liegen. Wir können sie nach Beispiel 23.3 (b) als abgeschlossene Kugeln

$$Q_i = \{x \in \mathbb{R}^n : \|x - a_i\|_\infty \le \delta_i\}$$

in der Maximumsnorm mit geeigneten Mittelpunkten $a_1,\dots,a_k$ und Radien $\delta_1,\dots,\delta_k$ schreiben. Dann gilt nach der Folgerung 26.19 aus dem Mittelwertsatz aber für alle $x \in Q_i$, dass $f(x) \in R_i$ mit

$$R_i = \{y \in \mathbb{R}^n : \|y - f(a_i)\|_\infty \le c\delta_i\}$$

(hierfür brauchen wir unsere Voraussetzung $N \subset Q \subset R$, da Folgerung 26.19 benötigt, dass mit a_i und x auch die gesamte Verbindungsstrecke $\overline{a_i x}$ in der betrachteten Menge liegt – was für die Nullmenge N nicht notwendig der Fall wäre). Damit wird $f(N)$ überdeckt durch die Würfel R_i mit Gesamtvolumen

$$\sum_{i=1}^{k} \operatorname{vol}(R_i) = \sum_{i=1}^{k} (2c\delta_i)^n = \sum_{i=1}^{k} c^n \operatorname{vol}(Q_i) < \varepsilon,$$

d. h. $f(N)$ ist eine Nullmenge.

2. Fall: N ist beliebig. Da D offen ist, können wir dann um jeden Punkt $a \in N$ eine offene Umgebung $U_{\delta_a}(a) \subset D$ in der Maximumsnorm finden. Betrachten wir die zugehörigen Würfel $Q_a = K_{\delta_a/2}(a)$ mit halber Kantenlänge, so liegen diese also auch in D, und die Mengen $\overset{\circ}{Q}_a$ ihrer inneren Punkte überdecken N. Wegen der vorausgesetzten Kompaktheit von N können wir daraus dann nach Satz 23.58 endlich viele Quader $Q_{a_1},\dots,Q_{a_m}$ auswählen, die auch schon N überdecken. Also ist dann

$$N = (N \cap Q_{a_1}) \cup \dots \cup (N \cap Q_{a_m}).$$

Da jede dieser Mengen $N \cap Q_{a_i}$ nun eine Nullmenge ist, die noch ganz in einem Quader $Q_{a_i} \subset D$ enthalten ist, ist nach dem 1. Fall also jede Menge $f(N \cap Q_{a_i})$ eine Nullmenge, und damit nach Bemerkung 29.10 (c) auch $f(N)$. □

Bemerkung 30.3. Die folgenden Beispiele zeigen, dass die in Satz 30.2 gemachten Voraussetzungen wirklich wesentlich sind: Die Aussage wird falsch, wenn …

(a) Start- und Zielbereich von f nicht die gleiche Dimension haben: Die Projektion $f\colon \mathbb{R}^2 \to \mathbb{R}$ auf die erste Koordinate bildet die kompakte Nullmenge $[0,1] \times \{0\}$ auf die Menge $[0,1] \subset \mathbb{R}$ mit Volumen 1 ab.

(b) f nur stetig, aber nicht stetig differenzierbar ist: Mit der Peano-Kurve aus Satz 24.43 erhält man eine stetige Abbildung $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, die die kompakte Nullmenge $[0,1] \times \{0\}$ auf das Einheitsquadrat $[0,1] \times [0,1]$ mit Volumen 1 abbildet.

(c) N nicht kompakt ist: Wir betrachten die Menge $N = \{\frac{1}{n} : n \in \mathbb{N}_{>0}\}$ und die stetig differenzierbare Abbildung $f\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$, $x \mapsto \frac{1}{x}$. Dann ist N nach Aufgabe 29.19 eine Nullmenge, aber $f(N) = \mathbb{N}_{>0}$ kann als unbeschränkte Menge natürlich keine Nullmenge sein.

Von unseren Koordinatentransformationen wollen wir im Folgenden voraussetzen, dass sie nicht nur stetig differenzierbar, sondern auch bijektiv mit stetig differenzierbarer Umkehrabbildung sind. Derartige Abbildungen werden als Diffeomorphismen bezeichnet – was man sich als Verschmelzung der Worte „differenzierbar“ und „Isomorphismus“ vorstellen sollte.

Definition 30.4 (Diffeomorphismen)**.** Eine Abbildung $f\colon D \to D'$ zwischen offenen Teilmengen D und D' von $\mathbb{R}^n$ heißt **Diffeomorphismus**, wenn f bijektiv ist, und sowohl f als auch die Umkehrabbildung $f^{-1}\colon D' \to D$ stetig differenzierbar sind.

Folgerung 30.5. *Es sei $f\colon D \to D'$ ein Diffeomorphismus zwischen offenen Teilmengen D und D' von $\mathbb{R}^n$. Dann gilt für jede kompakte Teilmenge $K \subset D$:*

(a) $f(\partial K) = \partial f(K)$.

(b) *Ist K messbar, so auch $f(K)$.*

Beweis.

(a) „$\subset$“: Es seien $a \in \partial K$ und $U \subset D'$ eine Umgebung von $f(a)$. Da K kompakt und damit nach Satz 23.51 (a) auch abgeschlossen ist, gilt dann auch $a \in K$. Nun ist $f^{-1}(U) \subset D$ nach Lemma 24.16 aber eine Umgebung von a und muss wegen $a \in K$ damit auch einen Punkt $b \notin K$ enthalten. Wenden wir darauf die bijektive Funktion f an, so erhalten wir die Punkte $f(a) \in f(K)$ und $f(b) \notin f(K)$ in U. Damit ist $f(a)$ ein Randpunkt von $f(K)$, d. h. es ist $f(\partial K) \subset \partial f(K)$.

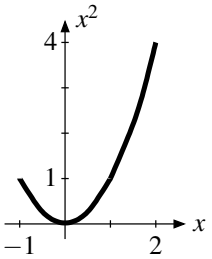
Die Richtung „$\supset$“ ergibt sich nun einfach durch Anwendung dieses Resultats auf die stetig differenzierbare Abbildung f^{-1}: Die Menge $K' := f(K)$ ist nach Satz 24.29 ebenfalls kompakt, also gilt nach der obigen Rechnung $f^{-1}(\partial K') \subset \partial f^{-1}(K')$, d. h. $f^{-1}(\partial f(K)) \subset \partial K$ und damit durch Anwenden von f auch $\partial f(K) \subset f(\partial K)$.

(b) Ist K messbar, so ist ∂K nach Folgerung 29.28 eine Nullmenge. Außerdem ist K nach Voraussetzung kompakt und damit nach Satz 23.51 (a) beschränkt, so dass ∂K nach Lemma 23.40 (e) und Satz 23.51 (b) ebenfalls abgeschlossen und beschränkt und damit kompakt ist. Also ist $f(\partial K)$ nach Satz 30.2 eine Nullmenge. Wegen (a) ist damit auch $\partial f(K)$ eine Nullmenge, woraus sich wieder mit Folgerung 29.28 die Messbarkeit von $f(K)$ ergibt. □

Bemerkung 30.6. Die Gleichung aus Folgerung 30.5 (a) sieht zwar sehr natürlich aus, dennoch sind von ihr für allgemeine (nicht notwendig bijektive) stetig differenzierbare Funktionen im Allgemeinen sogar beide Inklusionen falsch: Betrachten wir z. B. die Abbildung $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ und die kompakte Menge $K = [-1,2]$ mit Bild $f(K) = [0,4]$, so ist

$$f(\partial K) = f(\{-1,2\}) = \{1,4\} \quad \text{und} \quad \partial f(K) = \partial[0,4] = \{0,4\},$$

so dass hier keine dieser beiden Mengen in der anderen enthalten ist.



In Folgerung 30.5 (b) haben wir nun also gesehen, dass für einen Diffeomorphismus $f\colon D \to D'$ und eine kompakte messbare Menge $K \subset D$ auch $f(K) \subset D'$ messbar ist, so dass also das Volumen von $f(K)$ definiert ist und wir es mit dem von K vergleichen können. Wie bereits in der Einleitung zu diesem Kapitel erwähnt, beginnen wir dabei der Einfachheit halber zunächst mit dem Fall einer *linearen* Abbildung f.

Satz 30.7 (Volumenformel für lineare Isomorphismen)**.** *Es sei $f\colon \mathbb{R}^n \to \mathbb{R}^n$ ein linearer Isomorphismus, also $f\colon x \mapsto Ax$ für eine invertierbare Matrix $A \in \mathrm{GL}(n,\mathbb{R})$. Dann gilt für jede kompakte messbare Menge $K \subset \mathbb{R}^n$*

$$\mathrm{vol}(f(K)) = |\det A| \cdot \mathrm{vol}(K).$$

Beweis. Wir machen zunächst drei Reduktionsschritte, um den Satz auf eine letztlich sehr einfach zu zeigende Aussage zurückzuführen:

(1) Es genügt, die Ungleichung $\operatorname{vol}(f(K)) \leq |\det A| \cdot \operatorname{vol}(K)$ zu zeigen: Wenden wir diese Aussage dann nämlich auf die Umkehrfunktion $f^{-1}\colon x \mapsto A^{-1}x$ und die nach Satz 24.29 und Folgerung 30.5 ebenfalls kompakte und messbare Menge $f(K)$ an, so erhalten wir

$$\operatorname{vol}(K) = \operatorname{vol}(f^{-1}(f(K))) \leq |\det A^{-1}| \cdot \operatorname{vol}(f(K)),$$

nach Satz 18.6 (b) also auch die umgekehrte Ungleichung $\operatorname{vol}(f(K)) \geq |\det A| \cdot \operatorname{vol}(K)$.

(2) Es genügt auch, diese Ungleichung für Quader zu zeigen: Ist $K \subset \mathbb{R}^n$ dann eine allgemeine kompakte messbare Menge, so gibt es nach dem Riemannschen Integrabilitätskriterium aus Lemma 28.10 (b) für alle $\varepsilon \in \mathbb{R}_{>0}$ eine Zerlegung Z eines Quaders, der K enthält, mit

$$\operatorname{OS}(1_K, Z) = \sum_{Q \in \operatorname{TQ}(Z)} \operatorname{vol}(Q) \cdot \sup 1_K(Q) = \sum_{Q \in \operatorname{TQ}(Z) : Q \cap K \neq \emptyset} \operatorname{vol}(Q) < \operatorname{vol}(K) + \varepsilon.$$

Da diese Quader $Q \in \operatorname{TQ}(Z)$ mit $Q \cap K \neq \emptyset$ die Menge K überdecken, überdecken die Mengen $f(Q)$ für diese Quader aber auch die Menge $f(K)$, und wir erhalten aus Bemerkung 29.10 unter Verwendung der zu zeigenden Aussage für Quader

$$\operatorname{vol}(f(K)) \leq \sum_{Q \cap K \neq \emptyset} \operatorname{vol}(f(Q)) \leq \sum_{Q \cap K \neq \emptyset} |\det A| \cdot \operatorname{vol}(Q) < |\det A| \cdot (\operatorname{vol}(K) + \varepsilon).$$

Nehmen wir hier nun den Grenzwert für $\varepsilon \to 0$, so erhalten wir die zu zeigende Ungleichung $\operatorname{vol}(f(K)) \leq |\det A| \cdot \operatorname{vol}(K)$ auch für K.

(3) Weiterhin genügt es, die Aussage in dem Fall zu zeigen, wenn die Abbildungsmatrix A von f eine Elementarmatrix wie in Konstruktion 17.3 ist: Nach Lemma 17.17 ist jede invertierbare Matrix nämlich ein Produkt von solchen Elementarmatrizen, und wenn die zu zeigende Ungleichung für zwei Isomorphismen $f\colon x \mapsto Ax$ und $g\colon x \mapsto Bx$ gilt, dann auch für ihre Verkettung $g \circ f$ bzw. das Produkt BA wegen

$$\operatorname{vol}(g(f(K))) \leq |\det B| \cdot \operatorname{vol}(f(K)) \leq |\det B| \cdot |\det A| \cdot \operatorname{vol}(K) = |\det(BA)| \cdot \operatorname{vol}(K).$$

Es sei nun nach (2) also $K = Q = [a,b] \subset \mathbb{R}^n$ ein Quader; wir untersuchen die beiden Typen von Elementarmatrizen nach (3) gemäß Konstruktion 17.3 separat:

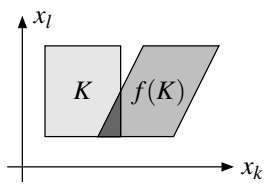
(a) Für die Elementarmatrix $A = F_k(\lambda)$ aus Konstruktion 17.3 (a), die die k-te Koordinate um den Faktor λ streckt, ist

$$\begin{aligned} f(Q) &= [a_1,b_1] \times \cdots \times [\lambda a_k, \lambda b_k] \times \cdots \times [a_n,b_n] \quad \text{für } \lambda > 0 \\ \text{bzw.} \quad f(Q) &= [a_1,b_1] \times \cdots \times [\lambda b_k, \lambda a_k] \times \cdots \times [a_n,b_n] \quad \text{für } \lambda < 0 \end{aligned}$$

und damit in beiden Fällen $\operatorname{vol}(f(Q)) = |\lambda| \cdot \operatorname{vol}(Q) = |\det A| \cdot \operatorname{vol}(Q)$.

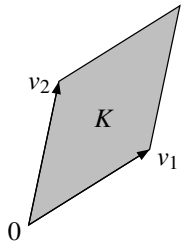
(b) Für die Elementarmatrix $A = F_{k,l}(\lambda)$ aus Konstruktion 17.3 (b), die das λ-fache der l-ten Koordinate auf die k-te addiert, ist $f(Q)$ ein Parallelotop wie im Bild rechts. Nach dem Prinzip von Cavalieri aus Beispiel 29.40 hat dieses das gleiche Volumen wie der Quader Q, und wir erhalten wie gewünscht

$$\operatorname{vol}(f(Q)) = 1 \cdot \operatorname{vol}(Q) = |\det A| \cdot \operatorname{vol}(Q). \qquad \square$$



Beispiel 30.8.

(a) Es sei K wie im Bild rechts ein von n linear unabhängigen Vektoren $v_1,\dots,v_n \in \mathbb{R}^n$ aufgespanntes Parallelotop
$$K = \{\lambda_1 v_1 + \dots + \lambda_n v_n : 0 \le \lambda_1,\dots,\lambda_n \le 1\}.$$
Offensichtlich ist K dann das Bild des Einheitswürfels
$$Q = \{\lambda_1 e_1 + \dots + \lambda_n e_n : 0 \le \lambda_1,\dots,\lambda_n \le 1\}$$
unter der linearen Abbildung mit Abbildungsmatrix $A = (\,v_1 \mid \cdots \mid v_n\,)$, die für alle $i = 1,\dots,n$ den i-ten Einheitsvektor e_i auf v_i abbildet. Nach Satz 30.7 gilt damit also $\operatorname{vol}(K) = |\det A| \cdot \operatorname{vol}(Q) = |\det A|$.



(b) Eine kompakte messbare Menge $K \subset \mathbb{R}^n$ wird durch die Streckung $f\colon \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto \lambda x$ um den Faktor $\lambda \in \mathbb{R}_{>0}$ nach Satz 30.7 auf die Menge $f(K) = \lambda K$ mit dem Volumen $\operatorname{vol}(f(K)) = \lambda^n \operatorname{vol}(K)$ abgebildet, da die Determinante der zugehörigen Abbildungsmatrix λE gleich λ^n ist.

(c) Ist $f\colon \mathbb{R}^n \to \mathbb{R}^n$ eine (bezüglich des Standardskalarprodukts) orthogonale Abbildung, also wie in Bemerkung 22.3 z. B. eine Drehung oder Spiegelung, so wird f nach Lemma 22.4 (a) durch eine orthogonale Matrix A beschrieben, deren Determinante nach Lemma 22.8 (a) Betrag 1 hat. In diesem Fall erhalten wir aus Satz 30.7 also für jede kompakte messbare Menge die anschaulich einleuchtende Aussage $\operatorname{vol}(f(K)) = \operatorname{vol}(K)$, also dass das Volumen unter Drehungen und Spiegelungen unverändert bleibt.

79

Im Rest dieses Abschnitts wollen wir diese Volumentransformationsformel nun noch auf den Fall eines Diffeomorphismus f verallgemeinern. Dazu müssen wir f wie in Kapitel 25 lokal an jedem Punkt durch die lineare Abbildung mit Abbildungsmatrix f' approximieren, wodurch die reine Multiplikation mit $|\det A|$ in Satz 30.7 zu einem Integral über $|\det f'(x)|$ wird. Wie im Reduktionsschritt (1) des obigen Satzes zeigen wir dabei zunächst nur eine Ungleichung – durch Anwenden auf die Umkehrfunktion werden wir in Bemerkung 30.13 (a) dann wieder sehen, dass sogar die Gleichheit gilt. Wir nähern uns der Aussage zunächst wieder über Würfel bzw. Quader und eine Hilfsabschätzung durch das Maximum von $|\det f'(x)|$ statt durch das Integral.

Lemma 30.9. *Es sei $f\colon D \to D'$ ein Diffeomorphismus zwischen offenen Teilmengen D und D' von $\mathbb{R}^n$. Dann gilt für jeden Würfel $Q \subset D$*
$$\operatorname{vol}(f(Q)) \le \max\{|\det f'(x)| : x \in Q\} \cdot \operatorname{vol}(Q).$$

Beweis. Wir verwenden in diesem Beweis durchgängig die Maximumsnorm für Vektoren in $\mathbb{R}^n$ und die davon induzierte Zeilensummennorm wie in Beispiel 24.39 (b) für Matrizen (wie z. B. die Ableitung $f'(x)$).

Es sei $\varepsilon \in \mathbb{R}_{>0}$ beliebig. Da f' auf dem kompakten Quader Q stetig und invertierbar ist, existiert das Maximum $M := \max\{\|f'(x)^{-1}\| : x \in Q\}$. Gleichzeitig ist f' nach Bemerkung 24.25 (a) auf Q auch gleichmäßig stetig, und somit gibt es ein $\delta \in \mathbb{R}_{>0}$ mit $\|f'(x) - f'(a)\| < \frac{\varepsilon}{M}$ für alle $a, x \in Q$ mit $\|x - a\| < \delta$.

Wir betrachten nun zunächst einen beliebigen Würfel $W = K_d(a) \subset Q$ mit einem Radius $d < \delta$. Wie in der Definition der Differenzierbarkeit in Kapitel 25 können wir dann für alle $x \in W$
$$f(x) = f(a) + f'(a)\cdot(x-a) + r(x) = f(a) + f'(a)\cdot\underbrace{\left(x - a + f'(a)^{-1}\cdot r(x)\right)}_{(2)} \tag{1}$$
mit einer Restfunktion r schreiben, für die wir aus der obigen Gleichung sofort $r'(x) = f'(x) - f'(a)$ und $r(a) = 0$ ablesen. Dabei können wir den Ausdruck (2) mit Hilfe des Mittelwertsatzes wie folgt

abschätzen:

$$\begin{aligned}
\|x-a+f'(a)^{-1}\cdot r(x)\| &\le \|x-a\| + \|f'(a)^{-1}\|\cdot\|r(x)\| && \text{(Folgerung 24.38)}\\
&\le \underbrace{\|x-a\|}_{\le d} + \underbrace{\|f'(a)^{-1}\|}_{\le M}\cdot\underbrace{\|r'|_{\overline{ax}}\|}_{\le \frac{\varepsilon}{M}}\cdot\underbrace{\|x-a\|}_{\le d} && \text{(Folgerung 26.19)}\\
&\le (1+\varepsilon)\,d.
\end{aligned}$$

Der Ausdruck (2) liegt also für alle $x \in W$ im Würfel $K_{(1+\varepsilon)d}(0)$ mit Volumen $(1+\varepsilon)^n \operatorname{vol}(W)$. Dieser Würfel wird in (1) durch die Multiplikation mit $f'(a)$ nach Satz 30.7 auf eine kompakte messbare Menge mit Volumen $|\det f'(a)|\cdot(1+\varepsilon)^n \operatorname{vol}(W)$ abgebildet. Da die Translation um $f(a)$ in (1) daran nichts ändert, ist also

$$\operatorname{vol}(f(W)) \le |\det f'(a)|\cdot(1+\varepsilon)^n \operatorname{vol}(W) \le \max\{|\det f'(x)| : x \in Q\}\cdot(1+\varepsilon)^n \operatorname{vol}(W).$$

Zerlegen wir nun den Gesamtwürfel Q in Teilwürfel W mit Radius $d < \delta$ und addieren diese Ungleichung für alle diese Teilwürfel auf, so erhalten wir daraus sofort

$$\operatorname{vol}(f(Q)) \le \max\{|\det f'(x)| : x \in Q\}\cdot(1+\varepsilon)^n \operatorname{vol}(Q),$$

woraus durch Grenzwertbildung $\varepsilon \to 0$ die Behauptung folgt. □

Bemerkung 30.10. Lemma 30.9 gilt in der Tat nicht nur für Würfel, sondern auch für einen beliebigen Quader $Q = [a,b] = [a_1,b_1]\times\cdots\times[a_n,b_n]$: Ein solcher Quader ist nämlich das Bild des Einheitswürfels $W = [0,1]^n$ unter dem Diffeomorphismus

$$g\colon \mathbb{R}^n \to \mathbb{R}^n,\ x \mapsto a + \operatorname{diag}(b_1-a_1,\dots,b_n-a_n)\cdot x$$

mit konstanter Ableitung $g' = \operatorname{diag}(b_1-a_1,\dots,b_n-a_n)$ und demnach $\det g' = \operatorname{vol}(Q)$, und damit liefert Lemma 30.9 angewendet auf den Diffeomorphismus $f\circ g$ und den Einheitswürfel W

$$\begin{aligned}
\operatorname{vol}(f(Q)) &= \operatorname{vol}((f\circ g)(W))\\
&\le \max\{|\det(f\circ g)'(x)| : x \in W\} && \text{(Lemma 30.9)}\\
&= \max\{|\det f'(g(x))|\cdot|\det g'(x)| : x \in W\} && \text{(Kettenregel, Satz 25.29)}\\
&= \max\{|\det f'(y)| : y \in Q\}\cdot\operatorname{vol}(Q) && (\det g' = \operatorname{vol}(Q)).
\end{aligned}$$

Satz 30.11 (Volumenformel für Diffeomorphismen)**.** *Es seien $f\colon D \to D'$ ein Diffeomorphismus zwischen offenen Teilmengen D und D' von $\mathbb{R}^n$ sowie $K \subset D$ eine kompakte messbare Menge. Dann gilt*

$$\operatorname{vol}(f(K)) \le \int_K |\det f'(x)|\,dx$$

(wobei wir in Bemerkung 30.13 (a) noch sehen werden, dass hier sogar die Gleichheit gilt).

Beweis. Da K kompakt und D offen ist, existiert nach Aufgabe 24.41 (b) bezüglich der Maximumsnorm der Abstand $\varepsilon := d(K,\mathbb{R}^n\setminus D) > 0$ von K zum Komplement von D. Die Menge $L := K + K_{\varepsilon/2}(0)$ liegt dann also ganz in D und ist nach Aufgabe 23.61 ebenfalls kompakt.

Es sei nun Z eine beliebige Zerlegung eines Quaders $[a,b] \supset K$, so dass die Seitenlängen aller Teilquader kleiner als $\frac{\varepsilon}{2}$ sind. Alle Teilquader $Q \in \operatorname{TQ}(Z)$ mit $Q \cap K \neq \emptyset$ liegen dann nach Konstruktion

in L und damit auch in D, so dass f dort also definiert ist. Damit gilt

$$\begin{aligned}
\operatorname{vol}(f(K)) &\le \sum_{Q\in \mathrm{TQ}(Z):Q\cap K\neq\emptyset} \operatorname{vol}(f(Q)) && \text{(diese } f(Q) \text{ überdecken } f(K)\text{)}\\
&\le \sum_{Q\in \mathrm{TQ}(Z):Q\cap K\neq\emptyset} \max\{|\det f'(x)| : x\in Q\}\cdot \operatorname{vol}(Q) && \text{(Bemerkung 30.10)}\\
&\overset{(1)}{\le} \sum_{Q\cap K\neq\emptyset} \max\{|\det f'(x)| : x\in Q\cap K\}\cdot \operatorname{vol}(Q)\\
&\quad + \sum_{Q\cap K\neq\emptyset,\, Q\cap(\mathbb{R}^n\setminus K)\neq\emptyset} \underbrace{\max\{|\det f'(x)| : x\in Q\}}_{\le \max\{|\det f'(x)| : x\in L\} =: c}\cdot \operatorname{vol}(Q)\\
&\overset{(2)}{\le} \mathrm{OS}(|\det f'|_K, Z) + c\cdot \mathrm{OS}(1_{\partial K}, Z).
\end{aligned}$$

Dabei haben wir in (1) in der ersten Zeile das Maximum zunächst nur über alle $x\in Q\cap K$ laufen lassen, um exakt die Obersumme der Einschränkung von $|\det f'|$ auf K zu erhalten. Für alle Teilquader $Q\in \mathrm{TQ}(Z)$ mit $Q\subset K$ ändert dies natürlich nichts; alle anderen haben wir in der zweiten Zeile unverändert kopiert, damit die Ungleichung wieder stimmt. Der zweite Summand in (2) ergibt sich daraus, dass jeder Teilquader $Q\in \mathrm{TQ}(Z)$, der sowohl K als auch $\mathbb{R}^n\setminus K$ schneidet, nach Aufgabe 24.23 (a) einen Punkt in ∂K enthalten muss und damit in der Obersumme der Indikatorfunktion $1_{\partial K}$ auftritt.

Bilden wir in dieser Ungleichung nun das Infimum über alle Zerlegungen, so erhalten wir wie behauptet

$$\operatorname{vol}(f(K)) \le \int_K |\det f'(x)|\,dx + c\cdot \operatorname{vol}(\partial K) \overset{29.28}{=} \int_K |\det f'(x)|\,dx. \qquad \square$$

30.B Transformation von Integralen

Wir haben jetzt alle notwendigen Vorarbeiten geleistet, um den in der Einleitung zu diesem Kapitel angekündigten Transformationssatz für mehrdimensionale Integrale zu beweisen.

Satz 30.12 (**Transformationssatz** für Integrale)**.** *Es seien D und D' beschränkte offene Teilmengen von $\mathbb{R}^n$, $f\colon D\to D'$ ein Diffeomorphismus und $g\colon D'\to\mathbb{R}$ eine Funktion, so dass sowohl g auf D' als auch $(g\circ f)\cdot|\det f'|$ auf D (beschränkt und) integrierbar sind. Dann gilt*

$$\int_{D'} g(y)\,dy = \int_D g(f(x))\cdot|\det f'(x)|\,dx.$$

Beweis. Zunächst einmal genügt es, den Satz für nicht-negative Funktionen $g\colon D'\to\mathbb{R}_{\ge 0}$ zu zeigen: Nach Folgerung 29.8 (b) können wir ihn im allgemeinen Fall dann nämlich separat auf die ebenfalls integrierbaren Funktionen g_+ und g_- anwenden und erhalten durch Subtraktion dieser beiden Transformationsformeln die gewünschte Aussage auch für g.

Außerdem reicht es wieder wie im Beweis von Satz 30.7, die Ungleichung

$$\int_{D'} g(y)\,dy \le \int_D g(f(x))\cdot|\det f'(x)|\,dx$$

zu zeigen: Anwenden dieser Ungleichung auf den Diffeomorphismus $f^{-1}\colon D'\to D$ und die Funktion $D\to\mathbb{R}$, $(g\circ f)\cdot|\det f'|$ liefert dann nämlich auch die andere Ungleichung

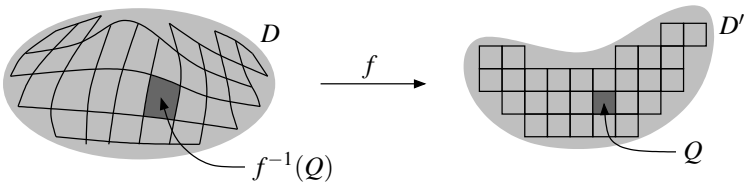
$$\int_D g(f(x))\cdot|\det f'(x)|\,dx \le \int_{D'} (g\circ f)(f^{-1}(y))\cdot|\det f'(f^{-1}(y))|\cdot|\det(f^{-1})'(y)|\,dy = \int_{D'} g(y)\,dy,$$

da $f'(f^{-1}(y))\cdot(f^{-1})'(y) = (f\circ f^{-1})'(y) = \mathrm{id}'(y) = E$ nach der Kettenregel aus Satz 25.29 gilt.

Dazu sei nun Z eine Zerlegung eines Quaders, der D' enthält. Dann folgt mit der Funktion $g_{D'}$ wie in Definition 29.1 (a)

$$\begin{aligned}
\mathrm{US}(g_{D'},Z) &= \sum_{Q\in\mathrm{TQ}(Z)} \inf g_{D'}(Q)\cdot\mathrm{vol}(Q)\\
&= \sum_{Q\in\mathrm{TQ}(Z):Q\subset D'} \inf g(Q)\cdot\mathrm{vol}(Q)\\
&= \sum_{Q\in\mathrm{TQ}(Z):Q\subset D'} \inf g(f(f^{-1}(Q)))\cdot\mathrm{vol}(f(f^{-1}(Q)))\\
&\le \sum_{Q\in\mathrm{TQ}(Z):Q\subset D'} \inf g(f(f^{-1}(Q)))\cdot\int_{f^{-1}(Q)}|\det f'(x)|\,dx \qquad \text{(Satz 30.11)}\\
&\le \sum_{Q\in\mathrm{TQ}(Z):Q\subset D'} \int_{f^{-1}(Q)} g(f(x))\cdot|\det f'(x)|\,dx \qquad \text{(Folgerung 29.7 (c))}\\
&\overset{(*)}{\le} \int_D g(f(x))\cdot|\det f'(x)|\,dx.
\end{aligned}$$

Dabei zeigt das Bild unten rechts die oben in den Summen auftretenden Teilquader Q der Zerlegung Z, die ganz in D' liegen und auf denen somit das Infimum von $g_{D'}$ größer als 0 sein kann. Beachte außerdem für $(*)$, dass für zwei solche Teilquader $Q,Q'\in\mathrm{TQ}(Z)$ mit $Q\cap Q'$ nach Satz 30.2 auch der Durchschnitt $f^{-1}(Q)\cap f^{-1}(Q')=f^{-1}(Q\cap Q')$ eine Nullmenge ist und die behauptete Ungleichung somit aus der Additivität des Integrals gemäß Folgerung 29.9 (a) folgt, da sich die Mengen $f^{-1}(Q)$ für die betrachteten Teilquader Q bis auf diese Nullmengen wie im Bild unten links disjunkt zu einer Teilmenge von D vereinigen.



Die rechte Seite der obigen Ungleichungskette ist also eine obere Schranke für alle betrachteten Untersummen von $g_{D'}$. Für das Supremum dieser Untersummen, also das gesuchte Integral, folgt somit wie behauptet

$$\int_{D'} g(y)\,dy \le \int_D g(f(x))\cdot|\det f'(x)|\,dx. \qquad \square$$

Bemerkung 30.13.

(a) Es seien $f\colon D\to D'$ ein Diffeomorphismus wie in Satz 30.12 sowie $K\subset D$ eine kompakte messbare Menge. Wenden wir den Transformationssatz dann auf die Indikatorfunktion $1_{f(K)}$ der nach Folgerung 30.5 (b) ebenfalls messbaren Menge $f(K)$ an, so erhalten wir die Volumentransformationsformel für Diffeomorphismen

$$\mathrm{vol}(f(K)) = \int_{D'} 1_{f(K)}(y)\,dy = \int_D 1_{f(K)}(f(x))\cdot|\det f'(x)|\,dx = \int_K |\det f'(x)|\,dx$$

und damit die Aussage, dass in Satz 30.11 sogar die Gleichheit gilt.

(b) Man kann zeigen, dass die Aussage von Satz 30.12 bereits gilt, wenn nur die Ausgangsfunktion g als integrierbar vorausgesetzt wird – die Integrierbarkeit der transformierten Funktion $(g\circ f)\cdot|\det f'|$ ist dann automatisch. Wir werden dies hier jedoch nicht beweisen, da dies noch etwas aufwändiger wäre und sich die Integrierbarkeit beider Funktionen in den meisten konkret gegebenen Anwendungsfällen ohnehin sofort aus dem Lebesgueschen Integrabilitätskriterium in Folgerung 29.30 ergibt.

(c) Im Fall $n=1$ ist $f'(x)$ im Transformationssatz eine 1×1-Matrix, so dass $\det f'(x)=f'(x)$ gilt. Die Formel aus Satz 30.12 sieht damit bis auf die Betragsstriche formal so aus wie die Substitutionsregel für eindimensionale Integrale aus Satz 12.31.

Diese Betragsstriche kommen daher, dass im Fall einer *monoton fallenden* Koordinatentransformation in Satz 12.31 (also wenn dort $f' < 0$ gilt) durch die Substitution der Grenzen im resultierenden Integral die Obergrenze kleiner ist als die Untergrenze – was gemäß Notation 12.15 genau das Vorzeichen des Integrals ändert. Dieses Vorzeichen wird in unserer mehrdimensionalen Integrationstheorie (in der es generell keine Unter- und Obergrenzen, sondern nur eine *Menge* als Integrationsbereich gibt) berücksichtigt, indem wir statt f' den Betrag von f' verwenden.

80

Beispiel 30.14 (Polarkoordinaten in $\mathbb{R}^2$)**.** Wir betrachten noch einmal die ebene Polarkoordinatenabbildung aus Satz 9.26 und Beispiel 27.4, die durch die Einschränkung

$$\begin{aligned} f\colon \quad \mathbb{R}_{>0} \times (0,2\pi) &\to \mathbb{R}^2 \setminus \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \geq 0 \right\} \\ \begin{pmatrix} r \\ \varphi \end{pmatrix} &\mapsto \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \end{aligned}$$

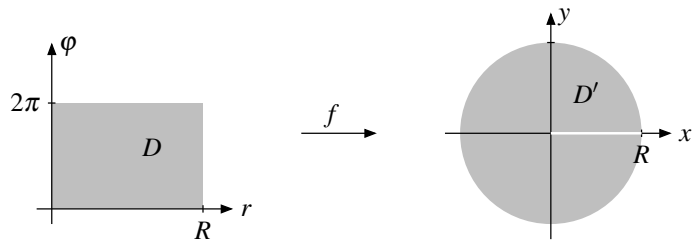
also das Herausnehmen der Punkte mit Winkel 0, zu einem Diffeomorphismus wird. Schränken wir die Abbildung nun weiter auf beschränkte offene Mengen D und D' ein, so dass $f|_D\colon D \to D'$ dort bijektiv und damit ebenfalls wieder ein Diffeomorphismus ist, besagt Satz 30.12 also

$$\int_{D'} g(x,y)\,d(x,y) = \int_D g(f(r,\varphi))\,r\,d(r,\varphi)$$

(sofern die beiden Integrale existieren), da wir in Beispiel 27.4 bereits die Determinante der Ableitungsmatrix $\det f'(r,\varphi) = r$ berechnet haben. Hier sind zwei konkrete Beispiele dafür.

(a) *(Kreisfläche)* Wir wollen den Flächeninhalt des (euklidischen) Kreises $K_R(0) \subset \mathbb{R}^2$ mit Radius $R \in \mathbb{R}_{>0}$ berechnen. Dazu verwenden wir wie im Bild unten Polarkoordinaten mit Start- bzw. Zielbereich

$$D = (0,R) \times (0,2\pi) \qquad \text{und} \qquad D' = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x^2+y^2 < R^2 \right\} \setminus \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \geq 0 \right\}.$$



Die Menge D' unterscheidet sich dabei vom gewünschten Kreis $K_R(0)$ nur durch eine Nullmenge, so dass es keine Rolle spielt, von welcher dieser beiden Mengen wir das Volumen berechnen. Aus dem gleichen Grund sind auch die resultierenden Integrale in Polarkoordinaten über die Bereiche $(0,R) \times (0,2\pi)$ und $[0,R] \times [0,2\pi]$ gleich. Derartige Abänderungen der Integrationsbereiche, um Mengen je nach Bedarf offen bzw. abgeschlossen zu machen, werden wir im Folgenden in der Regel durchführen, ohne dies jedesmal explizit zu erwähnen.

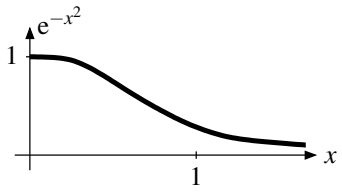
Wir erhalten für die gesuchte Kreisfläche also mit Hilfe des Transformationssatzes wie erwartet

$$\begin{aligned} \operatorname{vol}(K_R(0)) &= \int_{D'} 1\,d(x,y) = \int_D r\,d(r,\varphi) = \int_0^R \int_0^{2\pi} r\,d\varphi\,dr = 2\pi \int_0^R r\,dr = 2\pi \cdot \left[\frac{1}{2}r^2\right]_0^R \\ &= \pi R^2. \end{aligned}$$

(b) Das in der Praxis häufig vorkommende uneigentliche eindimensionale Integral

$$I := \int_0^\infty \mathrm{e}^{-x^2}\,dx$$

lässt sich nicht mit den herkömmlichen Methoden berechnen, da der Integrand keine Stammfunktion besitzt, die sich mit den bekannten speziellen Funktionen aus Kapitel 9 ausdrücken lässt.



In der Tat kann man Integrale über e^{-x^2} mit gegebenen Grenzen in $\mathbb{R}$ aus diesem Grund nur mit Näherungsverfahren berechnen. Das obige *uneigentliche* Integral I ist jedoch mit einem unerwarteten Trick durch Polarkoordinaten exakt berechenbar: Dazu quadrieren wir das Integral zunächst und erhalten

$$\begin{aligned}
I^2 &= \lim_{n\to\infty}\int_0^n \mathrm{e}^{-x^2}\,dx \cdot \lim_{n\to\infty}\int_0^n \mathrm{e}^{-y^2}\,dy \\
&= \lim_{n\to\infty}\left(\int_0^n \mathrm{e}^{-x^2}\,dx \cdot \int_0^n \mathrm{e}^{-y^2}\,dy\right) && \text{(Grenzwertsatz 5.13 (b))} \\
&= \lim_{n\to\infty}\int_{[0,n]\times[0,n]} \mathrm{e}^{-x^2-y^2}\,d(x,y) && \text{(Satz 28.16 von Fubini)} \\
&= \int_{\mathbb{R}^2_{\geq 0}} \mathrm{e}^{-x^2-y^2}\,d(x,y). && \text{(Definition 29.45 (a))}
\end{aligned}$$

Verwenden wir zur Berechnung dieses uneigentlichen zweidimensionalen Integrals über $\mathbb{R}^2_{\geq 0}$ nun die ausschöpfende Folge von Viertelkreisen

$$B_n := \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : x^2+y^2 \leq n^2 \text{ und } x,y \geq 0 \right\}$$

für $n \in \mathbb{N}$, so erhalten wir mit Polarkoordinaten wegen $x^2+y^2 = r^2$

$$I^2 = \lim_{n\to\infty}\int_{B_n} \mathrm{e}^{-x^2-y^2}\,d(x,y) = \lim_{n\to\infty}\int_0^n\int_0^{\pi/2} \mathrm{e}^{-r^2}\,r\,d\varphi\,dr = \frac{\pi}{2}\lim_{n\to\infty}\int_0^n \mathrm{e}^{-r^2}\,r\,dr.$$

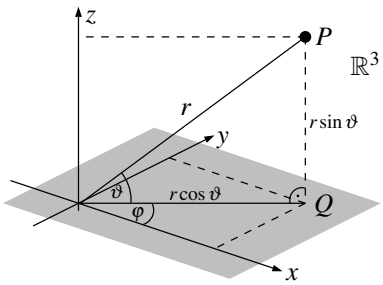
Im Gegensatz zum ursprünglichen Problem ist dieses Integral nun wegen des zusätzlichen Faktors r mit Hilfe einer Substitution für r^2 einfach zu lösen: Wir erhalten

$$I^2 = \frac{\pi}{2}\lim_{n\to\infty}\left[-\frac{1}{2}\mathrm{e}^{-r^2}\right]_0^n = -\frac{\pi}{4}\lim_{n\to\infty}\left(\mathrm{e}^{-n^2}-1\right) = \frac{\pi}{4}$$

und damit für das Ausgangsintegral

$$I = \int_0^\infty \mathrm{e}^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}.$$

Beispiel 30.15 (Kugelkoordinaten in $\mathbb{R}^3$)**.** Die zu Polarkoordinaten analoge Konstruktion im Dreidimensionalen sind die sogenannten *Kugelkoordinaten*. Hier beschreibt man einen Punkt $P \in \mathbb{R}^3$ wie im Bild rechts durch eine Abstandskoordinate r sowie zwei Winkelkoordinaten φ und ϑ: Ist Q die orthogonale Projektion von P auf die x-y-Ebene, so ist r der Abstand von P zum Nullpunkt, φ der Winkel zwischen Q und der x-Achse und ϑ der Winkel zwischen P und der x-y-Ebene.



Der Abstand von Q zum Nullpunkt ist dann $r\cos\vartheta$, und damit gilt für die Koordinaten der beteiligten Punkte

$$Q = \begin{pmatrix} r\cos\vartheta\cos\varphi \\ r\cos\vartheta\sin\varphi \\ 0 \end{pmatrix} \qquad \text{und} \qquad P = \begin{pmatrix} r\cos\vartheta\cos\varphi \\ r\cos\vartheta\sin\varphi \\ r\sin\vartheta \end{pmatrix}.$$

Die Kugelkoordinatenabbildung ist damit also

$$f\colon \begin{pmatrix} r \\ \varphi \\ \vartheta \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z \end{pmatrix} := \begin{pmatrix} r\cos\vartheta\cos\varphi \\ r\cos\vartheta\sin\varphi \\ r\sin\vartheta \end{pmatrix},$$

und sie wird offensichtlich bijektiv, wenn wir Start- und Zielbereich auf

$$D = \mathbb{R}_{>0} \times (0,2\pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \quad \text{bzw.} \quad D' = \mathbb{R}^3 \setminus \left\{ \begin{pmatrix} x \\ 0 \\ z \end{pmatrix} : x \geq 0, z \in \mathbb{R} \right\}$$

einschränken. Die Jacobi-Matrix von f ist damit

$$f' = \begin{pmatrix} \cos\vartheta\cos\varphi & -r\cos\vartheta\sin\varphi & -r\sin\vartheta\cos\varphi \\ \cos\vartheta\sin\varphi & r\cos\vartheta\cos\varphi & -r\sin\vartheta\sin\varphi \\ \sin\vartheta & 0 & r\cos\vartheta \end{pmatrix},$$

und ihre Determinante berechnet sich z. B. mit Beispiel 18.13 (b) einfach zu $\det f' = r^2\cos\vartheta$.

Wie in Beispiel 30.14 (a) können wir damit nun z. B. das Volumen der (euklidischen) Kugel $K_R(0)$ in $\mathbb{R}^3$ vom Radius R berechnen und erhalten wie erwartet

$$\begin{aligned} \operatorname{vol}(K_R(0)) &= \int_{K_R(0)} 1\, d(x,y,z) \\ &= \int_0^R \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} r^2 \cos\vartheta\, d\vartheta\, d\varphi\, dr \\ &= 2\int_0^R \int_0^{2\pi} r^2\, d\varphi\, dr \\ &= 4\pi \int_0^R r^2\, dr \\ &= \frac{4}{3}\pi R^3. \end{aligned}$$

Aufgabe 30.16. Berechne das Integral $\int_D (2x-y)\, d(x,y)$, wobei D das Parallelogramm mit den Eckpunkten $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} \frac{5}{3} \\ \frac{1}{3} \end{pmatrix}$, $\begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}$ und $\begin{pmatrix} \frac{4}{3} \\ -\frac{1}{3} \end{pmatrix}$ in $\mathbb{R}^2$ ist.

(Hinweis: Betrachte dazu den Diffeomorphismus $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ 2x-y \end{pmatrix}$.)

Aufgabe 30.17. Berechne das Integral $\displaystyle\int_D \frac{1}{(x^2+y^2)^2}\, d(x,y)$, wobei D das Gebiet ist, das von den vier Kurven $x^2+y^2 = 4x$, $x^2+y^2 = 8x$, $y = x$ und $y = 2x$ im Bereich $x \geq 0$ begrenzt wird.

Aufgabe 30.18. Es seien $A \subset \mathbb{R}$ eine kompakte messbare Menge und

$$M_A := \{((1-t^2)x, t) : x \in A \text{ und } t \in [0,1]\} \quad \subset \mathbb{R}^2.$$

(a) Skizziere M_A für den Fall $A = [1,2]$.

(b) Zeige, dass M_A messbar ist.

(c) Berechne $\operatorname{vol}(M_A)$ in Abhängigkeit von $\operatorname{vol}(A)$.

Aufgabe 30.19. Es seien $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ ein Diffeomorphismus und $K_r(a)$ die abgeschlossene Kugel in der Maximumsnorm um $a \in \mathbb{R}^3$ mit Radius $r \in \mathbb{R}_{>0}$. Zeige, dass

$$\lim_{r\to 0} \frac{\operatorname{vol}(f(K_r(a)))}{8r^3} = |\det f'(a)|.$$

81

Literatur

[B] A. Beutelspacher, *Lineare Algebra*, Vieweg-Verlag (2003)

[BF] M. Barner, F. Flohr, *Analysis 1*, de Gruyter Lehrbuch (2000)

[E] H.-D. Ebbinghaus et al., *Zahlen*, Springer-Verlag (1988)

[Fi] G. Fischer, *Lineare Algebra*, Vieweg-Verlag (2002)

[Fo1] O. Forster, *Analysis 1*, Vieweg-Verlag (2011)

[Fo2] O. Forster, *Analysis 2*, Vieweg-Verlag (2010)

[G] A. Gathmann, *Algebraische Strukturen*, Vorlesungsskript TU Kaiserslautern (2023), https://agag-gathmann.math.rptu.de/ags

[GK] G.-M. Greuel, T. Keilen, *Lineare Algebra I und II*, Vorlesungsskript TU Kaiserslautern (1999), www.math.uni-tuebingen.de/~keilen/download/LectureNotes/linearealgebra.pdf

[J] K. Jänich, *Lineare Algebra*, Springer-Verlag (2010)

[K1] K. Königsberger, *Analysis 1*, Springer-Verlag (2003)

[K2] K. Königsberger, *Analysis 2*, Springer-Verlag (2003)

[M] T. Markwig, *Grundlagen der Mathematik*, Vorlesungsskript TU Kaiserslautern (2011), www.math.uni-tuebingen.de/~keilen/download/LectureNotes/grundlagen11.pdf

Index

A^{-1} 193
A^* 283
A^T 187
A_b 262
A_b^B 263
$A^{B',B}$ 197
A_f 190
A_f^B 228
$A_f^{B,C}$ 194
A_s^B 268
$\mathrm{Abb}(M,K)$ 157
$\mathrm{Abb}(M,W)$ 157
Abbildung 15
  adjungierte 283
  alternierende 217
  bestimmt divergente 96
  bijektive 17
  bilineare 261
  diagonalisierbare 236
  differenzierbare 117, 339
  gleichmäßig stetige 101, 326
  identische 16
  injektive 17
  lineare 161
  lokal umkehrbare 366
  mehrfach differenzierbare 131, 355
  mehrfach partiell differenzierbare 355
  multilineare 217
  normale 284
  orthogonal diagonalisierbare 286
  orthogonale 279
  partiell differenzierbare 341
  selbstadjungierte 284
  stetig differenzierbare 131, 345, 355
  stetig partiell differenzierbare 344, 355
  stetige 91, 320
  surjektive 17
  total differenzierbare 339
  unitär diagonalisierbare 286
  unitäre 279
Abbildungsmatrix 190, 194
  eines Endomorphismus 228
abelsche Gruppe 24
abgeschlossene Bedingung 325
abgeschlossene Kugel 304
abgeschlossene Menge 309
abgeschlossenes Intervall 40
Abgeschlossenheit
  eines Unterraums 158
Ableitung 117, 339
  eines Polynoms 120, 163
  höhere 131, 353
  partielle 341
  totale 339
Abschluss 90, 311
absolut konvergente Reihe 79
Absolutbetrag 79
Abstand
  von Mengen 333
  von Punkten 303
abzählbar unendliche Menge 62
abzählbare Menge 62
Additionstheoreme 110
Additivität des Integrals 143, 382, 390
Additivität des Volumens 391
adjungierte Abbildung 283
adjungierte Matrix 283
Ähnlichkeit
  von Matrizen 228
Äquivalenz
  von Aussagen 8
  von Matrizen 199
  von Normen 302
Äquivalenzklasse 22
Äquivalenzrelation 22
affiner Unterraum 180
algebraische Vielfachheit 233
alternierende Abbildung 217
alternierende Reihe 78
angeordneter Körper 38
Antisymmetrie
  einer Relation 39
antisymmetrische Bilinearform 293
antisymmetrische Matrix 293
$\arccos x$ 113
archimedische Ordnung 44
$\arcsin x$ 113
$\arctan x$ 113
arithmetische Vielfachheit 233
Arkuskosinus 113
Arkussinus 113
Arkustangens 113
Assoziativität
  der Matrixmultiplikation 189
  der Skalarmultiplikation 155
  der Verkettung 19
  in Gruppen 24
aufgespannter Unterraum 168
Aussage 7
  äquivalente 8
  zusammengesetzte 8
Aussageform 7
ausschöpfende Folge 403
Austauschlemma 172
Austauschsatz von Steinitz 173
Axiom 6
Axiomensystem
  von Zermelo und Fraenkel 13

b_A 262
b_A^B 263
Banachraum 309
Banachscher Fixpunktsatz 327

Basis 169
Basisauswahl 171, 212
Basisergänzung 173, 212
Basiswechselmatrix 197, 215
Bernoullische Ungleichung 41
Berührpunkt 90, 311
beschränkte Folge 50, 55, 307
  nach oben 55
  nach unten 55
beschränkte Funktion 97
beschränkte Menge 42, 307
bestimmt divergent 96
bestimmt divergente Folge 58
bestimmt divergente Funktion 96
Betrag
  einer komplexen Zahl 66
  einer Zahl 40
bijektive Abbildung 17
Bild
  einer Matrix 191, 211
  einer Menge 18
  eines Elements 15
  eines Morphismus 164
Bilinearform 261
  antisymmetrische 293
  indefinite 264
  negativ definite 264
  negativ semidefinite 264
  positiv definite 264
  positiv semidefinite 264
  symmetrische 264
Binomialkoeffizienten 36
  verallgemeinerte 154
binomische Formel 37
$\mathrm{BLF}(V)$ 261
Blockdiagonalmatrix 247
Blockmatrixmultiplikation 190
Bogenmaß 109, 130
Bolzano-Weierstraß 315
  für $\mathbb{K}$ 72
  für $\mathbb{R}$ 60

$\mathbb{C}$ 65
$C^n(D)$ 131, 355
Cantor 12
Cantorsches Diagonalverfahren 63
Cauchy-Hadamard 84
Cauchy-Kriterium
  für Folgen 73
  für Reihen 79
Cauchy-Produkt 87
Cauchy-Schwarz-Ungleichung 269
Cauchyfolge
  in einem metrischen Raum 308
  in $\mathbb{K}$ 73
Cavalieri 402
Cayley-Hamilton 258
charakteristische Funktion 387
charakteristisches Polynom 232
$\cos x$ 109
Cramersche Regel 225

$\partial_i f$ 341
$\partial_v f$ 341

Definitheit
  einer Bilinearform 264
  einer Matrix 264, 268, 289
  einer Sesquilinearform 268
Definitionsmenge 15
$\deg f$ 33
Determinante
  Eindeutigkeit 219
  einer Matrix 216
  eines Endomorphismus 230
  Existenz 220
Determinantenproduktsatz 219
diagonalisierbare Abbildung 236
diagonalisierbare Matrix 236
Diagonalmatrix 236, 294
Diagonalverfahren
  von Cantor 63
dichte Teilmenge 44, 311
Diffeomorphismus 410
Differentialgleichung 254, 328
Differentialquotient 117
Differentialschreibweise 121
Differenzenquotient 117
differenzierbare Funktion 117, 339
Differenzmenge 14
$\dim V$ 174
Dimension
  einer Untermannigfaltigkeit 371
  eines Vektorraums 174
Dimensionsformel
  für direkte Summen 177
  für Durchschnitte 185
  für Matrizen 192
  für Morphismen 184
  für Produkte 176
  für Quotientenräume 182
  für Summen 185
direkte Summe 177
disjunkte Mengen 14
disjunkte Vereinigung 14
diskrete Metrik 304
Distributivität 26
  der Matrixmultiplikation 189
  der Skalarmultiplikation 155
divergente Folge
  in metrischen Räumen 305
  in $\mathbb{R}$ 47
Divergenz
  von Folgen 47, 305
  von Funktionen 91, 320
  von uneigentlichen Integralen 148, 404
Drehmatrix 231
Dreiecke
  kongruente 21
Dreiecksmatrix
  echte 224
  obere 224
  untere 224
Dreiecksungleichung
  einer Metrik 303
  einer Norm 270, 299
  für Integrale 142, 381, 390
  in $\mathbb{C}$ 68
  in $\mathbb{R}$ 40

nach unten 41
Dualraum 267
Durchschnitt
von Unterräumen 213

e 105
E 192
ε-Umgebung
reelle 47
ε-Umgebung 304
komplexe 72
e_i 169
E_n 192
echte Dreiecksmatrix 224
$\mathrm{Eig}(A,\lambda)$ 230
$\mathrm{Eig}(f,\lambda)$ 230
Eigenraum
einer Matrix 230
eines Endomorphismus 230
verallgemeinerter 244
Eigenvektor
einer Matrix 230
eines Endomorphismus 230
verallgemeinerter 244
Eigenwert
einer Matrix 230
eines Endomorphismus 230
Eigenwertkriterium
für Definitheit von Matrizen 289
Einheitsmatrix 192
Einheitsvektor 169
Einheitswurzeln 115
Einschachtelungssatz 54
Einschränkung 16
Element
einer Familie 167
einer Menge 12
inverses 24
linksinverses 24
linksneutrales 24
neutrales 24
elementare Spaltenumformung 206
elementare Zeilenumformung 205
Elementarmatrix 205
$\mathrm{End}(V)$ 228
endlich erzeugter Vektorraum 171
endlich-dimensionaler Vektorraum 174
endliche Familie 167
endliche Menge 13
Endomorphismus 228
adjungierter 283
diagonalisierbarer 236
normaler 284
orthogonal diagonalisierbarer 286
orthogonaler 279
selbstadjungierter 284
unitär diagonalisierbarer 286
unitärer 279
Entwicklungspunkt 132, 356
Entwicklungssatz von Laplace 223
erweiterte Koeffizientenmatrix 203
Erzeugendensystem 169
erzeugter Unterraum 168
euklidische Metrik 304
euklidische Norm 300
euklidischer Raum 265
Eulersche Zahl 105
$\exp x$ 84
Exponentialfunktion 84, 86, 105
Funktionalgleichung der 88
Extremum
globales 122, 352
isoliertes 122, 352
lokales 122, 352
unter Nebenbedingungen 373
Extremwertkriterium 135, 360

f_A 190
$f_A^{B,C}$ 194
f_D 387
$F_k(\lambda)$ 205
$F_{k,l}(\lambda)$ 205
Faktorraum 182, 213
Fakultät 35
Familie
endliche 167
linear abhängige 169
linear unabhängige 169
orthogonale 271
orthonormale 271
von Vektoren 167
fast alle 48
fast überall 394
nach Lebesgue 397
Feinheit
einer Zerlegung 145
Fibonacci-Folge 176
Fixpunktsatz
von Banach 327
Folge 47
ausschöpfende 403
beschränkte 50, 55, 307
bestimmt divergente 58
divergente 47, 305
geometrische 48, 74
konvergente 47, 305
monoton fallende 55
monoton steigende 55
monoton wachsende 55
nach oben beschränkte 55
nach unten beschränkte 55
reelle 47
rekursive 55
streng monotone 55
Umordnung 53
unbestimmt divergente 58
von Fibonacci 176
folgenkompakte Menge 315
Folgenkriterium
für Abgeschlossenheit 313
für Funktionsgrenzwerte 94, 321
für Stetigkeit 94, 321
Fraenkel 13
Frobenius-Norm 301
Fubini
für Normalbereiche 401
für Quader 383
Fundamentalsatz der Algebra 69

Funktion 15
alternierende 217
beschränkte 97
bestimmt divergente 96
charakteristische 387
differenzierbare 117, 339
gleichmäßig stetige 101, 326
implizite 368
integrierbare 140, 379, 389
lineare 161
Lipschitz-stetige 104
lokal umkehrbare 366
mehrfach differenzierbare 131, 355
mehrfach partiell differenzierbare 355
monotone 97
multilineare 217
partiell differenzierbare 341
rationale 95
stetig differenzierbare 131, 345, 355
stetig fortsetzbare 91, 320
stetig partiell differenzierbare 344, 355
stetige 91, 320
streng monotone 97
stückweise stetige 144
total differenzierbare 339
Funktionalgleichung
der Exponentialfunktion 88
der Logarithmusfunktion 107
Funktionenfolge 100
gleichmäßig konvergente 101, 327
punktweise konvergente 101, 327
Funktionswert 15

ganze Zahl 13
Gauß
Summenformel von 30
Verfahren von 207
Gaußklammer 44
geometrische Folge 48, 74
geometrische Reihe
endliche 35
unendliche 75
geometrische Vielfachheit 233
geordneter Körper 38
geordnetes Paar 14
Geradenpaar 293
$\mathrm{GL}(n,K)$ 192
gleichmächtige Mengen 62
gleichmäßige Konvergenz 101, 327
gleichmäßige Stetigkeit 101, 326
Gleichungssystem
lineares 203, 208
globales Extremum 122, 352
globales Maximum 122, 352
globales Minimum 122, 352
Grad
einer Polynomfunktion 31, 33
$\operatorname{grad} f$ 347
Gradient 347
Gram-Schmidtsches
Orthonormalisierungsverfahren 273
Gramsche Matrix 262, 263, 268
Graph 17
Grenzwert
einer Folge 305
einer Funktion 91, 320
einer komplexen Folge 72
einer reellen Folge 47
uneigentlicher 58, 96
Grenzwertsätze
für Folgen 51
für Funktionen 94, 323
in normierten Räumen 307
Gruppe 24
abelsche 24
kommutative 24
orthogonale 281
unitäre 281
Gruppenaxiome 24

$H(A,\lambda)$ 244
$H_r(A,\lambda)$ 244
Häufungspunkt 314
einer Folge 53
einer Menge 312
halboffenes Intervall 40
harmonische Reihe 76
alternierende 78
Hauptachse 292
Hauptachsentransformation 292
Hauptminor 277
Hauptminorenkriterium 277
Hauptraum 244
Hauptraumzerlegung 246
Hauptsatz
der Differential- und Integralrechnung 145
Hauptvektor 244
Hausdorff-Metrik 334
Heine-Borel 317
Hermitizität
einer Matrix 268
einer Sesquilinearform 268
Hesse-Matrix 356
Höhenlinie 348
$\mathrm{Hom}(V,W)$ 162
Homomorphiesatz 183
Homomorphismus 161
de l'Hôpital 128
Hurwitz-Kriterium 277, 289
Hyperbel 293

i 65
identische Abbildung 16
$\operatorname{Im} f$ 164
$\operatorname{Im} z$ 66
Imaginärteil 66
implizite Funktion 368
Indefinitheit
einer Bilinearform 264
einer Matrix 265
Indexverschiebung 29
Indikatorfunktion 387
Induktion 30
Induktionsanfang 30
Induktionsannahme 30
Induktionsschluss 30
Induktionsschritt 30
Induktionsvoraussetzung 30

Infimum 42
$\inf M$ 43
injektive Abbildung 17
innerer Punkt 311
Inneres
  einer Menge 311
Integrabilitätskriterium
  von Lebesgue 395, 398
  von Riemann 141, 380
Integral
  Additivität 143, 382, 390
  nach Lebesgue 140, 377
  nach Riemann 140, 377, 379
  Translationsinvarianz 392
  unbestimmtes 147
  uneigentliches 148, 404
Integralkriterium
  für Reihenkonvergenz 154
integrierbare Funktion 140, 379, 389
Intervall
  abgeschlossenes 40
  halboffenes 40
  kompaktes 40, 310
  offenes 40
  uneigentliches 40
Intervallschachtelung 58
invarianter Unterraum 183, 245
inverse Matrix 193, 214, 224
inverses Element 24
invertierbare Matrix 192
isolierter Punkt 116, 312
isoliertes Maximum 122
isoliertes Extremum 122, 352
isoliertes Maximum 352
isoliertes Minimum 122, 352
isomorph 166
Isomorphiesatz 185
Isomorphismus 166

$Jf(a)$ 342
$J_m(\lambda)$ 248
Jacobi-Matrix 342
Jordan-Messbarkeit 389
Jordan-Nullmenge 392
Jordanbasis 250
Jordanblock 249
Jordandiagramm 248
  Spaltenbedingung 248
  Zeilenbedingung 248
Jordanform 250
Jordankästchen 249
Jordansche Normalform 250

$\mathbb{K}$ 72
$K_r(a)$ 304
$\operatorname{Ker} f$ 164
Kern
  einer Matrix 191, 210
  eines Morphismus 164
Kettenregel 120, 349
Klasse 22
Koeffizientenmatrix
  erweiterte 203
Koeffizientenvergleich
  für Polynomfunktionen 33
  für Potenzreihen 104
Körper 26
  angeordneter 38
  geordneter 38
Körperaxiome 26
Körpererweiterung 65
kommutative Gruppe 24
Kommutativität 24
kompakte Menge 315
kompaktes Intervall 40, 310
Komplement
  einer Menge 309
  eines Unterraums 178, 213
  orthogonales 276
komplexe Konjugation 66
komplexe Zahl 65
komplexe Zahlenebene 66
Kongruenz 21
Kongruenzklasse 21
Konjugation 66
Kontraktion 327
Kontraposition 10
konvergente Folge
  in metrischen Räumen 305
  in $\mathbb{R}$ 47
Konvergenz
  absolute 79
  gleichmäßige 101, 327
  punktweise 101, 327
  von Folgen 47, 305
  von Funktionen 91, 320
  von uneigentlichen Integralen 148, 404
Konvergenzgebiet 85
Konvergenzkriterien
  Cauchy-Kriterium 73, 79
  Integralkriterium 154
  Leibniz-Kriterium 78
  Majorantenkriterium 81
  Minorantenkriterium 81
  Monotoniekriterium 55
  Quotientenkriterium 82
  Trivialkriterium 77
  Wurzelkriterium 82
Konvergenzradius 85
Koordinaten
  eines Vektors 170
Koordinatenabbildung 194
Koordinatenvektor 194
Kosinus 109
Kreisfläche 416
kritischer Punkt 123, 352
Kugel
  abgeschlossene 304
  offene 304
Kugelkoordinaten 417
Kugelvolumen 418

Länge
  eines Vektors 265
Lagrange-Multiplikator 373
Laplacescher Entwicklungssatz 223
Lebesgue-Integral 140, 377
Lebesgue-Nullmenge 397

Lebesguesches Integrabilitätskriterium
  für messbare Mengen 398
  für Quader 395
leere Summe 29
leeres Produkt 29
Leibniz-Kriterium 78
Leitkoeffizient
  einer Polynomfunktion 31
Lemma 19
Lemniskate 371
$\lim a_n$ 47
Limes inferior 60
Limes superior 60
$\liminf a_n$ 60
$\limsup a_n$ 60
$\operatorname{Lin} B$ 167
linear abhängig 169
linear unabhängig 169
lineare Abbildung 161
lineare Substitution 151
lineares Gleichungssystem 203, 208
Linearkombination 167
  nicht-triviale 169
linksinverses Element 24
linksneutrales Element 24
Lipschitz-Konstante 104
Lipschitz-Stetigkeit 104
$\ln x$ 106
$\log x$ 106
Logarithmus
  natürlicher 106
lokale Umkehrfunktion 366
lokales Extremum 122, 352
lokales Maximum 122, 352
lokales Minimum 122, 352

$\mu_{\mathrm{a}}(A,\lambda)$ 233
$\mu_{\mathrm{a}}(f,\lambda)$ 234
$\mu_{\mathrm{g}}(A,\lambda)$ 233
$\mu_{\mathrm{g}}(f,\lambda)$ 234
Mächtigkeit 62
Majorante 81
Majorantenkriterium 81
$\operatorname{Mat}(m \times n, K)$ 187
Matrix 187
  adjungierte 283
  ähnliche 228
  äquivalente 199
  antisymmetrische 293
  diagonalisierbare 236
  einer Bilinearform 262, 263
  einer linearen Abbildung 190, 194
  einer Sesquilinearform 268
  eines Basiswechsels 197, 215
  hermitesche 268
  indefinite 265
  inverse 193, 214, 224
  invertierbare 192
  negativ definite 265
  negativ semidefinite 265
  nilpotente 254
  normale 284
  orthogonal diagonalisierbare 287
  orthogonale 279
  positiv definite 264, 268
  positiv semidefinite 265
  pseudoinverse 298
  quadratische 187
  selbstadjungierte 284
  spezielle orthogonale 281
  spezielle unitäre 281
  symmetrische 264
  transponierte 187
  unitär diagonalisierbare 287
  unitäre 279
Matrixmultiplikation 188
Matrixnorm
  induzierte 332
Maximum
  einer Menge 42
  globales 122, 352
  isoliertes 122, 352
  lokales 122, 352
  Satz vom 97, 331
  zweier Zahlen 42
Maximumsnorm 300
$\max M$ 43
mehrfach differenzierbare Funktion 131, 355
mehrfach partiell differenzierbare Funktion 355
Menge 12
  abgeschlossene 309
  abzählbar unendliche 62
  abzählbare 62
  beschränkte 42, 307
  dichte 44, 311
  endliche 13
  folgenkompakte 315
  gleichmächtige 62
  kompakte 315
  leere 12
  messbare 389
  nach oben beschränkte 42
  nach unten beschränkte 42
  offene 309
  relativ offene 325
  überabzählbare 62
  überdeckungskompakte 316
messbare Menge 389
Metrik 303
  diskrete 304
  euklidische 304
  von Hausdorff 334
metrischer Raum 303
  vollständiger 309
Minimalpolynom
  einer Matrix 257
  eines Endomorphismus 259
Minimum
  einer Menge 42
  globales 122, 352
  isoliertes 122, 352
  lokales 122, 352
  Satz vom 97, 331
  zweier Zahlen 42
$\min M$ 43
Minor 277
Minorante 81
Minorantenkriterium 81

Mischfolge 53
Mittelwertsatz
der Differentialrechnung 123, 359
der Integralrechnung 145, 382
modulo 182
monoton fallend
für Folgen 55
für Funktionen 97
monoton steigend
für Folgen 55
für Funktionen 97
monoton wachsend
für Folgen 55
für Funktionen 97
monotone Funktion 97
Monotonie des Volumens 391
Monotoniekriterium 55
Morphismus 161
adjungierter 283
diagonalisierbarer 236
normaler 284
orthogonal diagonalisierbarer 286
orthogonaler 279
selbstadjungierter 284
unitär diagonalisierbarer 286
unitärer 279
Multi-Index 356
multilineare Abbildung 217
Multiplikation
von Matrizen 188
Multiplikator
von Lagrange 373
Multiplizität
algebraische 233
arithmetische 233
einer Nullstelle 33
geometrische 233

$\mathbb{N}$ 13
natürliche Zahl 13
Nebenbedingungen 373
Negation 10
negative Definitheit
einer Bilinearform 264
einer Matrix 265
negative Semidefinitheit
einer Bilinearform 264
einer Matrix 265
negative Zahl 38
neutrales Element 24
nicht-triviale Linearkombination 169
nilpotent 254
Norm 299
äquivalente 302
eines Vektors 265
euklidische 300
von Frobenius 301
Normalbereich 399
Normalbereiche
bezüglich einer Variablenordnung 400
normale Abbildung 284
normale Matrix 284
Normalform
bezüglich Äquivalenz 201
einer Abbildungsmatrix 201
normierte Polynomfunktion 31
normierter Raum 299
normierter Vektor 265
Nullfolge 51
Nullmatrix 188
Nullmenge
nach Jordan 392
nach Lebesgue 397
Nullstelle 31
Nullvektor 156
Nullvektorraum 156

$\mathrm{O}(n)$ 279
obere Dreiecksmatrix 224
obere Schranke 42
Oberintegral 139, 379
Obermenge 12
Obersumme 138, 378
offene Bedingung 325
offene Kugel 304
offene Menge 309
offene Überdeckung 316
offenes Intervall 40
$\mathrm{OI}(f)$ 139, 379
Ordnung
archimedische 44
auf einem Körper 38
auf einer Menge 39
einer Nullstelle 33
partielle 39
totale 39
orthogonal diagonalisierbare Abbildung 286
Orthogonalbasis 271
orthogonale Abbildung 279
orthogonale Familie 271
orthogonale Gruppe 281
orthogonale Matrix 279
orthogonale Projektion 272
orthogonale Vektoren 271
orthogonales Komplement 276
Orthonormalbasis 271
orthonormale Familie 271
Orthonormalisierungsverfahren
von Gram-Schmidt 273
$\mathrm{OS}(f,Z)$ 138, 378

π 111
p-Norm 301
p_A 257
Φ_B 194
Paar
geordnetes 14
Paradoxon
von Russell 13
Parallelogrammgleichung 302
Parameterintegral 346
Partialsummen 75
partiell differenzierbare Funktion 341
partielle Integration 149
partielle Ordnung 39
Partition
einer Menge 23
Pascalsches Dreieck 36

Peano-Kurve 334
Polarkoordinaten 114, 364, 416
Polynom 33
  charakteristisches 232
Polynomdivision 32
Polynomfunktion 31
  normierte 31
positive Definitheit
  einer Bilinearform 264
  einer Matrix 264, 268
  einer Sesquilinearform 268
positive Semidefinitheit
  einer Bilinearform 264
  einer Matrix 265
positive Zahl 38
Potenz 28, 107
Potenzmenge 14
Potenzreihe 84
Prinzip von Cavalieri 402
Produkt
  leeres 29
Produktintegration 149
Produktmenge 14
Produktregel 118, 348
Produktsatz
  für Determinanten 219
Produktzeichen 29
Projektion
  orthogonale 272
pseudoinverse Matrix 298
Punkt
  innerer 311
  isolierter 116, 312
  kritischer 123, 352
punktweise Konvergenz 101, 327

$\mathbb{Q}$ 13
QR-Zerlegung 282
Quader 377
quadratische Matrix 187
Quadratwurzel 56
Quantor 9
Quotientenkriterium 82
Quotientenraum 182, 213
Quotientenregel 119, 348

$\mathbb{R}$ 13
$R^n_{f,a}$ 133
Rand 312
Randextremum 123, 372
Randpunkt 312
Rang
  einer Matrix 199, 208
  eines Morphismus 199
rationale Funktion 95
rationale Zahl 13
Raum
  euklidischer 265
  metrischer 303
  normierter 299
  unitärer 269
  vollständiger 309
  wegzusammenhängender 326
$\operatorname{Re} z$ 66
Realteil 66
reduzierte Zeilenstufenform 206
reelle Zahl 13
reelle Folge 47
Reflexivität
  einer Relation 22, 39
Regel
  von Cramer 225
  von Sarrus 221
Reihe 75
  absolut konvergente 79
  alternierende 78
  geometrische 35, 75
  harmonische 76
  Umordnung 80
rekursive Folge 55
Relation 15
relativ offene Menge 325
Repräsentant
  einer Äquivalenzklasse 22
Restglied 133, 358
Restklasse 182
Richtungsableitung 341
Riemann-Integral 140, 377, 379, 389
Riemannsches Integrabilitätskriterium 141, 380
$\operatorname{rk} f$ 199
Rolle 123
Russell 13
Russellsches Paradoxon 13

s^B_A 268
Sarrus
  Regel von 221
Sattelpunkt 362
Satz
  über implizite Funktionen 368
  über lokale Umkehrfunktionen 366
  vom Maximum und Minimum 97, 331
  von Bolzano-Weierstraß 60, 72, 315
  von Cayley-Hamilton 258
  von Fubini 383, 401
  von Heine-Borel 317
  von Laplace 223
  von Rolle 123
  von Schwarz 354
  von Steinitz 173
Schnittmenge 14
Schranke
  obere 42
  untere 42
Schwarz
  Satz von 354
selbstadjungierte Abbildung 284
selbstadjungierte Matrix 284
Semidefinitheit
  einer Bilinearform 264
  einer Matrix 265
senkrecht 271
Sesquilinearform 268
  hermitesche 268
  positiv definite 268
$\sin x$ 109
Singulärwert 294
Singulärwertzerlegung 294

Sinus 109
Skalar 156
Skalarmultiplikation 155
Skalarprodukt 265, 269
SLF(V) 268
Spaltenbedingung
  im Jordandiagramm 248
Spaltenumformung 206
Spektralnorm 333
Spektralsatz
  für Endomorphismen 286
  für Matrizen 287
spezielle orthogonale Matrix 281
spezielle unitäre Matrix 281
Sprungstelle 144
Spur
  einer Matrix 229
  eines Endomorphismus 230
Stammfunktion 146
Standardbasis 169
Standardskalarprodukt 266, 269
Startmenge 15
Startraum 15
Steinitzscher Austauschsatz 173
stetig differenzierbare Funktion 131, 345, 355
stetig partiell differenzierbare Funktion 344, 355
stetige Fortsetzung 91, 320
stetige Funktion 91, 320
Stetigkeit 91, 320
  gleichmäßige 101, 326
Streichungsmatrix 220
streng monoton 97
streng monoton steigend
  für Folgen 55
streng monoton wachsend
  für Folgen 55
stückweise stetige Funktion 144
Stufe
  eines Hauptvektors 244
  eines verallgemeinerten Eigenraums 244
Stufenspalte 206
Subadditivität des Volumens 391
Submultiplikativität
  einer Matrixnorm 332
Substitutionsregel 150
  lineare 151
Summe
  direkte 177
  leere 29
  von Unterräumen 160, 213
Summenformel
  von Gauß 30
Summennorm 300
Summenzeichen 28
$\sup M$ 43
Supremum 42
  uneigentliches 46
Supremumsaxiom 43
Supremumsnorm 104, 301
surjektive Abbildung 17
Sylvester
  Trägheitssatz 290
Symmetrie
  einer Bilinearform 264
  einer Matrix 264
  einer Metrik 303
  einer Relation 22

$T_{f,a}$ 133
$T^n_{f,a}$ 132, 356
$\tan x$ 112
Tangens 112
Tangente 116, 346
Tangentialraum
  an einen Graphen 347
Tangentialvektor 346
Taylor-Formel 133, 358
  für Potenzreihen 131
Taylor-Polynom 132, 356
Taylor-Reihe 133
Teilfolge 53
Teilmenge 12
  echte 12
Teilquader
  einer Zerlegung 377
Teleskopreihe 76
Topologie 306
topologische Eigenschaft 306
total differenzierbare Funktion 339
totale Ableitung 339
totale Ordnung 39
Totalität
  einer Relation 39
Trägheitssatz von Sylvester 290
Transformationssatz
  für Integrale 414
Transitivität
  einer Relation 22, 39
Translationsinvarianz
  des Integrals 392
transponierte Matrix 187
trivialer Unterraum 159
Trivialkriterium 77

$U^\perp$ 276
$U_\varepsilon(a)$ 47
$U_\varepsilon(a)$ 72
$\mathrm{U}(n)$ 279
$U_r(a)$ 304
überabzählbare Menge 62
Überdeckung
  offene 316
überdeckungskompakte Menge 316
$\mathrm{UI}(f)$ 139, 379
Umgebung 304, 311
  komplexe 72
  reelle 47
Umkehrabbildung 19, 98
  lokale 366
Umkehrfunktion 19, 98
  lokale 366
Umordnung
  einer Folge 53
  einer Reihe 80
Umordnungssatz 80
unbestimmt divergente Folge 58
unbestimmtes Integral 147
uneigentlicher Grenzwert

von Folgen 58
von Funktionen 96
uneigentliches Integral 148, 404
uneigentliches Intervall 40
uneigentliches Supremum 46
Ungleichung
von Bernoulli 41
von Cauchy-Schwarz 269
unitär diagonalisierbare Abbildung 286
unitäre Abbildung 279
unitäre Gruppe 281
unitäre Matrix 279
unitärer Raum 269
untere Dreiecksmatrix 224
untere Schranke 42
Unterintegral 139, 379
Untermannigfaltigkeit 371
Unterraum 158
affiner 180
aufgespannter 168
erzeugter 168
invarianter 183, 245
trivialer 159
verschobener 180
Untersumme 138, 378
Untervektorraum 158
affiner 180
aufgespannter 168
erzeugter 168
trivialer 159
verschobener 180
Urbild
einer Menge 18
eines Elements 17
$\mathrm{US}(f,Z)$ 138, 378

V^* 267
Variable 7
Vektor 156
normierter 265
orthogonaler 271
Vektoraddition 155
Vektorraum 155
dualer 267
endlich erzeugter 171
endlich-dimensionaler 174
isomorpher 166
mit Skalarprodukt 269
normierter 299
Vektorraumhomomorphismus 161
Vektorraumisomorphismus 166
verallgemeinerte Binomialkoeffizienten 154
verallgemeinerter Eigenraum 244
verallgemeinerter Eigenvektor 244
Vereinigung 14
disjunkte 14
Vereinigungsmenge 14
Verfeinerung 138, 377
Verkettung 19
differenzierbarer Funktionen 120, 349
Verknüpfung 24
Verneinung 10
verschobener Unterraum 180
Vielfachheit
algebraische 233
arithmetische 233
einer Nullstelle 33
geometrische 233
$\mathrm{vol}(D)$ 389
vollständige Induktion 30
vollständiger Raum 309
Vollständigkeit 73
Volumen
Additivität 391
einer messbaren Menge 389
eines Quaders 377
Monotonie 391
Subadditivität 391
Volumenformel
für Diffeomorphismen 413
für Isomorphismen 410

Wahrheitstafel 8
Weg 326
wegzusammenhängender Raum 326
Wert
einer Funktion 15
Widerspruchsbeweis 10
Winkel 270
Wohldefiniertheit 181
Wohlordnung 44
Wurzel 56
einer Matrix 288
höhere 57, 98
Wurzelfunktion 98
Wurzelkriterium 82

χ_A 232
χ_f 234

$\mathbb{Z}$ 13
Zahl
ganze 13
komplex konjugierte 66
komplexe 65
natürliche 13
negative 38
positive 38
rationale 13
reelle 13
Zahlenebene
komplexe 66
Zeilenbedingung
im Jordandiagramm 248
Zeilenstufenform 206
reduzierte 206
Zeilensummennorm 333
Zeilenumformung 205
Zerlegung 377
eines Intervalls 138
Feinheit einer 145
induzierte 388
Zermelo 13
Zielmenge 15
Zielraum 15
Zwischenwertsatz 96, 326