Mathematik I für Informatik

Vorlesungsskript, Wintersemester 2026/27
Prof. Dr. Yann Disser
31. Oktober 2026



Inhaltsverzeichnis

1 Grundlagen **1**
- 1.1 Aussagen . . . 1
- 1.2 Logische Verknüpfungen . . . 2
- 1.3 Logisches Schließen und Beweise . . . 5
- 1.4 Mengen . . . 8
- 1.5 All- und Existenzquantor . . . 12
- 1.6 Relationen . . . 12
- 1.7 Abbildungen . . . 19

2 Zählen und Arithmetik **25**
- 2.1 Die natürlichen Zahlen . . . 25
- 2.2 Vollständige Induktion . . . 26
- 2.3 Endliche Mengen . . . 28
- 2.4 Addition . . . 30
- 2.5 Multiplikation . . . 33
- 2.6 Potenz . . . 37
- 2.7 Fakultät . . . 40
- 2.8 Binomialkoeffizienten . . . 42
- 2.9 Anwendungsbeispiel: Kryptographie . . . 44

3 Erweiterung des Zahlenraums **49**
- 3.1 Die ganzen Zahlen . . . 49
- 3.2 Gruppen . . . 51
- 3.3 Ringe . . . 62
- 3.4 Die rationalen Zahlen . . . 65
- 3.5 Körper . . . 67
- 3.6 Die reellen Zahlen . . . 71
- 3.7 Die komplexen Zahlen . . . 74
- 3.8 Die Zahlbereiche im Überblick . . . 79

4 Mehrdimensionale Räume **81**
- 4.1 Geometrie des $\mathbb{R}^n$ . . . 81
- 4.2 Vektorräume . . . 89
- 4.3 Untervektorräume . . . 91
- 4.4 Basen . . . 95
- 4.5 Dimension . . . 98
- 4.6 Skalarprodukte und Normen . . . 102
- 4.7 Orthonormalbasen . . . 107

4.8 Projektionen . . . 110
4.9 Affine Räume . . . 113

5 Lineare Abbildungen 115
5.1 Lineare Abbildungen . . . 115
5.2 Kern, Bild und Rang . . . 119
5.3 Matrizen . . . 124
5.4 Rechnen mit Matrizen . . . 129
5.5 Invertierbarkeit . . . 135
5.6 Basiswechsel . . . 139
5.7 Anwendungsbeispiel: Bildcodierung . . . 144

6 Lineare Gleichungssysteme und Eigenwerte 147
6.1 Lineare Gleichungssysteme . . . 147
6.2 Das Gauß-Verfahren . . . 151
6.3 Determinanten . . . 162
6.4 Eigenwerte, Eigenvektoren und Eigenräume . . . 169
6.5 Symmetrische Matrizen . . . 178

7 Wachstumsverhalten und Grenzwerte 187
7.1 Reelle Folgen . . . 187
7.2 Reihen . . . 198
7.3 Asymptotik . . . 206

1 Grundlagen

Mathematisches Denken ist ein wesentlicher Bestandteil informatischer Arbeit: Wir beschreiben Probleme, entwickeln Verfahren und untersuchen deren Eigenschaften und Grenzen. Dabei müssen wir selbstständig Lösungswege finden und begründen können, weshalb sie funktionieren. Mathematik verlangt, Begriffe zu klären, Voraussetzungen offenzulegen und aus ihnen korrekte Schlüsse zu ziehen. Sie schärft damit unsere Fähigkeit, Probleme systematisch zu bearbeiten und Argumente eigenständig zu beurteilen. Diese Selbstständigkeit im Denken und Urteilen ist weit über die Informatik hinaus von Bedeutung.

In dieser Veranstaltung entwickeln wir grundlegende mathematische Begriffe und Methoden. Dabei verfolgen wir, wie sich aus wenigen Voraussetzungen schrittweise weiterführende Einsichten gewinnen und zu begründeten Lösungen zusammenführen lassen. Die Voraussetzungen und die einzelnen Schritte unserer Argumentation sollen nachvollziehbar und überprüfbar sein. Wir beginnen mit den Grundbegriffen, auf denen ein Großteil der Mathematik aufbaut.

1.1 Aussagen

Mathematik beschäftigt sich mit der logischen Herleitung neuer Information aus bereits Bekanntem. Wissen wir beispielsweise

A: Der Einheitskreis hat Radius 1,

B: Hat ein Kreis Radius 1, so beträgt seine Fläche π,

so können wir daraus schließen

C: Die Fläche des Einheitskreises beträgt π.

Hierbei sind A, B, C jeweils *Aussagen*.

Definition 1.1. Eine *Aussage* ist ein Satz oder mathematischer Ausdruck, der entweder eindeutig wahr oder eindeutig falsch ist. Ist eine Aussage wahr, sagen wir, dass sie *gilt*.

Beispiel 1.2. *Die folgenden sind jeweils Aussagen:*

A_1: *Die Zahl 3 ist ungerade. (wahr)*

A_2: $3 + 5 = 9$. *(falsch)*

A_3: *Hat ein Kreis Radius* 3*, so beträgt seine Fläche* 9π*. (wahr)*

A_4: *Es regnet gerade in Madrid. (unbekannt)*

A_5: *Der Himmel ist blau. (wahr)*

A_6: *Jede natürliche Zahl ist gerade. (falsch)*

A_7: *Jede gerade natürliche Zahl größer 2 ist die Summe zweier Primzahlen. (unbekannt)*

***Keine** Aussagen sind hingegen:*

- *Guten Morgen.*
- *Besuchen Sie möglichst alle Vorlesungen!*
- $3x < 8$.

△

Wir erlauben auch Ausdrücke, die von Variablen abhängen.

Definition 1.3. Eine *Aussageform* ist ein Satz oder mathematischer Ausdruck mit einer oder mehreren Variablen, der durch die Belegung der Variablen zu einer Aussage wird.

Beispiel 1.4. *Aussageformen sind*

$B_1(x)$: *Die Zahl x ist gerade,*

$B_2(a, b)$: *Es gilt $a \leq b$.*

$B_3(r)$: *Hat ein Kreis Radius r, so ist seine Fläche πr^2.*

Damit ist $B_1(4)$ eine wahre Aussage und $B_1(5)$ eine falsche Aussage. Weiterhin gilt die Aussage $B_2(4, 5)$, während $B_2(5, 4)$ nicht gilt. Die Aussage $B_3(1)$ entspricht genau der Aussage B oben. △

1.2 Logische Verknüpfungen

Mithilfe logischer Verknüpfungen können wir aus einfachen Aussagen komplexere Aussagen zusammensetzen. Das Prinzip, aus einfachen Bausteinen durch bestimmte Operationen neue Objekte zu erzeugen, wird uns in dieser Vorlesung noch häufiger begegnen.

Definition 1.5. Die *Negation* einer Aussage P ist die Aussage

$\neg P$: Die Aussage P ist falsch.

Die Negation der Aussage A_2 aus dem vorigen Abschnitt erzeugt also beispielsweise die (wahre) Aussage „Die Aussage $3 + 5 = 9$ ist falsch.“

Definition 1.6. Die *Konjunktion* zweier Aussagen P, Q ist die Aussage

$P \wedge Q$: Die Aussage P ist wahr **und** die Aussage Q ist wahr.

Definition 1.7. Die *Disjunktion* zweier Aussagen P, Q ist die Aussage

$P \vee Q$: Die Aussage P ist wahr **oder** die Aussage Q ist wahr.

Eine Wahrheitstabelle enthält für jede mögliche Kombination der Wahrheitswerte der zugrunde liegenden Aussagen eine Zeile. Diese gibt auch den resultierenden Wahrheitswert der verknüpften Aussage an. Die kombinierte Wahrheitstabelle für Konjunktion und Disjunktion ist:

P	Q	$P \wedge Q$	$P \vee Q$	$P \oplus Q$
wahr	wahr	wahr	wahr	falsch
wahr	falsch	falsch	wahr	wahr
falsch	wahr	falsch	wahr	wahr
falsch	falsch	falsch	falsch	falsch

Bemerkung 1.8. Beachte, dass $P \vee Q$ auch wahr ist, wenn sowohl P als auch Q wahr sind. Die Verknüpfung, die diesen Fall ausschließt, ist in der Informatik ebenfalls wichtig und wird als *Kontravalenz* (bzw. exklusives Oder, XOR, $\oplus$ in obiger Tabelle) bezeichnet.

Bemerkung 1.9. Jede mögliche Verknüpfung von Aussagen lässt sich bereits durch Verknüpfung von Negation und entweder Konjunktion oder alternativ Disjunktion erzeugen (Übung). Für die Kontravalenz (XOR) gilt beispielsweise

$$P \oplus Q = \neg(P \wedge Q) \wedge \neg(\neg P \wedge \neg Q) = \neg(\neg P \vee Q) \vee \neg(P \vee \neg Q).$$

Für die Mathematik sind die folgenden Verknüpfungen zentral.

Definition 1.10. Die *Implikation* von Aussage P zu Aussage Q ist die Aussage

$P \Rightarrow Q$: Wenn P wahr ist, dann ist auch Q wahr.

Definition 1.11. Die *Äquivalenz* zweier Aussagen P, Q ist die Aussage

$P \Leftrightarrow Q$: Die Aussagen $P \Rightarrow Q$ und $Q \Rightarrow P$ sind beide wahr.

Es gibt viele gängige Umschreibungen dieser Verknüpfungen. Gilt $P \Rightarrow Q$, so sagen wir auch „Aus P folgt Q“ oder „P impliziert Q“ oder „P ist *hinreichende Bedingung* für Q“ oder „Q ist *notwendige Bedingung* für P“. Gilt $P \Leftrightarrow Q$, so sagen wir auch „P gilt genau dann, wenn Q gilt“ oder „P und Q sind äquivalent“ oder „P ist notwendig und hinreichend für Q“. Es lohnt sich hier ein Blick auf die Wahrheitstabelle, um Verwirrung zu vermeiden:

P	Q	$P \Rightarrow Q$	$P \Leftrightarrow Q$
wahr	wahr	wahr	wahr
wahr	falsch	falsch	falsch
falsch	wahr	wahr	falsch
falsch	falsch	wahr	wahr

Insbesondere fällt auf, dass aus einer falschen Aussage beliebige Aussagen folgen (lat. *„ex falso quodlibet“*). Um das plausibel zu machen, folgen wir einem Argument von Bertrand Russell (britischer Logiker, 1872–1970): Aus der falschen Aussage „$0 = 1$“ folgert er, dass er der Papst ist. Dazu addiert er 1 auf beiden Seiten, um die Aussage „$1 = 2$“ zu erhalten. Russell und der Papst sind zwei Personen. Wegen $2 = 1$ sind Russell und der Papst eine Person. Folglich ist Russell der Papst.

Mit Hilfe von Wahrheitstabellen können wir prüfen, dass zusammengesetzte logische Aussagen gelten. Zum Beispiel erhalten wir das Kontrapositionsgesetz, indem wir die Wahrheitstabelle für $((\neg Q) \Rightarrow (\neg P))$ mit der obigen für $P \Rightarrow Q$ vergleichen:

P	Q	$\neg Q$	$\neg P$	$\neg Q \Rightarrow \neg P$
wahr	wahr	falsch	falsch	wahr
wahr	falsch	wahr	falsch	falsch
falsch	wahr	falsch	wahr	wahr
falsch	falsch	wahr	wahr	wahr

Satz 1.12 (Kontrapositionsgesetz)**.** Sind P und Q Aussagen, dann gilt

$$(P \Rightarrow Q) \Leftrightarrow ((\neg Q) \Rightarrow (\neg P)).$$

Beispiel 1.13. *Betrachten wir die Aussagen*

P: *Jemand im Vorlesungssaal heißt Anna.*

Q: *Jemand in Darmstadt heißt Anna.*

Dann gilt $P \Rightarrow Q$, da sich der Vorlesungssaal in Darmstadt befindet. Die Aussage $\neg Q \Rightarrow \neg P$ aus dem Kontrapositionsgesetz gilt ebenfalls, also „Wenn niemand in Darmstadt Anna heißt, dann auch niemand im Vorlesungssaal.“ Die Implikation $\neg P \Rightarrow \neg Q$ gilt aber offensichtlich nicht, da es jemanden namens Anna in Darmstadt geben kann, ohne dass sich jemand namens Anna im Vorlesungsraum befindet. △

Die folgenden Zusammenhänge zwischen Negation, Konjunktion und Disjunktion lassen sich ebenso über die Wahrheitstafel nachweisen (Übung).

Satz 1.14 (De Morgan’sche Regeln)**.** Sind P und Q Aussagen, dann gelten

$$(\neg(P \vee Q)) \Leftrightarrow ((\neg P) \wedge (\neg Q))$$
$$\text{und } (\neg(P \wedge Q)) \Leftrightarrow ((\neg P) \vee (\neg Q)).$$

1.3 Logisches Schließen und Beweise

Wir haben soeben unsere ersten mathematischen Sätze (Satz 1.12 und 1.14) eingeführt. Ein *Satz* ist eine mathematische Aussage, oft in der Form einer Implikation oder Äquivalenz, für die nachgewiesen worden ist oder nachgewiesen werden kann, dass sie gilt (also wahr ist). Ein besonders wichtiger Satz wird *Theorem* genannt, ein Hilfssatz wird *Lemma*, und eine unmittelbare Konsequenz wird *Korollar* genannt. Der zweifelsfreie, schriftliche Nachweis eines Satzes heißt *Beweis*. Ein Beweis muss verständlich und überzeugend sein, sofern man die verwendeten Begriffe und Symbole kennt. Dazu ist es entscheidend, dass Verfasser:in und Leser:in des Beweises sich über die genaue Bedeutung der Begriffe einig sind. Definitionen, denen wir ebenfalls bereits begegnet sind, sind die exakten und unmissverständlichen Erklärungen der nötigen Begriffe.

Wir wenden uns jetzt dem Beweisen von mathematischen Aussagen durch logische Schlussfolgerungen zu. Wie gesagt, haben diese Aussagen in der Regel die Form einer Implikation oder einer Äquivalenz. Da Äquivalenzen aus zwei Implikationen zusammengesetzt sind, konzentrieren wir uns auf Implikationen der Form $P \Rightarrow Q$. Die Aussage P ist dabei die *Voraussetzung* und Q die *Behauptung* der Implikationsaussage. Im Fall, dass P falsch ist, gilt diese Implikation (*ex falso quodlibet*). Für den Beweis der Implikation genügt es also, den Fall zu betrachten, dass P gilt. In diesem Fall müssen wir nachweisen, dass auch Q gelten muss (siehe Wahrheitstabelle der Implikation).

Anders ausgedrückt starten wir also mit der Information, dass P wahr ist, und suchen eine Kette von Schlussfolgerungen, an deren Ende steht, dass auch Q wahr ist. Die üblichsten logischen Schlüsse sind dabei die folgenden (ihre Namen sind nicht wichtig):

- Gelten P und $(P \Rightarrow Q)$, dann auch Q. (Modus ponens)
- Gelten $(P \Rightarrow Q)$ und $\neg Q$, dann auch $\neg P$. (Modus tollens)
- Gelten $(P \vee Q)$ und $\neg P$, dann auch Q. (Elimination)
- Gelten P und Q, dann auch $P \wedge Q$.
- Gilt $P \wedge Q$, dann auch P (und Q).
- Gilt P, dann auch $P \vee Q$.

Der Modus tollens ist genau eine der Implikationen aus dem Kontrapositionsgesetz (Satz 1.12) und die anderen Schlüsse ergeben sich sofort aus den Wahrheitstafeln von Implikation, Konjunktion und Disjunktion. Auch wenn sie sofort ersichtlich sind, lohnt es sich, die Gültigkeit dieser Schlussregeln explizit an den Wahrheitstabellen nachzuvollziehen.

Beweisformen

Natürlich hängt die Form eines Beweises stark von der zu zeigenden Aussage ab und wir werden viele unterschiedliche Argumentationen sehen. Dabei treten bestimmte Grundmuster der Beweisführung und ihre Kombinationen immer wieder auf, die wir im Folgenden kurz erläutern.

Bereits aus der Schule kennen wir den *Beweis durch Umformung*, bei dem die Gleichheit zweier mathematischer Ausdrücke durch eine Kette von Gleichungen erklärt wird.

Analog zum Beweis durch Umformung zeigt der *direkte Beweis* eine Implikation durch eine einfache Kette von Schlussfolgerungen, also elementaren Implikationen. Ausgehend von der Annahme, dass die Voraussetzung

der zu zeigenden Implikation gilt, wird hergeleitet, dass auch die Behauptung gilt. Bevor wir ein Beispiel geben, führen wir eine Definition ein, um die Begrifflichkeiten zu klären.

Definition 1.15. Eine Zahl n heißt *gerade*, wenn eine ganze Zahl a existiert mit $n = 2a$, und *ungerade*, wenn eine ganze Zahl a existiert mit $n = 2a + 1$.

Insbesondere ist jede ganze Zahl entweder gerade oder ungerade, aber nicht beides.

Satz 1.16. Sind n, m gerade Zahlen, dann ist $n + m$ gerade.

Beweis (direkt). Seien n und m gerade Zahlen (wir nehmen an, die Voraussetzung gilt). Per Definition existieren also ganze Zahlen a, b mit $n = 2a$ und $m = 2b$. Dann gilt $n + m = 2a + 2b = 2(a + b)$. Also existiert $c \coloneqq a + b$ mit $n + m = 2c$.[1] Per Definition ist also $n + m$ gerade (Behauptung ist gezeigt). □

Ein *Beweis durch Äquivalenzumformung* zeigt eine Äquivalenz $P \Leftrightarrow Q$ (also zwei Implikationen auf einmal) durch eine Kette von Äquivalenzen. Zum Beispiel können wir so einen Beweis des Kontrapositionsgesetzes (Satz 1.12) formulieren.

Beweis von Satz 1.12. Seien P und Q Aussagen. Dann gilt

$$\begin{aligned}(P \Rightarrow Q) &\iff (\neg P \vee Q)\\ &\iff (Q \vee \neg P)\\ &\iff (\neg(\neg Q) \vee \neg P)\\ &\iff (\neg Q \Rightarrow \neg P).\end{aligned}$$

□

Der *Beweis durch Kontraposition* basiert auf eben diesem Kontrapositionsgesetz (Satz 1.12). Er beweist die Aussage $P \Rightarrow Q$ durch einen Beweis von $\neg Q \Rightarrow \neg P$, nimmt also an, Q gelte *nicht*, und folgert daraus, dass auch P *nicht* gilt. Wir geben zwei Beispiele.

Satz 1.17. Ist n^2 eine gerade Zahl, dann ist auch n gerade.

Beweis (Kontraposition). Sei n nicht gerade, also ungerade (wir nehmen an, die Behauptung gilt nicht). Per Definition existiert also eine ganze Zahl a mit $n = 2a+1$. Dann gilt $n^2 = (2a+1)^2 = 4a^2+4a+1 = 2(2a^2+2a)+1$. Also existiert $b \coloneqq 2a^2+2a$ mit $n^2 = 2b+1$. Per Definition ist also n^2 ungerade (Voraussetzung ist widerlegt). □

Der Beweis durch Widerspruch nimmt an, die zu zeigende Aussage $P \Rightarrow Q$ gelte nicht, also $\neg(P \Rightarrow Q)$, und leitet daraus eine absurde Aussage der Form $(C \wedge \neg C)$ her, wobei C eine beliebige Aussage ist. Offensichtlich kann $(C \wedge \neg C)$ nicht gelten, da C und $\neg C$ nicht beide wahr sein können (siehe Wahrheitstabelle). Also muss die anfängliche Annahme $\neg(P \Rightarrow Q)$ falsch gewesen sein. Also gilt $P \Rightarrow Q$.

Die Annahme $\neg(P \Rightarrow Q)$ ist gleichbedeutend mit $P \wedge \neg Q$, da dies der einzige Fall ist, in dem die Implikation falsch ist (siehe Wahrheitstabelle). Wir nehmen also sowohl P an, wie im direkten Beweis, als auch $\neg Q$, wie bei der Kontraposition. Die Herausforderung beim Widerspruchsbeweis ist, dass wir eine zusätzliche Aussage C finden müssen, die den Widerspruch bildet. Wir geben einen alternativen Beweis von Satz 1.17.

[1] Wir verwenden das Symbol „:=“, um Symbole zu definieren. Im Gegensatz dazu drückt das Symbol „=“ die Gleichheit bereits definierter Ausdrücke aus.

Beweis von Satz 1.17 (Widerspruch). Nehmen wir zum Widerspruch an, n^2 sei gerade, aber n sei ungerade (wir nehmen an, die Annahme gilt, die Behauptung aber nicht). Da n ungerade ist, existiert per Definition eine ganze Zahl a, so dass $n = 2a+1$. Dann ist $n^2 = (2a+1)^2 = 4a^2+4a+1 = 2(2a^2+2a)+1$, also ist n^2 ungerade. Das ist ein Widerspruch zur Annahme, n^2 sei gerade (die absurde Aussage $((n^2 \text{ ist gerade}) \wedge \neg(n^2 \text{ ist gerade}))$ ist bewiesen). □

Wir betrachten ein weiteres Beispiel, das auf die Pythagoräer zurückgeht. Diesmal zeigen wir eine Aussage ohne Annahme.

Satz 1.18. Die Zahl $\sqrt{2}$ kann nicht als Bruch dargestellt werden.

Beweis (Widerspruch). Angenommen, $\sqrt{2}$ könnte als Bruch dargestellt werden. Dann existieren also ganze Zahlen n und m, so dass $\frac{n}{m} = \sqrt{2}$ und $\frac{n}{m}$ so weit wie möglich gekürzt ist. Dann ist $2 = \frac{n^2}{m^2}$, also

$$n^2 = 2m^2. \tag{1.1}$$

Per Definition ist n^2 also eine gerade Zahl. Mit Satz 1.17 folgt, dass n ebenfalls gerade ist. Per Definition existiert also eine ganze Zahl a mit $n = 2a$. Eingesetzt in (1.1) ergibt sich $2m^2 = (2a)^2 = 4a^2$, also $m^2 = 2a^2$. Also ist m^2 ebenfalls gerade, mit Satz 1.17 also auch m. Aber dann lässt sich der Bruch $\frac{n}{m}$ kürzen – ein Widerspruch zur Wahl von n und m (wir haben gezeigt, dass die Aussage $\frac{n}{m}$ lässt sich kürzen sowohl falsch als auch wahr ist). □

Ein weiteres Beweisprinzip ist das der Fallunterscheidung, das wir implizit verwenden, wenn wir über die Wahrheitstabelle argumentieren. Es basiert auf der folgenden Schlussfolgerung, die es erlaubt, den Beweis einer Aussage $(P \vee Q) \Rightarrow R$ in die Fälle, dass P gilt und dass Q gilt, aufzutrennen.

Satz 1.19. Sind P, Q und R Aussagen, dann gilt

$$((P \Rightarrow R) \wedge (Q \Rightarrow R)) \Longrightarrow ((P \vee Q) \Rightarrow R)$$

Beweis. Zum Widerspruch nehmen wir an, $(P \Rightarrow R)$ und $(Q \Rightarrow R)$ gelten, aber $(P \vee Q) \Rightarrow R$ gilt nicht. Da $(P \vee Q) \Rightarrow R$ nicht gilt, gelten also $\neg R$ und $(P \vee Q)$ (Wahrheitstabelle). Nach dem Kontrapositionsgesetz (Satz 1.12) folgt aus $(P \Rightarrow R)$ und $\neg R$, dass $\neg P$ gilt. Per Elimination folgt aus $P \vee Q$ und $\neg P$, dass Q gilt. Mit Q und $(Q \Rightarrow R)$ folgt also R. Das ist ein Widerspruch zu $\neg R$. □

Zuletzt geben wir ein Beispiel eines Fallunterscheidungsarguments. Das Argument hat die Form $P \Rightarrow (A \vee B) \Rightarrow Q$, mit den folgenden Aussagen:

P: Es ist n eine ganze Zahl.

A: Die Zahl n ist gerade.

B: Die Zahl n ist ungerade.

Q: Die Zahl $n(n+1)$ ist gerade.

Die Fallunterscheidung besteht darin, den Beweis von $(A \vee B) \Rightarrow Q$ durch Beweise von $A \Rightarrow Q$ und $B \Rightarrow Q$ zu ersetzen, was wir gemäß Satz 1.19 dürfen.

Satz 1.20. Ist n eine ganze Zahl, so ist $n(n+1)$ gerade.

Beweis. Sei n eine ganze Zahl (P). Dann ist entweder n gerade oder n ungerade ($A \vee B$).

Fall 1: n ist gerade (A). Dann existiert per Definition eine ganze Zahl a mit $n = 2a$. Es folgt $n(n+1) = 2a(2a+1) = 2(2a^2+a) = 2b$ mit $b = 2a^2 + a$. Also ist $n(n+1)$ gerade. ($A \Rightarrow Q$ ist gezeigt)

Fall 2: n ist ungerade (B). Dann existiert per Definition eine ganze Zahl a mit $n = 2a+1$. Es folgt $n(n+1) = (2a+1)(2a+2) = 4a^2+4a+2a+2 = 2b$ mit $b = 2a^2+3a+1$. Also ist $n(n+1)$ gerade. ($B \Rightarrow Q$ ist gezeigt) □

1.4 Mengen

Der Begriff einer Menge ist fundamental für die Mathematik. Tatsächlich lassen sich alle anderen mathematischen Objekte auf Mengen zurückführen. In Kapitel 2 werden wir das am Beispiel der natürlichen Zahlen andeuten.

Anschaulich gesprochen ist eine *Menge* M eine ungeordnete Sammlung von *Elementen*. Wir schreiben $e_1, \dots, e_n \in M$ für die Aussage, dass die Elemente $e_1, \dots, e_n$ in der Menge M enthalten sind und $(e \notin M) \coloneqq \neg(e \in M)$ für die Aussage, dass e nicht Element der Menge M ist. Die *leere Menge* $\emptyset$ oder $\{\}$ ist die Menge, die keine Elemente enthält. Enthalten die Mengen $M_1, \dots, M_n$ nur Elemente, die auch in M enthalten sind, nennen wir sie *Teilmengen von* M und M *Obermenge* dieser Mengen und schreiben $M_1, \dots, M_n \subseteq M$ bzw. $M \supseteq M_1, \dots, M_n$. Zwei Mengen M, M' sind gleich, das heißt, es gilt $M' = M$, wenn sie dieselben Elemente enthalten, wenn also sowohl $M' \subseteq M$ als auch $M \subseteq M'$ gilt.

Wir verzichten hier bewusst auf eine formale Definition des Mengenbegriffs, denn diese stellt sich als überraschend schwierig heraus. Beispielsweise kann die Sammlung $\mathcal{S}$ aller Mengen, die sich nicht selbst enthalten, keine Menge sein. Denn sonst ergibt sich die Russell'sche Antinomie: Aus $\mathcal{S} \in \mathcal{S}$ folgt $\mathcal{S} \notin \mathcal{S}$ und aus $\mathcal{S} \notin \mathcal{S}$ folgt $\mathcal{S} \in \mathcal{S}$, in beiden Fällen ergibt sich also ein Widerspruch. Dieses scheinbare Paradox stürzte die Mathematik in eine Krise und führte zu einer sorgfältigen Untersuchung, welche Arten von Sammlungen als Mengen angesehen werden können.

Wie bereits bei der Verknüpfung logischer Aussagen beschränken wir uns auf bestimmte Erzeugungsweisen von Mengen, die solche Probleme jeweils vermeiden.

Aufzählung

Am einfachsten lässt sich eine Menge durch Aufzählung ihrer Elemente in der Form $\{\dots\}$ erzeugen. Zum Beispiel ist die Menge der natürlichen Zahlen kleiner 6 gegeben durch

$$\{0, 1, 2, 3, 4, 5\}$$

und die Menge der logischen Verknüpfungen aus Abschnitt 1.2 gegeben durch

$$\{\neg, \vee, \wedge, \Rightarrow, \Leftrightarrow\}.$$

Wir erlauben auch unendliche Aufzählungen, wie

$$\{\dots, -4, -2, 0, 2, 4, 6, \dots\}$$

für die geraden Zahlen. Besonders wichtig für uns sind die Mengen der *natürlichen Zahlen*

$$\mathbb{N} \coloneqq \{0, 1, 2, 3, \dots\}$$

und der *ganzen Zahlen*

$$\mathbb{Z} \coloneqq \{\dots, -2, -1, 0, 1, 2, \dots\}.$$

Aussonderung

Gegeben eine Grundmenge G, können wir eine neue Teilmenge durch Aussonderung von Elementen mit einer bestimmten Eigenschaft erzeugen. Die Schreibweise dafür ist

$$\{x \in G \mid E(x)\},$$

wobei $E(x)$ eine Aussageform mit Variable x ist.

So können wir zum Beispiel schreiben

$$\begin{aligned} \{0, 1, 2, 3, 4, 5\} &= \{n \in \mathbb{N} \mid n \leq 5\}, \\ \{\dots, -4, -2, 0, 2, 4, 6, \dots\} &= \{n \in \mathbb{Z} \mid n \text{ ist gerade}\}, \\ \mathbb{N} &= \{n \in \mathbb{Z} \mid n \geq 0\}. \end{aligned}$$

Zusätzlich zu den Grundmengen $\mathbb{N}$ und $\mathbb{Z}$ benötigen wir die Menge $\mathbb{R}$ der *reellen Zahlen*. Wir verwenden die Menge $\mathbb{R}$, ohne zu zeigen, wie sie sich mittels der Erzeugungsprinzipien dieses Abschnitts konstruieren lässt. Mit ihrer Hilfe können wir die Grundmenge $\mathbb{Q}$ der *rationalen Zahlen* erzeugen als

$$\mathbb{Q} \coloneqq \{x \in \mathbb{R} \mid x \text{ lässt sich schreiben als } \frac{a}{b} \text{ mit } a \in \mathbb{Z} \text{ und } b \in \mathbb{N},\, b > 0\}.$$

Vereinigung, Schnitt, Mengendifferenz

Mittels Aussonderung können wir weitere Erzeugungsoperationen für Mengen aus zwei gegebenen Mengen konstruieren (siehe Abb. 1.1).

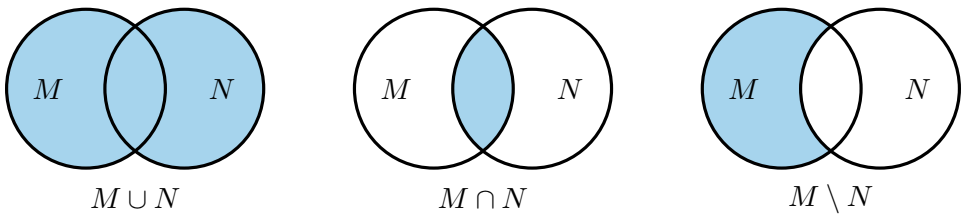


Abbildung 1.1: Illustration von Vereinigung, Schnitt und Mengendifferenz (Def. 1.21).

Definition 1.21. Seien M und N Teilmengen einer Grundmenge G. Dann ist

- $M \cup N \coloneqq \{x \in G \mid x \in M \vee x \in N\}$ die *Vereinigung* von M und N,
- $M \cap N \coloneqq \{x \in G \mid x \in M \wedge x \in N\}$ der *Schnitt* von M und N,
- $M \setminus N \coloneqq \{x \in G \mid x \in M \wedge x \notin N\}$ die *Mengendifferenz* von M und N.

Ist $M \cap N = \emptyset$, so heißen M und N *disjunkt*.

Dabei fällt auf, dass sich Schnitt und Mengendifferenz zwar unabhängig von der Grundmenge erzeugen lassen, wir aber eine explizite Grundmenge brauchen, um die Vereinigung durch Aussonderung zu erzeugen. Außerdem sei auf die Beziehung zwischen Vereinigung ($\cup$) und Disjunktion ($\vee$) sowie zwischen Schnitt ($\cap$) und Konjunktion ($\wedge$) hingewiesen.

Die folgende nützliche Definition gibt Beispiele für den Schnitt zweier Mengen.

Definition 1.22. Für $Z \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ definieren wir $Z_{\geq 0} \coloneqq Z \cap \{x \in \mathbb{R} \mid x \geq 0\}$.

Wir führen exemplarisch einen Beweis über die Beziehung zwischen Vereinigung und Schnitt vor.

Satz 1.23 (Distributivgesetz)**.** Für drei Mengen A, B, C gilt

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Beweis. Wir müssen zeigen, dass die linke Seite Teilmenge der rechten ist, und umgekehrt. Mit anderen Worten müssen wir für jedes Element $x \in A \cap (B \cup C)$ zeigen, dass auch $x \in (A \cap B) \cup (A \cap C)$ gilt, und umgekehrt.

Sei also zunächst $x \in A \cap (B \cup C)$. Dann ist $x \in A$ und $x \in B \cup C$, also $x \in B$ oder $x \in C$. Ist $x \in B$, dann folgt mit $x \in A$ zunächst $x \in A \cap B$ und damit $x \in (A \cap B) \cup (A \cap C)$. Ist $x \in C$, dann folgt mit $x \in A$ zunächst $x \in A \cap C$ und damit $x \in (A \cap B) \cup (A \cap C)$. In beiden Fällen gilt also $x \in (A \cap B) \cup (A \cap C)$.

Sei nun umgekehrt $x \in (A \cap B) \cup (A \cap C)$. Dann ist $x \in A \cap B$ oder $x \in A \cap C$. Ist $x \in A \cap B$, dann gilt $x \in A$ und $x \in B$. Insbesondere ist $x \in B \cup C$ und damit $x \in A \cap (B \cup C)$. Ist $x \in A \cap C$, dann gilt entsprechend $x \in A$ und $x \in C$, also ebenfalls $x \in A \cap (B \cup C)$. Somit ist auch die umgekehrte Teilmengenbeziehung gezeigt. □

Kartesisches Produkt

Das Erzeugungsprinzip der Aussonderung erlaubt es, beliebige Teilmengen einer gegebenen Grundmenge zu erzeugen. Wie aber erzeugen wir Mengen jenseits dieser Grundmengen, bzw. woher nehmen wir diese Grundmengen ursprünglich? Dazu brauchen wir einen Weg, neue Elemente aus bestehenden zu kombinieren. Die folgende Definition gibt uns diese Möglichkeit.

Definition 1.24. Ein *Tupel* ist eine Aneinanderreihung von endlich vielen Elementen in der Form

$$(a_1, a_2, \ldots, a_n) \qquad \text{mit } n \in \mathbb{N}.$$

Im Gegensatz zu einer Menge ist ein Tupel geordnet und kann das gleiche Element mehrfach enthalten.

Bemerkung 1.25 (für Interessierte). Ein Tupel $(a_1, a_2, \ldots, a_n)$ kann als Menge von Mengen ausgedrückt werden, indem wir es als eine Menge $\{(1, a_1), (2, a_2), \ldots, (n, a_n)\}$ von Tupeln mit zwei Elementen schreiben und jedes Tupel (i, a_i) jeweils durch die Menge $\{\{i\}, \{i, a_i\}\}$ ersetzen.

Wir führen ein weiteres Erzeugungsprinzip ein, das es erlaubt, Mengen von Tupeln zu konstruieren, wobei vorgegeben ist, aus welcher Grundmenge das Element an jeder Position des Tupels jeweils stammen soll.

Definition 1.26. Für Mengen $A_1, \ldots, A_n$ ist das *kartesische Produkt* $A_1 \times \cdots \times A_n$ eine Menge, die aus allen Tupeln $(a_1, a_2, \ldots, a_n)$ besteht, wobei $(a_1 \in A_1)$, $(a_2 \in A_2)$, $\ldots$, $(a_n \in A_n)$. Ist $A_1 = A_2 = \cdots = A_n = A$, so verwenden wir die Abkürzung $A^n \coloneqq A_1 \times \cdots \times A_n$, wobei $A^0 \coloneqq \{()\}$ die einelementige Menge ist, die aus dem leeren Tupel besteht.

Beispiel 1.27. *Die Menge der Bitstrings aus* 3 *Bits lässt sich ausdrücken als*

$$\{0,1\}^3 = \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}. \qquad \triangle$$

Beispiel 1.28. *Die Menge der gängigen Vorlesungsslots an der TU Darmstadt lässt sich erzeugen als*

$$\begin{aligned}
&\{\text{Mo}, \text{Di}, \text{Mi}, \text{Do}, \text{Fr}\} \times \{8{:}00, 9{:}50, 11{:}40, 13{:}30, 15{:}20, 17{:}10\} \\
&= \{(\text{Mo}, 8{:}00), (\text{Mo}, 9{:}50), (\text{Mo}, 11{:}40), (\text{Mo}, 13{:}30), (\text{Mo}, 15{:}20), (\text{Mo}, 17{:}10), \\
&\quad (\text{Di}, 8{:}00), (\text{Di}, 9{:}50), (\text{Di}, 11{:}40), (\text{Di}, 13{:}30), (\text{Di}, 15{:}20), (\text{Di}, 17{:}10), \\
&\quad (\text{Mi}, 8{:}00), (\text{Mi}, 9{:}50), (\text{Mi}, 11{:}40), (\text{Mi}, 13{:}30), (\text{Mi}, 15{:}20), (\text{Mi}, 17{:}10), \\
&\quad (\text{Do}, 8{:}00), (\text{Do}, 9{:}50), (\text{Do}, 11{:}40), (\text{Do}, 13{:}30), (\text{Do}, 15{:}20), (\text{Do}, 17{:}10), \\
&\quad (\text{Fr}, 8{:}00), (\text{Fr}, 9{:}50), (\text{Fr}, 11{:}40), (\text{Fr}, 13{:}30), (\text{Fr}, 15{:}20), (\text{Fr}, 17{:}10)\}. \qquad \triangle
\end{aligned}$$

Potenzmenge

Ein letztes wichtiges Konstruktionsprinzip ist die Erzeugung aller Teilmengen einer Menge. Während die Aussonderung eine Auswahl der Elemente vornimmt, machen wir jetzt alle möglichen Auswahlen zu Objekten der neuen Menge.

Definition 1.29. Die *Potenzmenge* $\mathcal{P}(M)$ einer Menge M ist die Menge aller Teilmengen von M, es gilt also $N \in \mathcal{P}(M)$ genau dann, wenn $N \subseteq M$.

Beispiel 1.30. *Es ist* $\mathcal{P}(\{0,1\}) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$. $\triangle$

Beispiel 1.31. *Die Menge aller möglichen Belegungspläne eines Raums an der TU Darmstadt für die entsprechende Auswahl an Zeiten ist*

$$\mathcal{P}(\{\text{Mo}, \text{Di}, \text{Mi}, \text{Do}, \text{Fr}\} \times \{8{:}00, 9{:}50, 11{:}40, 13{:}30, 15{:}20, 17{:}10\}). \qquad \triangle$$

1.5 All- und Existenzquantor

Wir werden häufig Aussagen über die Gesamtheit der Elemente einer Menge machen. Dafür sind die folgenden Schreibweisen praktisch.

Definition 1.32. Für eine Aussageform $E(x)$ und eine Menge M erzeugen der *Allquantor* $\forall$ und der *Existenzquantor* $\exists$ die Aussagen

$\forall x \in M\colon E(x)$: Für alle $x \in M$ gilt $E(x)$.

$\exists x \in M\colon E(x)$: Es existiert mindestens ein $x \in M$, für das $E(x)$ gilt.

Beispiel 1.33. *Wir erinnern uns an die Aussageformen*

$B_1(x)$: *Die Zahl x ist gerade.*

$B_3(r)$: *Hat ein Kreis Radius r, so ist seine Fläche πr^2.*

Damit ist $\forall n \in \mathbb{N}\colon B_1(n)$ die falsche Aussage A_6, dass jede natürliche Zahl gerade ist. Andererseits gilt offensichtlich $\exists n \in \mathbb{N}\colon B_1(n)$, denn es gibt gerade natürliche Zahlen. Ebenso gilt die Aussage $\forall r \in \{x \in \mathbb{R} \mid x > 0\}\colon B_3(r)$. $\triangle$

Die folgenden Zusammenhänge zwischen den Quantoren sind nützlich, um zu zeigen, dass eine Aussage mit Quantor *nicht* gilt.

Für eine Aussageform $E(x)$ und eine Menge M gelten

$$\neg(\forall x \in M\colon E(x)) \Leftrightarrow \exists x \in M\colon \neg E(x),$$
$$\neg(\exists x \in M\colon E(x)) \Leftrightarrow \forall x \in M\colon \neg E(x).$$

Wir wenden All- und Existenzquantor an, um die Notation für Vereinigung und Schnitt auf mehr als zwei Mengen zu erweitern.

Definition 1.34. Für eine Menge $\mathcal{M}$ von Teilmengen einer Grundmenge G definieren wir Vereinigung und Schnitt über $\mathcal{M}$ als

$$\textstyle\bigcup_{M\in\mathcal{M}} M \coloneqq \{x \in G \mid \exists M \in \mathcal{M}\colon x \in M\},$$
$$\textstyle\bigcap_{M\in\mathcal{M}} M \coloneqq \{x \in G \mid \forall M \in \mathcal{M}\colon x \in M\}.$$

Beachte, dass der leere Schnitt $\bigcap_{M\in\emptyset} M = G$ von der Grundmenge G abhängt.

1.6 Relationen

Die folgende Definition erlaubt es uns, die Elemente zweier Mengen in Beziehung zueinander zu setzen.

Definition 1.35. Eine *(binäre) Relation* zwischen den Mengen N, M ist eine Teilmenge $R \subseteq N \times M$. Ist $(x, y) \in R$ für $x \in N$ und $y \in M$, so schreiben wir xRy. Ist $N = M$, so bezeichnen wir R als *Relation auf der Menge* M (bzw. N).

In diesem Abschnitt betrachten wir ausschließlich Relationen auf einer Menge.

Definition 1.36. Wir werden in Kapitel 2 sehen, dass wir den vertrauten Vergleich von natürlichen Zahlen als die folgende Relation auffassen können:

$$\leq \; = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid \exists k \in \mathbb{N} \colon m = n + k\}.$$

Eine Relation R auf einer Menge M mit endlich vielen Elementen lässt sich gut durch einen *gerichteten Graphen* darstellen: Für jedes Element $a \in M$ zeichnen wir einen *Knoten* und wir zeichnen eine *gerichtete Kante* (also einen Pfeil) von dem Knoten eines Elements a zum Knoten eines Elements b, falls aRb, also $(a, b) \in R$ gilt.

Beispiel 1.37. *Sei* $M \coloneqq \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\} = \mathcal{P}(\{1, 2, 3\})$ *und betrachte die Relation*

$$S \coloneqq \{(A, B) \in M \times M \mid A \cap B \neq \emptyset\}$$

zwischen Mengen aus M*, die mindestens ein Element gemeinsam haben. Der Graph dieser Relation ist in Abbildung 1.2 dargestellt.* △

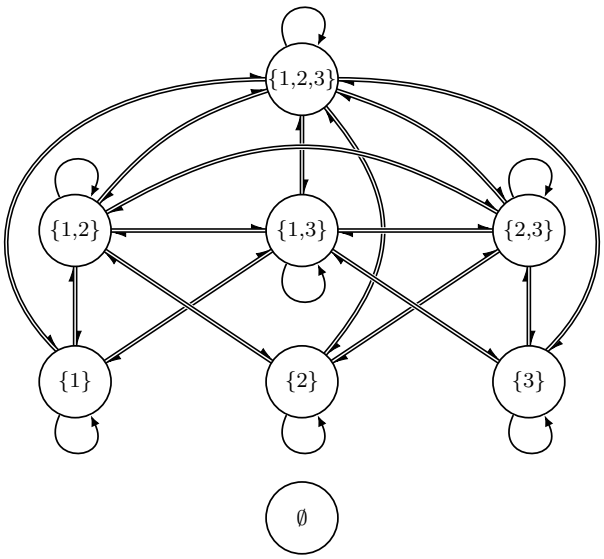


Abbildung 1.2: Gerichteter Graph der Relation S aus Beispiel 1.37.

Definition 1.38. Eine Relation R auf einer Menge X heißt

(i) *reflexiv*, falls $\forall x \in X : xRx$,

(ii) *symmetrisch*, falls $\forall x, y \in X : (xRy \Rightarrow yRx)$,

(iii) *antisymmetrisch*, falls $\forall x, y \in X : ((xRy \wedge yRx) \Rightarrow (x = y))$,

(iv) *transitiv*, falls $\forall x, y, z \in X : ((xRy \wedge yRz) \Rightarrow xRz)$.

Beispiel 1.39. *Die Relation $\leq$ auf $\mathbb{N}$ ist reflexiv, denn für jedes $n \in \mathbb{N}$ gilt $n = n + 0$ und damit $n \leq n$. Sie ist nicht symmetrisch, denn es gilt zwar $2 \leq 5$, aber nicht $5 \leq 2$. Die Relation ist antisymmetrisch, gelten nämlich $n \leq m$ und $m \leq n$, so gibt es per Definition $k, l \in \mathbb{N}$ mit $m = n + k$ und $n = m + l$. Daraus folgt $n = n + k + l$ und damit $k = l = 0$, also $n = m$. Schließlich ist* $\leq$ transitiv, *denn aus $n \leq m$ und $m \leq p$ folgt die Existenz von $k, l \in \mathbb{N}$ mit $m = n + k$ und $p = m + l$. Dann ist $p = n + (k + l)$ und wegen $k + l \in \mathbb{N}$ gilt $n \leq p$.* △

Beispiel 1.40. *Die Relation $< := \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid n \leq m \wedge n \neq m\}$ ist nicht reflexiv, nicht symmetrisch, aber transitiv. Sie ist auch antisymmetrisch, weil die Menge $\{(n, m) \in \mathbb{N} \times \mathbb{N} \mid n < m \wedge m < n\}$ leer ist und damit alle ihre Elemente (nämlich keine) die Antisymmetriebedingung erfüllen.* △

Auf einer endlichen Menge können wir die Eigenschaften einer Relation an ihrem Graphen ablesen: Sie ist reflexiv, wenn jeder Knoten mit sich selbst verbunden ist; symmetrisch, wenn es für jede Kante auch die entgegengesetzte Kante gibt; antisymmetrisch, wenn zwischen zwei verschiedenen Knoten nie Kanten in beiden Richtungen verlaufen; transitiv, wenn es jede Abkürzungskante zwischen Knoten gibt, zwischen denen ein gerichteter Weg verläuft.

Beispiel 1.41. *Die Relation S aus Beispiel 1.37 und Abb. 1.2 ist*

- *nicht reflexiv (wegen $\emptyset \cap \emptyset = \emptyset$), und im Graphen hat $\emptyset$ keine Kante zu sich selbst,*
- *symmetrisch (wegen $(A \cap B \neq \emptyset) \Leftrightarrow (B \cap A \neq \emptyset)$), und im Graphen treten Kanten in gegengerichteten Paaren auf,*
- *nicht antisymmetrisch (wegen beispielsweise $\{1\} S \{1, 2\}$ und $\{1, 2\} S \{1\}$, aber $\{1\} \neq \{1, 2\}$), und im Graphen treten gegengerichtete Kanten auf,*
- *nicht transitiv (wegen beispielsweise $\{1\} S \{1, 2\}$ und $\{1, 2\} S \{2\}$, aber nicht $\{1\} S \{2\}$), und im Graphen gibt es nicht jede Abkürzung (beispielsweise von $\{1\}$ nach $\{2\}$).* △

Ordnungsrelationen

Definition 1.42. Eine Relation R auf einer Menge M heißt *Ordnungsrelation*, falls R reflexiv, antisymmetrisch und transitiv ist. Man schreibt dann oft $\preccurlyeq := R$ und nennt $(M, \preccurlyeq)$ eine *partiell geordnete Menge*. Gilt zusätzlich $\forall x, y \in M : x \preccurlyeq y \vee y \preccurlyeq x$, so heißt $\preccurlyeq$ eine *Totalordnung* und $(M, \preccurlyeq)$ eine *total geordnete Menge*.

Beispiel 1.43. *Die Relation $\leq$ auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ ist jeweils eine Totalordnung.* △

Beispiel 1.44. *Für jede Menge M ist $\subseteq$ eine Ordnungsrelation auf $\mathcal{P}(M)$. Sie ist im Allgemeinen keine Totalordnung (zum Beispiel gilt weder $\{1\} \subseteq \{2\}$ noch $\{2\} \subseteq \{1\}$).* △

Beobachtung 1.45. Ist $(M, \preccurlyeq)$ partiell (total) geordnet, so auch $(N, \preccurlyeq \cap N^2)$ für $N \subseteq M$.
Wir schreiben kurz $(N, \preccurlyeq)$ für $(N, \preccurlyeq \cap N^2)$.

Definition 1.46. Sei $(M, \preccurlyeq)$ eine partiell geordnete Menge. Dann heißt $x \in M$

(i) *Maximum* von $(M, \preccurlyeq)$, falls $y \preccurlyeq x$ für alle $y \in M$,

(ii) *Minimum* von $(M, \preccurlyeq)$, falls $x \preccurlyeq y$ für alle $y \in M$.

Satz 1.47. Jede partiell geordnete Menge $(M, \preccurlyeq)$ hat höchstens ein Maximum und höchstens ein Minimum.

Beweis. Seien x und y jeweils Maxima oder jeweils Minima von $(M, \preccurlyeq)$. Dann gilt sowohl $x \preccurlyeq y$ als auch $y \preccurlyeq x$. Aus der Antisymmetrie der Ordnungsrelation $\preccurlyeq$ folgt damit $x = y$. Also kann es keine zwei verschiedenen Maxima oder Minima geben. □

Definition 1.48. Sofern sie existieren, bezeichnen wir das eindeutige Minimum einer partiell geordneten Menge $(M, \preccurlyeq)$ mit $\min M$ und ihr eindeutiges Maximum mit $\max M$.

Beispiel 1.49. *Sei $I := \{x \in \mathbb{R} \mid 0 < x \leq 1\}$. Dann hat die total geordnete Menge $(I, \leq)$ kein Minimum und das Maximum $\max I = 1$.* △

Beispiel 1.50. *Die total geordnete Menge $(\mathbb{N}, \leq)$ hat ein Minimum (nämlich 0), aber kein Maximum.* △

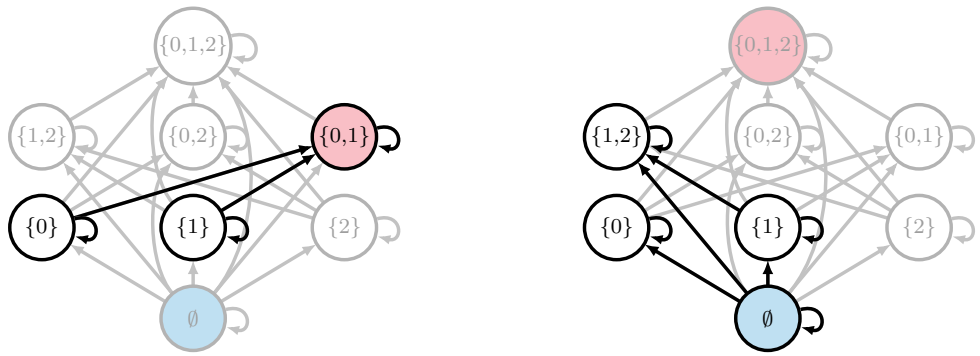


Abbildung 1.3: Die partiell geordneten Mengen aus den Beispielen 1.51 und 1.57 in $(\mathcal{P}(\{0,1,2\}), \subseteq)$; Infima sind hellblau und Suprema hellrot markiert. Elemente und Relationen der jeweiligen Teilordnung sind schwarz hervorgehoben, die übrigen sind ausgegraut.

Beispiel 1.51. *Die partiell geordnete Menge $(\mathcal{P}(\{0,1,2\}), \subseteq)$ hat ein Maximum und ein Minimum, hingegen hat $(\{\{0\},\{1\},\{0,1\}\}, \subseteq)$ ein Maximum, aber kein Minimum, und $(\{\emptyset,\{0\},\{1\},\{1,2\}\}, \subseteq)$ ein Minimum, aber kein Maximum. Die beiden partiell geordneten Mengen sind in Abbildung 1.3 dargestellt.* △

Definition 1.52. Sei $(M, \preccurlyeq)$ eine partiell geordnete Menge und $N \subseteq M$. Dann gilt:

(i) Ein Element $x \in M$ heißt *obere Schranke* von N in $(M, \preccurlyeq)$, falls $y \preccurlyeq x$ für alle $y \in N$.

(ii) Ein Element $x \in M$ heißt *untere Schranke* von N in $(M, \preccurlyeq)$, falls $x \preccurlyeq y$ für alle $y \in N$.

(iii) Die Menge N heißt *nach oben beschränkt* beziehungsweise *nach unten beschränkt*, falls sie eine obere beziehungsweise eine untere Schranke besitzt. Sie heißt *beschränkt*, falls sie nach oben und nach unten beschränkt ist.

Definition 1.53. Sei $(M, \preccurlyeq)$ eine partiell geordnete Menge und $N \subseteq M$. Sofern es jeweils existiert, heißt

(i) $\sup N := \min\{x \in M \mid x \text{ ist obere Schranke von } N \text{ in } (M, \preccurlyeq)\}$ das *Supremum* von N in $(M, \preccurlyeq)$,

(ii) $\inf N := \max\{x \in M \mid x \text{ ist untere Schranke von } N \text{ in } (M, \preccurlyeq)\}$ das *Infimum* von N in $(M, \preccurlyeq)$.

Merksatz:

Das Supremum ist die kleinste obere Schranke. Das Infimum ist die größte untere Schranke.

Beispiel 1.54. *In $(\mathbb{R}, \leq)$ hat die Menge $I := \{x \in \mathbb{R} \mid 0 < x \leq 1\}$ aus Beispiel 1.49 zwar kein Minimum, wohl aber untere Schranken, beispielsweise -7. Das Maximum der Menge der unteren Schranken und somit das Infimum ist $\inf I = 0 \notin I$.* △

Beispiel 1.55. *In $(\mathbb{Z}, \leq)$ hat die Menge $\mathbb{N}$ das Infimum $\inf \mathbb{N} = 0 = \min \mathbb{N}$, aber kein Supremum.* △

Beispiel 1.56. *In $(\mathbb{Q}, \leq)$ hat die Menge $\{x \in \mathbb{Q} \mid x \geq 0 \text{ und } x^2 < 2\}$ obere Schranken, zum Beispiel 2 oder 37, aber kein Supremum: Die Menge der oberen Schranken ist $\{q \in \mathbb{Q} \mid q \geq \sqrt{2}\}$ und diese Menge hat kein Minimum, denn $\sqrt{2} \notin \mathbb{Q}$ (Satz 1.18), aber wir können uns $\sqrt{2}$ mit rationalen Zahlen beliebig annähern.* △

Beispiel 1.57. *Wir betrachten die Mengen in $(\mathcal{P}(\{0,1,2\}), \subseteq)$ aus Beispiel 1.51 und Abbildung 1.3. Die Menge $A = \{\{0\}, \{1\}, \{0,1\}\}$ hat die oberen Schranken $\{0,1\}$ und $\{0,1,2\}$, also das Supremum $\sup A = \{0,1\} = \max A$, sowie das Infimum $\inf A = \emptyset \notin A$. Die Menge $B = \{\emptyset, \{0\}, \{1\}, \{1,2\}\}$ hat das Infimum $\inf B = \emptyset = \min B$ und das Supremum $\sup B = \{0,1,2\} \notin B$.* △

Bei den obigen Beispielen fällt auf, dass das Maximum immer auch das Supremum ist und das Minimum immer auch das Infimum. Die Umkehrungen gelten im Allgemeinen jedoch nicht.

Satz 1.58. Sei $(M, \preccurlyeq)$ eine partiell geordnete Menge und $N \subseteq M$. Es gilt:

(i) $\max N$ existiert genau dann, wenn $\sup N$ existiert und $\sup N \in N$. In diesem Fall ist $\max N = \sup N$.

(ii) $\min N$ existiert genau dann, wenn $\inf N$ existiert und $\inf N \in N$. In diesem Fall ist $\min N = \inf N$.

Beweis. Wir zeigen die erste Aussage, die zweite ist analog.

Existiert zunächst $m \coloneqq \max N$, so ist per Definition $m \in N$ und m eine obere Schranke von N. Ist $x \in M$ eine beliebige obere Schranke von N, so gilt $m \preccurlyeq x$, da $m \in N$. Somit ist $m \preccurlyeq x$ für alle oberen Schranken, also ist m das Minimum der Menge aller oberen Schranken von N. Damit existiert $\sup N$ und es gilt $\sup N = m \in N$.

Existiert umgekehrt $\sup N$ und gilt $\sup N \in N$, so folgt aus der Eigenschaft als obere Schranke, dass $y \preccurlyeq \sup N$ für alle $y \in N$ gilt. Damit ist $\sup N$ das Maximum von N und es gilt $\max N = \sup N$. □

Äquivalenzrelationen

Um mathematische Mengen zu konstruieren, ist es oft nützlich, große Mengen von Darstellungen zu generieren und dann gleichwertige Darstellungen zusammenzufassen. Beispielsweise können wir die rationalen Zahlen als die Menge aller Brüche konstruieren, indem wir gleichwertige Brüche wie $1/2, 2/4, 3/6, \dots$ jeweils als dasselbe Objekt betrachten. Für die Menge aller Winkel können wir Zahlen aus $\mathbb{R}$ zusammenfassen, die sich durch ganze Vielfache von 2π unterscheiden. Welche Darstellungen jeweils als dasselbe Objekt gelten sollen, legen wir durch eine Relation fest.

Definition 1.59. Eine Relation R auf einer Menge M heißt *Äquivalenzrelation*, falls sie reflexiv, symmetrisch und transitiv ist. Man schreibt dann oft $\sim \coloneqq R$.

Beispiel 1.60. *Die folgenden Relationen auf $M \coloneqq \{-1, 1, 2, 3, 4\}$ sind Äquivalenzrelationen (siehe Abbildung 1.4):*

$$
\begin{aligned}
\sim_1 &\coloneqq \{(x, y) \in M \times M \mid x = y\} \\
&= \{(-1,-1), (1,1), (2,2), (3,3), (4,4)\}, \\
\sim_2 &\coloneqq \{(x, y) \in M \times M \mid x \text{ und } y \text{ haben dieselbe Parität}\} \\
&= \big\{(-1,-1), (1,1), (2,2), (3,3), (4,4), (-1,1), (1,-1), (-1,3), (3,-1), (1,3), (3,1), (2,4), (4,2)\big\}, \\
\sim_3 &\coloneqq \{(x, y) \in M \times M \mid x \text{ und } y \text{ haben dasselbe Vorzeichen}\} \\
&= \big\{(-1,-1), (1,1), (2,2), (3,3), (4,4), (1,2), (2,1), (1,3), (3,1), (1,4), (4,1), \\
&\qquad (3,4), (4,3), (2,3), (3,2), (2,4), (4,2)\big\}, \\
\sim_4 &\coloneqq \{(x, y) \in M \times M \mid x \text{ und } y \text{ haben dieselbe Parität und dasselbe Vorzeichen}\} \\
&= \big\{(-1,-1), (1,1), (2,2), (3,3), (4,4), (1,3), (3,1), (2,4), (4,2)\big\}.
\end{aligned}
$$

△

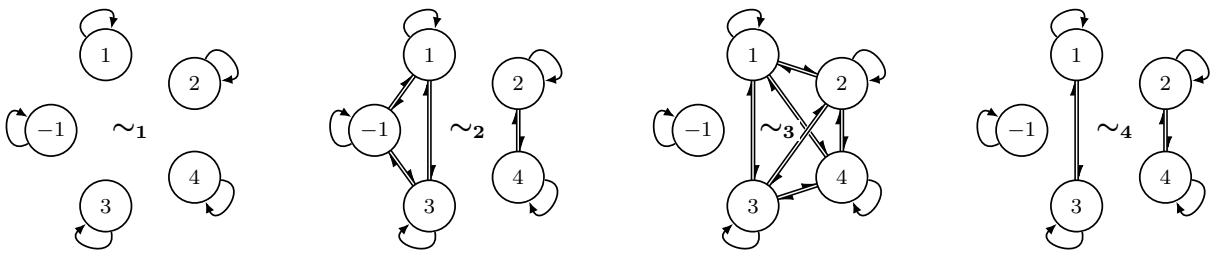


Abbildung 1.4: Die Graphen der Äquivalenzrelationen aus Beispiel 1.60.

Definition 1.61. Die *Äquivalenzklasse* von $a \in M$ bezüglich der Äquivalenzrelation $\sim$ auf einer Menge M ist

$$[a] \coloneqq \{x \in M \mid x \sim a\}.$$

Beispiel 1.62. *Für die Relationen aus Abbildung 1.4 sind die verschiedenen Äquivalenzklassen*

	$[-1]$	$[1]$	$[2]$	$[3]$	$[4]$
$\sim_1$	$\{-1\}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$
$\sim_2$	$\{-1,1,3\}$	$\{-1,1,3\}$	$\{2,4\}$	$\{-1,1,3\}$	$\{2,4\}$
$\sim_3$	$\{-1\}$	$\{1,2,3,4\}$	$\{1,2,3,4\}$	$\{1,2,3,4\}$	$\{1,2,3,4\}$
$\sim_4$	$\{-1\}$	$\{1,3\}$	$\{2,4\}$	$\{1,3\}$	$\{2,4\}$

△

Satz 1.63. Für jede Äquivalenzrelation $\sim$ auf einer Menge M gilt

(i) $a \in [a]$ für jedes $a \in M$, also insbesondere $[a] \neq \emptyset$,

(ii) $\bigcup_{a \in M} [a] = M$,

(iii) für alle $a, b \in M$ mit $[a] \neq [b]$ gilt $[a] \cap [b] = \emptyset$.

Beweis. Wegen der Reflexivität gilt $a \sim a$, also $a \in [a]$. Damit folgen die erste Aussage und die Inklusion $M \subseteq \bigcup_{a \in M} [a]$. Da andererseits jede Äquivalenzklasse per Definition eine Teilmenge von M ist, folgt auch die umgekehrte Inklusion und somit die zweite Aussage.

Die dritte Aussage zeigen wir mittels Kontraposition. Sei also $[a] \cap [b] \neq \emptyset$. Dann existiert ein $x \in [a] \cap [b]$, für das somit $x \sim a$ und $x \sim b$ gilt. Mit Symmetrie folgt $a \sim x$ und mit Transitivität dann $a \sim b$. Ist nun $y \in [a]$, so gilt $y \sim a \sim b$ und daher $y \in [b]$. Somit ist $[a] \subseteq [b]$. Analog folgt $[b] \subseteq [a]$, also $[a] = [b]$. □

Definition 1.64. Die Menge

$$M/\!\sim \coloneqq \{U \in \mathcal{P}(M) \mid \exists a \in M \colon U = [a]\}$$

aller Äquivalenzklassen heißt *Faktormenge* von M bezüglich $\sim$.

Beispiel 1.65. *Die Menge $\mathbb{Q}$ lässt sich aus $\mathbb{Z}$ konstruieren als die Faktormenge $B/\!\sim$ von $B \coloneqq \{(z,n) \in \mathbb{Z}^2 \mid n > 0\}$ bezüglich der Äquivalenzrelation $\sim \coloneqq \{((z_1,n_1),(z_2,n_2)) \in B \times B \mid z_1 n_2 = z_2 n_1\}$. Dabei entspricht B der Menge der Brüche mit Zähler z und Nenner n und Äquivalenzklassen bestehen aus allen Darstellungen desselben Bruchs. Die Relation $\sim$ ist offensichtlich reflexiv und symmetrisch. Transitivität gilt, da $(z_1,n_1) \sim (z_2,n_2) \sim (z_3,n_3)$ impliziert, dass*

$$n_2(z_1 n_3) = (z_1 n_2) n_3 = (z_2 n_1) n_3 = (z_2 n_3) n_1 = (z_3 n_2) n_1 = n_2(z_3 n_1),$$

wobei wir mit $n_2 > 0$ kürzen dürfen und so $z_1 n_3 = z_3 n_1$ bzw. $(z_1,n_1) \sim (z_3,n_3)$ erhalten. △

1.7 Abbildungen

Von besonderer Bedeutung in der Mathematik sind eindeutige Relationen, also Relationen $R \subseteq X \times Y$, bei denen jedes Element von X in Relation zu einem eindeutigen Element von Y steht. Im Gegensatz zum vorigen Abschnitt betrachten wir nun allgemeiner Relationen zwischen zwei (möglicherweise) verschiedenen Mengen X und Y.

Definition 1.66. Eine *Abbildung* ist eine Relation $f \subseteq X \times Y$ zwischen zwei Mengen X und Y, so dass für jedes Element $x \in X$ genau ein $f(x) \in Y$ existiert mit $(x, f(x)) \in f$. Wir definieren eine konkrete Abbildung f durch Angabe der *Definitionsmenge* X, der *Zielmenge* Y und des *Bildes* $f(x)$ jedes *Punktes* $x \in X$ in der Schreibweise

$$f\colon \begin{cases} X \to Y \\ x \mapsto f(x) \end{cases} \qquad \text{oder} \qquad f\colon X \to Y,\ x \mapsto f(x),$$

wobei die *Abbildungsvorschrift* $x \mapsto f(x)$ auch unspezifiziert bleiben kann. Abbildungen zwischen Zahlenräumen nennen wir auch *Funktionen*. Eine Abbildung $\diamond\colon X \times X \to X$ heißt *(binäre) Verknüpfung* auf X und wir schreiben $x \diamond y \coloneqq \diamond(x, y)$.

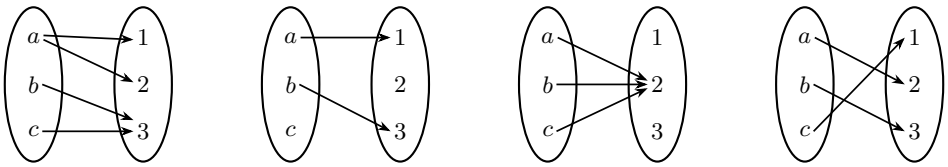


Abbildung 1.5: Die vier Relationen aus Beispiel 1.67.

Beispiel 1.67. *Wir betrachten die folgenden Relationen zwischen den Mengen* $\{a, b, c\}$ *und* $\{1, 2, 3\}$ *(siehe Abb. 1.5):*

- $\{(a, 1), (a, 2), (b, 3), (c, 3)\}$ *ist keine Abbildung, da* a *mehreren Elementen zugeordnet wird,*
- $\{(a, 1), (b, 3)\}$ *ist keine Abbildung, da* c *keinem Element zugeordnet wird,*
- $\{(a, 2), (b, 2), (c, 2)\}$ *ist eine Abbildung,*
- $\{(a, 2), (b, 3), (c, 1)\}$ *ist ebenfalls eine Abbildung.* △

Beispiel 1.68. *Die folgenden sind jeweils Abbildungen:*

- *die* Identität $\mathrm{id}_M\colon M \to M,\ x \mapsto x$ *(oder einfach* id*, wenn* M *aus dem Kontext klar ist) auf der Menge* M*,*
- *die Additionsfunktion* $+\colon Z \times Z \to Z$ *und Multiplikationsfunktion* $\cdot\colon Z \times Z \to Z$ *für* $Z \in \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ *sind jeweils Verknüpfungen auf* Z*,*
- *das* Polynom $f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ *von* Grad n *mit* $(a_n, \ldots, a_0) \in \mathbb{R}^{n+1}$.
- *die* kanonische Abbildung $\nu\colon M \to M/\!\sim,\ x \mapsto [x]$ *für eine Äquivalenzrelation* $\sim$ *auf der Menge* M. △

Definition 1.69. Für eine Abbildung $f\colon X \to Y$ ist

(i) das *Bild* von $A \subseteq X$ unter f die Menge $f(A) \coloneqq \{y \in Y \mid \exists x \in A\colon f(x) = y\}$,

(ii) das *Urbild* von $B \subseteq Y$ unter f die Menge $f^{-1}(B) \coloneqq \{x \in X \mid f(x) \in B\}$.

Beobachtung 1.70. Per Definition der Abbildung $f\colon X \to Y$ gilt $f^{-1}(Y) = X$. Andersherum gilt nur $f(X) \subseteq Y$, aber nicht immer Gleichheit.

Definition 1.71. Für eine Abbildung $g\colon X \to Y$, $x \mapsto g(x)$ und $A \subseteq X$ definieren wir die Schreibweise

$$\{g(a) \mid a \in A\} \coloneqq g(A).$$

Dabei geben wir die Abbildung nicht eigens an, wenn ihre Zuordnung aus der Schreibweise und die Mengen X und Y aus dem Zusammenhang klar sind.

Beispiel 1.72. *Es ist*

$$\{2a \mid a \in \{0,1,2\}\} = \{z \in \mathbb{N} \mid \exists a \in \{0,1,2\}\colon z = 2a\} = \{0,2,4\}$$
$$\{a+b \mid a \in \{0,1\},\, b \in \{0,2\}\} = \{z \in \mathbb{N} \mid \exists a \in \{0,1\}, b \in \{0,2\}\colon z = a+b\} = \{0,1,2,3\}$$

mit den impliziten Abbildungen $g\colon \mathbb{N} \to \mathbb{N}$, $z \mapsto 2z$ *und* $h\colon \mathbb{N}^2 \to \mathbb{N}$, $(a,b) \mapsto a+b$. △

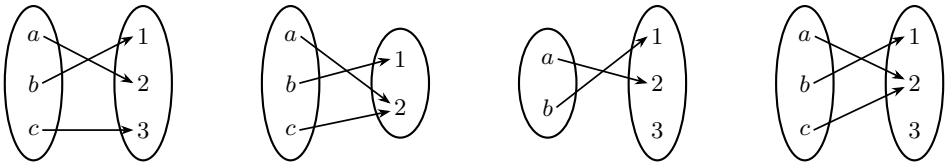


Abbildung 1.6: Von links nach rechts: bijektiv; surjektiv, aber nicht injektiv; injektiv, aber nicht surjektiv; weder surjektiv noch injektiv. Die erste Abbildung lässt sich durch Umdrehen der Pfeile umkehren.

Jede Relation $R \subseteq X \times Y$ lässt sich genauso als Relation $R' \subseteq Y \times X$ verstehen. Die Frage liegt nahe, unter welchen Umständen sich diese Symmetrie auch auf Abbildungen überträgt. Wollen wir eine Abbildung umkehren, müssen wir sicherstellen, dass die neue Abbildung wieder jedem Element mindestens und höchstens ein Element zuordnet. Die beiden resultierenden Bedingungen sind die Folgenden (siehe Abbildung 1.6).

Definition 1.73. Eine Abbildung $f\colon X \to Y$ heißt

(i) *surjektiv*, falls $f(X) = Y$,

(ii) *injektiv*, falls $\forall x_1, x_2 \in X\colon (x_1 \neq x_2) \Rightarrow (f(x_1) \neq f(x_2))$,

(iii) *bijektiv*, falls sie surjektiv und injektiv ist. Dann ist f eine *Bijektion* von X nach Y.

Wir vergewissern uns formal, dass Surjektivität und Injektivität die Eindeutigkeit garantieren.

Satz 1.74. Eine Abbildung $f\colon X \to Y$ ist genau dann bijektiv, wenn für jedes $y \in Y$ genau ein $x \in X$ mit $f(x) = y$ existiert.

Beweis. Sei f zunächst bijektiv. Wegen der Surjektivität von f existiert per Definition zu jedem $y \in Y$ mindestens ein $x \in X$ mit $f(x) = y$. Sind $x_1, x_2 \in X$ mit $f(x_1) = f(x_2) = y$, so folgt aus der Injektivität $x_1 = x_2$. Somit gibt es genau ein solches x.

Nehmen wir umgekehrt an, dass für jedes $y \in Y$ ein eindeutiges $x \in X$ mit $f(x) = y$ existiert. Dann ist $f(X) = Y$, also f surjektiv. Sind $x_1, x_2 \in X$ mit $f(x_1) = f(x_2) = y$, so sind x_1 und x_2 Urbilder desselben Elements $y \in Y$. Aus der Eindeutigkeit folgt $x_1 = x_2$, also ist f auch injektiv und damit bijektiv. □

Wir halten jetzt fest, dass sich bijektive Abbildungen umkehren lassen.

Definition 1.75. Die *Verkettung* zweier Abbildungen $f\colon X \to Y$ und $g\colon Y \to Z$ ist die Abbildung

$$g \circ f\colon X \to Z,\ x \mapsto (g \circ f)(x) \coloneqq g(f(x)).$$

Satz 1.76. Eine Abbildung $f\colon X \to Y$ ist genau dann bijektiv, wenn eine Abbildung $f^{-1}\colon Y \to X$ existiert mit

$$f^{-1} \circ f = \mathrm{id}_X \qquad \text{und} \qquad f \circ f^{-1} = \mathrm{id}_Y\,.$$

In diesem Fall ist f^{-1} eindeutig bestimmt.

Beweis. Angenommen, es existiert ein solches f^{-1}. Aus $f(x_1) = f(x_2)$ folgt $x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2$, also ist f injektiv. Für jedes $y \in Y$ gilt $y = f(f^{-1}(y))$, also ist f surjektiv und damit bijektiv.

Sei andersherum f bijektiv. Nach Satz 1.74 existiert dann für jedes $y \in Y$ genau ein zugehöriges $x_y \in X$ mit $f(x_y) = y$. Wir definieren die Abbildung $f^{-1}\colon Y \to X,\ y \mapsto x_y$. Für alle $y \in Y$ gilt per Definition $f(f^{-1}(y)) = f(x_y) = y$, also $f \circ f^{-1} = \mathrm{id}_Y$. Betrachten wir nun $x \in X$. Für $y \coloneqq f(x)$ ist $x_y = x$ und wir erhalten $f^{-1}(f(x)) = f^{-1}(y) = x_y = x$, also $f^{-1} \circ f = \mathrm{id}_X$.

Sei nun $g\colon Y \to X$ eine weitere Abbildung mit diesen Eigenschaften. Für jedes $y \in Y$ gilt dann $f(g(y)) = y = f(f^{-1}(y))$. Wegen der Injektivität von f folgt $g(y) = f^{-1}(y)$. Somit ist $g = f^{-1}$ und f^{-1} somit eindeutig. □

Definition 1.77. Die eindeutig bestimmte Abbildung f^{-1} aus Satz 1.76 heißt *Umkehrabbildung* (bzw. *Umkehrfunktion*) von f.

Beispiel 1.78. *Die Funktion $f\colon \mathbb{Z} \to \mathbb{Z}_{\geq 0},\ x \mapsto x^2$ ist weder surjektiv, wegen $2 \notin f(\mathbb{Z}) = \{0, 1, 4, 9, 16, \dots\}$, noch injektiv, wegen $f(1) = f(-1)$. Sie lässt sich auch nicht umkehren zu einer Funktion $f^{-1}\colon \mathbb{Z}_{\geq 0} \to \mathbb{Z}$, da $\sqrt{2} \notin \mathbb{Z}$ und da $f^{-1}(\{1\}) = \{-1, 1\}$ mehr als ein Element enthält und somit $f^{-1}(1)$ nicht eindeutig zugeordnet werden kann.* △

Satz 1.79. Ist f eine bijektive Abbildung, dann auch f^{-1} und es gilt $(f^{-1})^{-1} = f$.

Beweis. Wegen Surjektivität von f existiert zu jedem $y \in Y$ mindestens ein $x \in X$ mit $f(x) = y$ und wegen Injektivität auch höchstens eins. Also existiert genau ein x mit $f(x) = y$, bzw. mit $f^{-1}(y) = x$. Satz 1.74 angewendet auf f^{-1} ergibt, dass f^{-1} bijektiv ist.

Sei $g\colon Y \to X$, $y \mapsto f^{-1}(y)$. Dann gilt per Definition der Umkehrabbildung für alle $x \in X$, dass $g^{-1}(x) = y$ mit dem eindeutigen $y \in Y$, so dass $x = g(y)$. Also ist $(f^{-1})^{-1}(x) = y$ mit $f^{-1}(y) = x$ bzw. $f(x) = y$ für alle $x \in X$. Es folgt $(f^{-1})^{-1}(x) = f(x)$ für alle $x \in X$, also $(f^{-1})^{-1} = f$. □

Definition 1.80. Die *Einschränkung* von $f\colon X \to Y$ auf $A \subseteq X$ ist die Abbildung

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

Beispiel 1.81. *Die Funktion $f\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$, $x \mapsto x^2$ ist surjektiv, aber nicht injektiv, wegen $f(1) = f(-1)$. Ihre Einschränkung $f|_{\mathbb{R}_{\geq 0}}$ ist hingegen bijektiv mit der Umkehrfunktion $(f|_{\mathbb{R}_{\geq 0}})^{-1}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $x \mapsto \sqrt{x}$.* △

Satz 1.82. Sind $f\colon X \to X'$ und $g\colon Y \to Y'$ bijektiv, so auch die folgenden Abbildungen:

(i) falls $X' = Y$, die Verkettung $g \circ f\colon X \to Y'$;

(ii) für jede Teilmenge $A \subseteq X$ die Einschränkung $f|_A\colon A \to f(A)$;

(iii) falls $X \cap Y = \emptyset$ und $X' \cap Y' = \emptyset$, die Abbildung

$$h\colon (X \cup Y) \to (X' \cup Y'),\ x \mapsto \begin{cases} f(x), & \text{falls } x \in X, \\ g(x), & \text{falls } x \in Y; \end{cases}$$

(iv) die Abbildung

$$p\colon (X \times Y) \to (X' \times Y'),\ (x, y) \mapsto (f(x), g(y)).$$

Beweis. Wir wenden Satz 1.79 auf die jeweils gegebenen Bijektionen f und g an um f^{-1} und g^{-1} zu erhalten. Wiederum nach Satz 1.79 genügt es jeweils, eine Umkehrabbildung anzugeben.

(i) Eine Umkehrabbildung von $g \circ f$ ist $f^{-1} \circ g^{-1}$, denn $(f^{-1} \circ g^{-1}) \circ (g \circ f) = \mathrm{id}_X$ wegen $f^{-1}(g^{-1}(g(f(x)))) = f^{-1}(f(x)) = x$ und, analog, $(g \circ f) \circ (f^{-1} \circ g^{-1}) = \mathrm{id}_Y$.

(ii) Eine Umkehrabbildung von $f|_A\colon A \to f(A)$ ist die Einschränkung $f^{-1}|_{f(A)}\colon f(A) \to A$, denn $(f^{-1}|_{f(A)}) \circ (f|_A) = \mathrm{id}_A$ wegen $f^{-1}|_{f(A)}(f|_A(a)) = f^{-1}|_{f(A)}(f(a)) = f^{-1}(f(a)) = a$ für $a \in A$ und, analog, $(f|_A) \circ (f^{-1}|_{f(A)}) = \mathrm{id}_{f(A)}$.

(iii) Wegen der Disjunktheit der jeweiligen Mengen sind h und die Abbildung

$$k\colon (X' \cup Y') \to (X \cup Y),\ x \mapsto \begin{cases} f^{-1}(x), & \text{falls } z \in X', \\ g^{-1}(x), & \text{falls } z \in Y', \end{cases}$$

wohldefiniert. Die Abbildung k ist eine Umkehrabbildung von h, denn $k \circ h = \mathrm{id}_{X \cup Y}$ wegen $k(h(x)) = k(f(x)) = x$ für $x \in X$ und $k(h(x)) = k(g(x)) = x$ auch für $x \in Y$. Analog ist $h \circ k = \mathrm{id}_{X' \cup Y'}$.

(iv) Die Abbildung $q\colon (X' \times Y') \to (X \times Y)$, $(x', y') \mapsto (f^{-1}(x'), g^{-1}(y'))$ ist eine Umkehrabbildung von p, denn $q \circ p = \mathrm{id}_{X \times Y}$ wegen $(q \circ p)(x, y) = q(p(x, y)) = q((f(x), f(y)) = (x, y)$ und analog $p \circ q = \mathrm{id}_{X' \times Y'}$. □

2 Zählen und Arithmetik

In diesem Kapitel führen wir die Arithmetik, also das Rechnen mit natürlichen Zahlen, auf den Mengenbegriff zurück. Dabei werden wir sehen, wie sich die üblichen Rechenoperationen als Zählen der Elemente geeigneter Mengen verstehen lassen. Die aus der Schule bekannte Arithmetik dient uns so als Fallstudie für die mathematische Methode: Wir führen vertraute Begriffe auf präzise Definitionen zurück und leiten bekannte Rechenregeln daraus rigoros her.

Am Ende dieses Kapitels werden wir, auf der Basis weniger grundlegender Begriffe, eine beträchtliche Sammlung von Rechenregeln entwickelt haben. In späteren Kapiteln werden wir das gleiche Maß an Rigorosität nicht beibehalten können, aber stillschweigend darauf vertrauen, dass es prinzipiell möglich wäre.

2.1 Die natürlichen Zahlen

Durch Zählen suchen wir eine Antwort auf die Frage „Wie viele?“. Dabei zählen wir ab: „Null“, „Eins“, „Zwei“, „Drei“, ..., wobei wir jede Anzahl irgendwann erreichen können. Doch was genau tun wir, wenn wir zählen? Wie können wir diesen Prozess und die verwendeten Zahlen mathematisch präzise fassen?

Um zu zählen, brauchen wir Zahlen, wobei die verwendete Sprache („0“, „Null“, „zero“, „零“, etc.) dabei natürlich unerheblich ist. Wir brauchen einen Anfang (die Zahl Null) und eine eindeutige nächste Zahl für jede Zahl. Wir dürfen natürlich auch nicht „im Kreis“ zählen, dürfen also immer nur neue Zahlen erreichen — genauer dürfen wir die Zahl Null von keiner anderen Zahl erreichen und jede andere Zahl nur einmal. Außerdem sollten wir auch jede Zahl wirklich erreichen können.

Diese Bedingungen zusammen reichen als Grundlage der natürlichen Zahlen. Wir geben eine präzise Definition.

Definition 2.1. Das Tupel (N, n_0, succ) erfüllt die *Peano-Axiome* für die Menge N, wenn gilt:

P1. Es ist $n_0 \in N$. (Existenz der Null).

P2. Es ist $\mathrm{succ}\colon N \to N$ eine Abbildung. (Jedes Element $n \in N$ hat einen Nachfolger $\mathrm{succ}(n) \in N$).

P3. Für alle $n \in N$ gilt $\mathrm{succ}(n) \neq n_0$ (Die Null ist kein Nachfolger).

P4. Die Abbildung succ ist injektiv (Elemente haben jeweils unterschiedliche Nachfolger).

P5. Für alle $M \subseteq N$ mit $n_0 \in M$ und $\mathrm{succ}(M) \subseteq M$ gilt $M = N$
(Die Menge N ist kleinstmöglich, es gibt also keine überflüssigen Elemente).

Wir können jetzt die natürlichen Zahlen definieren und beweisen, dass sie die Peano-Axiome erfüllen.

Definition 2.2. Die Menge der *natürlichen Zahlen* ist definiert als

$$\mathbb{N} \coloneqq \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \dots\},$$

wobei jedes Element die Menge der vorigen Elemente ist. Wir verwenden die Abkürzungen $0 \coloneqq \emptyset$, $1 \coloneqq \{0\} = \{\emptyset\}$, $2 \coloneqq \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, $3 \coloneqq \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$, usw.

Definition 2.3. Per Definition ist $\subseteq$ eine Totalordnung auf $\mathbb{N}$ und wir schreiben $\leq \coloneqq \subseteq$ für diese. Wir definieren außerdem $< \coloneqq \{(a, b) \in \mathbb{N}^2 \mid a \in b\}$.

Satz 2.4. Das Tupel $(\mathbb{N}, 0, \mathrm{succ})$ mit $\mathrm{succ}\colon \mathbb{N} \to \mathbb{N}$, $n \mapsto (n \cup \{n\})$ erfüllt die Peano-Axiome.

Beweis. P1 gilt wegen $0 = \emptyset \in \mathbb{N}$.

P2 gilt, da jedes $n \in \mathbb{N}$ genau als die Menge der vorigen Elemente in der Aufzählung bei der Definition von $\mathbb{N}$ definiert ist. Damit ist $n \cup \{n\} = \mathrm{succ}(n)$ die Menge aller Elemente bis einschließlich n und damit das nächste Element der Aufzählung. Also ist $\mathrm{succ}(n) \in \mathbb{N}$.

P3 gilt, da für jedes $n \in \mathbb{N}$ gilt, dass $n \in \mathrm{succ}(n)$ und somit $\mathrm{succ}(n) \neq \emptyset$.

Für P4 seien $n, m \in \mathbb{N}$ mit $n \neq m$, so dass n vor m bei der Aufzählung in der Definition von $\mathbb{N}$ steht. Per Definition besteht $\mathrm{succ}(n)$ aus n und den in der Aufzählung davor liegenden Zahlen. Also ist $m \notin \mathrm{succ}(n)$. Andererseits ist $m \in \mathrm{succ}(m)$, also ist $\mathrm{succ}(n) \neq \mathrm{succ}(m)$. Damit ist succ injektiv.

Für P5 sei $M \subseteq \mathbb{N}$ mit $0 \in M$ und $\mathrm{succ}(M) \subseteq M$. Dann ist also $\mathrm{succ}(n) \in M$ für alle $n \in M$. Mit $0 \in M$ folgt dann $1 = \mathrm{succ}(0) \in M$, und so weiter. Somit enthält M alle Elemente der Aufzählung von $\mathbb{N}$, also gilt $\mathbb{N} \subseteq M$. Zusammen mit $M \subseteq \mathbb{N}$ folgt $M = \mathbb{N}$. □

Wir vergewissern uns noch, dass das fünfte Peano-Axiom tatsächlich überflüssige Elemente ausschließt.

Satz 2.5. Jedes $n \in \mathbb{N} \setminus \{0\}$ ist Nachfolger einer anderen Zahl, es existiert also $m \in \mathbb{N}$ mit $\mathrm{succ}(m) = n$.

Beweis. Setze $M \coloneqq \{0\} \cup \mathrm{succ}(\mathbb{N})$. Dann ist $0 \in M$ und $\mathrm{succ}(M) \subseteq \mathrm{succ}(\mathbb{N}) \subseteq M$. Nach Satz 2.4 und dem fünften Peano-Axiom ist also $M = \mathbb{N}$. Es folgt $\mathrm{succ}(\mathbb{N}) = \mathbb{N} \setminus \{0\}$, also ist jedes $n \in \mathbb{N} \setminus \{0\}$ im Bild von succ. □

2.2 Vollständige Induktion

Auf Basis der rigorosen Definition des vorigen Abschnitts wollen wir zentrale Eigenschaften der natürlichen Zahlen beweisen und Rechengesetze herleiten. Aber wie können wir Eigenschaften aller Zahlen der unendlichen Menge $\mathbb{N}$ beweisen? Offensichtlich können wir nicht jede Zahl einzeln betrachten.

Betrachten wir beispielsweise die naheliegende Eigenschaft, dass $n \neq \mathrm{succ}(n)$ für alle natürlichen Zahlen gilt. Wegen des dritten Peano-Axioms gilt $0 \neq \mathrm{succ}(0) = 1$. Wegen des vierten Peano-Axioms impliziert $0 \neq 1$ auch

$\mathrm{succ}(0) \neq \mathrm{succ}(1)$, also $1 \neq \mathrm{succ}(1) = 2$, was wiederum $2 \neq \mathrm{succ}(2)$ impliziert, und so weiter. Es ist plausibel, dass wir so $n \neq \mathrm{succ}(n)$ für jedes $n \in \mathbb{N}$ beweisen können, aber wie können wir das präzise fassen?

Ein anschauliches Argument ist das Folgende: Wäre $n = \mathrm{succ}(n)$ für $n \neq 0$, dann hätte keine andere Zahl n als Nachfolger (wegen Injektivität von succ) und wir könnten die Zahl n ausgehend von 0 nie mittels succ erreichen. Damit wäre n überflüssig, im Widerspruch zum fünften Peano-Axiom.

Wir zeigen, wie wir diese Überlegungen in ein wiederverwendbares formales Argument übersetzen können.

Satz 2.6 (vollständige Induktion)**.** Sei $E(n)$ eine Aussageform für $n \in \mathbb{N}$. Gelten der *Induktionsanfang* $E(0)$ und der *Induktionsschritt* $\forall n \in \mathbb{N}\colon (E(n) \Rightarrow E(\mathrm{succ}(n)))$, dann gilt $E(n)$ für alle $n \in \mathbb{N}$.

Beweis. Wir betrachten die Menge $M \coloneqq \{n \in \mathbb{N} \mid E(n)\}$ aller natürlichen Zahlen n, für die $E(n)$ gilt. Nach Annahme gilt $E(0)$, also ist $0 \in M$. Für alle $n \in M$ gilt per Definition $E(n)$, also nach dem Induktionsschritt auch $E(\mathrm{succ}(n))$. Zusammen mit $\mathrm{succ}(n) \in \mathbb{N}$ folgt $\mathrm{succ}(n) \in M$ für alle $n \in M$. Also haben wir $\mathrm{succ}(M) \subseteq M$, außerdem ist $M \subseteq \mathbb{N}$ und $0 \in M$. Nach Satz 2.4 erfüllt $(\mathbb{N}, 0, \mathrm{succ})$ das Peano-Axiom P5, also ist $M = \mathbb{N}$. Per Definition von M gilt somit $E(n)$ für alle $n \in \mathbb{N}$. □

Zur Veranschaulichung des Induktionsprinzips können wir uns die Aussagen $E(0), E(1), E(2), \ldots$ als eine Kette von Dominosteinen vorstellen, wobei die Induktionsannahme $E(n)$ dem Umfallen des n-ten Steins entspricht. Der Induktionsschritt entspricht der Tatsache, dass ein umfallender Stein das Umfallen des jeweils nächsten auslöst (also dass die Steine nah genug aneinander stehen). Der Induktionsanfang entspricht dem Umfallen des ersten Steins. Gelten Induktionsanfang und Induktionsschritt, sind wir überzeugt, dass alle Steine umfallen.

Mit vollständiger Induktion können wir unser anfängliches Beispiel sauber beweisen.

Satz 2.7. Für alle $n \in \mathbb{N}$ gilt $n \neq \mathrm{succ}(n)$.

Beweis. Wir beweisen per Induktion mit der Aussageform $E(n) \coloneqq (n \neq \mathrm{succ}(n))$.

Induktionsanfang: Die Aussage $E(0) = (0 \neq \mathrm{succ}(0))$ ist wahr, wegen des dritten Peano-Axioms und Satzes 2.4.

Induktionsschritt: Sei $n \in \mathbb{N}$, so dass die Induktionsannahme $E(n) = (n \neq \mathrm{succ}(n))$ gilt. Nach dem vierten Peano-Axiom und Satz 2.4 ist succ injektiv. Also impliziert $n \neq \mathrm{succ}(n)$, dass $\mathrm{succ}(n) \neq \mathrm{succ}(\mathrm{succ}(n))$. Mit anderen Worten impliziert also $E(n)$ auch $E(\mathrm{succ}(n))$.

Nach Satz 2.6 ist somit $E(n)$ für alle $n \in \mathbb{N}$ gezeigt. □

Beispiel 2.8. *Der Induktionsschritt alleine genügt offensichtlich nicht, um eine Aussage für alle $n \in \mathbb{N}$ zu zeigen: Wir haben soeben (Satz 2.7) gezeigt, dass die Aussage $E'(n) \coloneqq (n = \mathrm{succ}(n))$ für jedes $n \in \mathbb{N}$ falsch ist. Dennoch gilt $\forall n \in \mathbb{N}\colon E'(n) \Rightarrow E'(\mathrm{succ}(n))$, denn mit $n = \mathrm{succ}(n)$ ist auch $\mathrm{succ}(n) = \mathrm{succ}(\mathrm{succ}(n))$. Der Induktionsanfang $E'(0)$ gilt hier nicht.* △

Beobachtung 2.9. Die vollständige Induktion lässt sich vielfältig abwandeln:

(i) Wir können als Induktionsannahme $E(n) \coloneqq (\forall m \in \{0, 1, \dots, n\} \colon F(m))$ statt der Annahme $F(n)$ verwenden und den Induktionsschritt zeigen als $(\forall m \in \{0, 1, \dots, n\} \colon F(m)) \Rightarrow F(\mathrm{succ}(n))$, denn $E(n) \wedge F(\mathrm{succ}(n))$ impliziert $E(\mathrm{succ}(n)) = (\forall m \in \{0, 1, \dots, n, \mathrm{succ}(n)\} \colon F(m))$, so dass der Schritt für $E(n)$ gezeigt ist. Die Induktion liefert dann $E(n)$ für alle $n \in \mathbb{N}$, also auch $F(n)$ für alle $n \in \mathbb{N}$.

(ii) Wollen wir $F(n)$ nur für alle $n \in \mathbb{N}$ ab einem $n_0 \in \mathbb{N}$ zeigen, können wir die Induktionsannahme $E(n) \coloneqq F(m(n))$ verwenden, wobei $m(n) \in \mathbb{N}$ die n-te natürliche Zahl nach n_0 bezeichnet.

(iii) Wollen wir $F(m)$ für alle $m \in M$ mit $M \neq \mathbb{N}$ zeigen, genügt eine Bijektion $f \colon \mathbb{N} \to M$. Dann können wir die Induktionsannahme $E(n) \coloneqq F(f(n))$ verwenden. Haben wir $E(n)$ für alle $n \in \mathbb{N}$ gezeigt, dann auch $F(m) = E(f^{-1}(m))$ für alle $m \in M$, wegen Surjektivität von f^{-1}.

2.3 Endliche Mengen

Wir wenden uns jetzt dem Zählen zu. Dazu brauchen wir eine Menge, deren Elemente wir zählen wollen. Aber wie definieren wir die Anzahl Elemente einer Menge? Hier machen wir es uns zunutze, dass Zahlen bereits als Mengen definiert sind. Insbesondere wissen wir anschaulich bereits, dass die Zahlen aus einer wachsenden Anzahl an Elementen bestehen, angefangen bei der leeren Menge 0. Wir können also die Größe einer endlichen Menge durch einen Vergleich festlegen: Eine Menge hat n Elemente, wenn sie genauso viele Elemente wie die Menge n besitzt. Das „genauso viele“ können wir formal mittels einer Bijektion beschreiben, die die Elemente der beiden Mengen paarweise einander zuordnet.

Definition 2.10. Eine Menge M heißt *endlich*, wenn eine Bijektion $f \colon n \to M$ für ein $n \in \mathbb{N}$ existiert. Dann heißt $|M| \coloneqq n$ die *Kardinalität* von M.

Da jede natürliche Zahl eine Menge ist, die alle vorigen Zahlen enthält, ist die Kardinalität einer endlichen Menge damit eindeutig bestimmt. Per Definition ist insbesondere $|n| = n$. Das Zählen ist aus diesem Blickwinkel genau die schrittweise Konstruktion der Bijektion in Definition 2.10: Wir zählen „1“ und haben dabei das erste Element der Menge im Blick, dann zählen wir „2“ für das nächste Element, usw.

Beim Zählen ist unerheblich, wie die Elemente einer Menge bezeichnet sind.

Satz 2.11. Ist A eine endliche Menge und B eine Menge, für die eine Bijektion $h \colon A \to B$ existiert, so ist auch B endlich und es gilt $|B| = |A|$.

Beweis. Da A endlich ist, existiert eine Bijektion $f \colon |A| \to A$. Nach Satz 1.82 ist auch $h \circ f \colon |A| \to B$ bijektiv. Somit ist B endlich und nach Definition der Kardinalität gilt $|B| = |A|$. □

Satz 2.12. Ist M eine endliche Menge und $a \notin M$, dann ist auch $M \cup \{a\}$ endlich und $|M \cup \{a\}| = \mathrm{succ}(|M|)$.

Beweis. Da M endlich ist, existiert eine Bijektion $f\colon n \to M$ mit $n \coloneqq |M|$. Wir erweitern f zu einer Bijektion $\widetilde{f}\colon \operatorname{succ}(n) \to (M \cup \{a\})$, indem wir $\widetilde{f}(x) \coloneqq f(x)$ für alle $x \in n$ setzen und $\widetilde{f}(n) \coloneqq a$. Wegen $\operatorname{succ}(n) = n \cup \{n\}$ ist das die gewünschte Bijektion. □

Wir wenden vollständige Induktion an, um nachzuweisen, dass endliche Mengen nach den Erzeugungsprinzipien von Abschnitt 1.4 wieder endliche Mengen erzeugen.

Satz 2.13. Ist M eine endliche Menge und $N \subseteq M$, so ist auch N endlich.

Beweis. Wir zeigen zunächst per Induktion über n, dass jede Teilmenge $S \subseteq n$ endlich ist.

Induktionsanfang: Für $n = 0$ ist $S = \emptyset$ und damit endlich.

Induktionsschritt: Die Aussage gelte nun für ein $n \in \mathbb{N}$ und es sei $S \subseteq \operatorname{succ}(n)$. Dann ist $T \coloneqq S \cap n$ nach der Induktionsannahme endlich. Ist $n \notin S$, so gilt $S = T$. Andernfalls ist $S = T \cup \{n\}$ und wegen $n \notin T$ nach Satz 2.12 ebenfalls endlich. Also ist in beiden Fällen auch S endlich.

Sei nun $N \subseteq M$ und $f\colon |M| \to M$ bijektiv. Die Menge $S \coloneqq f^{-1}(N) \subseteq |M|$ ist nach dem soeben Bewiesenen endlich, und wegen $f(S) = N$ ist die Einschränkung $f|_S\colon S \to N$ nach Satz 1.82 bijektiv. Nach Satz 2.11 ist damit auch N endlich. □

Satz 2.14. Sind M und N endliche Mengen, so sind auch $M \cup N$, $M \cap N$, $M \setminus N$, $M \times N$, $\mathcal{P}(M)$ endlich.

Beweis. Übung. □

Satz 2.15. Jede nichtleere endliche total geordnete Menge $(M, \preccurlyeq)$ besitzt ein Minimum und ein Maximum.

Beweis. Wir beweisen per Induktion über n, dass jede total geordnete Menge der Kardinalität n entweder leer ist oder ein Minimum und ein Maximum besitzt.

Induktionsanfang: Für $n = 0$ ist die Menge leer.

Induktionsschritt: Die Aussage gelte nun für ein $n \in \mathbb{N}$, und sei $(M, \preccurlyeq)$ total geordnet mit $|M| = \operatorname{succ}(n)$. Per Definition existiert eine Bijektion $f\colon \operatorname{succ}(n) \to M$ und wir setzen $a \coloneqq f(n)$ und $A \coloneqq M \setminus \{a\}$. Die Einschränkung $f|_n\colon n \to A$ ist bijektiv, also gilt $|A| = n$. Ist A leer, so ist a sowohl Minimum als auch Maximum von $M = A \cup \{a\}$. Andernfalls besitzt A nach der Induktionsannahme ein Minimum u und ein Maximum v. Da $\preccurlyeq$ Totalordnung und insbesondere transitiv ist, gilt: Ist $a \preccurlyeq u$, dann ist a ein Minimum von M, sonst u; ist $a \preccurlyeq v$, dann ist v ein Maximum von M, sonst a. □

Satz 2.16. Jede nichtleere Teilmenge $N \subseteq \mathbb{N}$ besitzt ein Minimum.

Beweis. Da $N \neq \emptyset$, existiert ein $n \in N$. Setze $S \coloneqq N \cap \operatorname{succ}(n)$, so dass $n \in S$, also $S \neq \emptyset$. Als Teilmenge der endlichen Menge $\operatorname{succ}(n)$ ist S nach Satz 2.13 endlich und besitzt nach Satz 2.15 ein Minimum m bezüglich der Totalordnung $\leq = \subseteq$. Sei nun $n' \in N$. Da $\leq$ Totalordnung ist, gilt entweder $n' \leq n$ oder $n \leq n'$. Im ersten Fall ist $n' \in \{n\}$ (also $n' = n$) oder $n' \in n$ (also $n' < n$), also $n' \in \operatorname{succ}(n)$ und damit $n' \in S$ wegen $n' \in N$ – wegen Minimalität von m ist dann $m \leq n'$. Im zweiten Fall gilt mit Minimalität von m und $n \in S$, dass $m \leq n$, was mit Transitivität auch $m \leq n'$ impliziert. Somit ist m in beiden Fällen ein Minimum von N. □

2.4 Addition

Arithmetik, also das Rechnen mit natürlichen Zahlen, ist nichts weiter als eine Sammlung von Techniken, die das Zählen von Mengen beschleunigen. Die erste solche Technik, der man im Leben begegnet, ist in der Regel die Addition. Mit ihr können wir die Kardinalität der Vereinigung zweier endlicher Mengen bestimmen, deren Kardinalität wir bereits kennen, ohne neu zählen zu müssen. Beachte, dass die Vereinigung zweier endlicher Mengen nach Satz 2.14 wieder endlich ist.

Wir können also die Addition über die Vereinigung zweier Mengen definieren. Die Definition betrachtet die Vereinigung der Zahlenmengen, die den Anzahlen der zwei Mengen entsprechen. Wir müssen bei der Vereinigung zweier Zahlenmengen allerdings vorsichtig sein, denn Zahlenmengen zu verschiedenen natürlichen Zahlen (außer 0) verwenden die gleichen Elemente: Zum Beispiel ist $4 \cup 7 = 7$ wegen $4 = \{0, 1, 2, 3\}$ und $7 = \{0, 1, 2, 3, 4, 5, 6\}$. Damit jedes Element bei der Vereinigung erhalten bleibt, markieren wir die Elemente der zwei Mengen als $4 \times \{0\} = \{(0, 0), (1, 0), (2, 0), (3, 0)\}$ und $7 \times \{1\} = \{(0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1)\}$. Wegen $0 \neq 1$ ist die Vereinigung dieser Mengen dann

$$\{(0, 0), (1, 0), (2, 0), (3, 0), (0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1)\},$$

was zwar nicht genau der Zahlenmenge 11 entspricht, aber die richtige Kardinalität hat.

Definition 2.17. Wir definieren die *Addition* auf $\mathbb{N}$ als die Verknüpfung

$$+ \colon \mathbb{N}^2 \to \mathbb{N},\ (a, b) \mapsto |(a \times \{0\}) \cup (b \times \{1\})|$$

und nennen $a + b$ die *Summe* von $a \in \mathbb{N}$ und $b \in \mathbb{N}$.

Insbesondere gilt $a + 0 = a$ für alle $a \in \mathbb{N}$.

Wir stellen den Zusammenhang zwischen den Abbildungen succ und $+$ her.

Satz 2.18 (Nachfolgerregel)**.** Für alle $a, b \in \mathbb{N}$ gilt

$$\mathrm{succ}(a) + b = \mathrm{succ}(a + b).$$

Insbesondere ist $\mathrm{succ}(b) = \mathrm{succ}(0 + b) = \mathrm{succ}(0) + b = 1 + b$.

Beweis. Mit $\mathrm{succ}(a) = a \cup \{a\}$ und Satz 2.12 erhalten wir

$$\begin{aligned} \mathrm{succ}(a) + b &= |(\mathrm{succ}(a) \times \{0\}) \cup (b \times \{1\})| \\ &= |(a \times \{0\}) \cup (b \times \{1\}) \cup (\{a\} \times \{0\})| \\ &= \mathrm{succ}(|(a \times \{0\}) \cup (b \times \{1\})|) \\ &= \mathrm{succ}(a + b). \end{aligned}$$

□

Mit der Addition ist auch die Grundlage für die Beschreibung der Relation $\leq$ in Beispiel 1.36 geschaffen.

Satz 2.19. Für die Relation $\leq$ auf $\mathbb{N}$ und $a, b \in \mathbb{N}$ gilt

$$a \leq b \qquad \Longleftrightarrow \qquad \exists k \in \mathbb{N} \colon b = a + k.$$

Beweis. Übung. □

Wir zeigen, dass die Zahl k in Satz 2.19 jeweils eindeutig bestimmt ist.

Satz 2.20 (Kürzungsregel)**.** Für alle $a, k, \ell \in \mathbb{N}$ gilt

$$a + k = a + \ell \quad \Longrightarrow \quad k = \ell.$$

Beweis. Übung. □

Per Definition der Addition besteht ein direkter Zusammenhang zwischen der Vereinigung von Zahlenmengen und der Summe ihrer Kardinalitäten. Dieser Zusammenhang lässt sich auf allgemeine endliche Mengen erweitern. Wir fordern, dass die Mengen disjunkt sind und kommen so um die Markierungen $\times\{0\}$ und $\times\{1\}$ herum.

Satz 2.21. Für disjunkte endliche Mengen A und B (d.h., $A \cap B = \emptyset$) gilt

$$|A \cup B| = |A| + |B|.$$

Beweis. Aus Bijektionen $f \colon |A| \to A$ und $g \colon |B| \to B$ erhalten wir die Bijektionen $f_0 \colon (|A| \times \{0\}) \to A$, $(x, 0) \mapsto f(x)$ und $g_1 \colon (|B| \times \{1\}) \to B$, $(y, 1) \mapsto g(y)$. Wegen der Disjunktheit der jeweiligen Definitions- und Zielmengen liefert Satz 1.82 eine Bijektion $h \colon ((|A| \times \{0\}) \cup (|B| \times \{1\})) \to (A \cup B)$. Nach Satz 2.11 und der Definition der Addition folgt

$$|A \cup B| = |(|A| \times \{0\}) \cup (|B| \times \{1\})| = |A| + |B|.$$ □

Aus dem Zusammenhang mit der Vereinigung erhalten wir unmittelbar die folgenden Eigenschaften.

Korollar 2.22. Für alle $n, m, r \in \mathbb{N}$ gilt

(i) $n + m = m + n$ (Kommutativität),

(ii) $(n + m) + r = n + (m + r)$ (Assoziativität).

Wegen der Assoziativität können wir $n + m + r \coloneqq (n + m) + r$ schreiben, ohne dass das zu Mehrdeutigkeit führt.

Wir definieren zuletzt Notation für Vielfachsummen.

Definition 2.23. Seien $n_0, n \in \mathbb{N}$ und sei $a \colon \{n_0, n_0 + 1, \ldots, n\} \to Z$ eine Abbildung, wobei Z eine Menge ist, für die eine Addition $+ \colon Z^2 \to Z$ definiert ist. Dann schreiben wir

$$\sum_{i=n_0}^{n} a(i) \coloneqq a(n_0) + a(n_0 + 1) + \cdots + a(n).$$

Subtraktion

Für festes $a \in \mathbb{N}$ ist die Abbildung $f_a \colon \mathbb{N} \to \mathbb{N}$, $n \mapsto n + a$ im Allgemeinen nicht surjektiv, beispielsweise liegt 2 nicht im Bild von f_5. Insbesondere hat f_a im Allgemeinen keine Umkehrabbildung. Wir definieren die Subtraktion als eine partielle Umkehrung der Addition.

Definition 2.24. Sei $R \coloneqq \{(b, a) \in \mathbb{N}^2 \mid a \leq b\}$. Wir definieren die *Subtraktion* auf $\mathbb{N}$ als die Abbildung

$$- \colon R \to \mathbb{N},\ (b, a) \mapsto k,$$

wobei $k \in \mathbb{N}$ die nach Satz 2.19 und Satz 2.20 eindeutig bestimmte Zahl ist, so dass $b = a + k$.
Wir schreiben die *Differenz* von $b \in \mathbb{N}$ und $a \in \mathbb{N}$ als

$$b - a \coloneqq -(b, a).$$

Insbesondere gilt $a - 0 = a$ für alle $a \in \mathbb{N}$. Außerdem ist die Subtraktion tatsächlich eine bedingte Umkehrung der Addition.

Beobachtung 2.25. Für alle $a, b, c \in \mathbb{N}$ mit $a \leq c$ gilt

$$a + b = c \quad \iff \quad c - a = b.$$

Die Subtraktion tritt auf natürliche Weise beim Zählen von Differenzmengen auf.

Satz 2.26. Für endliche Mengen A und B mit $A \subseteq B$ gilt

$$|B \setminus A| = |B| - |A|.$$

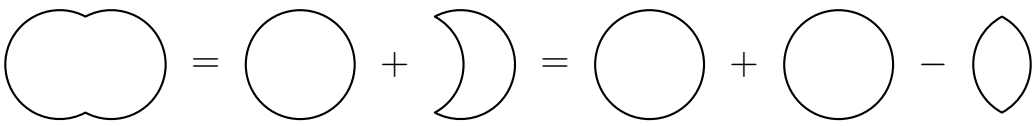
Beweis. Die Mengen A und $B \setminus A$ sind disjunkt und es gilt $B = A \cup (B \setminus A)$. Nach Satz 2.21 folgt $|B| = |A| + |B \setminus A|$. Insbesondere ist $|A| \leq |B|$. Mit Beobachtung 2.25 folgt unmittelbar $|B \setminus A| = |B| - |A|$. □

Die folgende Formel für die Vereinigung zweier beliebiger Mengen ist anschaulich klar.

Satz 2.27. Für endliche Mengen A und B gilt

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Beweis. Wir geben einen visuellen Beweis und überlassen die Formalisierung als Übungsaufgabe.



□

2.5 Multiplikation

Wir kennen die Multiplikation als Technik zum Zählen mehrerer identischer Zeilen oder Spalten. Ein Schachbrett besteht beispielsweise aus den Zeilen 1–8 und den Spalten A–H, also lässt sich die Menge der Felder als kartesisches Produkt $\{1,2,3,4,5,6,7,8\} \times \{\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D},\mathrm{E},\mathrm{F},\mathrm{G},\mathrm{H}\}$ schreiben. Die Kardinalität dieser Menge definiert das Produkt der Kardinalitäten der Zeilen- und Spaltenmengen.

Definition 2.28. Wir definieren die *Multiplikation* auf $\mathbb{N}$ als die Verknüpfung

$$\cdot\colon \mathbb{N}^2 \to \mathbb{N},\ (a,b) \mapsto |a \times b|$$

und nennen $ab \coloneqq a \cdot b$ das *Produkt* von $a \in \mathbb{N}$ und $b \in \mathbb{N}$.

Insbesondere gilt $a \cdot 1 = a$ für alle $a \in \mathbb{N}$.

Wir stellen den Zusammenhang zwischen den Abbildungen succ und $\cdot$ her.

Satz 2.29 (Nachfolgerregel)**.** Für alle $a, b \in \mathbb{N}$ gilt

$$a \cdot (b+1) = a \cdot \mathrm{succ}(b) = a \cdot b + a.$$

Beweis. Da die Mengen $a \times b$ und $a \times \{b\}$ disjunkt sind, gilt mit $\mathrm{succ}(b) = b \cup \{b\}$, Satz 2.21 und $|a \times \{b\}| = |a|$:

$$\begin{aligned} a \cdot \mathrm{succ}(b) &= |a \times \mathrm{succ}(b)| \\ &= |(a \times b) \cup (a \times \{b\})| \\ &= |a \times b| + |a \times \{b\}| \\ &= a \cdot b + a. \end{aligned}$$

□

Wir beweisen wieder eine Kürzungsregel.

Satz 2.30 (Kürzungsregel)**.** Für alle $a, b, c \in \mathbb{N}$ mit $a \neq 0$ gilt

$$a \cdot b = a \cdot c \quad \Longrightarrow \quad b = c.$$

Beweis. Übung. □

Wie bei der Addition können wir den Zusammenhang zwischen dem kartesischen Produkt und der Multiplikation auf beliebige endliche Mengen erweitern.

Satz 2.31. Für endliche Mengen A und B gilt

$$|A \times B| = |A| \cdot |B|.$$

Beweis. Aus Bijektionen $f\colon |A| \to A$ und $g\colon |B| \to B$ erhalten wir nach Satz 1.82 eine Bijektion $h\colon |A| \times |B| \to A \times B$. Mit Satz 2.11 und der Definition der Multiplikation folgt $|A \times B| = \big||A| \times |B|\big| = |A| \cdot |B|$. □

Aus dem Zusammenhang mit dem kartesischen Produkt erhalten wir die folgenden Eigenschaften.

Satz 2.32. Für alle $n, m, r \in \mathbb{N}$ gilt

(i) $n \cdot m = m \cdot n$ (Kommutativität),

(ii) $(n \cdot m) \cdot r = n \cdot (m \cdot r)$ (Assoziativität),

(iii) $n \cdot (m + r) = (n \cdot m) + (n \cdot r)$ (Distributivität).

Beweis. Übung. □

Wegen der Assoziativität können wir $nmr \coloneqq n \cdot m \cdot r \coloneqq (n \cdot m) \cdot r$ schreiben, ohne dass das zu Mehrdeutigkeit führt. Außerdem verwenden wir die Konvention, dass Multiplikation stärker bindet als Addition („Punkt vor Strich“), also $n + mr \coloneqq n + (m \cdot r)$.

Analog zum Summenzeichen $\sum$ definieren wir eine Notation für Vielfachprodukte.

Definition 2.33. Seien $n_0, n \in \mathbb{N}$ und sei $a\colon \{n_0, n_0 + 1, \dots, n\} \to Z$ eine Abbildung, wobei Z eine Menge ist, für die eine Multiplikation $\cdot\colon Z^2 \to Z$ definiert ist. Dann schreiben wir

$$\prod_{i=n_0}^{n} a(i) \coloneqq a(n_0) \cdot a(n_0 + 1) \cdot \dots \cdot a(n).$$

Wir beweisen wichtige Verbindungen zwischen Summe und Produkt.

Satz 2.34. Für alle $n, a \in \mathbb{N}$ mit $n \geq 1$ gilt $\sum_{i=1}^{n} a = a \cdot n$.

Beweis. Wir beweisen die Aussage per Induktion über $n \geq 1$. Für $n = 1$ folgt die Aussage aus $a \cdot 1 = a$. Gilt sie für $n \in \mathbb{N}$, dann ist

$$\sum_{i=1}^{n+1} a = a + \sum_{i=1}^{n} a = a + a \cdot n = a \cdot (n + 1),$$

wobei wir die Nachfolgerregel verwendet haben. Also gilt sie auch für $\operatorname{succ}(n) = n + 1$. □

Satz 2.35 (*Summenformel*)**.** Für alle $n \in \mathbb{N}$ mit $n \geq 1$ gilt $\sum_{i=1}^{n} 2i = n(n + 1)$.

Beweis. Wir beweisen die Aussage per Induktion über $n \geq 1$. Für $n = 1$ gilt $\sum_{i=1}^{1} 2i = 2 = 1 \cdot (1+1)$. Die Aussage gelte nun für ein $n \geq 1$. Dann ist

$$\begin{aligned}\sum_{i=1}^{n+1} 2i &= \sum_{i=1}^{n} 2i + 2(n+1) \\ &= n(n+1) + 2(n+1) \\ &= (n+1)(n+2),\end{aligned}$$

wobei der letzte Schritt Distributivität und Kommutativität verwendet. Damit gilt die Formel auch für $n+1$. □

Division

Für festes $b \in \mathbb{N} \setminus \{0\}$ ist die Abbildung $f_b \colon \mathbb{N} \to \mathbb{N},\ q \mapsto b \cdot q$ im Allgemeinen nicht surjektiv; beispielsweise liegt 3 nicht im Bild von f_2. Insbesondere hat f_b im Allgemeinen keine Umkehrabbildung. Wir definieren die Division als eine partielle Umkehrung der Multiplikation.

Definition 2.36. Eine Zahl $a \in \mathbb{N}$ heißt durch $b \in \mathbb{N} \setminus \{0\}$ *teilbar*, wenn die Aussage $b \mid a$ gilt, mit

$$b \mid a \coloneqq (\exists q \in \mathbb{N} \colon a = b \cdot q).$$

Definition 2.37. Sei $D \coloneqq \{(a,b) \in \mathbb{N}^2 \mid b \neq 0 \text{ und } b \mid a\}$. Wir definieren die *Division* auf $\mathbb{N}$ als die Abbildung

$$/ \colon D \to \mathbb{N},\ (a,b) \mapsto q,$$

wobei $q \in \mathbb{N}$ die nach Definition 2.36 und Satz 2.30 eindeutig bestimmte Zahl mit $a = b \cdot q$ ist. Wir schreiben den *Quotienten* von a und b als

$$\frac{a}{b} \coloneqq a/b \coloneqq /(a,b).$$

Insbesondere gilt $a/1 = a$ für alle $a \in \mathbb{N}$. Außerdem ist die Division tatsächlich eine bedingte Umkehrung der Multiplikation.

Beobachtung 2.38. Für alle $a, b, q \in \mathbb{N}$ mit $b \neq 0$ und $b \mid a$ gilt

$$b \cdot q = a \quad \iff \quad \frac{a}{b} = q.$$

Die Division tritt auf natürliche Weise beim Zählen gleich großer Äquivalenzklassen auf.

Satz 2.39. Sei M eine endliche Menge und $\sim$ eine Äquivalenzrelation auf M. Besitzen alle Äquivalenzklassen dieselbe Kardinalität $b \in \mathbb{N} \setminus \{0\}$, so gilt

$$b \mid |M| \qquad \text{und} \qquad |M/{\sim}| = \frac{|M|}{b}.$$

Beweis. Aus $M/\!\sim \subseteq \mathcal{P}(M)$ und den Sätzen 2.13 und 2.14 folgt die Endlichkeit von $M/\!\sim$. Die Äquivalenzklassen zerlegen M damit in paarweise disjunkte Mengen $N_1, N_2, \ldots, N_{|M/\sim|}$ mit jeweils b Elementen. Wiederholte Anwendung von Satz 2.21 ergibt

$$|M| = \left| \bigcup_{N \in M/\sim} N \right| = \sum_{i=1}^{|M/\sim|} |N_i| = \sum_{i=1}^{|M/\sim|} b = b \cdot |M/\!\sim|.$$

Somit ist $b \mid |M|$, und nach Definition der Division gilt $|M/\!\sim| = |M|/b$. □

Beispiel 2.40. *Sei* $M = \{2,3,4,5,6,7,8,9,10,\mathrm{B},\mathrm{D},\mathrm{K},\mathrm{A}\} \times \{\clubsuit,\spadesuit,\heartsuit,\diamondsuit\}$ *die Menge der* 52 *Karten eines französischen Blatts. Wir betrachten die Äquivalenzrelation* $\sim := \{((a,p),(b,q)) \in M^2 \mid a = b\}$*, zwei Karten sind also äquivalent, wenn sie denselben Rang (zum Beispiel* D *oder* 7*) haben. Die Faktormenge* $M/\!\sim$ *besteht aus*

$$|M/\!\sim| = \frac{52}{4} = 13$$

Mengen der Form $\{(a,\clubsuit),(a,\spadesuit),(a,\heartsuit),(a,\diamondsuit)\}$. △

Division mit Rest

Auch wenn eine Zahl b eine Zahl a nicht teilt, können wir a in ein Vielfaches von b und einen Rest zerlegen.

Satz 2.41 (Division mit Rest)**.** Für alle $a, b \in \mathbb{N}$ mit $b \neq 0$ existieren eindeutig bestimmte Zahlen $q, r \in \mathbb{N}$ mit

$$a = q \cdot b + r \qquad \text{und} \qquad r < b.$$

Beweis. Die Menge $Q := \{k \in \mathbb{N} \mid k \cdot b \leq a\}$ ist nichtleer, denn $0 \cdot b = 0 \leq a$ und somit $0 \in Q$. Wegen $b \neq 0$ gilt $b = 1 + c$ für ein $c \in \mathbb{N}$ (Satz 2.5). Also ist $Q \subseteq (a+1)$, denn für jedes $k \in Q$ gilt

$$k \leq k + k \cdot c = k \cdot (1+c) = k \cdot b \leq a.$$

Nach Satz 2.13 ist $Q \subseteq (a+1)$ also endlich. Nach Satz 2.15 besitzt die total geordnete Menge $(Q, \leq)$ ein Maximum $q := \max Q$.

Wegen $q \in Q$ gilt $q \cdot b \leq a$ und wir können die Differenz $r := a - q \cdot b$ bilden. Nach Beobachtung 2.25 gilt dann $a = q \cdot b + r$. Außerdem ist $r < b$, denn andernfalls gäbe es ein $s \in \mathbb{N}$ mit $r = b + s$ (Satz 2.19) und

$$a = q \cdot b + r = q \cdot b + b + s = (q+1) \cdot b + s,$$

also $(q+1) \cdot b \leq a$ und damit $q + 1 \in Q$, im Widerspruch zur Maximalität von q.

Es bleibt die Eindeutigkeit zu zeigen. Seien dazu $q_1, q_2 \in \mathbb{N}$ mit $q_1 \leq q_2$ und $r_1, r_2 \in \mathbb{N}$ mit $r_1 < b$ und $r_2 < b$ und $a = q_1 b + r_1 = q_2 b + r_2$. Angenommen, es wäre $q_1 < q_2$. Nach Satz 2.19 existiert dann ein $d \in \mathbb{N} \setminus \{0\}$ mit $q_2 = q_1 + d$. Mit Distributivität erhalten wir

$$q_1 b + r_1 = a = q_2 b + r_2 = (q_1 + d) \cdot b + r_2 = q_1 b + db + r_2.$$

Die Kürzungsregel (Satz 2.20) ergibt $r_1 = db + r_2$. Wegen $d \neq 0$ existiert ein $e \in \mathbb{N}$ mit $d = 1 + e$ (Satz 2.5). Somit ist

$$r_1 = db + r_2 = (1+e)b + r_2 = b + eb + r_2 \geq b,$$

im Widerspruch zu $r_1 < b$. Also ist $q_1 = q_2$ und nach der Kürzungsregel (Satz 2.20) auch $r_1 = r_2$. □

Definition 2.42. Für $a, b \in \mathbb{N}$ mit $b \neq 0$ seien $q, r \in \mathbb{N}$ die nach Satz 2.41 eindeutig bestimmten Zahlen. Wir definieren den *ganzzahligen Quotienten* und den *Rest* durch die Abbildungen

$$\operatorname{div}\colon \mathbb{N} \times (\mathbb{N} \setminus \{0\}) \to \mathbb{N},\ (a, b) \mapsto q,$$
$$\operatorname{mod}\colon \mathbb{N} \times (\mathbb{N} \setminus \{0\}) \to \mathbb{N},\ (a, b) \mapsto r,$$

und schreiben

$$a \operatorname{div} b \coloneqq \operatorname{div}(a, b) \qquad \text{und} \qquad a \bmod b \coloneqq \operatorname{mod}(a, b).$$

Beobachtung 2.43. Für alle $a, b \in \mathbb{N}$ mit $b \neq 0$ gilt

$$a = (a \operatorname{div} b) \cdot b + (a \bmod b).$$

Außerdem gilt $b \mid a$ genau dann, wenn $a \bmod b = 0$, und in diesem Fall ist $a \operatorname{div} b = a/b$.

Die folgenden Regeln erlauben es, Zwischenergebnisse einer Rechnung bereits durch ihre Reste zu ersetzen.

Satz 2.44. Für alle $a, b, n \in \mathbb{N}$ mit $n \neq 0$ gilt

$$(a + b) \bmod n = \big((a \bmod n) + (b \bmod n)\big) \bmod n,$$
$$(a \cdot b) \bmod n = \big((a \bmod n) \cdot (b \bmod n)\big) \bmod n.$$

Beweis. Übung. □

Beispiel 2.45. *Wir berechnen den Rest von* $(9 + 6) \cdot 23 + 705$ *bei Division durch* 7*. Mit den Rechenregeln für Reste können wir bereits alle Zwischenergebnisse durch ihre Reste ersetzen:*

$$\begin{aligned}\big((9 + 6) \cdot 23 + 705\big) \bmod 7 &= \big(((2 + 6) \bmod 7) \cdot 2 + 5\big) \bmod 7 \\ &= (1 \cdot 2 + 5) \bmod 7 = 0.\end{aligned}$$

Durch elementare Rechnungen haben wir also nachgewiesen, dass $(9 + 6) \cdot 23 + 705$ *durch* 7 *teilbar ist.* △

2.6 Potenz

Das Zählen von Zeichenketten („Strings“) ist in der Informatik von besonderer Bedeutung. Aus mathematischer Sicht entspricht das dem Zählen von Tupeln: Die Menge aller Zeichenketten fester Länge $k \in \mathbb{N}$ aus Zeichen eines Alphabets A lässt sich als die kartesische Produktmenge A^k beschreiben. So ist die Menge aller Kombinationen eines Kombinationszahlenschlosses mit 4 Stellen beispielsweise $\{0, 1, \ldots, 9\}^4 = \{(0, 0, 0, 0), (0, 0, 0, 1), \ldots, (9, 9, 9, 9)\}$. Wir können diese Menge aber auch als die Menge aller Abbildungen $f\colon \{0, 1, 2, 3\} \to \{0, 1, \ldots, 9\}$ verstehen: Dabei ordnet f jeder Stelle $i \in \{0, 1, 2, 3\}$ eine Zahl $f(i) \in \{0, 1, \ldots, 9\}$ zu. Mit den Zahlen 10 und 4 als Mengen können wir die Menge aller Kombinationen also sowohl als 10^4 als auch als $\{f \subseteq (4 \times 10) \mid f\colon 4 \to 10\}$ auffassen.

Wir halten diese enge Beziehung zwischen der Menge von Abbildungen und der kartesischen Produktmenge fest.

Definition 2.46. Die Menge aller Abbildungen von B nach A bezeichnen wir mit

$$\operatorname{abb}(B, A) \coloneqq \{f \subseteq (B \times A) \mid f\colon B \to A\}.$$

Satz 2.47. Für jede Menge A und jede endliche Menge B existiert eine Bijektion

$$\operatorname{abb}(B, A) \to A^{|B|}.$$

Ist auch A endlich, so ist also $\operatorname{abb}(B, A)$ endlich und $|\operatorname{abb}(B, A)| = |A^{|B|}|$.

Beweis. Die Aussage gilt für $B = \emptyset$, da beide Seiten genau ein Element haben (leere Abbildung und leeres Tupel).

Sei nun $B \neq \emptyset$ und wähle eine Bijektion $g\colon |B| \to B$. Dann hat die Abbildung

$$\Phi\colon \operatorname{abb}(B, A) \to A^{|B|}, \quad f \mapsto \big(f(g(0)), f(g(1)), \dots, f(g(|B| - 1))\big)$$

die Umkehrabbildung

$$\Phi^{-1}\colon A^{|B|} \to \operatorname{abb}(B, A), \quad (a_0, \dots, a_{|B|-1}) \mapsto (h\colon B \to A,\, b \mapsto a_{g^{-1}(b)})$$

und ist somit bijektiv (Satz 1.76). Um das für eine Richtung zu sehen, betrachte

$$\begin{aligned}\Phi(\Phi^{-1}((a_0, \dots, a_{|B|-1}))) &= \Phi(h\colon B \to A,\, b \mapsto a_{g^{-1}(b)}) \\ &= (h(g(0)), h(g(1)), \dots, h(g(|B| - 1))) \\ &= (a_{g^{-1}(g(0))}, a_{g^{-1}(g(1))}, \dots, a_{g^{-1}(g(|B|-1))}) \\ &= (a_0, \dots, a_{|B|-1}).\end{aligned}$$

Ist A endlich, so ist nach der Definition von A^0 und wiederholter Anwendung von Satz 2.14 auch $A^{|B|}$ endlich. Die Behauptung folgt nun mit Satz 2.11. □

Damit ist insbesondere $\operatorname{abb}(n, a)$ für $n, a \in \mathbb{N}$ endlich und wir können die Potenzierung wie folgt über eine Anzahl Abbildungen definieren.

Definition 2.48. Wir definieren die *Potenzierung* auf $\mathbb{N}$ als die Verknüpfung

$$\operatorname{pot}\colon \mathbb{N}^2 \to \mathbb{N},\, (a, n) \mapsto |\operatorname{abb}(n, a)|$$

und schreiben die Potenz mit *Basis* $a \in \mathbb{N}$ und *Exponenten* $n \in \mathbb{N}$ als

$$a^n \coloneqq \operatorname{pot}(a, n).$$

Insbesondere gilt $a^0 = 1$ für alle $a \in \mathbb{N}$, also auch $0^0 = 1$. Denn unabhängig von a gibt es genau eine Abbildung von der leeren Menge 0 nach a, nämlich die leere Abbildung.

Wir erweitern den Zusammenhang zwischen Abbildungen und kartesischen Produkten wieder auf beliebige endliche Mengen.

Satz 2.49. Für endliche Mengen A und B gilt

$$|\operatorname{abb}(B, A)| = |A|^{|B|} = |A^{|B|}|.$$

Beweis. Per Definition der Kardinalität existieren Bijektionen $f\colon |A| \to A$ und $g\colon |B| \to B$. Die Abbildung

$$\Phi\colon \operatorname{abb}(|B|, |A|) \to \operatorname{abb}(B, A), \quad h \mapsto f \circ h \circ g^{-1},$$

hat die Umkehrabbildung $\Phi^{-1}\colon \operatorname{abb}(B, A) \to \operatorname{abb}(|B|, |A|)$, $k \mapsto f^{-1} \circ k \circ g$ und ist somit bijektiv (Satz 1.76). Daher gilt nach den Sätzen 2.11 und 2.47, dass

$$|\operatorname{abb}(B, A)| = |\operatorname{abb}(|B|, |A|)| = |A|^{|B|} = |A^{|B|}|. \qquad \square$$

Wie die Multiplikation die wiederholte Addition desselben Summanden beschreibt, beschreibt die Potenzierung die wiederholte Multiplikation desselben Faktors.

Satz 2.50. Für alle $a, n \in \mathbb{N}$ mit $n \geq 1$ gilt

$$a^n = \prod_{i=1}^{n} a.$$

Beweis. Übung. $\square$

Die gezählten Abbildungsmengen lassen sich auf natürliche Weise zerlegen und umordnen. Daraus folgen die Potenzgesetze.

Satz 2.51 (Potenzgesetze)**.** Für alle $a, b, m, n \in \mathbb{N}$ gilt

$$a^{m+n} = a^m \cdot a^n, \qquad (a^m)^n = a^{m \cdot n}, \qquad (a \cdot b)^n = a^n \cdot b^n.$$

Beweis. Wir verwenden im Folgenden die Sätze 2.49, 2.11 und 2.31. Durch Zerlegen eines $(m+n)$-Tupels an der m-ten Stelle erhalten wir die Bijektion

$$\begin{aligned} \Phi_1\colon a^{m+n} &\to a^m \times a^n, \\ (x_1, \ldots, x_{m+n}) &\mapsto ((x_1, \ldots, x_m), (x_{m+1}, \ldots, x_{m+n})). \end{aligned}$$

Für $m = 0$ beziehungsweise $n = 0$ ist das entsprechende Tupel leer. Daher gilt

$$a^{m+n} = |a^{m+n}| = |a^m \times a^n| = |a^m| \cdot |a^n| = a^m \cdot a^n.$$

Die Abbildung

$$\Phi_2\colon \operatorname{abb}(n, \operatorname{abb}(m, a)) \to \operatorname{abb}(m \times n, a), \quad f \mapsto ((j, i) \mapsto f(i)(j))$$

ist bijektiv; ihre Umkehrabbildung bildet $F\colon m \times n \to a$ auf $i \mapsto (j \mapsto F(j, i))$ ab. Der Vergleich ihrer Kardinalitäten ergibt

$$(a^m)^n = |\operatorname{abb}(n, \operatorname{abb}(m, a))| = |\operatorname{abb}(m \times n, a)| = a^{m \cdot n}.$$

Schließlich ist die Abbildung

$$\Phi_3 \colon (a \times b)^n \to a^n \times b^n,\ ((x_1, y_1), \ldots, (x_n, y_n)) \mapsto ((x_1, \ldots, x_n), (y_1, \ldots, y_n))$$

bijektiv. Somit folgt

$$(a \cdot b)^n = |(a \times b)^n| = |a^n \times b^n| = |a^n| \cdot |b^n| = a^n \cdot b^n. \qquad \square$$

Die Produktregel für Reste überträgt sich auf Potenzen.

Satz 2.52 (Potenzregel für Reste)**.** Für alle $a, k, n \in \mathbb{N}$ mit $n \neq 0$ gilt

$$a^k \bmod n = (a \bmod n)^k \bmod n.$$

Beweis. Übung. $\square$

Auch das Zählen von Teilmengen lässt sich auf das Zählen von Abbildungen zurückführen: Für jedes Element entscheiden wir, ob es zu einer Teilmenge gehört oder nicht.

Satz 2.53. Für die Potenzmenge einer endlichen Menge M gilt

$$|\mathcal{P}(M)| = 2^{|M|}.$$

Beweis. Jeder Teilmenge $A \subseteq M$ ordnen wir ihre Abbildung $\chi_A \colon M \to 2$ mit

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases}$$

zu, wobei wir $2 = \{0, 1\}$ verwenden. Umgekehrt bestimmt jede Abbildung $\chi \colon M \to 2$ die Teilmenge $\{x \in M \mid \chi(x) = 1\}$. Damit ist $A \mapsto \chi_A$ eine Bijektion von $\mathcal{P}(M)$ nach $\operatorname{abb}(M, 2)$. Mit Satz 2.49 folgt

$$|\mathcal{P}(M)| = |\operatorname{abb}(M, 2)| = 2^{|M|}. \qquad \square$$

2.7 Fakultät

Analog zum Zählen von Zeichenketten ist das Zählen von Reihenfolgen oder Sortierungen in der Informatik wichtig. Die Menge aller Sortierungen eines Tupels $(a_0, a_1, \ldots, a_{n-1})$ lässt sich ausdrücken als die Menge aller Tupel $(b_0, \ldots, b_{n-1}) \in \{0, \ldots, n-1\}^n$, die kein Element doppelt verwenden, denn jede Sortierung des Tupels kann als $(a_{b_0}, a_{b_1}, \ldots, a_{b_{n-1}})$ mit paarweise unterschiedlichen b_i geschrieben werden. Analog zur Bijektion zwischen Tupeln und Abbildungen kann jedes Tupel ohne Wiederholungen als eine Bijektion verstanden werden. Wir zählen also Bijektionen von der Menge n auf sich selbst. Solche Bijektionen nennen wir Permutationen, um ihre Interpretation als Umsortierungen zu betonen.

Definition 2.54. Die Menge der *Permutationen* auf M ist

$$\operatorname{perm}(M) \coloneqq \{\sigma \in \operatorname{abb}(M, M) \mid \sigma \text{ ist bijektiv}\}.$$

Wegen $\operatorname{perm}(n) \subseteq \operatorname{abb}(n,n)$ und Satz 2.13 ist $\operatorname{perm}(n)$ endlich. Wir können damit die Fakultät als die Anzahl Umordnungen einer Menge wie folgt definieren.

Definition 2.55. Wir definieren die Fakultät auf $\mathbb{N}$ als die Abbildung

$$\operatorname{fak}\colon \mathbb{N} \to \mathbb{N},\ n \mapsto |\operatorname{perm}(n)|$$

und schreiben die Fakultät der Zahl $n \in \mathbb{N}$ als

$$n! \coloneqq \operatorname{fak}(n).$$

Insbesondere gilt $0! = 1$, denn die leere Abbildung ist die einzige Bijektion von 0 nach 0.

Satz 2.56. Für eine endliche Menge M gilt

$$|\operatorname{perm}(M)| = |M|!.$$

Beweis. Übung. □

Anschaulich lässt sich die Anzahl Reihenfolgen einer festen Menge von n Elementen auch wie folgt berechnen: Wir haben n Möglichkeiten für die Wahl des ersten Elements, $n-1$ für die des zweiten, und so weiter, bis schließlich nur noch eine Möglichkeit für das letzte Element besteht. Damit können wir unsere Wahlmöglichkeiten als Menge der Form $A_1 \times \cdots \times A_n$ darstellen, wobei A_i die Wahlmöglichkeiten im Schritt i bezeichnet. Intuitiv gilt $|A_i| = n - i + 1$, womit die folgende Formel plausibel ist.

Satz 2.57. Für alle $n \in \mathbb{N}$ mit $n \geq 1$ gilt

$$n! = \prod_{i=1}^{n} i.$$

Beweis. Wir beweisen die Aussage per Induktion über $n \geq 1$. Für $n = 1$ folgt die Behauptung wegen $\operatorname{perm}(1) = \{\mathrm{id}_1\}$, also $1! = |\operatorname{perm}(1)| = 1 = \prod_{i=1}^{1} i$.

Die Aussage gelte nun für ein $n \geq 1$. Wir lesen eine Permutation $\sigma \in \operatorname{perm}(n+1)$ als Reihenfolge $(\sigma(0), \ldots, \sigma(n))$, in der die Elemente von $n+1 = \{0, \ldots, n\}$ ausgewählt werden. Für jede Wahl $x \in n+1$ definieren wir die Bijektion

$$h_x\colon n \to (n+1) \setminus \{x\}, \quad h_x(j) \coloneqq \begin{cases} j, & j < x, \\ j+1, & j \geq x. \end{cases}$$

Die verbleibenden Elemente $\{0,1,\ldots,n\} \setminus \{x\} = \{h_x(0), h_x(1), \ldots, h_x(n-1)\}$ sind also so nummeriert, dass $h_x^{-1}\colon (n+1) \setminus \{x\} \to n$ jedem Element eine eindeutige Zahl aus $\{0,1,\ldots,n-1\}$ zuordnet. Damit ist

$$\pi_\sigma\colon n \to n,\ i \mapsto h_{\sigma(0)}^{-1}(\sigma(i+1))$$

für jedes $\sigma \in \operatorname{perm}(n+1)$ eine Permutation aus $\operatorname{perm}(n)$.

Die Abbildung

$$\Phi\colon \operatorname{perm}(n+1) \to ((n+1) \times \operatorname{perm}(n)), \quad \sigma \mapsto (\sigma(0), \pi_\sigma),$$

hat die Umkehrabbildung

$$\Phi^{-1}\colon ((n+1)\times \mathrm{perm}(n)) \to \mathrm{perm}(n+1), \quad (x,\pi)\mapsto \sigma_{x,\pi}$$

mit

$$\sigma_{x,\pi}\colon (n+1)\to(n+1), \quad i\mapsto \begin{cases} x, & \text{für } i=0,\\ h_x(\pi(i-1)), & \text{für } i\geq 1,\end{cases}$$

und somit ist Φ bijektiv (Satz 1.76).

Nach Satz 2.11, Satz 2.31 und der Induktionsannahme folgt

$$(n+1)! = (n+1)\cdot n! = (n+1)\cdot \prod_{i=1}^{n} i = \prod_{i=1}^{n+1} i. \qquad \square$$

2.8 Binomialkoeffizienten

Wir haben in Satz 2.53 bereits gesehen, dass sich die Menge aller Teilmengen einer endlichen Menge mittels der Potenz zählen lässt. Häufig wollen wir nur die Teilmengen einer festen Größe zählen, also die Anzahl Möglichkeiten, eine feste Anzahl unterschiedlicher Elemente aus einer Menge auszuwählen. Im Gegensatz zum vorigen Abschnitt soll dabei die Reihenfolge der Auswahl keine Rolle spielen.

Definition 2.58. Die Menge aller k-elementigen Teilmengen einer Menge M mit $k\in\mathbb{N}$ schreiben wir als

$$\mathrm{tm}_k(M) \coloneqq \{A\subseteq M \mid A \text{ ist endlich und } |A|=k\}.$$

Da $\mathrm{tm}_k(n)\subseteq\mathcal{P}(n)$ gilt und $\mathcal{P}(n)$ nach Satz 2.14 endlich ist, ist nach Satz 2.13 auch $\mathrm{tm}_k(n)$ endlich. Wir können damit den Binomialkoeffizienten über die Anzahl Teilmengen fester Größe definieren.

Definition 2.59. Wir definieren den Binomialkoeffizienten „n über k“ als die Verknüpfung

$$\mathrm{bin}\colon \mathbb{N}^2\to\mathbb{N},\ (n,k)\mapsto |\mathrm{tm}_k(n)|$$

und schreiben

$$\binom{n}{k} \coloneqq \mathrm{bin}(n,k).$$

Für $k>n$ ist $\mathrm{tm}_k(n)=\emptyset$ und daher $\binom{n}{k}=0$. Außerdem gilt $\binom{n}{0}=1$ für alle $n\in\mathbb{N}$, denn die leere Menge ist die einzige null-elementige Teilmenge von n.

Wir zeigen wieder, dass sich die Definition des Binomialkoeffizienten auf beliebige endliche Mengen überträgt.

Satz 2.60. Für eine endliche Menge M und $k\in\mathbb{N}$ gilt

$$|\mathrm{tm}_k(M)| = \binom{|M|}{k}.$$

Beweis. Übung. □

Wir können eine Auswahl von k Elementen zählen, indem wir zunächst alle Elemente anordnen und anschließend die Reihenfolge innerhalb der ausgewählten und der übrigen Elemente vergessen. Dadurch erhalten wir einen Zusammenhang zwischen Binomialkoeffizienten und Fakultäten.

Satz 2.61. Für alle $n, k \in \mathbb{N}$ mit $k \leq n$ gilt

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Beweis. Für jedes $A \in \mathrm{tm}_k(n)$ sei

$$\mathcal{P}_A \coloneqq \{\sigma \in \mathrm{perm}(n) \mid \sigma(k) = A\}$$

die Menge der bijektiven Abbildungen $\sigma\colon n \to n$, die die Elemente von $k \subseteq n$ auf die Teilmenge A abbilden.

Wegen $|A| = |k| = k$ und $|n \setminus A| = |n \setminus k| = n - k$ existieren Bijektionen $f\colon k \to A$ und $g\colon n \setminus k \to n \setminus A$. Wir können Permutationen von k und $n \setminus k$ eindeutig zu einer Permutation in $\mathcal{P}_A$ zusammensetzen, indem wir ihre Werte mit f bzw. g umbenennen. Andersherum können wir jede Permutation in $\mathcal{P}_A$ eindeutig in zwei solche Permutationen aufteilen, indem wir sie auf k beziehungsweise $n \setminus k$ einschränken und die Umbenennungen mit f^{-1} beziehungsweise g^{-1} rückgängig machen. Mit anderen Worten, es existiert eine Bijektion $\mathrm{perm}(k) \times \mathrm{perm}(n-k) \to \mathcal{P}_A$ (Übung: gib eine explizite Bijektion an).

Nach den Sätzen 2.11 und 2.31 sowie der Definition der Fakultät gilt damit

$$|\mathcal{P}_A| = |\mathrm{perm}(k) \times \mathrm{perm}(n-k)| = k!\,(n-k)!.$$

Jede Permutation $\sigma \in \mathrm{perm}(n)$ gehört zu genau einer der Mengen $\mathcal{P}_A$, nämlich zu $\mathcal{P}_{\sigma(k)}$. Diese Mengen sind daher paarweise disjunkt und zerlegen $\mathrm{perm}(n)$. Da es $\binom{n}{k}$ mögliche Mengen A gibt, folgt durch wiederholte Anwendung von Satz 2.21

$$n! = \binom{n}{k} k!\,(n-k)!.$$

Da die Identität eine Permutation jeder endlichen Zahlenmenge ist, gilt $k! \neq 0$ und $(n-k)! \neq 0$. Nach Definition der Division folgt daraus

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$ □

Wir leiten weitere Rechengesetze her.

Satz 2.62 (Pascalsche Regel)**.** Für alle $n, k \in \mathbb{N}$ mit $k \geq 1$ gilt

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

Beweis. Wir zerlegen $\mathrm{tm}_k(n+1)$ in die Mengen

$$\mathcal{A}_0 \coloneqq \{A \in \mathrm{tm}_k(n+1) \mid n \notin A\},$$
$$\mathcal{A}_1 \coloneqq \{A \in \mathrm{tm}_k(n+1) \mid n \in A\}.$$

Die Abbildung $A \mapsto A$ ist eine Bijektion von $\mathcal{A}_0$ nach $\mathrm{tm}_k(n)$. Die Abbildung $A \mapsto A \setminus \{n\}$ ist eine Bijektion von $\mathcal{A}_1$ nach $\mathrm{tm}_{k-1}(n)$; ihre Umkehrabbildung ist $B \mapsto B \cup \{n\}$. Nach Satz 1.76 ist damit auch diese Abbildung bijektiv. Da $\mathcal{A}_0$ und $\mathcal{A}_1$ disjunkt sind, folgt mit Satz 2.21 und Satz 2.11

$$\binom{n+1}{k} = |\mathrm{tm}_k(n+1)| = |\mathcal{A}_0 \cup \mathcal{A}_1| = |\mathcal{A}_0| + |\mathcal{A}_1| = |\mathrm{tm}_k(n)| + |\mathrm{tm}_{k-1}(n)| = \binom{n}{k} + \binom{n}{k-1}. \qquad \square$$

Der folgende Satz ergibt sich mit vollständiger Induktion nach Satz 2.6 aus der Pascalschen Regel in Satz 2.62, den Rechenregeln für die Multiplikation in Satz 2.32 und den Potenzgesetzen in Satz 2.51.

Satz 2.63 (Binomischer Lehrsatz)**.** Für alle $a, b, n \in \mathbb{N}$ gilt

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

Beweis. Übung. $\square$

Korollar 2.64. Für alle $n \in \mathbb{N}$ gilt

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Beweis. Setzen wir in Satz 2.63 $a = b = 1$, erhalten wir

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^k 1^{n-k} = \sum_{k=0}^{n} \binom{n}{k}. \qquad \square$$

2.9 Anwendungsbeispiel: Kryptographie

In diesem Kapitel haben wir einen beachtlichen Fundus von Rechenregeln aus wenigen Grundlagen abgeleitet. Wir wollen diese Rechenregeln jetzt anwenden, um das mathematische Fundament des Public-Key-Verschlüsselungsverfahrens RSA (nach seinen Schöpfern Rivest, Shamir und Adleman) zu legen. Das Verfahren basiert wesentlich auf der Annahme, dass es rechnerisch schwierig ist, große Zahlen in ein Produkt aus Primzahlen zu zerlegen. Dabei ist anzumerken, dass kein effizienter Algorithmus für Primzahlfaktorisierung bekannt ist, seine Existenz aber ein zentrales offenes Problem der Komplexitätstheorie bleibt.

Definition 2.65. Eine Zahl $p \in \mathbb{N}$ heißt *Primzahl*, wenn $p > 1$ und p keine Teiler außer 1 und p hat, also $\forall z \in \{2, \ldots, p\} \setminus \{p\} \colon z \nmid p$.

Ausgangspunkt des RSA-Verfahrens ist die Wahl zweier verschiedener (großer) Primzahlen $p, q \in \mathbb{N}$ und die Berechnung von

$$n := p \cdot q \qquad \text{und} \qquad t := (p-1)(q-1).$$

Eine zentrale Beobachtung für das Verfahren ist, dass p und q die einzigen nichttrivialen Teiler von n sind. Zusammen mit der Komplexitätsannahme zur Primzahlfaktorisierung heißt das, dass wir die Zahl n öffentlich machen können, ohne damit einen (effizienten) Rückschluss auf p und q und damit auf t zu erlauben.

Satz 2.66 (Primzahllemma)**.** Für jede Primzahl p und alle $a, b \in \mathbb{N}$ gilt

$$p \mid a \cdot b \quad \Longrightarrow \quad p \mid a \text{ oder } p \mid b.$$

Beweis. Es sei $p \nmid a$. Wir zeigen $p \mid b$. Die Menge $M := \{x \in \mathbb{N} \setminus \{0\} \mid p \text{ teilt } x \cdot a\}$ ist nichtleer, wegen $p \in M$. Nach Satz 2.16 besitzt M also ein Minimum $m := \min M$. Außerdem ist $1 \notin M$, da $p \nmid a$.

Die Division von p durch m liefert nach Satz 2.41 Zahlen $q, r \in \mathbb{N}$ mit $p = q \cdot m + r$ und $r < m$. Da $p \mid p \cdot a$ und wegen $m \in M$ auch $p \mid q \cdot m \cdot a$ gilt, folgt $p \mid r \cdot a$. Wäre $r \neq 0$, so wäre also $r \in M$ im Widerspruch zu $r < m$ und der Minimalität von m. Somit ist $r = 0$ und daher $m \mid p$. Weil p Primzahl ist und wegen $1 \notin M$ auch $m \neq 1$ gilt, folgt $m = p$.

Die Division von b durch p liefert nun $s, t \in \mathbb{N}$ mit $b = s \cdot p + t$ und $t < p = m$. Aus $p \mid a \cdot b$ und $p \mid a \cdot s \cdot p$ folgt $p \mid a \cdot t$. Wäre $t \neq 0$, so wäre $t \in M$ im Widerspruch zu $t < m$. Also ist $t = 0$ und damit $p \mid b$. □

Wir können damit den folgenden Satz formulieren.

Satz 2.67 (Kleiner Satz von Fermat)**.** Ist p eine Primzahl und $p \nmid a$, so gilt

$$a^{p-1} \bmod p = 1.$$

Beweis. Nach Satz 2.66 und da $p \nmid a$ und p Primzahl ist, gilt $p \nmid a \cdot x$ für alle $x \in \mathbb{N}$ mit $0 < x < p$. Damit ist die Abbildung

$$f \colon p \setminus \{0\} \to p \setminus \{0\}, \quad x \mapsto (a \cdot x) \bmod p$$

wohldefiniert.

Für $x, y \in p \setminus \{0\}$ mit $x \leq y$ und $f(x) = f(y)$ folgt aus der Gleichheit der Reste $p \mid a \cdot (y - x)$ und mit dem Satz 2.66 wegen $p \nmid a$ also $p \mid (y - x)$. Da $y - x < p$, ist $x = y$. Damit ist f injektiv und damit eine Permutation, das heißt, f ordnet die Zahlen $1, \ldots, p-1$ lediglich um. Daher gilt

$$\prod_{x=1}^{p-1} ((a \cdot x) \bmod p) = \prod_{x=1}^{p-1} x \tag{2.1}$$

und es folgt

$$\begin{aligned}(p-1)! \bmod p &\overset{2.57}{=} \left(\prod_{x=1}^{p-1} x\right) \bmod p \overset{(2.1)}{=} \left(\prod_{x=1}^{p-1} ((a \cdot x) \bmod p)\right) \bmod p \\ &\overset{2.44}{=} \left(\prod_{x=1}^{p-1} (a \cdot x)\right) \bmod p = \left(\prod_{x=1}^{p-1} a\right) \left(\prod_{x=1}^{p-1} x\right) \bmod p \overset{2.50,2.57}{=} \left(a^{p-1}(p-1)!\right) \bmod p.\end{aligned}$$

Wegen Gleichheit der Reste beider Seiten folgt also $p \mid (p-1)!\,(a^{p-1}-1)$. Keiner der Faktoren $1, \ldots, p-1$ ist durch p teilbar; wiederholte Anwendung von Satz 2.66 ergibt daher $p \nmid (p-1)!$. Eine weitere Anwendung desselben Satzes liefert $p \mid (a^{p-1}-1)$ und somit $a^{p-1} \bmod p = 1$. □

Das RSA-Verfahren bestimmt Zahlen $e, d \in \mathbb{N} \setminus \{0\}$ mit $e \cdot d \bmod t = 1$ (das ist effizient möglich) und veröffentlicht dann (n, e) als den *öffentlichen Schlüssel*, während die Zahl d geheim bleibt. Eine Nachricht (kodiert als Tupel aus natürlichen Zahlen) kann mittels des öffentlichen Schlüssels verschlüsselt werden, indem auf jede Zahl m der Nachricht die Funktion

$$f\colon \mathbb{N} \to \mathbb{N}, \quad m \mapsto m^e \bmod n$$

angewandt wird. Nur wer die geheime Information d besitzt, kann eine verschlüsselte Zahl c durch Anwenden von

$$g\colon \mathbb{N} \to \mathbb{N}, \quad c \mapsto c^d \bmod n$$

dechiffrieren. Wir zeigen, dass dieses Verfahren funktioniert.

Satz 2.68 (Korrektheit von RSA)**.** Mit den vorangehenden Bezeichnungen und für alle $m \in \mathbb{N}$, $m < n$, gilt

$$g(f(m)) = m.$$

Beweis. Es gilt

$$g(f(m)) = (f(m))^d \bmod n = (m^e \bmod n)^d \bmod n \overset{2.52}{=} (m^e)^d \bmod n \overset{2.51}{=} m^{ed} \bmod n. \tag{2.2}$$

Wegen der Wahl von e und d existiert $k \in \mathbb{N}$ mit $ed = 1 + k(p-1)(q-1)$, also

$$m^{ed} = m^{1+k(p-1)(q-1)} \overset{2.51}{=} m \cdot \left(m^{p-1}\right)^{k(q-1)}.$$

Ist $p \mid m$, so folgt unmittelbar $m^{ed} \bmod p = 0 = m \bmod p$. Andernfalls liefert Satz 2.67 $m^{p-1} \bmod p = 1$ und es folgt

$$\left(m^{p-1}\right)^{k(q-1)} \bmod p \overset{2.52}{=} \left(m^{p-1} \bmod p\right)^{k(q-1)} \bmod p = 1 \bmod p$$

und damit ebenfalls

$$m^{ed} \bmod p \overset{2.44}{=} ((m \bmod p) \cdot \left(\left(m^{p-1}\right)^{k(q-1)} \bmod p\right)) \bmod p = m \bmod p.$$

Dasselbe Argument mit q anstelle von p ergibt auch $m^{ed} \bmod q = m \bmod q$.

Da $ed \geq 1$ gilt, ist $m^{ed} \geq m$. Somit können wir $x \coloneqq m^{ed} - m \in \mathbb{N}$ setzen. Damit ist $p \mid x$ (Satz 2.44) und es gibt ein $l \in \mathbb{N}$ mit $x = p \cdot l$. Da die Primzahlen p und q verschieden sind, gilt $q \nmid p$. Aus $q \mid (x = p \cdot l)$ folgt daher mit Satz 2.66 auch $q \mid l$. Somit gilt $(p \cdot q) \mid (p \cdot l = x)$. Also haben m^{ed} und m denselben Rest bei Division durch $n = p \cdot q$. Wegen $m < n$ ist $m \bmod n = m$, also

$$g(f(m)) \overset{(2.2)}{=} m^{ed} \bmod n = m \bmod n = m.$$ □

Beispiel 2.69. *Für ein überschaubares Beispiel gehen wir von kleinen Primzahlen $p = 5$ und $q = 11$ aus und berechnen $n = p \cdot q = 55$ und $t = (p-1)(q-1) = 40$, und wählen $e = 3$ und $d = 27$. Wie gefordert ist dann*

$$(e \cdot d) \bmod t = 81 \bmod 40 = 1.$$

Die Nachricht $m = 12$ wird verschlüsselt zu

$$c = m^e \bmod n = 12^3 \bmod 55 = (144 \cdot 12) \bmod 55 = (288 \cdot 6) \bmod 55 \overset{2.44}{=} (13 \cdot 6) \bmod 55 = 78 \bmod 55 = 23.$$

Für die Entschlüsselung nutzen wir die Rechenregeln, die wir in diesem Kapitel hergeleitet haben. Wir machen die Nebenrechnungen

$$\begin{aligned}
11 \cdot 23 \bmod 55 &= (23 \cdot 2 + 23 \cdot 3 \cdot 3) \bmod 55 = (46 + 14 \cdot 3) \bmod 55 = 33 \\
12 \cdot 23 \bmod 55 &= (11 \cdot 23 + 23) \bmod 55 = (33 + 23) \bmod 55 = 1 \\
23^2 \bmod 55 &= (11 \cdot 23 + 12 \cdot 23) \bmod 55 = (33 + 1) \bmod 55 = 34 \\
23^3 \bmod 55 &= (34 \cdot 23) \bmod 55 = (23^2 + 11 \cdot 23) \bmod 55 = (34 + 33) \bmod 55 = 12 \\
23^4 \bmod 55 &= 12 \cdot 23 \bmod 55 = 276 \bmod 55 = 1
\end{aligned}$$

und erhalten die entschlüsselte Nachricht

$$c^d \bmod 55 = 23^{27} \bmod 55 = \left(\left(23^4\right)^6 \cdot 23^3\right) \bmod 55 = 1 \cdot 12 \bmod 55 = 12 = m. \qquad \triangle$$

3 Erweiterung des Zahlenraums

Im vorigen Kapitel haben wir die natürlichen Zahlen eingeführt und ihre Rechenoperationen auf das Zählen zurückgeführt. Dabei haben wir gesehen, dass Subtraktion und Division nur eingeschränkt möglich sind. Wir erweitern nun schrittweise den Zahlenraum um Differenzen und Quotienten und gelangen so zu den ganzen und rationalen Zahlen. Mit den reellen Zahlen nehmen wir auch Grenzwerte auf und schließen damit die Lücken auf der Zahlengeraden. Schließlich erweitern wir den Zahlenraum zu den komplexen Zahlen, um auch sämtliche Nullstellen von Polynomen zu erfassen.

Dabei beschreiben wir die wiederkehrenden Rechenstrukturen abstrakt. Wir untersuchen Teilmengen, auf denen diese Strukturen erhalten bleiben, und Abbildungen, die mit den Rechenoperationen verträglich sind. Das ermöglicht eine gemeinsame Behandlung der verschiedenen Zahlbereiche und macht dieselben Methoden auch darüber hinaus anwendbar: etwa auf die Verknüpfung von Bitfolgen durch XOR oder auf Drehungen und Spiegelungen digitaler Bilder.

3.1 Die ganzen Zahlen

Die Subtraktion auf $\mathbb{N}$ haben wir in Definition 2.24 als partielle Umkehrung der Addition definiert. Die Differenz $b - a$ existiert dort nur, wenn $a \leq b$ gilt. Für das Zählen ist diese Einschränkung unproblematisch: Eine Anzahl kann nicht negativ sein. Zahlen können jedoch auch Änderungen eines Bestands beschreiben. Verringert sich beispielsweise ein Zählerstand von 8 auf 5, so möchten wir die Änderung durch eine Zahl x mit $8 + x = 5$ beziehungsweise $x = 5 - 8$ beschreiben. Eine solche Zahl existiert in $\mathbb{N}$ nicht. Das führt uns zur Menge $\mathbb{Z}$ der ganzen Zahlen, die wir durch Erweiterung des Zahlenraums $\mathbb{N}$ um alle Differenzen erhalten.

Formal können wir jede Zahl der Form $a - b$ durch ein Tupel $(a, b) \in \mathbb{N}^2$ darstellen, wobei wir Tupel (a, b) und (c, d) zusammenfassen, die dieselbe Differenz beschreiben, für die also $a + d = b + c$ gilt. Mit anderen Worten ist jede ganze Zahl eine Äquivalenzklasse von Tupeln, die dieselbe Differenz beschreiben. Beachte hierbei, dass wir bewusst die Subtraktion vermeiden, da wir nicht $a \geq b$ bzw. $c \geq d$ voraussetzen können.

Definition 3.1. Die Menge der *ganzen Zahlen* ist die Faktormenge

$$\mathbb{Z} := \mathbb{N}^2/\sim,$$

wobei die Äquivalenzrelation $\sim$ gegeben ist durch

$$\sim := \{((a, b), (c, d)) \in \mathbb{N}^2 \times \mathbb{N}^2 \mid a + d = b + c\}.$$

Für $a \in \mathbb{N}$ identifizieren wir $a = [(a, 0)]$ und $-a = [(0, a)]$.

Jede ganze Zahl lässt sich auf diese Weise als natürliche Zahl oder als entgegengesetzte natürliche Zahl schreiben. Wir erhalten die vertraute Darstellung

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

Definition 3.2. Wir definieren die *Addition* auf $\mathbb{Z}$ als die Verknüpfung

$$+\colon \mathbb{Z}^2 \to \mathbb{Z},\ ([(a,b)],[(c,d)]) \mapsto [(a+c, b+d)].$$

Die Summe ist unabhängig von der Wahl der Repräsentanten und setzt die Addition auf $\mathbb{N}$ fort, denn $a + c = [(a,0)] + [(c,0)] = [(a+c, 0+0)] = a + c$ für $a, c \in \mathbb{N}$. Insbesondere beschreibt die Addition auf $\mathbb{Z}$ das Hintereinanderausführen zweier Änderungen und lässt sich, im Gegensatz zur Addition auf $\mathbb{N}$, umkehren.

Definition 3.3. Wir definieren die *Subtraktion* auf $\mathbb{Z}$ als die Verknüpfung

$$-\colon \mathbb{Z}^2 \to \mathbb{Z}, \qquad (a,b) \mapsto T_a^{-1}(b),$$

mit der Bijektion

$$T_a\colon \mathbb{Z} \to \mathbb{Z}, \qquad x \mapsto x + a.$$

Wir definieren außerdem

$$-b \coloneqq T_b^{-1}(0) = 0 - b.$$

Beobachtung 3.4. Für alle $a, b \in \mathbb{Z}$ und $c, d \in \mathbb{N}$ gilt

$$-a = 0 - a, \qquad -[(c,d)] = [(d,c)] \qquad \text{und} \qquad b - a = b + (-a).$$

Die Multiplikation auf $\mathbb{N}$ lässt sich als wiederholte Addition verstehen. Entsprechend bedeutet die Multiplikation einer ganzen Zahl mit $n \in \mathbb{N}$, dieselbe Änderung n-mal auszuführen. Die Multiplikation mit $-n$ bedeutet, stattdessen die entgegengesetzte Änderung n-mal auszuführen.

Definition 3.5. Wir definieren die *Multiplikation* auf $\mathbb{Z}$ als die Verknüpfung

$$\cdot\colon \mathbb{Z}^2 \to \mathbb{Z},\ ([(a,b)],[(c,d)]) \mapsto [(ac+bd, ad+bc)].$$

Auch das Produkt ist unabhängig von der Wahl der Repräsentanten und setzt die Multiplikation auf $\mathbb{N}$ fort, denn $a \cdot c = [(a,0)] \cdot [(c,0)] = [(a \cdot c + 0 \cdot 0, a \cdot 0 + 0 \cdot c)] = [(a \cdot c, 0)] = a \cdot c$ für $a, c \in \mathbb{N}$. Auch gelten für alle $z = [(a,b)] \in \mathbb{Z}$ und $q = [(c,d)] \in \mathbb{Z}$ die vertrauten Vorzeichenregeln

$$(-z)q = [(b,a)] \cdot [(c,d)] = [(ad+bc, ac+bd)] = -[(ac+bd, ad+bc)] = -(zq)$$

und

$$(-z)(-q) = [(b,a)] \cdot [(d,c)] = [(bd+ac, bc+ad)] = zq.$$

Wir erwähnen noch weitere Fortsetzungen.

Definition 3.6. Wir setzen Ordnung, Teilbarkeit, ganzzahligen Quotienten, Rest und Potenzierung auf $\mathbb{Z}$ wie folgt fort.

(i) Wir definieren die Ordnungsrelation

$$\leq := \{([(a,b)],[(c,d)]) \in \mathbb{Z}^2 \mid a+d \leq b+c\}.$$

(ii) Eine Zahl $z \in \mathbb{Z}$ heißt durch $q \in \mathbb{Z} \setminus \{0\}$ *teilbar*, wenn die Aussage $q \mid z$ gilt, mit

$$q \mid z := (\exists k \in \mathbb{Z} \colon z = k \cdot q).$$

(iii) Für $z \in \mathbb{Z}$ und $n \in \mathbb{N} \setminus \{0\}$ und die (eindeutig bestimmten, analog zu Satz 2.41) Zahlen $q \in \mathbb{Z}$ und $r \in \mathbb{N}$ mit $z = qn + r$ und $r < n$ definieren wir

$$z \operatorname{div} n := q \qquad \text{und} \qquad z \bmod n := r.$$

(iv) Für $a \in \mathbb{Z}$ definieren wir die Potenz durch $a^0 := 1$ und, für $n \in \mathbb{N} \setminus \{0\}$,

$$a^n := \prod_{i=1}^{n} a.$$

3.2 Gruppen

Im vorigen Abschnitt haben wir die Konstruktion der ganzen Zahlen damit motiviert, Gleichungen der Form $8 + x = 5$ lösen zu können. Das haben wir mit der Umkehrbarkeit der Addition verbunden, denn anschaulich lässt sich die Gleichung lösen, indem wir -8 auf beiden Seiten addieren. So lässt sich tatsächlich jede Gleichung der Form $a + x = b$ eindeutig lösen. Wir betrachten den zugrunde liegenden Beweis genauer.

Satz 3.7. Für alle $a, b \in \mathbb{Z}$ besitzt die Gleichung $a + x = b$ genau eine Lösung $x \in \mathbb{Z}$. Diese ist $x = b - a$.

Beweis. Die Zahl $x = (-a) + b$ ist eine Lösung, denn wegen Assoziativität der Addition gilt

$$a + x = a + ((-a) + b) = (a + (-a)) + b = 0 + b = b.$$

Ist $y \in \mathbb{Z}$ eine weitere Lösung, so gilt wegen $y = 0 + y$, wegen $(-a) + a = 0$ und wegen Assoziativität der Addition

$$y = 0 + y = ((-a) + a) + y = (-a) + (a + y) = (-a) + b = x,$$

also ist die Lösung x eindeutig.

Wegen der Kommutativität der Addition können wir die Lösung in die Form $x = b + (-a) = b - a$ bringen. □

Für die Existenz und Eindeutigkeit der Lösung haben wir die Umkehrbarkeit der Addition verwendet, also dass die Abbildung $z \mapsto z + a$ die Umkehrung $z \mapsto z + (-a)$ besitzt, so dass wir $a + (-a) = 0 + a + (-a) = 0$ haben. Dabei und für $y = y + 0$ haben wir verwendet, dass die Addition der Zahl 0 keine Wirkung hat. Zusätzlich haben wir die Assoziativität der Addition verwendet.

Verknüpfungen mit diesen Eigenschaften treten auch außerhalb von Zahlenräumen auf. Beispielsweise hatten wir in Abschnitt 2.7 bereits gesehen, dass sich die Umsortierungen eines Tupels als Permutationen beschreiben lassen. Insbesondere ist Umsortierung eine Bijektion und lässt sich also rückgängig machen (Satz 1.79). Die Verkettung von Abbildungen ist assoziativ (wegen $f \circ (g \circ h) = (x \mapsto f(g(h(x)))) = (f \circ g) \circ h$) und die Identität ist eine Permutation ohne Wirkung. Analog zum Beweis von Satz 3.7 folgt damit, dass wir von jeder Reihenfolge a eines Tupels zu jeder anderen Reihenfolge b durch Anwendung einer eindeutigen Umsortierung gelangen können. Mit anderen Worten hat $a \circ \pi = b$ eine eindeutige Lösung π.

Dieses Argument lässt sich auf vielfältige Zusammenhänge anwenden. Die folgende Abstraktion erlaubt es uns, das Argument losgelöst vom konkreten Zusammenhang zu führen.

Definition 3.8. Ein Tupel $(G, *)$ aus einer Menge $G \neq \emptyset$ und einer Verknüpfung $*\colon G^2 \to G$ heißt *Gruppe*, falls

(a) für alle $a, b, c \in G$ gilt, dass $a * (b * c) = (a * b) * c$ (*Assoziativität*),

(n) ein *neutrales Element* $e \in G$ existiert mit $e * a = a$ für alle $a \in G$,

(i) zu jedem $a \in G$ ein *inverses Element* $b \in G$ existiert mit $b * a = e$.

Eine Gruppe heißt *abelsch*, falls zusätzlich

(k) für alle $a, b \in G$ gilt, dass $a * b = b * a$ (*Kommutativität*).

Im Beweis von Satz 3.7 haben wir auch $y + 0 = y$ und $a + (-a) = 0$ verwendet, also Anwendung des neutralen und inversen Elements von rechts. Wir zeigen, dass das auch ohne Kommutativität für jede Gruppe möglich ist und dass das neutrale und inverse Element jeweils eindeutig sind.

Satz 3.9. In jeder Gruppe ist das neutrale Element e eindeutig bestimmt und jedes Element $a \in G$ besitzt genau ein inverses Element $b \in G$. Außerdem gilt $a * e = a$ und $a * b = e$.

Beweis. Seien $a, b \in G$ mit $b * a = e$ und sei $c \in G$ ein inverses Element zu b, also $c * b = e$. Zunächst gilt

$$a * b \overset{\text{(n)}}{=} (e * a) * b \overset{\text{(i)}}{=} ((c * b) * a) * b \overset{\text{(a)}}{=} (c * (b * a)) * b \overset{\text{(i)}}{=} (c * e) * b \overset{\text{(a)}}{=} c * (e * b) \overset{\text{(n)}}{=} c * b \overset{\text{(i)}}{=} e.$$

Bezeichnen wir die damit bewiesene rechte Inversenbedingung mit (i'), so folgt außerdem

$$a * e \overset{\text{(i)}}{=} a * (b * a) \overset{\text{(a)}}{=} (a * b) * a \overset{\text{(i')}}{=} e * a \overset{\text{(n)}}{=} a.$$

Diese rechte Neutralitätsbedingung bezeichnen wir mit (n'). Seien nun $e, e' \in G$ neutrale Elemente. Mit der gerade bewiesenen zweiten Neutralitätsbedingung ist $e \overset{\text{(n)}}{=} e' * e \overset{\text{(n')}}{=} e'$. Seien schließlich $b, c \in G$ inverse Elemente zu $a \in G$. Mit den bereits bewiesenen zweiten Bedingungen und der Assoziativität folgt

$$b \overset{\text{(n')}}{=} b * e \overset{\text{(i')}}{=} b * (a * c) \overset{\text{(a)}}{=} (b * a) * c \overset{\text{(i)}}{=} e * c \overset{\text{(n)}}{=} c. \qquad \square$$

Wegen Satz 3.9 sprechen wir fortan von *dem* neutralen Element e und bezeichnen das inverse Element zu $a \in G$ mit a^{-1}. Wird die Verknüpfung als Addition geschrieben, bezeichnen wir das neutrale Element üblicherweise mit 0 und das inverse Element zu a mit $-a$.

Beispiel 3.10. *Es ist $(\mathbb{Z}, +)$ eine abelsche Gruppe: Ihr neutrales Element ist 0 und das inverse Element zu $a \in \mathbb{Z}$ ist $-a$.* △

Beispiel 3.11. *Es ist $(\{-1, 1\}, \cdot)$ eine abelsche Gruppe: Ihr neutrales Element ist 1 und jedes Element ist zu sich selbst invers.* △

Beispiel 3.12. *Es ist $(\mathcal{P}(M), \cup)$ mit $M \neq \emptyset$ **keine** Gruppe. Die Vereinigung ist zwar assoziativ (beide Klammerungen enthalten genau die Elemente, die in mindestens einer der drei Mengen liegen) und besitzt auf $\mathcal{P}(M)$ das neutrale Element $\emptyset$, doch keine nichtleere Teilmenge $A \subseteq M$ hat ein inverses Element, wegen $A \cup B \neq \emptyset$ für alle $B \subseteq M$. Die Gleichung $A \cup X = B$ ist nicht immer lösbar (z.B. $A = \{1, 2\}$ und $B = \{2\}$) und kann mehrere Lösungen X zulassen (z.B. für $A = \{2\}$ und $B = \{1, 2\}$ sind $X = \{1\}$ und $X = \{1, 2\}$ Lösungen). Anschaulich verliert die Vereinigung Information, denn aus $A \cup X$ können wir nicht auf X zurückschließen, auch wenn wir A kennen – mit anderen Worten ist $X \mapsto A \cup X$ nicht umkehrbar bzw. bijektiv.* △

Beispiel 3.13. *Es ist $(\mathrm{perm}(M), \circ)$ für jede Menge M eine Gruppe: Nach Satz 1.82 ist die Verkettung zweier Permutationen wieder eine Permutation. Das neutrale Element ist id_M. Nach Satz 1.76 besitzt jede Permutation f die Umkehrabbildung f^{-1} als inverses Element. Die Assoziativität folgt unmittelbar aus der Definition der Verkettung. Die Gruppe ist aber im Allgemeinen nicht abelsch, denn die Reihenfolge von Umsortierungen ist wichtig: Betrachte zum Beispiel für $M = \{a, b, c\}$ die Umsortierungen $f\colon M \to M$, $a \mapsto a, b \mapsto c, c \mapsto b$ (also die Vertauschung $b \leftrightarrow c$) und $g\colon M \to M$, $a \mapsto b, b \mapsto c, c \mapsto a$ (also die zyklische Umsortierung $a \to b \to c \to a$), dann ist $f \circ g = (a \mapsto c, b \mapsto b, c \mapsto a)$ (also $a \leftrightarrow c$), aber $g \circ f = (a \mapsto b, b \mapsto a, c \mapsto c)$ (also $a \leftrightarrow b$).* △

Beispiel 3.14. *Für festes $n \in \mathbb{N} \setminus \{0\}$ ist die* Bitgruppe

$$B_n \coloneqq (\{0, 1\}^n, \oplus)$$

eine abelsche Gruppe (siehe auch Bemerkung 1.8), wobei

$$(x_1, \ldots, x_n) \oplus (y_1, \ldots, y_n) \coloneqq ((x_1 + y_1) \bmod 2, \ldots, (x_n + y_n) \bmod 2).$$

Die Rechenregeln aus Satz 2.44 zeigen die Assoziativität und Kommutativität. Das Tupel $(0, 0, \ldots, 0)$ ist das neutrale Element, und jedes Tupel ist sein eigenes inverses Element. Diese Gruppe modelliert die Manipulation von Bits in der Binärdarstellung von Daten. △

Beispiel 3.15. *Sei $n \in \mathbb{N} \setminus \{0\}$. Für $a \in \mathbb{Z}$ nennen wir*

$$[a]_n \coloneqq \{b \in \mathbb{Z} \mid b \bmod n = a \bmod n\}$$

die Restklasse *von a modulo n und verwenden für die Menge dieser Restklassen die Schreibweise*

$$\mathbb{Z}/n\mathbb{Z} \coloneqq \{[a]_n \mid a \in \mathbb{Z}\} = \{[0]_n, \ldots, [n-1]_n\}.$$

Wir definieren eine Addition auf dieser Menge durch

$$[a]_n + [b]_n \coloneqq [a + b]_n.$$

Diese Definition ist unabhängig von der Wahl der Repräsentanten: Sind $[a]_n = [a']_n$ und $[b]_n = [b']_n$, so unterscheiden sich a und a' sowie b und b' jeweils um ein Vielfaches von n. Das gilt dann auch für $a + b$ und $a' + b'$, also ist $[a + b]_n = [a' + b']_n$. Die Assoziativität und Kommutativität folgen aus den entsprechenden Eigenschaften der Addition auf $\mathbb{Z}$. Das neutrale Element ist $[0]_n$, und das inverse Element zu $[a]_n$ ist $[-a]_n$. Somit ist $(\mathbb{Z}/n\mathbb{Z}, +)$ eine abelsche Gruppe. △

Wir verallgemeinern jetzt Satz 3.7. Außerdem erhalten wir eine Verallgemeinerung von Satz 2.20 auf Gruppen, also insbesondere für die Gruppe $(\mathbb{Z}, +)$.

Satz 3.16. Sei $(G, *)$ eine Gruppe und seien $a, b, c \in G$. Dann gelten die folgenden Aussagen.

(i) Die Gleichungen $a * x = b$ und $x * a = b$ besitzen jeweils genau eine Lösung. Diese sind

$$x = a^{-1} * b \qquad \text{bzw.} \qquad x = b * a^{-1}.$$

(ii) Es gelten die Kürzungsregeln

$$(a * b = a * c) \Rightarrow b = c \qquad \text{und} \qquad (b * a = c * a) \Rightarrow b = c.$$

(iii) Für das inverse Element gilt

$$(a * b)^{-1} = b^{-1} * a^{-1}.$$

Beweis. (i): Es gilt

$$a * (a^{-1} * b) \overset{\text{(a)}}{=} (a * a^{-1}) * b \overset{3.9}{=} e * b \overset{\text{(n)}}{=} b,$$

also ist $a^{-1} * b$ eine Lösung der ersten Gleichung. Ist $y \in G$ eine beliebige Lösung, so folgt

$$y \overset{\text{(n)}}{=} e * y \overset{\text{(i)}}{=} (a^{-1} * a) * y \overset{\text{(a)}}{=} a^{-1} * (a * y) = a^{-1} * b.$$

Für die zweite Gleichung gilt entsprechend

$$(b * a^{-1}) * a \overset{\text{(a)}}{=} b * (a^{-1} * a) \overset{\text{(i)}}{=} b * e \overset{3.9}{=} b.$$

Ist $y * a = b$, so folgt

$$y \overset{3.9}{=} y * e \overset{3.9}{=} y * (a * a^{-1}) \overset{\text{(a)}}{=} (y * a) * a^{-1} = b * a^{-1}.$$

Damit ist auch diese Lösung eindeutig.

(ii): Gilt $a * b = a * c$, so sind b und c Lösungen derselben Gleichung $a * x = a * b$ und daher nach (i) gleich. Die zweite Kürzungsregel folgt entsprechend aus der eindeutigen Lösbarkeit von $x * a = b * a$.

(iii): Schließlich gilt

$$(a * b) * (b^{-1} * a^{-1}) \overset{\text{(a)}}{=} a * ((b * b^{-1}) * a^{-1}) \overset{3.9}{=} a * (e * a^{-1}) \overset{\text{(n)}}{=} a * a^{-1} \overset{3.9}{=} e,$$
$$(b^{-1} * a^{-1}) * (a * b) \overset{\text{(a)}}{=} b^{-1} * ((a^{-1} * a) * b) \overset{\text{(i)}}{=} b^{-1} * (e * b) \overset{\text{(n)}}{=} b^{-1} * b \overset{\text{(i)}}{=} e.$$

Nach der Eindeutigkeit des inversen Elements aus Satz 3.9 ist daher $(a * b)^{-1} = b^{-1} * a^{-1}$. □

Satz 3.16 (i) zeigt insbesondere, dass in jeder Gruppe Abbildungen der Form $x \mapsto a * x$ und $x \mapsto x * a$ jeweils bijektiv sind. Satz 3.16 (iii) zeigt, dass wir Verknüpfungen dieser Abbildungen in umgekehrter Reihenfolge rückgängig machen müssen. Natürlich ist das nur relevant, wenn die Gruppe nicht abelsch ist, wie für die Umsortierungen aus Beispiel 3.13.

Satz 3.17. Für jede Gruppe $(G, *)$ mit neutralem Element e gilt $e^{-1} = e$ und $(a^{-1})^{-1} = a$ für alle $a \in G$.

Beweis. Wegen $e * e = e$ ist e sein eigenes inverses Element, also $e^{-1} = e$. Nach Satz 3.9 gilt $g * g^{-1} = e$. Somit ist g inverses Element zu g^{-1}. Die Eindeutigkeit des inversen Elements (Satz 3.9) ergibt $(g^{-1})^{-1} = g$. □

Die folgende Konvention für leere Summen ist praktisch.

Definition 3.18. In einer additiv geschriebenen Gruppe $(G, +)$ definieren wir die *leere Summe* $\sum_{j=i}^{m}$ mit $i, m \in \mathbb{N}$ und $m < i$, also eine Summe ohne Summanden, als das neutrale Element 0.

Untergruppen

Wir wollen uns jetzt mit der inneren Struktur von Gruppen beschäftigen, genauer mit der Frage, welche Teilmengen bezüglich der Gruppenverknüpfung wieder eine Gruppe bilden.

Definition 3.19. Eine Teilmenge $U \subseteq G$ heißt *Untergruppe einer Gruppe* $(G, *)$, falls $(U, *) \coloneqq (U, *|_{U^2})$ eine Gruppe ist.

Damit eine Teilmenge U eine Untergruppe sein kann, muss die Einschränkung der Verknüpfung $*|_{U^2}$ wieder eine Verknüpfung auf U sein, die Verknüpfung zweier Elemente aus U muss also jeweils wieder in U liegen. Wir sagen dann, dass U *abgeschlossen* bezüglich $*$ ist. Zusätzlich muss U ein neutrales Element sowie alle inversen Elemente der eigenen Elemente enthalten. Die Assoziativität der Gruppe überträgt sich hingegen direkt.

Beispiel 3.20. *Für $n \in \mathbb{N}$ ist*

$$n\mathbb{Z} \coloneqq \{k \cdot n \mid k \in \mathbb{Z}\}$$

eine Untergruppe von $(\mathbb{Z}, +)$: Für alle $kn, k'n \in n\mathbb{Z}$ ist $kn + k'n = (k + k')n \in n\mathbb{Z}$, also ist $n\mathbb{Z}$ abgeschlossen bezüglich $+$. Außerdem ist $0 \in n\mathbb{Z}$ und für jedes $kn \in n\mathbb{Z}$ ist auch $-kn \in n\mathbb{Z}$. Insbesondere bilden die geraden Zahlen $2\mathbb{Z} = \{\dots, -6, -4, -2, 0, 2, 4, 6, \dots\}$ eine Untergruppe von $(\mathbb{Z}, +)$. △

Beispiel 3.21. *Wenn wir die natürlichen Zahlen $\mathbb{N}$ mit der Menge $N = \{z \in \mathbb{Z} \mid z \geq 0\}$ identifizieren, ist diese Menge **keine** Untergruppe von $(\mathbb{Z}, +)$: Zwar ist $\mathbb{N}$ abgeschlossen unter $+$ und enthält das neutrale Element 0, nicht aber alle Inversen.* △

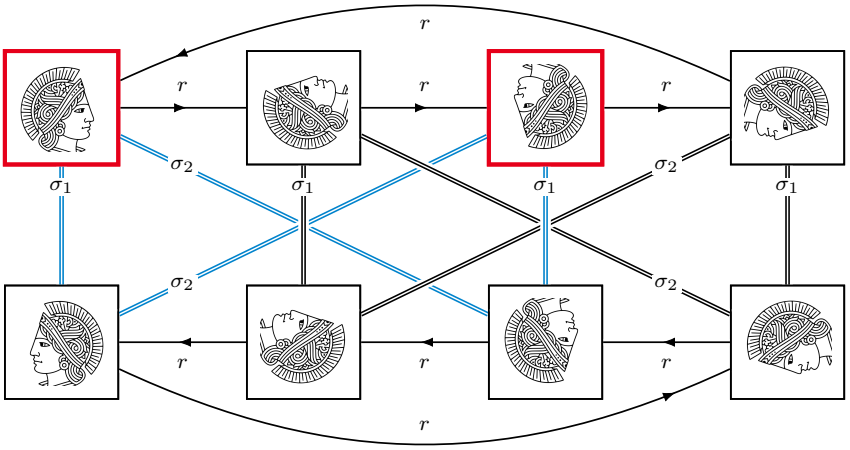


Abbildung 3.1: Die Drehungen und Spiegelungen aus den Beispielen 3.22 und 3.23. Jede Kombination von Rotationen ergibt eine Rotation, aber nicht jede Kombination von Spiegelungen ergibt eine Spiegelung. Beispielsweise können wir die rot markierten Bilder durch die Kombinationen $\sigma_1 \circ \sigma_2$ und $\sigma_2 \circ \sigma_1$ ineinander überführen, was einer Rotation, aber keiner Spiegelung entspricht. Die erzeugte Untergruppe $\langle\{\sigma_1, \sigma_2\}\rangle$ aus Beispiel 3.30 ist blau dargestellt.

Beispiel 3.22. *Wir können Rotationen eines Bildes um Vielfache von* $90°$ *als Permutationen* $\pi \in \mathrm{perm}(\{1,2,3,4\})$ *beschreiben (siehe Abb. 3.1), wobei wir die Ecken des Bildes gegen den Uhrzeigersinn mit* $1, 2, 3, 4$ *bezeichnen: Jede Drehung ist eindeutig dadurch beschrieben, wie die Ecken aufeinander abgebildet werden. Eine Drehung um* $90°$ *ist dann durch die zyklische Permutation* $r\colon 1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 4, 4 \mapsto 1$ *beschrieben, die Drehung um* $180°$ *durch* $r \circ r = (1 \mapsto 3, 2 \mapsto 4, 3 \mapsto 1, 4 \mapsto 2) = (r \circ r)^{-1}$ *und die Drehung um* $270°$ *durch* $r \circ r \circ r = (1 \mapsto 4, 2 \mapsto 1, 3 \mapsto 2, 4 \mapsto 3) = r^{-1}$. *Damit ist*

$$R \coloneqq \{\mathrm{id}, r, r \circ r, r \circ r \circ r\}$$

Untergruppe der Gruppe $(\mathrm{perm}(\{1,2,3,4\}), \circ)$ *(siehe Beispiel 3.13): Die Menge ist abgeschlossen unter* $\circ$*, da jede Kombination von Drehungen wieder eine Drehung ergibt, sie enthält das neutrale Element* id *sowie das Inverse jeder Drehung.* △

Beispiel 3.23. *Analog zu Beispiel 3.22 können wir die Spiegelung eines Bildes entlang der zwei Achsen beschreiben als die Permutationen* $\sigma_1\colon 1 \mapsto 4, 2 \mapsto 3, 3 \mapsto 2, 4 \mapsto 1$ *und* $\sigma_2\colon 1 \mapsto 2, 2 \mapsto 1, 3 \mapsto 4, 4 \mapsto 3$ *(siehe Abb. 3.1). Dann ist* $\sigma_1 = \sigma_1^{-1}$ *und* $\sigma_2 = \sigma_2^{-1}$*, also enthält die Menge*

$$S \coloneqq \{\mathrm{id}, \sigma_1, \sigma_2\}$$

das neutrale Element der Gruppe $(\mathrm{perm}(\{1,2,3,4\}), \circ)$ *sowie alle Inversen von Elementen in* S*. Allerdings ist* S *nicht abgeschlossen unter* $\circ$*, da beispielsweise* $\sigma_1 \circ \sigma_2 = (1 \mapsto 3, 2 \mapsto 4, 3 \mapsto 1, 4 \mapsto 2) = r \circ r \notin S$*, denn die Spiegelung um zwei Achsen ergibt eine Drehung um* $180°$ *und keine Spiegelung. Also ist* S ***keine*** *Untergruppe.* △

Der folgende Satz fasst die Bedingungen, abgeschlossen zu sein sowie ein neutrales Element und alle Inversen zu enthalten, in einer handlichen Bedingung zusammen.

Satz 3.24 (Untergruppenkriterium)**.** Sei $(G, *)$ eine Gruppe. Eine Teilmenge $U \subseteq G$ ist genau dann eine Untergruppe von $(G, *)$, wenn $U \neq \emptyset$ und $a * b^{-1} \in U$ für alle $a, b \in U$ gilt.

Beweis. Sei zunächst U eine Untergruppe von G. Die Existenz des neutralen Elements e_U von $(U, *)$ impliziert $U \neq \emptyset$. Bezeichnet e_G das neutrale Element von $(G, *)$, so gilt für jedes $u \in U \subseteq G$, dass

$$e_U * u = u = e_G * u.$$

Mit der Kürzungsregel aus Satz 3.16 (ii) folgt $e_U = e_G$. Ist b' inverses Element von $b \in U$ in der Gruppe $(U, *)$, dann gilt $b' * b = e_U = e_G$, also ist b' auch inverses Element von b in $(G, *)$. Wegen der Eindeutigkeit aus Satz 3.9 ist $b' = b^{-1}$, also $b^{-1} \in U$. Da U abgeschlossen bezüglich $*$ ist, bzw. $*|_{U^2} \colon U^2 \to U$ eine Verknüpfung ist, liegt $a * b^{-1} \in U$.

Sei nun umgekehrt $U \neq \emptyset$ und es gelte $a * b^{-1} \in U$ für alle $a, b \in U$. Wegen Assoziativität der Gruppe $(G, *)$ ist auch $*|_{U^2}$ assoziativ. Wegen $U \neq \emptyset$ können wir ein $u \in U$ wählen. Nach Annahme mit $a = b = u$ ist $u * u^{-1} \in U$, also enthält U das neutrale Element $e = u * u^{-1} \in U$ von $(G, *)$.

Für jedes $u \in U$ ist also $u^{-1} = e * u^{-1}$. Die Annahme mit $a = e$ und $b = u$ impliziert $u^{-1} \in U$, also enthält U das inverse Element jedes Elements in U.

Für alle $u, v \in U$ ist also auch $v^{-1} \in U$. Die Annahme mit $a = u$ und $b = v^{-1}$ impliziert $u * v = u * (v^{-1})^{-1} \in U$, wobei wir Satz 3.17 verwendet haben. Somit ist U abgeschlossen unter $*$, bzw. $*|_{U^2}$ ist eine Verknüpfung auf U. Damit haben wir alle Gruppeneigenschaften von $(U, *|_{U^2})$ nachgewiesen, also ist U eine Untergruppe. □

Satz 3.25. Sei $\mathcal{U}$ eine nichtleere Menge von Untergruppen einer Gruppe G. Dann ist auch $\bigcap_{U \in \mathcal{U}} U$ eine Untergruppe von G.

Beweis. Übung. □

Ist eine Teilmenge einer Gruppe keine Untergruppe, können wir sie um alle Elemente ergänzen, die für eine Untergruppe notwendig sind. Formal betrachten wir dazu alle Untergruppen, welche die gegebene Teilmenge enthalten.

Definition 3.26. Sei $(G, *)$ eine Gruppe. Für $M \subseteq G$ definieren wir die von M *erzeugte Untergruppe* durch

$$\langle M \rangle_G \coloneqq \langle M \rangle \coloneqq \bigcap \{U \subseteq G \mid U \text{ ist Untergruppe von } G \text{ und } M \subseteq U\}.$$

Satz 3.27. Die Menge $\langle M \rangle$ ist die kleinste Untergruppe von G, die M enthält, für jede solche Untergruppe U gilt also $\langle M \rangle \subseteq U$ und $\langle M \rangle = M$, falls M bereits eine Untergruppe ist.

Beweis. Übung. □

Beispiel 3.28. *Für $n \in \mathbb{N}$ ist $\langle \{n\} \rangle_{(\mathbb{Z},+)} = n\mathbb{Z}$ (siehe Beispiel 3.20), denn jede Untergruppe von $(\mathbb{Z}, +)$, die n enthält, enthält wegen der Abgeschlossenheit und der Existenz inverser Elemente auch alle ganzzahligen Vielfachen von n.* △

Beispiel 3.29. *Für die Permutation r aus Beispiel 3.22, die einer Rotation um $90°$ entspricht, gilt $\langle\{r\}\rangle = \langle\{r\}\rangle_{(\mathrm{perm}(\{1,2,3,4\}),\circ)} = R$: Da sich jede Drehung in R durch mehrfache Anwendung von r erreichen lässt, erfordert Abgeschlossenheit, dass $R \subseteq U$ für jede Untergruppe U gilt, die r enthält. Damit ist nach Definition 3.26 $R \subseteq \langle\{r\}\rangle$. Da R Untergruppe ist, folgt mit Satz 3.27 auch $\langle\{r\}\rangle \subseteq R$, also $\langle\{r\}\rangle = R$.* △

Beispiel 3.30. *Für die Menge S aus Beispiel 3.23 gilt $\langle S\rangle = S \cup \{r \circ r\}$ (siehe Abb. 3.1): Durch Spiegelung entlang der Achsen können wir eine Rotation um $180°$ erreichen. Nehmen wir die entsprechende Permutation $r \circ r$ in die Menge auf, ist sie abgeschlossen unter $\circ$. Da jedes Element von $S \cup \{r \circ r\}$ sein eigenes Inverses ist, ist damit die Bedingung aus Satz 3.24 erfüllt und $S \cup \{r \circ r\}$ somit eine Untergruppe. Analog zu Beispiel 3.29 ist damit $\langle S\rangle = S \cup \{r \circ r\}$, da jede Untergruppe, die S enthält, auch $r \circ r$ enthalten muss.* △

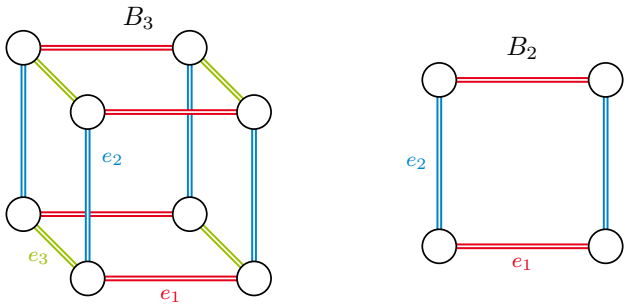


Abbildung 3.2: Die Graphen der Bitgruppen B_3 und B_2 aus Beispiel 3.31; die Farben entsprechen den Erzeugern e_1, e_2, e_3.

Beispiel 3.31. *Sei B_n die Bitgruppe aus Beispiel 3.14. Für $i \in \{1, \ldots, n\}$ sei $e_i \in \{0,1\}^n$ die Bitfolge, deren i-te Stelle 1 und deren übrige Stellen 0 sind. Dann ist $B_n = \langle\{e_1, \ldots, e_n\}\rangle$, denn jede Bitfolge $x = (x_1, \ldots, x_n)$ entsteht aus dem neutralen Element $e_0 := (0, \ldots, 0)$, indem wir genau diejenigen e_i mit $x_i = 1$ durch $\oplus$ verknüpfen (siehe Abb. 3.2). Beispielsweise, für B_3, ist $(0,1,1) = e_0 \oplus e_2 \oplus e_3 = (0,0,0) \oplus (0,1,0) \oplus (0,0,1) = (0 \oplus 0 \oplus 0, 0 \oplus 1 \oplus 0, 0 \oplus 0 \oplus 1)$.* △

Gruppenhomomorphismen

Der Wert abstrakter Strukturen wie Gruppen liegt insbesondere darin, dass wir dieselbe Struktur in sehr unterschiedlichen Zusammenhängen wiedererkennen und Erkenntnisse zwischen diesen Zusammenhängen übertragen können. Wir wollen jetzt präzisieren, was es heißt, dass zwei Gruppen trotz verschiedenartiger Elemente dieselbe Struktur besitzen.

Betrachten wir dazu noch einmal die Bitgruppe B_n (Beispiele 3.14 und 3.31). Intuitiv tritt B_2 als Substruktur von B_3 auf (siehe Abb. 3.2). Genauer entspricht B_2 der erzeugten Untergruppe

$$U := \langle\{e_1, e_2\}\rangle = \{(x_1, x_2, 0) \mid x_1, x_2 \in \{0,1\}\}.$$

Aber auch die aus den Spiegelungen erzeugte Untergruppe $\langle S\rangle = \langle\{\sigma_1, \sigma_2\}\rangle$ aus Beispiel 3.30 hat dieselbe Struktur: Sie hat vier Elemente und die Erzeuger σ_1 und σ_2 sind jeweils ihre eigenen Inversen.

Um zu formalisieren, dass B_2 als Substruktur von B_3 auftritt, geben wir eine Abbildung $\varphi\colon B_2 \to B_3$ an, wobei diese Abbildung strukturerhaltend sein sollte: Ist $a \oplus b = c$, dann sollte auch $\varphi(a) \oplus \varphi(b) = \varphi(c)$ gelten. Eine solche Abbildung ist beispielsweise $(x_1, x_2) \mapsto (x_1, x_2, 0)$.

Analog formalisieren wir, dass $\langle S\rangle$ dieselbe Struktur wie die Untergruppe U besitzt, indem wir eine Bijektion $\psi\colon \langle S\rangle \to U$ angeben, sodass für $\pi\circ\sigma = \nu$ auch $\psi(\pi)\oplus\psi(\sigma) = \psi(\nu)$ gilt. Eine solche Bijektion ist beispielsweise $(\mathrm{id} \mapsto e_0,\ \sigma_1 \mapsto e_1,\ \sigma_2 \mapsto e_2,\ r\circ r \mapsto e_1 \oplus e_2)$.

Wir führen Begriffe für solche strukturerhaltenden Abbildungen ein. Wir erlauben dabei auch Abbildungen, die Informationen „vergessen“, wie zum Beispiel $p\colon B_3 \to B_3,\ (x_1, x_2, x_3) \mapsto (x_1, x_2, 0)$. Obwohl p nicht injektiv ist, bleibt die Struktur erhalten, denn für $a \oplus b = c$ gilt auch $p(a) \oplus p(b) = p(c)$.

Definition 3.32. Eine Abbildung $f\colon G \to H$ heißt *(Gruppen-)Homomorphismus* zwischen den Gruppen $(G, *)$ und $(H, \diamond)$, falls

$$f(a * b) = f(a) \diamond f(b)$$

für alle $a, b \in G$ gilt. Ein bijektiver Gruppenhomomorphismus heißt *(Gruppen-)Isomorphismus*. Zwei Gruppen heißen *isomorph*, wenn zwischen ihnen ein Isomorphismus existiert. Ein Isomorphismus von G nach G heißt *(Gruppen-)Automorphismus*.

Beispiel 3.33. *Für jede Gruppe $(G, *)$ ist id_G ein Gruppenautomorphismus: Die Abbildung ist bijektiv und für alle $a, b \in G$ gilt* $\mathrm{id}_G(a * b) = a * b = \mathrm{id}_G(a) * \mathrm{id}_G(b)$. △

Beispiel 3.34. *Für zwei Gruppenhomomorphismen $f\colon (G, *) \to (H, \diamond)$ und $g\colon (H, \diamond) \to (K, \bullet)$ ist auch $g \circ f$ ein Gruppenhomomorphismus: Für alle $a, b \in G$ gilt*

$$(g \circ f)(a * b) = g(f(a) \diamond f(b)) = g(f(a)) \bullet g(f(b)) = (g \circ f)(a) \bullet (g \circ f)(b).$$ △

Beispiel 3.35. *Für einen Gruppenisomorphismus $f\colon (G, *) \to (H, \diamond)$ ist auch f^{-1} ein Gruppenisomorphismus: Als Umkehrabbildung ist f^{-1} bijektiv und für $x = f(a)$ und $y = f(b)$ gilt*

$$f^{-1}(x \diamond y) = f^{-1}(f(a) \diamond f(b)) = f^{-1}(f(a * b)) = a * b = f^{-1}(x) * f^{-1}(y).$$ △

Beispiel 3.36. *Für die Gruppe $\langle S\rangle$ aus Beispiel 3.30 ist*

$$\alpha\colon \langle S\rangle \to \langle S\rangle, \qquad f \mapsto r \circ f \circ r^{-1},$$

(also $\alpha(\mathrm{id}) = \mathrm{id}$, $\alpha(\sigma_1) = \sigma_2$, $\alpha(\sigma_2) = \sigma_1$, $\alpha(r \circ r) = r \circ r$) ein Gruppenautomorphismus: Die Abbildung ist bijektiv und für alle $f, g \in \langle S\rangle$ gilt

$$\alpha(f \circ g) = r \circ f \circ g \circ r^{-1} = (r \circ f \circ r^{-1}) \circ (r \circ g \circ r^{-1}) = \alpha(f) \circ \alpha(g).$$ △

Beispiel 3.37. *Für $n \in \mathbb{N} \setminus \{0\}$ ist die Abbildung*

$$\pi_n\colon (\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +), \quad a \mapsto [a]_n,$$

die jeder ganzen Zahl ihre Restklasse modulo n zuordnet (siehe Beispiel 3.15), ein Gruppenhomomorphismus: Für alle $a, b \in \mathbb{Z}$ gilt

$$\pi_n(a + b) = [a + b]_n = [a]_n + [b]_n = \pi_n(a) + \pi_n(b).$$ △

Satz 3.38. Für einen Gruppenhomomorphismus $f\colon G \to H$ bezüglich der Gruppen $(G, *)$ und $(H, \diamond)$ mit den neutralen Elementen e_G und e_H gilt:

(i) Es ist $f(e_G) = e_H$.

(ii) Für jedes $g \in G$ gilt $f(g^{-1}) = f(g)^{-1}$.

(iii) Das Bild bildet eine Untergruppe $(f(G), \diamond)$ von H.

(iv) Ist G abelsch, so ist auch $f(G)$ abelsch.

Beweis. (i): Aus der Homomorphieeigenschaft folgt

$$f(e_G) = f(e_G * e_G) = f(e_G) \diamond f(e_G).$$

Andererseits ist $f(e_G) = f(e_G) \diamond e_H$. Mit der Kürzungsregel aus Satz 3.16 (ii) folgt daher $f(e_G) = e_H$.

(ii): Für $g \in G$ ist

$$f(g^{-1}) \diamond f(g) = f(g^{-1} * g) = f(e_G) = e_H.$$

Wegen der Eindeutigkeit des inversen Elements (Satz 3.9) ist somit $f(g^{-1}) = f(g)^{-1}$.

(iii): Nach Teil (i) gilt $e_H = f(e_G) \in f(G)$, also ist $f(G) \neq \emptyset$. Für $x, y \in f(G)$ existieren $a, b \in G$ mit $x = f(a)$ und $y = f(b)$, und mit Teil (ii) gilt

$$x \diamond y^{-1} = f(a) \diamond f(b)^{-1} = f(a) \diamond f(b^{-1}) = f(a * b^{-1}) \in f(G).$$

Nach dem Untergruppenkriterium aus Satz 3.24 ist $f(G)$ daher eine Untergruppe von H.

(iv): Ist G abelsch und sind $x = f(a) \in f(G)$ und $y = f(b) \in f(G)$, so gilt

$$x \diamond y = f(a) \diamond f(b) = f(a * b) = f(b * a) = f(b) \diamond f(a) = y \diamond x.$$ □

Beispiel 3.39. *Wir greifen unser anfängliches Beispiel für die Bitgruppe B_3 auf. Für den Homomorphismus $p\colon B_3 \to B_3$, $(x_1, x_2, x_3) \mapsto (x_1, x_2, 0)$ gilt $p((0,0,0)) = (0,0,0)$ (Satz 3.38 (i)) und $p((x_1, x_2, x_3))^{-1} = (x_1, x_2, 0) = p((x_1, x_2, x_3)) = p((x_1, x_2, x_3)^{-1})$ (Satz 3.38 (ii)). Wir hatten bereits festgestellt, dass das Bild $p(B_3) = U$ eine Untergruppe ist, im Einklang mit Satz 3.38 (iii). Da B_3 abelsch ist (Beispiel 3.14), ist die Untergruppe gemäß Satz 3.38 (iv) ebenfalls abelsch.* △

Während das Bild beschreibt, welche Elemente im Ziel erreicht werden, soll der Kern den Informationsverlust eines Homomorphismus erfassen. So haben etwa $a = (1,0,0)$ und $b = (1,0,1)$ unter p dasselbe Bild $p(a) = a = p(b)$. Die Elemente unterscheiden sich um die Lösung $x = (0,0,1)$ der Gleichung $a \oplus x = b$, während sich ihr Bild nicht unterscheidet. Wegen der Homomorphieeigenschaft ist das gleichbedeutend mit $p(a) \oplus p(x) = p(a \oplus x) = p(b) = p(a) \oplus (0,0,0)$, also mit der Kürzungsregel (Satz 3.16 (ii)) mit $p(x) = (0,0,0)$. Tatsächlich ist $p((0,0,1)) = (0,0,0)$. Wir können also die Menge der Elemente, die mit der Verknüpfung keinen Unterschied im Bild machen, wie folgt erfassen.

Definition 3.40. Der *Kern* eines Gruppenhomomorphismus $f\colon G \to H$ bezüglich der Gruppen $(G, *)$ und $(H, \diamond)$ mit den neutralen Elementen e_G und e_H ist

$$\ker(f) \coloneqq \{g \in G \mid f(g) = e_H\}.$$

Wir halten eine zentrale Eigenschaft des Kerns fest.

Satz 3.41. Für jeden Gruppenhomomorphismus $f\colon G \to H$ bezüglich der Gruppen $(G, *)$ und $(H, \diamond)$ ist $(\ker(f), *)$ eine Untergruppe von G.

Beweis. Nach Satz 3.38 (i) gilt $f(e_G) = e_H$ für die neutralen Elemente e_G und e_H der Gruppen, also ist auch $e_G \in \ker(f)$. Damit ist der Kern nicht leer. Für $a, b \in \ker(f)$ gilt mit Satz 3.38 (ii)

$$f(a * b^{-1}) = f(a) \diamond f(b^{-1}) = f(a) \diamond f(b)^{-1} = e_H \diamond e_H^{-1} = e_H.$$

Somit ist $a * b^{-1} \in \ker(f)$, und die Behauptung folgt aus dem Untergruppenkriterium (Satz 3.24). □

Der Kern misst damit genau den Informationsverlust eines Gruppenhomomorphismus: Enthält er außer dem neutralen Element kein weiteres Element, so kann der Homomorphismus keine zwei verschiedenen Elemente miteinander identifizieren. Dies liefert ein einfaches Kriterium für Injektivität.

Satz 3.42. Ein Gruppenhomomorphismus $f\colon G \to H$ bezüglich der Gruppen $(G, *)$ und $(H, \diamond)$ mit den neutralen Elementen e_G und e_H ist genau dann injektiv, wenn $\ker(f) = \{e_G\}$.

Beweis. Sei zunächst f injektiv. Für $a \in \ker(f)$ ist dann nach Satz 3.38 (i) $f(a) = e_H = f(e_G)$ und folglich $a = e_G$. Da stets $e_G \in \ker(f)$ gilt, ist also $\ker(f) = \{e_G\}$.

Sei umgekehrt $\ker(f) = \{e_G\}$ und seien $a, b \in G$ mit $f(a) = f(b)$. Nach Satz 3.38 (ii) gilt dann

$$f(a * b^{-1}) = f(a) \diamond f(b)^{-1} = f(a) \diamond f(a)^{-1} = e_H,$$

also $a * b^{-1} \in \ker(f)$ und damit $a * b^{-1} = e_G$. Es folgt $a = (a * b^{-1}) * b = e_G * b = b$. Somit ist f injektiv. □

Beispiel 3.43. *Für den Restklassen-Homomorphismus π_n aus Beispiel 3.37 gilt $\ker(\pi_n) = n\mathbb{Z}$, der Kern ist also genau die Menge $n\mathbb{Z}$ aus Beispiel 3.20, denn*

$$a \in \ker(\pi_n) \iff \pi_n(a) = [a]_n = [0]_n \iff a \bmod n = 0 \iff a \in n\mathbb{Z}.$$

△

Beispiel 3.44. *Die Abbildung*

$$\phi_4\colon (\mathbb{Z}, +) \to (\mathbb{Z}, +), \qquad k \mapsto 4k,$$

ist ein Gruppenhomomorphismus, denn für alle $k, \ell \in \mathbb{Z}$ gilt $\phi_4(k + \ell) = 4(k + \ell) = 4k + 4\ell = \phi_4(k) + \phi_4(\ell)$. Sein Bild ist die Untergruppe $\phi_4(\mathbb{Z}) = 4\mathbb{Z}$ und sein Kern ist die triviale Untergruppe $\ker(\phi_4) = \{0\}$. △

Beispiel 3.45. *Die Abbildung*

$$f\colon (\mathbb{Z}/4\mathbb{Z}, +) \to (\mathbb{Z}/4\mathbb{Z}, +), \quad [n]_4 \mapsto [2n]_4$$

ist ein Gruppenhomomorphismus: Für alle $n, m \in \mathbb{Z}$ gilt

$$f([n]_4 + [m]_4) = f([n + m]_4) = [2(n + m)]_4 = [2n + 2m]_4 = f([n]_4) + f([m]_4).$$

Außerdem ist $f([0]_4) = [0]_4$, $f([1]_4) = [2]_4$, $f([2]_4) = [4]_4 = [0]_4$ und $f([3]_4) = [6]_4 = [2]_4$, also ist $\ker(f) = \{[0]_4, [2]_4\}$ und wir können uns vergewissern, dass $\{[0]_4, [2]_4\}$ tatsächlich eine Untergruppe von $(\mathbb{Z}/4\mathbb{Z}, +)$ ist. △

3.3 Ringe

Bisher haben wir mit dem Gruppenbegriff aus Definition 3.8 insbesondere die Struktur der Addition auf $\mathbb{Z}$ untersucht. Wir interessieren uns auf $\mathbb{Z}$ jedoch außerdem für die Multiplikation (Abschnitt 3.1), die mit der Addition zusammenwirkt. Um dieses Zusammenspiel unabhängig von den konkreten ganzen Zahlen untersuchen zu können, führen wir nun den Begriff des Rings ein.

Definition 3.46. Ein Tupel $(R, +, \cdot)$ aus einer Menge $R \neq \emptyset$ und zwei Verknüpfungen $+, \cdot \colon R^2 \to R$ heißt *Ring*, falls

(g) das Tupel $(R, +)$ eine abelsche Gruppe ist,

(m) für alle $a, b, c \in R$ gilt, dass $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (*Assoziativität der Multiplikation*),

(d) für alle $a, b, c \in R$ gilt, dass $a \cdot (b + c) = a \cdot b + a \cdot c$ und $(a + b) \cdot c = a \cdot c + b \cdot c$ (Distributivität).

Das neutrale Element der Gruppe $(R, +)$ heißt *Nullelement* und wird mit 0 bezeichnet. Das additiv inverse Element zu $a \in R$ bezeichnen wir mit $-a$. Außerdem schreiben wir meistens ab statt $a \cdot b$.
Ein Ring $(R, +, \cdot)$ heißt *Ring mit Eins*, falls

$$\exists 1 \in R \colon \forall a \in R \colon 1 \cdot a = a \cdot 1 = a.$$

Das eindeutig bestimmte Element 1 heißt dann *Einselement*. Der Ring heißt *kommutativ*, falls $ab = ba$ für alle $a, b \in R$ gilt.

Die Eindeutigkeit des Einselements folgt wie für das neutrale Element einer Gruppe: Sind $1, 1' \in R$ Einselemente, so ist $1 = 1 \cdot 1' = 1'$. Um Null- und Einselemente mehrerer Ringe unterscheiden zu können, schreiben wir auch 0_R und 1_R.

Beispiel 3.47. *Die in Abschnitt 3.1 definierten Verknüpfungen machen $(\mathbb{Z}, +, \cdot)$ zu einem kommutativen Ring mit Nullelement 0 und Einselement 1. Dagegen ist $(\mathbb{N}, +, \cdot)$ kein Ring, da $(\mathbb{N}, +)$ keine Gruppe ist.* △

Aus der Definition des Rings folgen insbesondere die aus $\mathbb{Z}$ bekannten Vorzeichenregeln.

Satz 3.48. Sei $(R, +, \cdot)$ ein Ring und seien $a, b \in R$. Dann gilt

$$0 \cdot a = a \cdot 0 = 0, \qquad (-a)b = a(-b) = -(ab), \qquad (-a)(-b) = ab.$$

Beweis. Wegen Distributivität (d) gilt $a0 = a(0 + 0) = a0 + a0$. Da auch $a0 = a0 + 0$ ist, folgt mit der Kürzungsregel aus Satz 3.16 (ii) für die Gruppe $(R, +)$, dass $a0 = 0$. Analog folgt $0a = 0$.

Weiter ist, wegen Distributivität der Gruppe $(R, +)$,

$$(-a)b + ab = ((-a) + a)b = 0b = 0.$$

Somit ist $(-a)b$ das Inverse zu ab bezüglich der Addition $+$ und daher $(-a)b = -(ab)$. Die Gleichung $a(-b) = -(ab)$ folgt analog.

Aus der soeben bewiesenen zweiten Gleichung erhalten wir

$$(-a)(-b) = -((-a)b) = -(-(ab)) = ab. \qquad \square$$

Wir greifen nun die Restklassengruppe aus Beispiel 3.15 wieder auf und ergänzen auf derselben Menge eine Multiplikation.

Definition 3.49. Für $n \in \mathbb{N} \setminus \{0\}$ definieren wir auf $\mathbb{Z}/n\mathbb{Z}$ die Multiplikation

$$[a]_n \cdot [b]_n \coloneqq [ab]_n.$$

Die Menge $\mathbb{Z}/n\mathbb{Z}$ mit der Addition aus Beispiel 3.15 und dieser Multiplikation heißt *Restklassenring modulo* n.

Satz 3.50. Die Multiplikation aus Definition 3.49 ist unabhängig von der Wahl der Repräsentanten wohldefiniert. Mit Addition und Multiplikation ist $\mathbb{Z}/n\mathbb{Z}$ ein kommutativer Ring mit Nullelement $[0]_n$ und Einselement $[1]_n$.

Beweis. Seien $[a]_n = [a']_n$ und $[b]_n = [b']_n$. Dann existieren $k, \ell \in \mathbb{Z}$ mit $a - a' = kn$ und $b - b' = \ell n$. Wegen

$$ab - a'b' = a(b - b') + b'(a - a') = (a\ell + b'k)n$$

unterscheiden sich auch ab und $a'b'$ um ein Vielfaches von n. Somit ist $[ab]_n = [a'b']_n$ und die Multiplikation wohldefiniert.

Nach Beispiel 3.15 ist $(\mathbb{Z}/n\mathbb{Z}, +)$ eine abelsche Gruppe. Die weiteren Ringaxiome folgen aus den entsprechenden Rechenregeln auf $\mathbb{Z}$. Beispielsweise gilt für alle $a, b, c \in \mathbb{Z}$

$$([a]_n \cdot [b]_n) \cdot [c]_n = [(ab)c]_n = [a(bc)]_n = [a]_n \cdot ([b]_n \cdot [c]_n)$$

und

$$[a]_n \cdot ([b]_n + [c]_n) = [a(b + c)]_n = [ab + ac]_n = [ab]_n + [ac]_n.$$

Das zweite Distributivgesetz und die Kommutativität folgen analog. Außerdem ist $[1]_n \cdot [a]_n = [a]_n \cdot [1]_n = [a]_n$ für alle $a \in \mathbb{Z}$. $\square$

Beispiel 3.51. *Die Multiplikationstafeln der Restklassenringe $\mathbb{Z}/4\mathbb{Z}$ und $\mathbb{Z}/5\mathbb{Z}$ sind*

$\cdot$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[0]_4$	$[0]_4$	$[0]_4$	$[0]_4$	$[0]_4$
$[1]_4$	$[0]_4$	$[1]_4$	$[2]_4$	$[3]_4$
$[2]_4$	$[0]_4$	$[2]_4$	$[0]_4$	$[2]_4$
$[3]_4$	$[0]_4$	$[3]_4$	$[2]_4$	$[1]_4$

$\cdot$	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$	$[0]_5$
$[1]_5$	$[0]_5$	$[1]_5$	$[2]_5$	$[3]_5$	$[4]_5$
$[2]_5$	$[0]_5$	$[2]_5$	$[4]_5$	$[1]_5$	$[3]_5$
$[3]_5$	$[0]_5$	$[3]_5$	$[1]_5$	$[4]_5$	$[2]_5$
$[4]_5$	$[0]_5$	$[4]_5$	$[3]_5$	$[2]_5$	$[1]_5$

In $\mathbb{Z}/4\mathbb{Z}$ gilt insbesondere $[2]_4 \cdot [2]_4 = [0]_4$, obwohl $[2]_4 \neq [0]_4$. In $\mathbb{Z}/5\mathbb{Z}$ tritt dieses Verhalten nicht auf. $\triangle$

Das Beispiel zeigt, dass in einem Ring das Produkt zweier von Null verschiedener Elemente gleich Null sein kann. Das steht einer Kürzungsregel nach Satz 2.30 entgegen, denn $a \cdot b = 0 = a \cdot 0$ würde mit einer Kürzungsregel auch $b = 0$ implizieren. Wir benennen dieses Hindernis für die Kürzungsregel und kommen später darauf zurück.

Definition 3.52. Sei $(R, +, \cdot)$ ein kommutativer Ring. Ein Element $a \in R \setminus \{0\}$ heißt *Nullteiler*, falls

$$\exists b \in R \setminus \{0\} \colon ab = 0.$$

Satz 3.53. Sei $(R, +, \cdot)$ ein kommutativer Ring und sei $a \in R \setminus \{0\}$. Das Element a ist genau dann kein Nullteiler, wenn die Abbildung $M_a \colon R \to R,\ x \mapsto ax$ injektiv ist.

Beweis. Sei zunächst a kein Nullteiler und seien $b, c \in R$ mit $M_a(b) = M_a(c)$. Dann ist $ab = ac$ und daher $a(b - c) = 0$. Da $a \neq 0$ kein Nullteiler ist, folgt $b - c = 0$ und somit $b = c$. Also ist M_a injektiv.

Sei umgekehrt M_a injektiv und sei $b \in R$ mit $ab = 0$. Wegen Satz 3.48 ist $M_a(b) = 0 = M_a(0)$. Aus der Injektivität folgt $b = 0$, also ist a kein Nullteiler. □

Satz 3.54. Ist $n \in \mathbb{N}$ mit $n > 1$ keine Primzahl, so besitzt der Restklassenring $\mathbb{Z}/n\mathbb{Z}$ Nullteiler.

Beweis. Ist $n > 1$ keine Primzahl, existieren per Definition $r, s \in \mathbb{N}$ mit $n = rs$ und $1 < r, s < n$. Daher sind $[r]_n \neq [0]_n$ und $[s]_n \neq [0]_n$, aber

$$[r]_n \cdot [s]_n = [rs]_n = [n]_n = [0]_n.$$

Somit sind $[r]_n$ und $[s]_n$ Nullteiler. □

Ringhomomorphismen

Bei Gruppen haben wir strukturerhaltende Abbildungen mit dem Begriff des Gruppenhomomorphismus aus Definition 3.32 erfasst. Wir betrachten Homomorphismen im Folgenden nur zwischen Ringen mit Eins. Eine strukturerhaltende Abbildung zwischen Ringen soll sowohl die Addition als auch die Multiplikation erhalten. Dass die additive Bedingung allein nicht genügt, zeigt der Gruppenhomomorphismus $\phi_4 \colon \mathbb{Z} \to \mathbb{Z}$ aus Beispiel 3.44: Es ist $\phi_4(1 \cdot 1) = 4$, aber $\phi_4(1) \cdot \phi_4(1) = 16$.

Definition 3.55. Eine Abbildung $f \colon R \to S$ heißt *(Ring-)Homomorphismus* zwischen den Ringen mit Eins $(R, +, \cdot)$ und $(S, \oplus, \odot)$, falls für alle $a, b \in R$ gilt, dass

$$f(a + b) = f(a) \oplus f(b), \qquad f(a \cdot b) = f(a) \odot f(b), \qquad f(1_R) = 1_S.$$

Ein bijektiver Ringhomomorphismus heißt *(Ring-)Isomorphismus*. Zwei Ringe mit Eins heißen *isomorph*, wenn zwischen ihnen ein Ringisomorphismus existiert. Ein Ringisomorphismus von R nach R heißt *(Ring-)Automorphismus*.

Beispiel 3.56. *Wie bei Gruppen ist für jeden Ring mit Eins R die Identität id_R ein Ringautomorphismus. Für Ringhomomorphismen $f\colon R \to S$ und $g\colon S \to T$ ist $g \circ f$ ein Ringhomomorphismus, und für einen Ringisomorphismus f ist auch f^{-1} ein Ringisomorphismus. Die Argumente aus den Beispielen 3.33, 3.34 und 3.35 werden dazu jeweils auf beide Verknüpfungen angewendet. Außerdem bleibt die Eins erhalten:*

$$\mathrm{id}_R(1_R) = 1_R, \qquad (g \circ f)(1_R) = g(1_S) = 1_T, \qquad f^{-1}(1_S) = 1_R. \qquad \triangle$$

Beobachtung 3.57. Jeder Ringhomomorphismus $f\colon R \to S$ ist insbesondere ein Gruppenhomomorphismus zwischen den additiven Gruppen $(R, +)$ und $(S, \oplus)$. Nach Satz 3.38 gilt daher insbesondere $f(0_R) = 0_S$ und $f(-a) = -f(a)$ für alle $a \in R$.

Beispiel 3.58. *Für $n \in \mathbb{N} \setminus \{0\}$ ist die bereits in Beispiel 3.37 betrachtete Abbildung*

$$\pi_n\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}, \qquad a \mapsto [a]_n,$$

ein surjektiver Ringhomomorphismus. Neben der Addition erhält sie die Multiplikation, denn für alle $a, b \in \mathbb{Z}$ gilt

$$\pi_n(ab) = [ab]_n = [a]_n \cdot [b]_n = \pi_n(a) \cdot \pi_n(b),$$

und $\pi_n(1) = [1]_n$. Die Abbildung ist surjektiv, da jede Restklasse $[a]_n = \pi_n(a)$ ist. $\triangle$

Wie bei Gruppen ist das Bild unter einem Homomorphismus wieder eine Struktur derselben Art.

Satz 3.59. Sei $f\colon R \to S$ ein Ringhomomorphismus bezüglich der Ringe $(R, +, \cdot)$ und $(S, \oplus, \odot)$. Dann ist $(f(R), \oplus, \odot) = (f(R), \oplus|_{f(R)^2}, \odot|_{f(R)^2})$ ein Ring mit dem Einselement 1_S.

Beweis. Übung. □

3.4 Die rationalen Zahlen

Die Division auf $\mathbb{N}$ haben wir in Definition 2.37 als partielle Umkehrung der Multiplikation definiert. Auch nach der Erweiterung zu den ganzen Zahlen in Abschnitt 3.1 lässt sich die Multiplikation nicht immer umkehren: Die Gleichung $2x = 1$ besitzt keine Lösung $x \in \mathbb{Z}$. Solange wir nur ganze Anzahlen und Änderungen betrachten, ist diese Einschränkung unproblematisch. Wollen wir jedoch eine Einheit in zwei gleich große Anteile zerlegen, benötigen wir gerade eine Zahl x mit $2x = 1$. Das führt uns zur Menge $\mathbb{Q}$ der rationalen Zahlen, die wir durch Erweiterung des Zahlenraums $\mathbb{Z}$ um alle Quotienten erhalten.

Für die Abbildung

$$M_2\colon \mathbb{Z} \to \mathbb{Z}, \qquad x \mapsto 2x,$$

fehlt dabei nur die Surjektivität: Sie ist injektiv, aber die Zahl 1 liegt nicht in ihrem Bild. Dies unterscheidet die fehlenden Quotienten vom Problem der Nullteiler, bei denen nach Satz 3.53 bereits die Injektivität und damit die Kürzungsregel scheitert.

Formal können wir jede Zahl der Form $\frac{a}{b}$ durch ein Tupel $(a, b) \in \mathbb{Z} \times (\mathbb{N} \setminus \{0\})$ darstellen, wobei wir Tupel (a, b) und (c, d) zusammenfassen, die denselben Quotienten beschreiben, für die also $ad = bc$ gilt. Mit anderen Worten ist jede rationale Zahl eine Äquivalenzklasse von Tupeln, die denselben Quotienten beschreiben. Beachte hierbei, dass wir bewusst die Division vermeiden, da wir nicht $b \mid a$ und $d \mid c$ voraussetzen können.

Definition 3.60. Die Menge der *rationalen Zahlen* ist die Faktormenge

$$\mathbb{Q} \coloneqq (\mathbb{Z} \times (\mathbb{N} \setminus \{0\}))/\sim,$$

wobei die Äquivalenzrelation $\sim$ gegeben ist durch

$$\sim \coloneqq \{((a,b),(c,d)) \in (\mathbb{Z} \times (\mathbb{N} \setminus \{0\}))^2 \mid ad = bc\}.$$

Für $a \in \mathbb{Z}$ identifizieren wir $a = [(a,1)]$ und für $a \in \mathbb{Z}$ und $b \in \mathbb{N} \setminus \{0\}$ verwenden wir die vertraute Schreibweise

$$\frac{a}{b} \coloneqq [(a,b)].$$

Unmittelbar aus der Definition folgt die bekannte Regel

$$\frac{a}{b} = \frac{c}{d} \quad \Longleftrightarrow \quad ad = bc.$$

Definition 3.61. Wir setzen Addition, Subtraktion, Multiplikation, Division und Ordnung auf $\mathbb{Q}$ wie folgt fort. Für $[(a,b)], [(c,d)] \in \mathbb{Q}$ definieren wir

(i) die Addition durch

$$[(a,b)] + [(c,d)] \coloneqq [(ad + bc, bd)],$$

(ii) die Subtraktion durch

$$[(a,b)] - [(c,d)] \coloneqq [(ad - bc, bd)],$$

(iii) die Multiplikation durch

$$[(a,b)] \cdot [(c,d)] \coloneqq [(ac, bd)],$$

(iv) für $c \neq 0$ die Division durch

$$\frac{[(a,b)]}{[(c,d)]} \coloneqq \begin{cases} [(ad,bc)], & \text{falls } c > 0, \\ [(-ad,-bc)], & \text{falls } c < 0, \end{cases}$$

(v) die Ordnungsrelation durch

$$[(a,b)] \leq [(c,d)] \quad \Longleftrightarrow \quad ad \leq bc.$$

Beobachtung 3.62. Für alle $q, r \in \mathbb{Q}$ besitzt die Gleichung $q + x = r$ die eindeutige Lösung $x = r - q$. Ist zusätzlich $q \neq 0$, so besitzt die Gleichung $q \cdot x = r$ die eindeutige Lösung $x = \frac{r}{q}$.

Beobachtung 3.63. Die Verknüpfungen und die Ordnungsrelation aus Definition 3.61 sind unabhängig von der Wahl der Repräsentanten wohldefiniert. Sie setzen Addition, Subtraktion, Multiplikation und Ordnung auf $\mathbb{Z}$ sowie die Division aus Definition 2.37 fort. Mit Addition und Multiplikation ist $\mathbb{Q}$ ein kommutativer Ring mit Nullelement $0 = [(0,1)]$ und Einselement $1 = [(1,1)]$. Außerdem besitzt $\mathbb{Q}$ keine Nullteiler.

3.5 Körper

Die rationalen Zahlen haben wir so konstruiert, dass nicht nur die Addition, sondern auch die Multiplikation mit einem von Null verschiedenen Faktor umkehrbar ist. Wir lösen nun diese Eigenschaft vom konkreten Zahlenraum $\mathbb{Q}$ und fassen sie in einem Strukturbegriff zusammen.

Definition 3.64. Ein *Körper* ist ein kommutativer Ring $(K, +, \cdot)$ mit Eins, für den $(K \setminus \{0_K\}, \cdot)$ eine (abelsche) Gruppe ist.

Definition 3.65. Sei $(K, +, \cdot)$ ein Körper. Für $a, b \in K$ mit $b \neq 0_K$ definieren wir die Division durch

$$\frac{a}{b} \coloneqq a \cdot b^{-1}.$$

Für $a \in K$ definieren wir die ganzzahligen Potenzen für $z \in \mathbb{Z}$ durch

$$a^z \coloneqq \begin{cases} \prod_{i=1}^{z} a, & \text{falls } z > 0, \\ 1_K, & \text{falls } z = 0, \\ \prod_{i=1}^{-z} a^{-1}, & \text{falls } z < 0 \text{ und } a \neq 0_K. \end{cases}$$

Beobachtung 3.66. Ein Körper $(K, +, \cdot)$ besitzt per Definition keine Nullteiler, da $K \setminus \{0_K\}$ unter der Multiplikation abgeschlossen ist. Außerdem gilt $0_K \neq 1_K$, da das Einselement zur multiplikativen Gruppe $(K \setminus \{0_K\}, \cdot)$ gehört.

Satz 3.67. In einem Körper $(K, +, \cdot)$ besitzt die Gleichung $a \cdot x = b$ für $a, b \in K$ mit $a \neq 0_K$ die eindeutige Lösung $x = a^{-1}b$.

Beweis. Für $x = a^{-1}b$ gilt

$$ax = a(a^{-1}b) = (aa^{-1})b = 1_K b = b.$$

Ist auch $y \in K$ eine Lösung, so folgt $a(x - y) = ax - ay = b - b = 0_K$. Da $(K, +, \cdot)$ keine Nullteiler besitzt, ist also $x - y = 0_K$ und damit $x = y$. □

Beispiel 3.68. *Die rationalen Zahlen bilden den Körper $(\mathbb{Q}, +, \cdot)$: Nach Beobachtung 3.63 sind sie ein kommutativer Ring mit Eins und für $q \in \mathbb{Q} \setminus \{0\}$ ist die nach Definition 3.61 definierte rationale Zahl $\frac{1}{q}$ das multiplikative Inverse von q. Jede Gleichung der Form $ax = b$ mit $a \neq 0$ hat die eindeutige Lösung $x = \frac{b}{a} \in \mathbb{Q}$ (Beobachtung 3.62).* △

Satz 3.69. Für $n \in \mathbb{N} \setminus \{0\}$ ist der Restklassenring $\mathbb{Z}/n\mathbb{Z}$ genau dann ein Körper, wenn n eine Primzahl ist.

Beweis. Für $n = 1$ gilt $[0]_1 = [1]_1$, also ist $\mathbb{Z}/1\mathbb{Z}$ kein Körper. Ist $n > 1$ keine Primzahl, so besitzt $\mathbb{Z}/n\mathbb{Z}$ nach Satz 3.54 Nullteiler und kann daher nach Satz 3.67 kein Körper sein.

Sei nun p eine Primzahl. Nach Satz 3.50 ist $\mathbb{Z}/p\mathbb{Z}$ ein kommutativer Ring mit Eins, und wegen $p \nmid 1$ gilt $[0]_p \neq [1]_p$. Sei $[a]_p \neq [0]_p$ mit $a \in \{1, \ldots, p-1\}$. Dann ist $p \nmid a$, und der kleine Satz von Fermat (Satz 2.67) liefert $a^{p-1} \bmod p = 1$. Mit der Multiplikation aus Definition 3.49 folgt

$$[a]_p \cdot [a^{p-2}]_p = [a^{p-1}]_p = [1]_p.$$

Somit besitzt jedes von Null verschiedene Element ein multiplikatives Inverses, und $\mathbb{Z}/p\mathbb{Z}$ ist ein Körper. □

Beispiel 3.70. *Wir greifen die Multiplikationstafeln aus Beispiel 3.51 auf. Im Körper $\mathbb{Z}/5\mathbb{Z}$ sind die multiplikativen Inversen der von Null verschiedenen Elemente durch*

$$[1]_5^{-1} = [1]_5, \qquad [2]_5^{-1} = [3]_5, \qquad [3]_5^{-1} = [2]_5, \qquad [4]_5^{-1} = [4]_5$$

gegeben. Im Ring $\mathbb{Z}/4\mathbb{Z}$ besitzt dagegen $[2]_4$ kein multiplikatives Inverses, denn für jedes $a \in \mathbb{Z}$ ist $[2]_4 \cdot [a]_4$ entweder $[0]_4$ oder $[2]_4$, aber nie $[1]_4$. △

Körperhomomorphismen

In den folgenden Abschnitten erweitern wir den Zahlbereich schrittweise von $\mathbb{Q}$ zu $\mathbb{R}$ und $\mathbb{C}$. Um dabei präzise ausdrücken zu können, in welchem Sinn ein Zahlbereich in einem größeren vorkommt und seine Rechenstruktur bewahrt, verwenden wir den Begriff des Ringhomomorphismus aus Definition 3.55 für Körper.

Definition 3.71. Ein *Körperhomomorphismus* ist ein Ringhomomorphismus zwischen zwei Körpern. Ein bijektiver Körperhomomorphismus heißt *(Körper-)Isomorphismus*. Zwei Körper heißen isomorph, wenn zwischen ihnen ein Körperisomorphismus existiert. Ein Isomorphismus eines Körpers auf sich selbst heißt *Körperautomorphismus*.

Der Beweis des folgenden Satzes illustriert die Definition.

Satz 3.72. Jeder Körperhomomorphismus $f\colon \mathbb{Q} \to \mathbb{Q}$ zwischen $(\mathbb{Q}, +, \cdot)$ und $(\mathbb{Q}, +, \cdot)$ ist die Identität.

Beweis. Zunächst ist $f(1) = 1$, da f Ringhomomorphismus ist. Da f Gruppenhomomorphismus bezüglich $+$ ist, gilt für jedes $n \in \mathbb{N} \setminus \{0\}$ daher

$$f(n) = f(1 + \cdots + 1) = f(1) + \cdots + f(1) = n$$

und nach Satz 3.38 (i) auch $f(0) = 0$. Zusammen folgt für $n \in \mathbb{N} \setminus \{0\}$ aus $0 = f(0) = f(n + (-n)) = f(n) + f(-n)$, dass $f(-n) = -f(n) = -n$. Somit gilt $f(a) = a$ für alle $a \in \mathbb{Z}$.

Für $q = \frac{a}{b} \in \mathbb{Q}$ gilt also

$$b\, f(q) = f(b) f(q) = f(bq) = f(a) = a.$$

Nach Satz 3.67 besitzt diese Gleichung genau eine Lösung, nämlich $f(q) = \frac{a}{b} = q$. □

Bei einem Ringhomomorphismus können verschiedene Elemente miteinander identifiziert werden. Bei Körperhomomorphismen ist dies nicht möglich: Da jedes von Null verschiedene Element ein multiplikatives Inverses besitzt und das Einselement erhalten bleibt, kann kein solches Element auf Null abgebildet werden. Ein Körperhomomorphismus bildet seinen Ausgangskörper daher ohne Informationsverlust in den Zielkörper ab.

Satz 3.73. Jeder Körperhomomorphismus ist injektiv.

Beweis. Sei $f\colon K \to L$ ein Körperhomomorphismus, und sei $c \in K$ mit $f(c) = 0_L$. Wäre $c \neq 0_K$, so folgte

$$1_L = f(1_K) = f(cc^{-1}) = f(c)f(c^{-1}) = 0_L,$$

im Widerspruch zur Definition eines Körpers. Also gilt

$$f(c) = 0_L \quad \Longrightarrow \quad c = 0_K.$$

Sind nun $a, b \in K$ mit $f(a) = f(b)$, so ist $f(a-b) = f(a) - f(b) = 0_L$ und folglich $a - b = 0_K$. Damit ist $a = b$ und f injektiv. □

Ein Körperhomomorphismus $f\colon K \to L$ ist bezüglich der Addition bereits ein Gruppenhomomorphismus und erfüllt daher $f(0_K) = 0_L$. Nach Satz 3.73 ist f injektiv; aus $a \neq 0_K$ folgt daher $f(a) \neq 0_L$, sodass sich f auch auf die multiplikativen Gruppen einschränken lässt.

Beobachtung 3.74. Ein Körperhomomorphismus zwischen zwei Körpern $(K, +, \cdot)$ und $(L, \oplus, \odot)$ induziert Gruppenhomomorphismen zwischen den abelschen Gruppen $(K, +)$ und $(L, \oplus)$ sowie $(K \setminus \{0_K\}, \cdot)$ und $(L \setminus \{0_L\}, \odot)$.

Analog zu den Sätzen 3.38 (iii) und 3.59 ist das Bild eines Körperhomomorphismus wieder ein Körper.

Satz 3.75. Ist $f\colon K \to L$ ein Körperhomomorphismus zwischen zwei Körpern $(K, +, \cdot)$ und $(L, \oplus, \odot)$, dann ist $(f(K), \oplus, \odot) \coloneqq (f(K), \oplus|_{f(K)^2}, \odot|_{f(K)^2})$ ein Körper.

Beweis. Übung. □

Ein Körperhomomorphismus $f\colon K \to L$ erlaubt uns daher, den Körper $(K, +, \cdot)$ mit seinem Bild $f(K)$ in L zu identifizieren. In diesem Sinn können wir $(K, +, \cdot)$ als Teil des größeren Körpers $(L, \oplus, \odot)$ auffassen.

Angeordnete Körper

Die Körperstruktur erfasst Rechenoperationen auf $\mathbb{Q}$, nicht aber seine Ordnungsrelation. Wir ergänzen die Definition um eine mit Addition und Multiplikation verträgliche Ordnung.

Definition 3.76. Ein Körper $(K, +, \cdot)$ mit einer Totalordnung $\leq$ auf K heißt *angeordneter Körper*, falls für alle $a, b, c \in K$ gilt

$$\begin{aligned} a \leq b &\quad\Longrightarrow\quad a + c \leq b + c, \\ a \leq b \text{ und } 0_K \leq c &\quad\Longrightarrow\quad a \cdot c \leq b \cdot c. \end{aligned}$$

Wir schreiben dann $(K, +, \cdot, \leq)$ für den angeordneten Körper.

Beispiel 3.77. *Die rationalen Zahlen bilden mit den Verknüpfungen und der Ordnungsrelation aus Definition 3.61 den angeordneten Körper* $(\mathbb{Q}, +, \cdot, \leq)$. △

Wir zeigen noch, dass sich die Ordnung eines Körpers so verhält, wie wir es von $(\mathbb{Q}, +, \cdot, \leq)$ gewohnt sind.

Satz 3.78. Sei $(K, +, \cdot, \leq)$ ein angeordneter Körper. Dann gilt:

(i) Für alle $a \in K$ mit $a > 0_K$ gilt $-a < 0_K$.

(ii) Für alle $a \in K$ gilt $a^2 \geq 0_K$. Insbesondere ist $1_K > 0_K$.

Beweis. Zu (i): Aus $0_K < a$ folgt durch Addition von $-a$ zunächst $-a \leq 0_K$. Da $-a = 0_K$ die Gleichheit $a = 0_K$ zur Folge hätte, gilt sogar $-a < 0_K$.

Zu (ii): Wegen der Totalordnung gilt für jedes $a \in K$ entweder $0_K \leq a$ oder $a \leq 0_K$. Im ersten Fall folgt aus der Verträglichkeit der Ordnung mit der Multiplikation

$$0_K = 0_K \cdot a \leq a \cdot a = a^2.$$

Im zweiten Fall ergibt die Addition von $-a$ zunächst $0_K \leq -a$ und damit ebenso

$$0_K \leq (-a)^2 = a^2.$$

Da $1_K = 1_K^2 \geq 0_K$ und $1_K \neq 0_K$ gilt, ist schließlich $1_K > 0_K$. □

Definition 3.79. Ein Körperhomomorphismus $f\colon K \to L$ zwischen zwei angeordneten Körpern heißt *ordnungserhaltend*, wenn für alle $a, b \in K$ gilt:

$$a \leq b \quad\Longrightarrow\quad f(a) \leq f(b).$$

3.6 Die reellen Zahlen

Mit den rationalen Zahlen können wir alle Gleichungen der Form $ax = b$ mit $a, b \in \mathbb{Q}$ und $a \neq 0$ lösen. Dennoch besitzt $\mathbb{Q}$ Lücken, die sichtbar werden, wenn wir Größen nicht nur beliebig genau annähern, sondern auch ihren Grenzwert als Zahl erfassen wollen.

Betrachten wir zum Beispiel die Funktion $h\colon \mathbb{Q} \to \mathbb{Q}$, $t \mapsto 2 - t^2$ als vereinfachtes Modell einer Fallkurve, mit anfänglicher Höhe $h(0) = 2$ und Fallbeschleunigung $g = 2$. Die rationalen Zeitpunkte, zu denen sich die Kurve noch über dem Boden (Höhe 0) befindet, sind

$$S \coloneqq \{t \in \mathbb{Q} \mid t \geq 0 \text{ und } t^2 < 2\}. \tag{3.1}$$

Diese Menge ist nichtleer ($0 \in S$) und besitzt nach Beispiel 1.56 obere Schranken, aber kein Supremum. Die rationalen Zeitpunkte vor dem Aufprall lassen sich also beliebig genau eingrenzen, aber ihre Grenze ist keine rationale Zahl.

Wir suchen eine Zahlenmenge, die diese Lücke schließt. Ähnlich wie bei den natürlichen Zahlen (Abschnitt 2.1) stellen wir Axiome auf, die von der Zahlenmenge erfüllt sein sollen.

Definition 3.80. Eine Struktur $(R, +, \cdot, \leq)$ erfüllt die *Axiome der reellen Zahlen*, wenn gilt:

R1. Es ist $(R, +, \cdot, \leq)$ ein angeordneter Körper.

R2. Jede nichtleere, nach oben beschränkte Teilmenge von R besitzt ein Supremum in R. Dieses Axiom heißt *Vollständigkeitsaxiom*. Ein angeordneter Körper, der es erfüllt, heißt *vollständig*.

Das Vollständigkeitsaxiom lässt sich äquivalent für untere Schranken und Infima ausdrücken: Ist A nichtleer und nach unten beschränkt, so ist $-A \coloneqq \{-a \mid a \in A\}$ nichtleer und nach oben beschränkt; das Negative seines Supremums ist das Infimum von A.

Beobachtung 3.81. In jedem vollständigen angeordneten Körper besitzt jede nichtleere, nach unten beschränkte Teilmenge ein Infimum.

Wir nennen (ohne Beweis) das fundamentale Resultat, dass die Struktur aller vollständigen angeordneten Körper dieselbe ist.

Satz 3.82. Es existiert ein vollständiger angeordneter Körper. Je zwei vollständige angeordnete Körper sind durch genau einen ordnungserhaltenden Körperisomorphismus isomorph. Es existiert außerdem jeweils ein ordnungserhaltender Körperhomomorphismus ι von $\mathbb{Q}$ zu jedem dieser Körper.

Wir können die Menge der reellen Zahlen also losgelöst von einer konkreten Konstruktion definieren.

Definition 3.83. Wir wählen einen beliebigen vollständigen angeordneten Körper $(\mathbb{R}, +, \cdot, \leq)$. Seine Elemente heißen *reelle Zahlen*. Wir identifizieren jede rationale Zahl $q \in \mathbb{Q}$ mit ihrem Bild $\iota(q) \in \mathbb{R}$ unter dem Körperhomomorphismus ι aus Satz 3.82.

Bemerkung 3.84. Eine mögliche konkrete Konstruktion von $\mathbb{R}$ verwendet Mengen ähnlicher Form wie S in (3.1). Dabei wird jede rationale Zahl $q \in \mathbb{Q}$ mit der Menge $\{q' \in \mathbb{Q} \mid q' < q\}$ identifiziert und, beispielsweise, $\sqrt{2}$ mit der Menge $\{q \in \mathbb{Q} \mid q < 0 \text{ oder } q^2 < 2\}$. Die Addition zweier reeller Zahlen r, s ist dann $r + s := \{p + p' \mid p \in r,\, p' \in s\}$ und die Ordnung wird fortgesetzt als $r \leq s \Leftrightarrow r \subseteq s$; die Multiplikation ist komplizierter.

Definition 3.85. Seien $a, b \in \mathbb{R}$ mit $a \leq b$. Wir definieren die *Intervalle*

$$(a,b) := \{x \in \mathbb{R} \mid a < x < b\}, \quad [a,b] := \{x \in \mathbb{R} \mid a \leq x \leq b\},$$
$$(a,b] := \{x \in \mathbb{R} \mid a < x \leq b\}, \quad [a,b) := \{x \in \mathbb{R} \mid a \leq x < b\}.$$

Das Intervall (a,b) heißt offen, $[a,b]$ heißt abgeschlossen und $(a,b]$ sowie $[a,b)$ heißen halboffen. Für $a \in \mathbb{R}$ definieren wir außerdem die unbeschränkten Intervalle

$$(a,\infty) := \{x \in \mathbb{R} \mid a < x\}, \qquad [a,\infty) := \{x \in \mathbb{R} \mid a \leq x\},$$
$$(-\infty,a) := \{x \in \mathbb{R} \mid x < a\}, \quad (-\infty,a] := \{x \in \mathbb{R} \mid x \leq a\},$$
$$(-\infty,\infty) := \mathbb{R}.$$

Dabei sind $-\infty$ und ∞ keine reellen Zahlen, sondern lediglich Bestandteile dieser Schreibweise.

Wir halten noch fest (ohne Beweis), dass sich reelle Zahlen beliebig durch rationale Zahlen annähern lassen. Man sagt, $\mathbb{Q}$ liegt *dicht* in $\mathbb{R}$.

Satz 3.86. Für alle $x, y \in \mathbb{R}$ mit $x < y$ existiert $q \in \mathbb{Q}$ mit $x < q < y$.

Die Vollständigkeit von $\mathbb{R}$ garantiert die Existenz von Lösungen von Potenzgleichungen, wie $x^2 = 2$ aus unserem Eingangsbeispiel. Wir geben den folgenden Satz ohne Beweis.

Satz 3.87. Für alle $a \in [0,\infty)$ und $n \in \mathbb{N} \setminus \{0\}$ besitzt die Gleichung $x^n = a$ genau eine Lösung $x \in [0,\infty)$.

Definition 3.88. Für $a \in [0,\infty)$ und $n \in \mathbb{N} \setminus \{0\}$ definieren wir die *n-te Wurzel* $\sqrt[n]{a} := x$, wobei $x \in [0,\infty)$ die eindeutige Lösung von $x^n = a$ aus Satz 3.87 ist.

Damit können wir die Potenz auf rationale Exponenten erweitern.

Definition 3.89. Für $a \in (0,\infty)$ und $q \in \mathbb{Q}$ wählen wir $m \in \mathbb{Z}$ und $n \in \mathbb{N} \setminus \{0\}$ mit $q = \frac{m}{n}$ und definieren die *rationale Potenz* durch

$$a^q := \sqrt[n]{a^m}.$$

Satz 3.90. Die rationale Potenz ist unabhängig von der Wahl von m und n mit $q = \frac{m}{n}$. Für $a, b \in (0,\infty)$ und $p, q \in \mathbb{Q}$ gelten außerdem die Potenzgesetze

$$a^{p+q} = a^p a^q, \qquad (a^p)^q = a^{pq}, \qquad (ab)^q = a^q b^q.$$

Beweis. Sei zunächst $q = \frac{m}{n} = \frac{k}{l}$ mit $m, k \in \mathbb{Z}$ und $n, l \in \mathbb{N} \setminus \{0\}$. Für

$$x \coloneqq \sqrt[n]{a^m} \qquad \text{und} \qquad y \coloneqq \sqrt[l]{a^k}$$

gilt wegen $ml = kn$

$$x^{nl} = (x^n)^l = a^{ml} = a^{kn} = (y^l)^n = y^{ln}.$$

Da $x, y \geq 0$ sind, folgt aus der Eindeutigkeit in Satz 3.87 die Gleichheit $x = y$. Die Definition ist also unabhängig von der Darstellung von q als Bruch.

Die Potenzgesetze folgen auf dieselbe Weise aus den Potenzgesetzen für ganzzahlige Exponenten und der Eindeutigkeit der Wurzeln. Schreiben wir beispielsweise $p = \frac{m}{n}$ und $q = \frac{k}{l}$, so sind a^{p+q} und $a^p a^q$ positiv, und ihre (nl)-ten Potenzen sind jeweils gleich a^{ml+kn}. Somit sind sie gleich. Die beiden übrigen Regeln folgen analog. □

Beispiel 3.91. *Nach Definition 3.89 gilt beispielsweise*

$$16^{3/4} = \sqrt[4]{16^3} = 8 \qquad \textit{und} \qquad 16^{-1/2} = \sqrt{16^{-1}} = \frac{1}{4}. \qquad \triangle$$

Definition 3.92. Der *Betrag* von $x \in \mathbb{R}$ ist

$$|x| \coloneqq \begin{cases} x & \text{falls } x \geq 0, \\ -x & \text{falls } x < 0 \end{cases}$$

und die Funktion $x \mapsto |x|$ heißt *Betragsfunktion*.

Satz 3.93. Für alle $x, y \in \mathbb{R}$ gilt

$$\begin{aligned} |x| &\geq 0, & |x| &= |-x| = \sqrt{x^2}, \\ -|x| &\leq x \leq |x|, & |xy| &= |x| \cdot |y|, \\ |x| &= 0 \Leftrightarrow x = 0, & |x + y| &\leq |x| + |y|. \end{aligned}$$

Die letzte Ungleichung heißt *Dreiecksungleichung*.

Beweis. Nichtnegativität, Symmetrie, die Schranken für x und die Aussage über $|x| = 0$ folgen unmittelbar aus der Definition. Für die Produktregel stellen wir fest, dass $|xy|$ und $|x| \cdot |y|$ nichtnegativ sind und dass

$$|xy|^2 = (xy)^2 = x^2 y^2 = |x|^2 |y|^2 = (|x| \cdot |y|)^2.$$

Nach der Eindeutigkeit der nichtnegativen Quadratwurzel in Satz 3.87 sind sie daher gleich. Ebenso folgt $|x| = \sqrt{x^2}$ aus $|x| \geq 0$ und $|x|^2 = x^2$. Für die Dreiecksungleichung folgt aus den Schranken für x und y

$$x + y \leq |x| + |y| \qquad \text{und} \qquad -(x + y) \leq |x| + |y|.$$

Da $|x + y|$ je nach Vorzeichen gleich $x + y$ oder $-(x + y)$ ist, folgt die Dreiecksungleichung. □

3.7 Die komplexen Zahlen

Nach Satz 3.78 sind Quadrate in einem angeordneten Körper nichtnegativ. Daher besitzt die Gleichung $x^2 + 1 = 0$ keine Lösung in $\mathbb{R}$. Wir erweitern den Zahlbereich ein letztes Mal, um auch diese Gleichung lösen zu können.

Auch wenn physikalische Ausgangsdaten und Rechenergebnisse reell sind, kann die Erweiterung des Zahlenraums jenseits von $\mathbb{R}$ Rechnungen vereinfachen. Ein einfaches Beispiel aus der Signalverarbeitung: Die Abfolge von Signalen (zum Beispiel in Form elektrischer Spannungen)

$$a_0 = 1, a_1 = 0, a_2 = -1, a_3 = 0, a_4 = 1, a_5 = 0, a_6 = -1, a_7 = 0, \dots$$

lässt sich elegant als $a_n = \frac{1}{2} \cdot (q^n + (-q)^n)$ beschreiben, wobei q eine Lösung von $q^2 + 1 = 0$ ist. Eine Beschreibung ohne eine solche Zahl q ist weniger elegant, zum Beispiel mittels einer Fallunterscheidung: Für alle $k \in \mathbb{N}$ ist $a_{4k} = 1$, $a_{4k+2} = -1$ und $a_{2k+1} = 0$. Die elegantere Beschreibung mit q ergibt dieselbe Folge, denn

$$\begin{aligned}
a_{4k} &= \frac{1}{2} \cdot (q^{4k} + (-1)^{4k} \cdot q^{4k}) = \frac{1}{2}((-1)^{2k} + 1 \cdot (-1)^{2k}) = \frac{1}{2}(1^k + 1^k) = 1, \\
a_{4k+2} &= \frac{1}{2} \cdot (q^{4k+2} + (-1)^{4k+2} \cdot q^{4k+2}) = a_{4k} \cdot q^2 = -1, \\
a_{2k+1} &= \frac{1}{2} \cdot (q^{2k+1} + (-1)^{2k+1} \cdot q^{2k+1}) = \frac{1}{2} \cdot ((-1)^k \cdot q - (-1)^k \cdot q) = 0.
\end{aligned}$$

Wir konstruieren jetzt einen Zahlenraum, der die Zahl q enthält. Anders als bei den reellen Zahlen ist die Konstruktion einfach genug, um sie hier konkret anzugeben.

Definition 3.94. Die Menge der *komplexen Zahlen* ist definiert als

$$\mathbb{C} \coloneqq \mathbb{R}^2.$$

Wir identifizieren jede reelle Zahl $x \in \mathbb{R}$ mit $(x, 0) \in \mathbb{C}$. Komplexe Zahlen der Form $(x, 0) \in \mathbb{C}$ heißen *reell* und komplexe Zahlen der Form $(0, y) \in \mathbb{C}$ heißen *imaginär*.

Definition 3.95. Für $(x_1, y_1), (x_2, y_2) \in \mathbb{C}$ setzen wir die Addition und Multiplikation von $\mathbb{R}$ auf $\mathbb{C}$ fort durch

$$\begin{aligned}
(x_1, y_1) + (x_2, y_2) &\coloneqq (x_1 + x_2, y_1 + y_2), \\
(x_1, y_1) \cdot (x_2, y_2) &\coloneqq (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2).
\end{aligned}$$

Wir überprüfen nun, dass die so definierten Verknüpfungen die Körperaxiome erfüllen.

Satz 3.96. Mit den Verknüpfungen aus Definition 3.95 ist $(\mathbb{C}, +, \cdot)$ ein Körper mit Nullelement $(0, 0)$ und Einselement $(1, 0)$. Für $(x, y) \neq (0, 0)$ gilt

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

Beweis. Seien $z = (a, b), w = (c, d), t = (e, f) \in \mathbb{C}$. Für die Addition gilt

$$\begin{aligned} z + w &= (a + c, b + d) = (c + a, d + b) = w + z, \\ (z + w) + t &= ((a + c) + e, (b + d) + f) = (a + (c + e), b + (d + f)) = z + (w + t), \\ z + (0, 0) &= z, \\ z + (-a, -b) &= (0, 0), \end{aligned}$$

also ist $(\mathbb{C}, +)$ eine abelsche Gruppe.

Die Multiplikation ist kommutativ und assoziativ, denn

$$\begin{aligned} zw &= (ac - bd, ad + bc) = (ca - db, cb + da) = wz, \\ (zw)t &= ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e) \\ &= (ace - bde - adf - bcf, acf - bdf + ade + bce) \\ &= (ace - adf - bcf - bde, acf + ade + bce - bdf) \\ &= (a(ce - df) - b(cf + de), a(cf + de) + b(ce - df)) \\ &= z(wt). \end{aligned}$$

Weiter ist $(1, 0)$ das Einselement, denn

$$z \cdot (1, 0) = (a \cdot 1 - b \cdot 0, a \cdot 0 + b \cdot 1) = (a, b) = z.$$

Für das Distributivgesetz gilt

$$\begin{aligned} z(w + t) &= (a(c + e) - b(d + f), a(d + f) + b(c + e)) \\ &= (ac - bd + ae - bf, ad + bc + af + be) \\ &= (ac - bd, ad + bc) + (ae - bf, af + be) \\ &= zw + zt. \end{aligned}$$

Das andere Distributivgesetz folgt aus der Kommutativität der Multiplikation. Damit ist $(\mathbb{C}, +, \cdot)$ ein kommutativer Ring mit Eins.

Ist $(x, y) \neq (0, 0)$, so gilt $x^2 + y^2 > 0$: Ist $x \neq 0$, so ist wegen der Nullteilerfreiheit auch $x^2 \neq 0$ und daher nach Satz 3.78 sogar $x^2 > 0$. Für $x = 0$ folgt entsprechend $y \neq 0$ und $y^2 > 0$. Damit ist der angegebene Ausdruck definiert und es gilt

$$(x, y) \cdot \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2} \right) = \left(\frac{x^2}{x^2 + y^2} - \frac{-y^2}{x^2 + y^2}, \frac{xy}{x^2 + y^2} + \frac{x \cdot (-y)}{x^2 + y^2} \right) = \left(\frac{x^2 + y^2}{x^2 + y^2}, 0 \right) = (1, 0).$$

Also besitzt jede von Null verschiedene komplexe Zahl ein multiplikatives Inverses.

Da $0 \neq 1$ in $\mathbb{R}$ gilt, sind auch $(0, 0)$ und $(1, 0)$ verschieden. Das Einselement gehört daher zu $\mathbb{C} \setminus \{0\}$. Diese Menge ist unter der Multiplikation abgeschlossen: Für $z, w \neq 0$ würde aus $zw = 0$ nämlich $w = z^{-1}(zw) = z^{-1}0 = 0$ folgen, ein Widerspruch. Auch das Inverse eines von Null verschiedenen Elements ist wegen $zz^{-1} = 1 \neq 0$ von Null verschieden. Damit ist $(\mathbb{C} \setminus \{0\}, \cdot)$ eine abelsche Gruppe und $(\mathbb{C}, +, \cdot)$ ein Körper. □

Definition 3.97. Wir definieren die *imaginäre Einheit* durch $\mathrm{i} \coloneqq (0, 1) \in \mathbb{C}$. Für $z = (x, y) \in \mathbb{C}$ definieren wir den *Realteil* und *Imaginärteil* von z als

$$\mathrm{Re}(z) \coloneqq x \qquad \text{und} \qquad \mathrm{Im}(z) \coloneqq y.$$

Beobachtung 3.98. Es gilt $\mathrm{i}^2 = -1$ und für jedes $z \in \mathbb{C}$ existieren eindeutig bestimmte $x, y \in \mathbb{R}$ mit $z = x + y\mathrm{i}$.

Die eindeutige Darstellung $z = x + y\mathrm{i}$ aus Beobachtung 3.98 heißt *kartesische Darstellung* von z. Wegen $\mathrm{i}^2 = -1$ können wir in dieser Darstellung nach den vertrauten Rechenregeln rechnen.

Satz 3.99. Für $z = x + y\mathrm{i}$ und $w = u + v\mathrm{i}$ gilt

$$\begin{aligned} z + w &= (x + u) + (y + v)\mathrm{i}, \\ zw &= (xu - yv) + (xv + yu)\mathrm{i}. \end{aligned}$$

Beweis. Die erste Gleichheit folgt direkt aus der Definition der Addition. Für die zweite gilt

$$(x + y\mathrm{i})(u + v\mathrm{i}) = xu + xv\mathrm{i} + yu\mathrm{i} + yv\mathrm{i}^2 = (xu - yv) + (xv + yu)\mathrm{i}. \qquad \square$$

Beispiel 3.100. *Nach Satz 3.99 gilt*

$$\begin{aligned} (2 + 3\mathrm{i}) + (1 - 5\mathrm{i}) &= 3 - 2\mathrm{i}, \\ (2 + 3\mathrm{i})(1 - 5\mathrm{i}) &= 2 - 10\mathrm{i} + 3\mathrm{i} - 15\mathrm{i}^2 = 17 - 7\mathrm{i}. \end{aligned}$$

Außerdem sind i *und* $-\mathrm{i}$ *Lösungen der Gleichung* $x^2 + 1 = 0$. $\triangle$

Satz 3.101. Es existiert keine Totalordnung auf $\mathbb{C}$, die $(\mathbb{C}, +, \cdot)$ zu einem angeordneten Körper macht.

Beweis. In einem angeordneten Körper sind nach Satz 3.78 alle Quadrate nichtnegativ und es gilt $1 > 0$, also $-1 < 0$. Wegen $\mathrm{i}^2 = -1$ wäre in $\mathbb{C}$ jedoch das Quadrat i^2 negativ. $\square$

Konjugation und Betrag

Definition 3.102. Die zu $z \in \mathbb{C}$ *konjugiert komplexe Zahl* und der *Betrag* von z sind

$$\overline{z} \coloneqq \mathrm{Re}(z) - \mathrm{Im}(z)\mathrm{i} \qquad \text{und} \qquad |z| \coloneqq \sqrt{(\mathrm{Re}(z))^2 + (\mathrm{Im}(z))^2}.$$

Der komplexe Betrag ist dabei eine Fortsetzung des reellen Betrags, denn nach Satz 3.93 gilt $|(x, 0)| = \sqrt{x^2 + 0^2} = \sqrt{x^2} = |x|$.

Satz 3.103. Für alle $z, w \in \mathbb{C}$ gilt

$$\begin{aligned} \overline{\overline{z}} &= z, & \overline{z + w} &= \overline{z} + \overline{w}, & \overline{zw} &= \overline{z}\,\overline{w}, \\ z\overline{z} &= |z|^2, & |zw| &= |z|\,|w|, & |z + w| &\leq |z| + |w|. \end{aligned}$$

Außerdem ist $|z| = 0$ genau dann, wenn $z = 0$ ist. Für $z \neq 0$ gilt

$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

Beweis. Die Regeln für die Konjugation folgen unmittelbar aus Satz 3.99. Für $z = x + y\mathrm{i}$ gilt außerdem

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

Da Quadrate nichtnegativ sind, ist $x^2 + y^2 = 0$ genau dann, wenn $x = y = 0$ gilt. Dies zeigt die Aussage über $|z| = 0$ und für $z \neq 0$ folgt

$$z\frac{\overline{z}}{|z|^2} = \frac{z\overline{z}}{|z|^2} = 1.$$

Mit den Regeln für die Konjugation erhalten wir

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

Da beide Seiten vor dem Quadrieren nichtnegativ sind, folgt aus der Eindeutigkeit in Satz 3.87 die Gleichheit $|zw| = |z|\,|w|$.

Seien $z = x + y\mathrm{i}$ und $w = u + v\mathrm{i}$. Aus

$$(|z|\,|w|)^2 - (xu + yv)^2 = (xv - yu)^2 \geq 0$$

folgt $xu + yv \leq |z|\,|w|$: Für $xu + yv \leq 0$ ist dies unmittelbar klar; andernfalls sind beide Seiten nichtnegativ und wir können ihre Quadrate vergleichen. Daher gilt

$$\begin{aligned} |z + w|^2 &= (x + u)^2 + (y + v)^2 \\ &= |z|^2 + 2(xu + yv) + |w|^2 \\ &\leq (|z| + |w|)^2. \end{aligned}$$

Da $|z + w|$ und $|z| + |w|$ nichtnegativ sind, folgt $|z + w| \leq |z| + |w|$. □

Beispiel 3.104. *Für $z = 3 + 4\mathrm{i}$ gilt*

$$\overline{z} = 3 - 4\mathrm{i}, \qquad |z| = \sqrt{3^2 + 4^2} = 5 \qquad \textit{und} \qquad z^{-1} = \frac{3 - 4\mathrm{i}}{25}. \qquad \triangle$$

Satz 3.105. Die Konjugation $z \mapsto \overline{z}$ ist ein nichttrivialer Körperautomorphismus von $(\mathbb{C}, +, \cdot)$. Eine komplexe Zahl z ist genau dann reell, wenn $\overline{z} = z$ gilt.

Beweis. Nach Satz 3.103 erhält die Konjugation Addition und Multiplikation; außerdem gilt $\overline{1} = 1$. Wegen $\overline{\overline{z}} = z$ ist sie bijektiv und damit ein Körperautomorphismus. Sie ist nicht die Identität, denn $\overline{\mathrm{i}} = -\mathrm{i} \neq \mathrm{i}$. Für $z = x + y\mathrm{i}$ gilt schließlich $\overline{z} = z$ genau dann, wenn $y = 0$, also $z = x \in \mathbb{R}$ gilt. □

Algebraische Abgeschlossenheit

Wir haben gesehen, dass die komplexen Zahlen eine Lösung von $x^2 + 1 = 0$ zulassen. Natürlich überträgt sich das auch auf Gleichungen ähnlicher Form. Wir wollen jetzt genauer erfassen, welche Gleichungen über $\mathbb{C}$ zusätzlich lösbar sind. Wir führen zunächst Polynome über beliebigen Körpern ein.

Definition 3.106. Seien $(K, +, \cdot)$ ein Körper, $n \in \mathbb{N}$ und $a_0, \dots, a_n \in K$ mit $a_n \neq 0$ oder $n = 0$. Der Ausdruck

$$\sum_{k=0}^{n} a_k X^k$$

heißt *Polynom über* $(K, +, \cdot)$ mit der zugehörigen *Polynomfunktion* $p\colon K \to K$, $X \mapsto \sum_{k=0}^{n} a_k X^k$.
Die *Koeffizienten* des Polynoms sind $a_0, \dots, a_n$ und sein *Grad* ist n, sofern $n \neq 0$ oder $a_0 \neq 0$. Zwei Polynome sind gleich, wenn ihre Koeffizienten übereinstimmen. Eine Zahl $c \in K$ heißt *Nullstelle* des Polynoms, wenn $p(c) = 0$ gilt.

Beobachtung 3.107. Zwei verschiedene Polynome können dieselbe zugehörige Polynomfunktion haben: Über $(\mathbb{Z}/2\mathbb{Z}, +, \cdot)$ sind die Polynome X^2 und X nicht gleich, beschreiben aber dieselbe Funktion.

Wir können jetzt die Eigenschaft, Lösungen von $x^2 + 1 = 0$ zu besitzen, auf die Existenz von Nullstellen von Polynomen verallgemeinern.

Definition 3.108. Ein Körper K heißt *algebraisch abgeschlossen*, falls jedes Polynom über K von positivem Grad eine Nullstelle in K besitzt.

Der vollständige angeordnete Körper $\mathbb{R}$ ist somit nicht algebraisch abgeschlossen, während $\mathbb{C}$ nicht angeordnet werden kann (Satz 3.101). Dafür ist $\mathbb{C}$ algebraisch abgeschlossen (ohne Beweis).

Satz 3.109 (Fundamentalsatz der Algebra)**.** Der Körper $(\mathbb{C}, +, \cdot)$ ist algebraisch abgeschlossen. Insbesondere hat jede Polynomfunktion die Form

$$p(X) = a \prod_{j=1}^{n} (X - z_j)$$

mit $n \in \mathbb{N}$ und $a, z_1, \dots, z_n \in \mathbb{C}$. Wir sagen, dass das Polynom in seine *Linearfaktoren* zerfällt.

Beispiel 3.110. *Das Polynom* $p(X) = X^2 + 1$ *besitzt keine reelle Nullstelle. Über* $\mathbb{C}$ *gilt dagegen*

$$X^2 + 1 = (X - \mathrm{i})(X + \mathrm{i}).$$

Seine Nullstellen sind also i *und* $-\mathrm{i}$. △

3.8 Die Zahlbereiche im Überblick

Die von uns konstruierten Zahlbereiche sind als Mengen zunächst nicht automatisch ineinander enthalten. Bei jeder Erweiterung haben wir den vorhergehenden Zahlbereich jedoch mit seinem Bild unter einer injektiven, strukturerhaltenden Abbildung identifiziert. Konkret identifizieren wir

- eine natürliche Zahl $n \in \mathbb{N}$ mit der ganzen Zahl $[(n, 0)] \in \mathbb{Z}$ aus Abschnitt 3.1,
- eine ganze Zahl $z \in \mathbb{Z}$ mit der rationalen Zahl $[(z, 1)] \in \mathbb{Q}$ aus Definition 3.60,
- eine rationale Zahl $q \in \mathbb{Q}$ mit der reellen Zahl $\iota(q) \in \mathbb{R}$ aus Abschnitt 3.6 und
- eine reelle Zahl $r \in \mathbb{R}$ mit der komplexen Zahl $(r, 0) \in \mathbb{C}$ aus Definition 3.94.

Mit den genannten Identifikationen können wir also die Zahlbereiche so auffassen, dass

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

Im Folgenden identifizieren wir der Übersichtlichkeit halber Gruppen und (angeordnete) Körper mit den zugrunde liegenden Mengen, wenn die jeweiligen Verknüpfungen und gegebenenfalls die Ordnung aus dem Zusammenhang hervorgehen. So kürzen wir beispielsweise ab und schreiben $\mathbb{Z}$ für $(\mathbb{Z}, +)$ und $\mathbb{R}$ für $(\mathbb{R}, +, \cdot, \leq)$.

4 Mehrdimensionale Räume

Die Dimension eines Raums erfasst seine Anzahl an Freiheitsgraden. Zur Beschreibung eines Punkts in der Ebene benötigen wir beispielsweise zwei reelle Koordinaten, daher ist die Menge der Punkte in der Ebene $\mathbb{R}^2$ anschaulich zweidimensional. Andererseits sind $\mathbb{Z}$ und $\mathbb{Q}$ als Teilmengen von $\mathbb{R}$ anschaulich eindimensional, obwohl wir diese Mengen ebenfalls über (Äquivalenzklassen von) Tupeln definiert haben (siehe Abschnitte 3.1 und 3.4), da jede Zahl durch ihre Position auf der Zahlengeraden, also durch einen Freiheitsgrad, bestimmt ist. Die Menge der komplexen Zahlen hatten wir als $\mathbb{C} = \mathbb{R}^2$ definiert (siehe Abschnitt 3.7), ihre Dimension ist 1 oder 2, je nachdem, ob wir komplexe oder reelle Freiheitsgrade betrachten.

Beispiele für höherdimensionale Räume sind der dreidimensionale physikalische Raum $\mathbb{R}^3$ und der Raum $(\mathbb{Q} \cap [0,1])^{w \cdot h}$ aller Graustufenbilder, die w Pixel breit und h Pixel hoch sind. Die Freiheitsgrade einer solchen Grafik sind die Graustufenwerte jedes Pixels; bei 256 Graustufen können wir diese beispielsweise durch die rationalen Zahlen $\frac{0}{255}, \frac{1}{255}, \ldots, \frac{255}{255}$ darstellen. Ein weiteres Beispiel eines n-dimensionalen Raums ist der Raum aller Polynome vom Grad höchstens $n-1$, wobei jeder Freiheitsgrad der Wahl eines Koeffizienten entspricht (siehe Abschnitt 3.7).

Wir beginnen mit der Geometrie des konkreten Raums $\mathbb{R}^n$, insbesondere also des zwei- und dreidimensionalen Raums. Wie im vorigen Kapitel abstrahieren wir im Anschluss und führen allgemeine Vektorräume ein, die unsere Anschauung aus der Geometrie vom Zahlenraum $\mathbb{R}$ lösen. Das erlaubt uns, auch unendlichdimensionale Räume wie den Raum $\mathrm{abb}(\mathbb{R}, \mathbb{R})$ aller Funktionen $f\colon \mathbb{R} \to \mathbb{R}$ einzufangen (siehe Abschnitt 4.2): Hier benötigen wir unendlich viele Freiheitsgrade für die Wahl des Funktionswerts $f(x)$ für alle $x \in \mathbb{R}$.

4.1 Geometrie des $\mathbb{R}^n$

Elemente des $\mathbb{R}^n$ sind Tupel mit n Einträgen, die wir schlicht durch Angabe ihrer Einträge beschreiben können. Anschaulich interpretieren wir ein solches Tupel als eine Verschiebung, wobei wir n senkrechte Koordinatenachsen zugrunde legen und jeden Eintrag als Verschiebung entlang der jeweiligen Achse auffassen. Wenn wir den Ursprung als Referenzpunkt wählen, können wir eine Verschiebung auch als einen Punkt, also als eine Verschiebung des Ursprungs, auffassen.

Definition 4.1. Ein (*n-dimensionaler*) *Vektor* mit $n \in \mathbb{N} \setminus \{0\}$ ist ein Element

$$\boldsymbol{x} = (x_1, \ldots, x_n)^\top \coloneqq \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n,$$

wobei wir Vektoren als Spalten schreiben. Die Einträge $x_1, \ldots, x_n$ heißen *Komponenten* von $\boldsymbol{x}$. Wir setzen $\mathbf{0} \coloneqq (0, \ldots, 0)^\top \in \mathbb{R}^n$ und $-\boldsymbol{x} \coloneqq (-x_1, \ldots, -x_n)^\top$.

Definition 4.2. Für Vektoren $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ und einen *Skalar* $\lambda \in \mathbb{R}$ definieren wir *Vektoraddition* und *Skalarmultiplikation* komponentenweise durch

$$\boldsymbol{x} + \boldsymbol{y} \coloneqq \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \qquad \lambda \boldsymbol{x} \coloneqq \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Weitere Rechenregeln folgen komponentenweise aus den Rechenregeln in $\mathbb{R}$.

Beobachtung 4.3. Mit der Vektoraddition ist $\mathbb{R}^n$ eine abelsche Gruppe. Das neutrale Element ist $\mathbf{0}$, das additive Inverse von $\boldsymbol{x}$ ist $-\boldsymbol{x}$. Für alle $\lambda, \mu \in \mathbb{R}$ und $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ gilt außerdem

$$\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda \boldsymbol{x} + \lambda \boldsymbol{y}, \qquad (\lambda + \mu)\boldsymbol{x} = \lambda \boldsymbol{x} + \mu \boldsymbol{x},$$
$$\lambda(\mu \boldsymbol{x}) = (\lambda \mu)\boldsymbol{x}, \qquad 1\boldsymbol{x} = \boldsymbol{x}.$$

Wir erhalten daher auch eine Vektorsubtraktion, wobei der Vektor $\boldsymbol{q} - \boldsymbol{p}$ eine Verschiebung vom Punkt $\boldsymbol{p}$ zum Punkt $\boldsymbol{q}$ beschreibt (denn $\boldsymbol{p} + (\boldsymbol{q} - \boldsymbol{p}) = \boldsymbol{q}$).

Definition 4.4. Der *j-te Einheitsvektor* $\boldsymbol{e}^{(j)} \in \{0, 1\}^n$ für $j \in \{1, \ldots, n\}$ ist definiert durch seine Einträge

$$e_i^{(j)} \coloneqq \begin{cases} 1, & \text{für } i = j, \\ 0, & \text{für } i \neq j. \end{cases}$$

Jeder Vektor $\boldsymbol{x} \in \mathbb{R}^n$ lässt sich also eindeutig schreiben als

$$\boldsymbol{x} = x_1 \boldsymbol{e}^{(1)} + \cdots + x_n \boldsymbol{e}^{(n)}.$$

Geraden und Ebenen

Mit der Anschauung, dass die Dimension einer Menge ihrer Anzahl an Freiheitsgraden entspricht, können wir Geraden und Ebenen als ein- und zweidimensionale Teilmengen von $\mathbb{R}^n$ beschreiben, indem wir ihre Position sowie ihre Freiheitsgrade angeben.

Definition 4.5. Seien $n \in \mathbb{N} \setminus \{0\}$ und $\boldsymbol{p}, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ mit $\boldsymbol{u} \neq \mathbf{0}$. Dann sind

$$\{\boldsymbol{p} + t\boldsymbol{u} \mid t \in \mathbb{R}\} \subseteq \mathbb{R}^n \qquad \text{bzw.} \qquad \{\boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v} \mid s, t \in \mathbb{R}\} \subseteq \mathbb{R}^n$$

eine *Gerade* bzw. eine *Ebene* mit *Stützvektor* $\boldsymbol{p}$ und *Richtungsvektor*en $\boldsymbol{u}$ und $\boldsymbol{v}$, wobei für den Richtungsvektor $\boldsymbol{v}$ einer Ebene $\boldsymbol{v} \neq \lambda \boldsymbol{u}$ für alle $\lambda \in \mathbb{R}$ gelten muss.

Die Einschränkung $\boldsymbol{v} \neq \lambda \boldsymbol{u}$ bei der Ebene verhindert, dass wir eine Gerade mit Richtungsvektor $\boldsymbol{u}$ beschreiben. Allgemeiner sind die Beschreibungen von Geraden und Ebenen nicht eindeutig, da wir Richtungsvektoren beliebig skalieren und andere Stützvektoren auswählen können.

Der folgende Satz zeigt, dass wir die Freiheitsgrade einer Geraden und einer Ebene auch durch die Angabe von Punkten fixieren können, die sie enthalten sollen.

Satz 4.6. Durch zwei verschiedene Punkte $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ verläuft genau eine Gerade, nämlich $\{\boldsymbol{p}+t(\boldsymbol{q}-\boldsymbol{p}) \mid t \in \mathbb{R}\}$. Durch drei Punkte $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r} \in \mathbb{R}^n$, die nicht auf einer Geraden liegen, verläuft genau eine Ebene, nämlich $\{\boldsymbol{p} + s(\boldsymbol{q} - \boldsymbol{p}) + t(\boldsymbol{r} - \boldsymbol{p}) \mid s, t \in \mathbb{R}\}$.

Beweis. Wegen $\boldsymbol{q} - \boldsymbol{p} \neq \boldsymbol{0}$ ist die angegebene Menge eine Gerade; sie enthält $\boldsymbol{p}$ und $\boldsymbol{q}$. Enthält eine weitere Gerade $g = \{\boldsymbol{a} + t\boldsymbol{v} \mid t \in \mathbb{R}\}$ beide Punkte, so gibt es $\alpha \neq \beta$ mit $\boldsymbol{p} = \boldsymbol{a} + \alpha\boldsymbol{v}$ und $\boldsymbol{q} = \boldsymbol{a} + \beta\boldsymbol{v}$. Daher ist $\boldsymbol{q} - \boldsymbol{p} = (\beta - \alpha)\boldsymbol{v}$. Für jeden Parameter $t \in \mathbb{R}$ gilt daher

$$\boldsymbol{a} + t\boldsymbol{v} = \boldsymbol{p} + \frac{t - \alpha}{\beta - \alpha}(\boldsymbol{q} - \boldsymbol{p}).$$

Da $\frac{t-\alpha}{\beta-\alpha}$ alle reellen Werte durchläuft, ist g genau die im Satz angegebene Menge.

Bei drei nicht auf einer Geraden liegenden Punkten sind $\boldsymbol{q} - \boldsymbol{p}$ und $\boldsymbol{r} - \boldsymbol{p}$ von null verschieden und keine Vielfachen voneinander. Die angegebene Menge ist daher eine Ebene und enthält die drei Punkte. Jede weitere Ebene durch sie lässt sich zunächst mit Stützvektor $\boldsymbol{p}$ schreiben, indem wir die Parameterwerte dieses Punkts von den ursprünglichen Parametern abziehen. Sind $\boldsymbol{u}, \boldsymbol{v}$ ihre Richtungsvektoren, so gilt

$$\boldsymbol{q} - \boldsymbol{p} = a\boldsymbol{u} + b\boldsymbol{v}, \qquad \boldsymbol{r} - \boldsymbol{p} = c\boldsymbol{u} + d\boldsymbol{v}.$$

Es ist $ad - bc \neq 0$: Andernfalls gäbe es wegen $(a, b) \neq (0, 0)$ eine Zahl $\lambda \in \mathbb{R}$ mit $c = \lambda a$ und $d = \lambda b$. Dann wäre $\boldsymbol{r} - \boldsymbol{p} = \lambda(\boldsymbol{q} - \boldsymbol{p})$, im Widerspruch zur Voraussetzung. Durch Eliminieren erhalten wir nun

$$\boldsymbol{u} = \frac{d(\boldsymbol{q} - \boldsymbol{p}) - b(\boldsymbol{r} - \boldsymbol{p})}{ad - bc}, \qquad \boldsymbol{v} = \frac{a(\boldsymbol{r} - \boldsymbol{p}) - c(\boldsymbol{q} - \boldsymbol{p})}{ad - bc}.$$

Jeder Punkt der weiteren Ebene lässt sich daher als $\boldsymbol{p} + s(\boldsymbol{q} - \boldsymbol{p}) + t(\boldsymbol{r} - \boldsymbol{p})$ mit $s, t \in \mathbb{R}$ schreiben. Umgekehrt zeigen die Gleichungen für $\boldsymbol{q} - \boldsymbol{p}$ und $\boldsymbol{r} - \boldsymbol{p}$, dass jeder so dargestellte Punkt zur weiteren Ebene gehört. Die beiden Ebenen stimmen also überein. □

Insbesondere können zwei unterschiedliche Geraden maximal einen Schnittpunkt besitzen, da zwei Punkte bereits eine Gerade bestimmen. Genauso können sich zwei unterschiedliche Ebenen maximal in einer Geraden schneiden.

Beispiel 4.7. *Wir berechnen den Schnittpunkt*

$$\{\boldsymbol{x}^*\} = \left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + r \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \;\middle|\; r \in \mathbb{R} \right\} \cap \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + s \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\}$$

einer Geraden mit einer Ebene (Abbildung 4.1).

Für den Schnittpunkt müssen $r, s, t \in \mathbb{R}$ existieren mit

$$\begin{pmatrix} 0 + r \cdot 1 \\ 1 + r \cdot 1 \\ 1 + r \cdot 1 \end{pmatrix} = \boldsymbol{x}^* = \begin{pmatrix} 0 + s \cdot 1 + t \cdot 1 \\ 0 + s \cdot 1 + t \cdot 2 \\ 1 + s \cdot 0 + t \cdot 2 \end{pmatrix}.$$

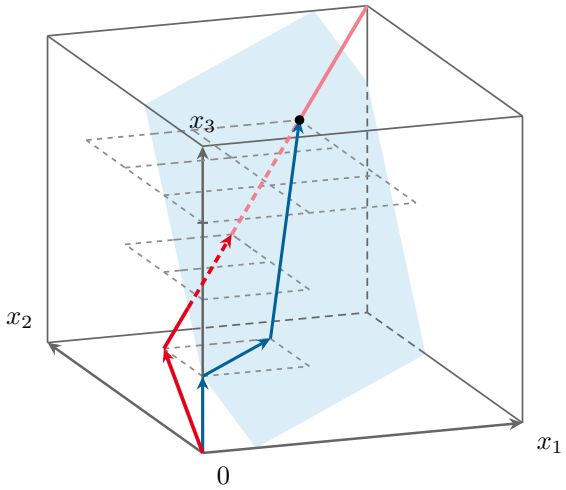


Abbildung 4.1: Gerade (rot) und Ebene (blau) aus Beispiel 4.7 mit ihren Stütz- und Richtungsvektoren sowie ihrem Schnittpunkt.

Umformung ergibt das Gleichungssystem

$$\begin{aligned} s + t - r &= 0, \\ s + 2t - r &= 1, \\ 2t - r &= 0. \end{aligned}$$

Subtrahieren wir die erste Gleichung von der zweiten, erhalten wir $t = 1$*. Einsetzen in die dritte Gleichung ergibt* $r = 2$ *und anschließend in die erste* $s = 1$*. Der einzige Schnittpunkt ist somit* $\boldsymbol{x}^* = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$. △

Wir bemerken, dass wir das Gleichungssystem in Beispiel 4.7 durch die schrittweise Elimination von Unbekannten gelöst haben. In Abschnitt 6.2 entwickeln wir daraus ein allgemeines Lösungsverfahren.

Bemerkung 4.8. Beim *Raytracing* in der Computergrafik rendern wir eine dreidimensionale Szene, indem wir Sehstrahlen durch die Bildpunkte verfolgen, um den Farbwert des Bildpunkts zu ermitteln. Für ebene Oberflächen bestimmen wir dazu zunächst den Schnitt einer Geraden mit einer Ebene, wie in Beispiel 4.7; anschließend prüfen wir, ob der Schnittpunkt auf dem vorwärts gerichteten Strahl und innerhalb der dargestellten Fläche liegt.

Länge und Skalarprodukt

Um aus Verschiebungen Abstände zu bestimmen, müssen wir Vektoren eine Länge zuordnen. In der Ebene legt der Satz des Pythagoras für $\boldsymbol{x} \in \mathbb{R}^2$ die Länge $\sqrt{x_1^2 + x_2^2}$ nahe. Für einen Vektor $\boldsymbol{x} \in \mathbb{R}^3$ im Raum können wir seine Länge in der x_1-x_2-Ebene mit der Länge in x_3-Richtung kombinieren und erhalten so durch zweimalige Anwendung des Satzes von Pythagoras

$$\sqrt{\sqrt{x_1^2 + x_2^2}^2 + x_3^2} = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

Dieser Längenbegriff verallgemeinert sich genauso für $\mathbb{R}^n$ mit beliebigem $n \in \mathbb{N} \setminus \{0\}$.

Definition 4.9. Für $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ definieren wir das *Standardskalarprodukt* und die *euklidische Länge* durch

$$(\boldsymbol{x} \mid \boldsymbol{y}) \coloneqq \sum_{j=1}^{n} x_j y_j \qquad \text{und} \qquad \|\boldsymbol{x}\| \coloneqq \sqrt{(\boldsymbol{x} \mid \boldsymbol{x})} = \sqrt{\sum_{j=1}^{n} x_j^2}.$$

Der *Abstand* von $\boldsymbol{x}$ zu $\boldsymbol{y}$ ist $\|\boldsymbol{x} - \boldsymbol{y}\|$.

Das Skalarprodukt liefert einen Skalar statt eines Vektors und erlaubt uns, die folgenden Eigenschaften der Länge kompakt auszudrücken.

Beobachtung 4.10. Für $n = 1$ stimmt die euklidische Länge mit dem reellen Betrag überein, und unter der Identifikation von $\mathbb{R}^2$ mit $\mathbb{C}$ mit dem komplexen Betrag. Für alle $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$, $\lambda, \mu \in \mathbb{R}$ und $i \in \{1, \dots, n\}$ gilt

$$\begin{aligned} \|\boldsymbol{x}\| &\geq 0, & \|\boldsymbol{x}\| = 0 &\Leftrightarrow \boldsymbol{x} = \boldsymbol{0}, \\ \|\boldsymbol{e}^{(i)}\| &= 1, & (\boldsymbol{x} \mid \boldsymbol{x}) &> 0 \quad \text{für } \boldsymbol{x} \neq \boldsymbol{0}, \\ (\boldsymbol{x} \mid \boldsymbol{y}) &= (\boldsymbol{y} \mid \boldsymbol{x}), & \|\lambda \boldsymbol{x}\| &= |\lambda| \, \|\boldsymbol{x}\|. \end{aligned}$$

Außerdem gelten die Rechenregeln

$$\begin{aligned} (\lambda \boldsymbol{x} + \mu \boldsymbol{y} \mid \boldsymbol{z}) &= \lambda (\boldsymbol{x} \mid \boldsymbol{z}) + \mu (\boldsymbol{y} \mid \boldsymbol{z}), \\ (\boldsymbol{x} \mid \lambda \boldsymbol{y} + \mu \boldsymbol{z}) &= \lambda (\boldsymbol{x} \mid \boldsymbol{y}) + \mu (\boldsymbol{x} \mid \boldsymbol{z}), \\ \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 - 2(\boldsymbol{x} \mid \boldsymbol{y}) + \|\boldsymbol{y}\|^2. \end{aligned}$$

Beweis. Bis auf die letzte Identität folgen alle Aussagen unmittelbar aus den Definitionen, den Rechenregeln in $\mathbb{R}$ und der Betragsformel $|x| = \sqrt{x^2}$ aus Satz 3.93. Die letzte Identität erhalten wir durch Anwendung der obigen Rechenregeln mit $\lambda = 1$, $\mu = -1$:

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= (\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x} - \boldsymbol{y}) \\ &= (\boldsymbol{x} \mid \boldsymbol{x} - \boldsymbol{y}) - (\boldsymbol{y} \mid \boldsymbol{x} - \boldsymbol{y}) \\ &= (\boldsymbol{x} \mid \boldsymbol{x}) - (\boldsymbol{x} \mid \boldsymbol{y}) - (\boldsymbol{y} \mid \boldsymbol{x}) + (\boldsymbol{y} \mid \boldsymbol{y}) \\ &= \|\boldsymbol{x}\|^2 - 2(\boldsymbol{x} \mid \boldsymbol{y}) + \|\boldsymbol{y}\|^2. \end{aligned}$$

□

Definition 4.11. Zwei Vektoren $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ heißen *orthogonal*, wenn die Aussage $(\boldsymbol{x} \perp \boldsymbol{y}) \coloneqq ((\boldsymbol{x} \mid \boldsymbol{y}) = 0)$ gilt. Ein Vektor $\boldsymbol{n} \in \mathbb{R}^3 \setminus \{\boldsymbol{0}\}$ heißt *Normalenvektor* einer Ebene, wenn er zu ihren Richtungsvektoren orthogonal ist.

Anschaulich bedeutet Orthogonalität in Ebene und Raum, dass zwei von null verschiedene Vektoren senkrecht aufeinander stehen. Der Nullvektor ist nach der Definition zu jedem Vektor orthogonal, obwohl wir ihm keine Richtung zuordnen.

Beispiel 4.12. *Die Vektoren* $(1, 1)^\top$ *und* $(1, -1)^\top$ *sind orthogonal, denn ihr Skalarprodukt ist* $1 - 1 = 0$. △

Beispiel 4.13. *Die Ebene aus Beispiel 4.7 hat den Stützvektor* $\boldsymbol{p} = (0,0,1)^\top$ *und die Richtungsvektoren* $\boldsymbol{u} = (1,1,0)^\top$, $\boldsymbol{v} = (1,2,2)^\top$. *Der Vektor* $\boldsymbol{n} = (2,-2,1)^\top$ *ist ein Normalenvektor, denn*

$$(\boldsymbol{u} \mid \boldsymbol{n}) = 2 - 2 = 0, \qquad (\boldsymbol{v} \mid \boldsymbol{n}) = 2 - 4 + 2 = 0.$$

Jeder Punkt $\boldsymbol{x}$ *der Ebene erfüllt daher* $(\boldsymbol{x} \mid \boldsymbol{n}) = (\boldsymbol{p} \mid \boldsymbol{n}) = 1$, *also*

$$2x_1 - 2x_2 + x_3 = 1.$$

Umgekehrt liegt jeder Punkt $\boldsymbol{x}$, *der diese Gleichung erfüllt, auf der Ebene: Mit* $t = (x_3 - 1)/2$ *und* $s = x_1 - t$ *gilt* $\boldsymbol{x} = \binom{0}{0}{1} + s\boldsymbol{u} + t\boldsymbol{v}$, *denn aus der Gleichung folgt* $x_2 = x_1 + t$. *Die Gleichung beschreibt somit dieselbe Ebene.* △

Die Darstellung aus Beispiel 4.13 können wir auch mit mehr als drei Koordinaten betrachten. Dafür führen wir einen eigenen Begriff ein.

Definition 4.14. Für $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ und $b \in \mathbb{R}$ heißt die Menge

$$H \coloneqq \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x} \mid \boldsymbol{a}) = b\}$$

eine *Hyperebene* in $\mathbb{R}^n$. Der Vektor $\boldsymbol{a}$ heißt ein *Normalenvektor* von H.

Die Gleichung lautet ausgeschrieben $a_1x_1 + \cdots + a_nx_n = b$. Weil mindestens ein a_j von null verschieden ist, können wir nach x_j auflösen und erhalten $x_j = \frac{b}{a_j} - \sum_{i=1}^{j-1} \frac{a_i}{a_j} x_i - \sum_{i=j+1}^{n} \frac{a_i}{a_j} x_i$. Wir können also alle anderen Koordinaten frei wählen und dann ein x_j erhalten, so dass der resultierende Punkt auf der Hyperebene liegt. Anschaulich hat die Hyperebene also $n-1$ Freiheitsgrade. Wir werden in Abschnitt 4.9 sehen, dass ihre Dimension tatsächlich $n-1$ beträgt. Im $\mathbb{R}^3$ sind Ebenen und Hyperebenen dasselbe. In höheren Räumen müssen wir unterscheiden: Eine Ebene hat weiterhin zwei frei wählbare Parameter, eine Hyperebene dagegen $n-1$ frei wählbare Koordinaten.

Für zwei Punkte $\boldsymbol{x}, \boldsymbol{y} \in H$ gilt $(\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{a}) = (\boldsymbol{x} \mid \boldsymbol{a}) - (\boldsymbol{y} \mid \boldsymbol{a}) = b - b = 0$. Der Normalenvektor steht also senkrecht auf jedem Verbindungsvektor zwischen Punkten der Hyperebene. Die Ebene aus Beispiel 4.13 ist eine Hyperebene, weil sie genau aus den Punkten besteht, die $2x_1 - 2x_2 + x_3 = 1$ erfüllen; hier sind also $\boldsymbol{a} = (2,-2,1)^\top$ und $b = 1$.

Wir suchen jetzt den Punkt einer Hyperebene, der einem gegebenen Punkt am nächsten kommt. Dazu gehen wir vom gegebenen Punkt aus in Normalenrichtung, bis wir die Hyperebene treffen.

Satz 4.15. Seien $H = \{\boldsymbol{y} \in \mathbb{R}^n \mid (\boldsymbol{y} \mid \boldsymbol{a}) = b\}$ eine Hyperebene mit $\boldsymbol{a} \neq \boldsymbol{0}$ und $\boldsymbol{x} \in \mathbb{R}^n$. Dann ist

$$\min\{\|\boldsymbol{x} - \boldsymbol{p}\| \mid \boldsymbol{p} \in H\} = \frac{|(\boldsymbol{x} \mid \boldsymbol{a}) - b|}{\|\boldsymbol{a}\|}.$$

Beweis. Wir setzen

$$\alpha \coloneqq ((\boldsymbol{x} \mid \boldsymbol{a}) - b)/\|\boldsymbol{a}\|^2 \tag{4.1}$$

und $\boldsymbol{p} \coloneqq \boldsymbol{x} - \alpha\boldsymbol{a}$. Dann ist $\boldsymbol{p} \in H$, denn

$$(\boldsymbol{p} \mid \boldsymbol{a}) = (\boldsymbol{x} \mid \boldsymbol{a}) - \alpha\|\boldsymbol{a}\|^2 = b.$$

Für jeden Punkt $\boldsymbol{y} \in H$ gilt $(\boldsymbol{p} - \boldsymbol{y} \mid \boldsymbol{a}) = b - b = 0$. Also sind $\boldsymbol{p} - \boldsymbol{y}$ und $\boldsymbol{x} - \boldsymbol{p} = \alpha \boldsymbol{a}$ orthogonal zueinander. Damit ist

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 \overset{4.10}{=} \|\boldsymbol{x} - \boldsymbol{p}\|^2 + \|\boldsymbol{p} - \boldsymbol{y}\|^2 \overset{3.78}{\geq} \|\boldsymbol{x} - \boldsymbol{p}\|^2,$$

also wird das Minimum durch $\boldsymbol{p}$ angenommen. Der Wert des Minimums ist damit

$$\|\boldsymbol{x} - \boldsymbol{p}\| = \|\alpha \boldsymbol{a}\| \overset{4.10}{=} |\alpha| \, \|\boldsymbol{a}\| \overset{(4.1)}{=} \frac{|(\boldsymbol{x} \mid \boldsymbol{a}) - b|}{\|\boldsymbol{a}\|}. \qquad \square$$

Beispiel 4.16. *Die Ebene aus Beispiel 4.13 ist die Hyperebene* $H = \{\boldsymbol{y} \in \mathbb{R}^3 \mid (\boldsymbol{y} \mid \boldsymbol{n}) = 1\}$ *mit* $\boldsymbol{n} = (2, -2, 1)^\top$. *Wegen* $\|\boldsymbol{n}\| = 3$ *erhalten wir mit Satz 4.15 den kleinsten Abstand zum Ursprung:*

$$\min\{\|\boldsymbol{p}\| \mid \boldsymbol{p} \in H\} = \frac{|(\boldsymbol{0} \mid \boldsymbol{n}) - 1|}{\|\boldsymbol{n}\|} = \frac{1}{3}.$$

Der im Beweis konstruierte nächstgelegene Punkt ist

$$\boldsymbol{p} = \boldsymbol{0} - \frac{0 - 1}{9} \boldsymbol{n} = \frac{1}{9} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}. \qquad \triangle$$

Bemerkung 4.17. Eine Hyperebene $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{n} \mid \boldsymbol{x}) = b\}$ trennt die Punkte außerhalb von H danach, ob $(\boldsymbol{n} \mid \boldsymbol{x}) < b$ oder $(\boldsymbol{n} \mid \boldsymbol{x}) > b$ gilt. Dieses Prinzip wird auch zur Bildklassifikation verwendet, etwa zur Unterscheidung von Aufnahmen gesunden und kranken Gewebes. Ein neuronales Netz bildet ein Bild $\boldsymbol{a}$ zunächst auf einen Merkmalsvektor $f(\boldsymbol{a}) \in \mathbb{R}^n$ ab. Die Klassifikation kann dann danach erfolgen, auf welcher Seite einer gelernten Hyperebene $f(\boldsymbol{a})$ liegt. Beim Training werden die Abbildung f und die Hyperebene anhand bereits zugeordneter Beispielbilder angepasst.

Das Kreuzprodukt in $\mathbb{R}^3$

Zu den beiden Richtungsvektoren einer Ebene im dreidimensionalen Raum suchen wir einen Normalenvektor. Das folgende Produkt liefert dafür eine direkte Formel. Anders als beim Skalarprodukt ist das Ergebnis wieder ein Vektor.

Definition 4.18. Für $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ definieren wir das *Kreuzprodukt* durch

$$\boldsymbol{x} \times \boldsymbol{y} \coloneqq \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

Satz 4.19. Für alle $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ gilt

$$(\boldsymbol{x} \times \boldsymbol{y}) \perp \boldsymbol{x}, \qquad (\boldsymbol{x} \times \boldsymbol{y}) \perp \boldsymbol{y}, \qquad \boldsymbol{y} \times \boldsymbol{x} = -(\boldsymbol{x} \times \boldsymbol{y}).$$

Das Kreuzprodukt ist genau dann der Nullvektor, wenn $\boldsymbol{x} = \boldsymbol{0}$ ist oder $\boldsymbol{y} = \lambda \boldsymbol{x}$ für ein $\lambda \in \mathbb{R}$ gilt.

Beweis. Beim Ausmultiplizieren heben sich alle Terme paarweise auf:

$$(\boldsymbol{x} \times \boldsymbol{y} \mid \boldsymbol{x}) = (x_2y_3 - x_3y_2)x_1 + (x_3y_1 - x_1y_3)x_2 + (x_1y_2 - x_2y_1)x_3 = 0.$$

Der Vorzeichenwechsel beim Vertauschen folgt unmittelbar aus der Definition; damit folgt auch die Orthogonalität zu $\boldsymbol{y}$. Ist ein Faktor null oder ein Vielfaches des anderen, verschwinden alle Komponenten des Kreuzprodukts. Ist umgekehrt $\boldsymbol{x} \times \boldsymbol{y} = \boldsymbol{0}$ und $x_i \neq 0$, so gilt $x_iy_j = x_jy_i$ für alle j, also $\boldsymbol{y} = (y_i/x_i)\boldsymbol{x}$. □

Für die Richtungsvektoren $\boldsymbol{x}, \boldsymbol{y}$ einer Ebene ist $\boldsymbol{x} \times \boldsymbol{y}$ daher ein Normalenvektor dieser Ebene. Teilen wir ihn durch seine Länge, erhalten wir einen Normalenvektor der Länge 1. Die Reihenfolge legt fest, welche der beiden Normalenrichtungen wir erhalten: Es gilt etwa $\boldsymbol{e}^{(1)} \times \boldsymbol{e}^{(2)} = \boldsymbol{e}^{(3)}$, aber $\boldsymbol{e}^{(2)} \times \boldsymbol{e}^{(1)} = -\boldsymbol{e}^{(3)}$.

Beobachtung 4.20. Für $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$ ist $\|\boldsymbol{u} \times \boldsymbol{v}\|$ der Flächeninhalt des von ihnen aufgespannten Parallelogramms. Für $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ ist der Flächeninhalt entsprechend

$$|u_1v_2 - u_2v_1|.$$

Beweis. Wir schreiben $\boldsymbol{u} = a\boldsymbol{e}$ mit $a \coloneqq \|\boldsymbol{u}\|$ und einem Vektor $\boldsymbol{e}$ der Länge 1 und zerlegen $\boldsymbol{v}$ in einen Anteil entlang der Grundseite und einen dazu senkrechten Anteil:

$$\boldsymbol{v} = b\boldsymbol{e} + \boldsymbol{w}, \qquad b \coloneqq (\boldsymbol{v} \mid \boldsymbol{e}), \qquad \boldsymbol{w} \coloneqq \boldsymbol{v} - b\boldsymbol{e}.$$

Dann ist $(\boldsymbol{w} \mid \boldsymbol{e}) = 0$, und die Höhe des Parallelogramms ist $\|\boldsymbol{w}\|$. Einsetzen in die Koordinatenformel ergibt $\boldsymbol{u} \times \boldsymbol{v} = a(\boldsymbol{e} \times \boldsymbol{w})$. Durch Ausmultiplizieren erhalten wir

$$\|\boldsymbol{e} \times \boldsymbol{w}\|^2 = \|\boldsymbol{e}\|^2\|\boldsymbol{w}\|^2 - (\boldsymbol{e} \mid \boldsymbol{w})^2 = \|\boldsymbol{w}\|^2.$$

Somit ist $\|\boldsymbol{u} \times \boldsymbol{v}\| = a\|\boldsymbol{w}\|$, also Grundseite mal Höhe. Für die zweidimensionale Formel ergänzen wir beide Vektoren um eine dritte Koordinate 0. Ihr Kreuzprodukt ist $(0, 0, u_1v_2 - u_2v_1)^\top$; seine Länge liefert die angegebene Formel. □

Mit einer dritten Kante entsteht ein *Spat*, also ein von drei Kantenvektoren aufgespannter räumlicher Körper. Sein Volumen erhalten wir ebenfalls aus dem Kreuzprodukt, indem wir den Flächeninhalt der Grundfläche mit der Höhe multiplizieren.

Beobachtung 4.21. Für $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$ ist das Volumen des von ihnen aufgespannten Spats

$$|(\boldsymbol{u} \times \boldsymbol{v} \mid \boldsymbol{w})|.$$

Beweis. Wir wählen einen Normalenvektor $\boldsymbol{e}$ der Länge 1 zur Grundebene so, dass $\boldsymbol{u} \times \boldsymbol{v} = \|\boldsymbol{u} \times \boldsymbol{v}\|\boldsymbol{e}$ gilt. Nach Satz 4.15 ist der Abstand von $\boldsymbol{w}$ zur Grundebene $|(\boldsymbol{w} \mid \boldsymbol{e})|$. Mit Beobachtung 4.20 ergibt Grundfläche mal Höhe daher

$$\|\boldsymbol{u} \times \boldsymbol{v}\|\,|(\boldsymbol{w} \mid \boldsymbol{e})| = |(\boldsymbol{w} \mid \boldsymbol{u} \times \boldsymbol{v})| = |(\boldsymbol{u} \times \boldsymbol{v} \mid \boldsymbol{w})|.$$ □

Beispiel 4.22. *Für die Richtungsvektoren der Ebene aus Beispiel 4.13 erhalten wir*

$$\boldsymbol{u} \times \boldsymbol{v} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}.$$

Das ist genau der dort verwendete Normalenvektor $\boldsymbol{n}$. Das von den beiden Richtungsvektoren aufgespannte Parallelogramm hat daher den Flächeninhalt $\|\boldsymbol{n}\| = \sqrt{4+4+1} = 3$.

Ergänzen wir die dritte Kante $\boldsymbol{w} \coloneqq (2,0,1)^\top$, so hat der Spat nach Beobachtung 4.21 das Volumen

$$|(\boldsymbol{u} \times \boldsymbol{v} \mid \boldsymbol{w})| = |(\boldsymbol{n} \mid \boldsymbol{w})| = |2 \cdot 2 - 2 \cdot 0 + 1 \cdot 1| = 5.$$

Seine Höhe ist nach Satz 4.15 $|(\boldsymbol{w} \mid \boldsymbol{n})|/\|\boldsymbol{n}\| = \frac{5}{3}$. Mit der Grundfläche 3 ergibt sich ebenfalls das Volumen $3 \cdot \frac{5}{3} = 5$. △

4.2 Vektorräume

Wie schon in Kapitel 3 wollen wir die Rechnung mit Vektoren im $\mathbb{R}^n$ aus Abschnitt 4.1 wieder abstrahieren, um sie auch auf andere mathematische Objekte anwenden zu können. Zum einen wollen wir nicht verlangen, dass Komponenten eines Vektors reell oder überhaupt Zahlen sind, zum anderen wollen wir nicht einmal davon ausgehen, dass Vektoren als Tupel von Koordinaten beschrieben sind.

Wir abstrahieren zunächst die Addition und die Multiplikation mit Skalaren und greifen erst später (Abschnitte 4.6 und 4.9) Hyperebenen und den Längenbegriff auf.

Definition 4.23. Ein Tupel $(V, +, \cdot)$ aus einer Menge $V \neq \emptyset$, einer Verknüpfung $+\colon V^2 \to V$ und einer Abbildung $\cdot\colon K \times V \to V$ heißt *Vektorraum* über einem Körper K, falls die folgenden Bedingungen gelten:

(g) Es ist $(V, +)$ eine abelsche Gruppe.

(d1) Für alle $\lambda \in K$ und $\boldsymbol{v}, \boldsymbol{w} \in V$ gilt $\lambda(\boldsymbol{v} + \boldsymbol{w}) = \lambda\boldsymbol{v} + \lambda\boldsymbol{w}$.

(d2) Für alle $\lambda, \mu \in K$ und $\boldsymbol{v} \in V$ gilt $(\lambda + \mu)\boldsymbol{v} = \lambda\boldsymbol{v} + \mu\boldsymbol{v}$.

(a) Für alle $\lambda, \mu \in K$ und $\boldsymbol{v} \in V$ gilt $\lambda(\mu\boldsymbol{v}) = (\lambda\mu)\boldsymbol{v}$.

(e) Für alle $\boldsymbol{v} \in V$ gilt $1_K\boldsymbol{v} = \boldsymbol{v}$.

Die Elemente von V heißen *Vektoren*, die Elemente von K *Skalare*. Die Verknüpfung $+$ heißt *Vektoraddition*, die Abbildung $\cdot$ heißt *Skalarmultiplikation*; wir schreiben meist $\lambda\boldsymbol{v}$ statt $\lambda \cdot \boldsymbol{v}$. Den *Nullvektor*, also das neutrale Element von $(V, +)$, bezeichnen wir mit $\mathbf{0}_V$, das additive Inverse von $\boldsymbol{v}$ mit $-\boldsymbol{v}$. Wir setzen $\boldsymbol{v} - \boldsymbol{w} \coloneqq \boldsymbol{v} + (-\boldsymbol{w})$.

Ist aus dem Zusammenhang klar, welcher Körper und welcher Vektorraum gemeint sind, schreiben wir 0 und 1 statt 0_K und 1_K sowie $\mathbf{0}$ statt $\mathbf{0}_V$. Die Skalarmultiplikation ist im Unterschied zur Multiplikation in einem Ring oder Körper im Allgemeinen die Verknüpfung zweier verschiedenartiger Objekte.

Wir können nun die vertrauten Koordinatenräume einordnen und zugleich den Zahlenraum der Komponenten wechseln.

Beispiel 4.24. *Für $n \in \mathbb{N} \setminus \{0\}$ ist $(\mathbb{R}^n, +, \cdot)$ mit Vektoraddition und Skalarmultiplikation ein Vektorraum über $\mathbb{R}$ (siehe Beobachtung 4.3). Analog ist $(K^n, +, \cdot)$ mit komponentenweiser Addition und Skalarmultiplikation ein Vektorraum über jedem Körper $(K, +, \cdot)$. Insbesondere bilden so $\mathbb{Q}^n$ und $(\mathbb{Z}/2\mathbb{Z})^n$ Vektorräume über den entsprechenden Körpern.* △

Beispiel 4.25. *Wir können $(\mathbb{C}, +, \cdot)$ sowohl als Vektorraum über $\mathbb{R}$ als auch als Vektorraum über $\mathbb{C}$ betrachten. In beiden Fällen sind die Vektoren komplexe Zahlen und die Vektoraddition ist die Addition in $\mathbb{C}$. Der Unterschied liegt darin, welche Skalare wir zulassen: Über $\mathbb{R}$ multiplizieren wir die Vektoren nur mit reellen Zahlen, über $\mathbb{C}$ auch mit beliebigen komplexen Zahlen. Beispielsweise ist i in beiden Fällen ein Vektor, aber nur im zweiten Fall auch ein zulässiger Skalar. Die Vektorraumaxiome folgen jeweils aus den Rechenregeln des Körpers der komplexen Zahlen (Satz 3.96).* △

Beispiel 4.26. *Bezeichne $K[X]$ die Menge aller Polynome über einem Körper K (siehe Definition 3.106). Zwei Polynome addieren wir, indem wir die Koeffizienten gleicher Potenzen addieren; zur Multiplikation mit einem Skalar aus K multiplizieren wir jeden Koeffizienten mit diesem Skalar. Fehlende Koeffizienten behandeln wir als null. Nach der Rechnung lassen wir verschwindende höchste Terme weg; sind alle Koeffizienten null, erhalten wir das Nullpolynom. Beispielsweise gilt in $\mathbb{R}[X]$*

$$(1 + X) + (2 - X + X^2) = 3 + X^2, \qquad 2(1 + X) = 2 + 2X.$$

Mit diesen Operationen bildet $K[X]$ einen Vektorraum über K: Der Nullvektor ist das Nullpolynom, und das additive Inverse erhalten wir durch Negieren aller Koeffizienten. Die Vektorraumaxiome folgen aus den Körperrechenregeln für die einzelnen Koeffizienten. △

Beispiel 4.27. *Die Menge $\mathbb{Z}^2$ ist mit der komponentenweisen Addition zwar eine abelsche Gruppe, bildet aber keinen reellen Vektorraum mit der üblichen Skalarmultiplikation: Es ist $\boldsymbol{e}^{(1)} \in \mathbb{Z}^2$, jedoch $\frac{1}{2}\boldsymbol{e}^{(1)} \notin \mathbb{Z}^2$. Auch die Bildcodierungen $(\mathbb{Q} \cap [0, 1])^N$ aus der Einleitung dieses Kapitels bilden für $N \geq 1$ keinen reellen Vektorraum. Beispielsweise ist $\boldsymbol{e}^{(1)}$ eine zulässige Codierung, aber $2\boldsymbol{e}^{(1)}$ nicht. Für die Vektorrechnung können wir die Bildcodierungen im umgebenden reellen Vektorraum $\mathbb{R}^N$ betrachten.* △

Bisher waren unsere Vektoren durch endlich viele Zahlen oder Koeffizienten beschrieben. Die Axiome verlangen jedoch nur, dass sich die Elemente addieren und mit Skalaren multiplizieren lassen.

Beispiel 4.28. *Für eine Menge M und einen Körper $(K, +, \cdot)$ ist der Raum $(\mathrm{abb}(M, K), +, \cdot)$ aller Abbildungen von M nach K mit* punktweiser Addition und Skalarmultiplikation *ein Vektorraum über K. Für $\boldsymbol{f}, \boldsymbol{g} \in \mathrm{abb}(M, K)$ und $\lambda \in K$ definieren wir diese Operationen durch*

$$(\boldsymbol{f} + \boldsymbol{g})(x) \coloneqq \boldsymbol{f}(x) + \boldsymbol{g}(x), \qquad (\lambda \boldsymbol{f})(x) \coloneqq \lambda \boldsymbol{f}(x) \qquad \text{für alle } x \in M.$$

Der Nullvektor ist die konstante Nullfunktion $x \mapsto 0_K$, das additive Inverse von $\boldsymbol{f}$ ist die Funktion $x \mapsto -\boldsymbol{f}(x)$. Wir nennen $(\mathrm{abb}(M, K), +, \cdot)$ einen Funktionenraum.

Zwei Abbildungen mit gleicher Definitions- und Zielmenge sind gleich, wenn ihre Werte an jeder Stelle übereinstimmen. Daher folgen die Axiome aus den Rechenregeln in K. Die Addition ist insbesondere assoziativ und kommutativ; die angegebene Nullfunktion und die negierten Funktionen machen $(\mathrm{abb}(M, K), +)$ zu einer abelschen Gruppe. Beispielsweise gilt für die Distributivität und jedes $x \in M$

$$(\lambda(\boldsymbol{f} + \boldsymbol{g}))(x) = \lambda(\boldsymbol{f}(x) + \boldsymbol{g}(x)) \overset{(d)}{=} \lambda \boldsymbol{f}(x) + \lambda \boldsymbol{g}(x) = (\lambda \boldsymbol{f} + \lambda \boldsymbol{g})(x).$$ △

Beispiel 4.29. *Wir betrachten die Menge*

$$\mathcal{F} \coloneqq \{\boldsymbol{f} \in \operatorname{abb}(\mathbb{R}, \mathbb{R}) \mid \exists a, b \in \mathbb{R} \colon \forall t \in \mathbb{R} \colon \boldsymbol{f}(t) = a + bt\}$$

aller reellen Funktionen dieser Form. Mit den punktweisen Operationen aus Beispiel 4.28 ist auch $\mathcal{F}$ ein Vektorraum über $\mathbb{R}$. Sind nämlich $\boldsymbol{f}(t) = a + bt$ und $\boldsymbol{g}(t) = c + dt$, so gilt für jedes $t \in \mathbb{R}$ und $\lambda \in \mathbb{R}$

$$\begin{aligned}(\boldsymbol{f} + \boldsymbol{g})(t) &= (a + c) + (b + d)t,\\ (\lambda \boldsymbol{f})(t) &= \lambda a + (\lambda b)t.\end{aligned}$$

Die Ergebnisse liegen also wieder in $\mathcal{F}$. Auch die Nullfunktion und mit jeder Funktion $\boldsymbol{f}$ ihr additives Inverses $-\boldsymbol{f}$ gehören zu $\mathcal{F}$. Die übrigen Vektorraumaxiome gelten bereits im Funktionenraum aus Beispiel 4.28. △

Zum Abschluss halten wir Rechenregeln fest, die unabhängig von der Art der Vektoren in jedem Vektorraum gelten.

Satz 4.30. Für alle $\boldsymbol{v} \in V$ und $\lambda \in K$ in einem Vektorraum über K gilt

$$0\boldsymbol{v} = \boldsymbol{0}, \quad \lambda \boldsymbol{0} = \boldsymbol{0}, \quad (-1)\boldsymbol{v} = -\boldsymbol{v}.$$

Außerdem gilt $\lambda \boldsymbol{v} = \boldsymbol{0}$ genau dann, wenn $\lambda = 0$ oder $\boldsymbol{v} = \boldsymbol{0}$ ist.

Beweis. Übung. □

4.3 Untervektorräume

Bei den Untergruppen in Definition 3.19 haben wir untersucht, auf welchen Teilmengen einer Gruppe die eingeschränkte Verknüpfung wieder eine Gruppe ergibt. Für Vektorräume stellen wir dieselbe Frage; hier muss zusätzlich die Skalarmultiplikation berücksichtigt werden.

Im gesamten Abschnitt sei K ein Körper und $(V, +, \cdot)$ ein Vektorraum über K.

Definition 4.31. Ein Vektorraum $(U, +_U, \cdot_U)$ über K heißt *Unter(vektor)raum* von $(V, +, \cdot)$, wenn $U \subseteq V$ gilt und $+_U = +|_{U^2}$ sowie $\cdot_U = \cdot|_{K \times U}$ Einschränkungen der Verknüpfungen von $(V, +, \cdot)$ sind. Wir schreiben dann vereinfachend $(U, +, \cdot) \coloneqq (U, +_U, \cdot_U)$.

Die Ergebnisse der Addition und der Skalarmultiplikation müssen also wieder in U liegen. Die Rechenregeln brauchen wir dagegen nicht alle neu zu prüfen: Sie gelten bereits in $(V, +, \cdot)$. Wie beim Untergruppenkriterium aus Satz 3.24 können wir die nötigen Bedingungen zusammenfassen.

Satz 4.32 (Unterraumkriterium)**.** Für eine Teilmenge $U \subseteq V$ ist $(U, +, \cdot)$ genau dann ein Untervektorraum von $(V, +, \cdot)$, wenn $U \neq \emptyset$ ist und $\lambda \boldsymbol{u} + \mu \boldsymbol{w} \in U$ für alle $\boldsymbol{u}, \boldsymbol{w} \in U$ und $\lambda, \mu \in K$ gilt.

Beweis. Ist $(U, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$, so ist U nicht leer und unter Addition und Skalarmultiplikation abgeschlossen. Damit gilt auch die angegebene Bedingung.

Sei umgekehrt die Bedingung erfüllt. Für $\boldsymbol{u}, \boldsymbol{w} \in U$ gilt im Vektorraum $(V, +, \cdot)$

$$\boldsymbol{u} - \boldsymbol{w} \overset{4.30}{=} \boldsymbol{u} + (-1)\boldsymbol{w} \overset{\text{(e)}}{=} 1\boldsymbol{u} + (-1)\boldsymbol{w}.$$

Nach der Bedingung mit $\lambda = 1_K$ und $\mu = -1_K$ folgt $\boldsymbol{u} - \boldsymbol{w} \in U$.

Nach dem Untergruppenkriterium aus Satz 3.24 ist $(U, +)$ daher eine Untergruppe von $(V, +)$ und insbesondere abelsch. Für $\lambda \in K$ und $\boldsymbol{u} \in U$ gilt außerdem $\lambda\boldsymbol{u} = \lambda\boldsymbol{u} + 0_K\boldsymbol{u} \in U$. Somit können wir die Skalarmultiplikation auf U einschränken. Die übrigen Vektorraumaxiome aus Definition 4.23 gelten für die Elemente von U, weil sie in $(V, +, \cdot)$ gelten. Somit ist $(U, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$. □

Beispiel 4.33. *Jeder Vektorraum $(V, +, \cdot)$ besitzt die Untervektorräume $(\{\mathbf{0}\}, +, \cdot)$ und $(V, +, \cdot)$.* △

Beispiel 4.34. *Der Vektorraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.29 ist ein Untervektorraum des Funktionenraums $(\mathrm{abb}(\mathbb{R}, \mathbb{R}), +, \cdot)$ aus Beispiel 4.28: Es gilt $\mathcal{F} \subseteq \mathrm{abb}(\mathbb{R}, \mathbb{R})$, und die Operationen auf $\mathcal{F}$ sind die Einschränkungen der punktweisen Operationen. Die Abgeschlossenheit haben wir dort bereits nachgerechnet.* △

Beispiel 4.35. *Die Menge $\mathcal{P}_K \subseteq \mathrm{abb}(K, K)$ aller Polynomfunktionen über K enthält die Nullfunktion und ist unter punktweiser Addition und Skalarmultiplikation abgeschlossen: Dabei werden die Koeffizienten gleicher Potenzen addiert bzw. alle Koeffizienten mit dem Skalar multipliziert. Nach Satz 4.32 ist $(\mathcal{P}_K, +, \cdot)$ daher ein Untervektorraum von $(\mathrm{abb}(K, K), +, \cdot)$. Der Vektorraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.29 ist auch ein Untervektorraum von $(\mathcal{P}_\mathbb{R}, +, \cdot)$.* △

Beispiel 4.36. *Geraden, Ebenen und Hyperebenen (Definitionen 4.5 und 4.14) der Form*

$$G = \{t\boldsymbol{u} \mid t \in \mathbb{R}\} \subseteq \mathbb{R}^n, \qquad E = \{s\boldsymbol{u} + t\boldsymbol{v} \mid s, t \in \mathbb{R}\} \subseteq \mathbb{R}^n, \qquad H = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x} \mid \boldsymbol{a}) = 0\}$$

enthalten jeweils den Ursprung und bilden Untervektorräume von $(\mathbb{R}^n, +, \cdot)$. Dagegen bilden Geraden, Ebenen und Hyperebenen, die den Ursprung nicht enthalten, mit den eingeschränkten Operationen keine Vektorräume, da ihnen der Nullvektor fehlt. △

Wie schon für die erzeugten Untergruppen (Definition 3.26) verfolgen wir auch für Untervektorräume wieder das Prinzip der Hüllenbildung: Für eine gegebene Menge von Vektoren in V suchen wir die kleinste in V enthaltene Obermenge, die einen Untervektorraum bildet.

Definition 4.37. Für $M \subseteq V$ sei $\mathcal{U}$ die Menge aller Teilmengen $U \subseteq V$, für die jeweils $(U, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$ ist und $M \subseteq U$ gilt. Die *lineare Hülle* von M ist

$$\langle M \rangle \coloneqq \bigcap_{U \in \mathcal{U}} U.$$

Wir nennen $(\langle M \rangle, +, \cdot)$ den von M *erzeugten Untervektorraum*. Gilt $\langle M \rangle = V$, so heißt M ein *Erzeugendensystem* von $(V, +, \cdot)$. Für endlich viele Vektoren schreiben wir kurz $\langle \boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)} \rangle \coloneqq \langle \{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)}\} \rangle$.

Wir müssen natürlich noch zeigen, dass die lineare Hülle tatsächlich einen Untervektorraum bildet. Wir beobachten dazu, dass wir, wie beim Schnitt von Untergruppen in Satz 3.25, die Einschränkungen mehrerer Untervektorräume kombinieren können.

Satz 4.38. Sei $\mathcal{U} \neq \emptyset$ eine Menge von Teilmengen $U \subseteq V$, für die jeweils $(U, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$ ist. Dann ist auch $\left(\bigcap_{U \in \mathcal{U}} U, +, \cdot\right)$ ein Untervektorraum von $(V, +, \cdot)$. Insbesondere ist $(\langle M \rangle, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$.

Beweis. Übung. □

Beispiel 4.39. *Die Ebenen*

$$E_1 \coloneqq \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_1 + x_2 = 0\}, \qquad E_2 \coloneqq \{\boldsymbol{x} \in \mathbb{R}^3 \mid x_2 + x_3 = 0\}$$

bilden mit den eingeschränkten Operationen Untervektorräume von $(\mathbb{R}^3, +, \cdot)$ (Beispiel 4.36). Im Schnitt gelten beide Gleichungen, also $x_2 = -x_1$ und $x_3 = x_1$. Damit ist

$$E_1 \cap E_2 = \{t\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \mid t \in \mathbb{R}\}.$$

Der Schnitt ist also eine Gerade, die den Ursprung enthält und nach Beispiel 4.36 wieder einen Untervektorraum von $(\mathbb{R}^3, +, \cdot)$ bildet, im Einklang mit Satz 4.38. △

Beobachtung 4.40. Die Vereinigung der zugrunde liegenden Mengen zweier Untervektorräume bildet im Allgemeinen keinen Untervektorraum: Die beiden Koordinatenachsen $\{\begin{pmatrix} t \\ 0 \end{pmatrix} \mid t \in \mathbb{R}\}$ und $\{\begin{pmatrix} 0 \\ t \end{pmatrix} \mid t \in \mathbb{R}\}$ bilden Untervektorräume von $(\mathbb{R}^2, +, \cdot)$, nicht aber ihre Vereinigung $\{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 = 0 \text{ oder } x_2 = 0\}$, denn sie enthält $\boldsymbol{e}^{(1)}$ und $\boldsymbol{e}^{(2)}$, aber nicht deren Summe $\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Wir haben in Beispiel 4.36 gesehen, dass Geraden und Ebenen mit Stützvektor $\mathbf{0}$ Untervektorräume bilden. Diese Untervektorräume werden durch die jeweiligen Richtungsvektoren erzeugt, in dem Sinne, dass sich jeder Punkt durch eine Kombination der Richtungsvektoren erreichen lässt. Wir verallgemeinern dieses Prinzip auf eine beliebige Anzahl Vektoren.

Definition 4.41. Die *Linearkombination* der Vektoren $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)} \in V$ für $m \in \mathbb{N}$ mit den *Koeffizienten* $\lambda_1, \ldots, \lambda_m \in K$ ist der Vektor

$$\sum_{j=1}^{m} \lambda_j \boldsymbol{v}^{(j)}.$$

Nach Satz 4.38 bildet die lineare Hülle mit den eingeschränkten Operationen einen Untervektorraum. Wir zeigen nun, dass sie sich durch Linearkombinationen beschreiben lässt.

Satz 4.42. Für jede Teilmenge $M \subseteq V$ besteht $\langle M \rangle$ genau aus den Linearkombinationen endlich vieler Vektoren aus M. Insbesondere gilt $\langle \emptyset \rangle = \{\mathbf{0}\}$.

Beweis. Sei L die Menge aller Linearkombinationen endlich vieler Vektoren aus M, einschließlich der leeren Linearkombination. Dann ist $\mathbf{0} \in L$ und $M \subseteq L$, denn $\boldsymbol{v} = 1_K \boldsymbol{v}$ für jeden Vektor $\boldsymbol{v} \in M$. Sind

$$\boldsymbol{x} = \sum_{j=1}^{r} a_j \boldsymbol{v}^{(j)}, \qquad \boldsymbol{y} = \sum_{j=1}^{s} b_j \boldsymbol{w}^{(j)}$$

zwei solche Linearkombinationen, so gilt für alle $\lambda, \mu \in K$

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} = \sum_{j=1}^{r} (\lambda a_j) \boldsymbol{v}^{(j)} + \sum_{j=1}^{s} (\mu b_j) \boldsymbol{w}^{(j)} \in L.$$

Nach Satz 4.32 ist $(L, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$. Da $M \subseteq L$ gilt, tritt L im definierenden Schnitt aus Definition 4.37 auf. Daher ist $\langle M \rangle \subseteq L$. Umgekehrt ist $L \subseteq \langle M \rangle$, weil die Menge $\langle M \rangle$ mit den Vektoren aus M auch alle ihre Linearkombinationen enthält. Somit gilt $\langle M \rangle = L$. Für $M = \emptyset$ bleibt nur die leere Linearkombination, also gilt $\langle \emptyset \rangle = \{\mathbf{0}\}$. □

Beispiel 4.43. *Für $n \in \mathbb{N} \setminus \{0\}$ ist jeder Vektor $\boldsymbol{x} \in \mathbb{R}^n$ eine Linearkombination der Einheitsvektoren:*

$$\boldsymbol{x} = x_1 \boldsymbol{e}^{(1)} + \cdots + x_n \boldsymbol{e}^{(n)}.$$

Also gilt $\mathbb{R}^n = \langle \boldsymbol{e}^{(1)}, \ldots, \boldsymbol{e}^{(n)} \rangle$, und $\{\boldsymbol{e}^{(1)}, \ldots, \boldsymbol{e}^{(n)}\}$ ist ein Erzeugendensystem von $(\mathbb{R}^n, +, \cdot)$. In $\mathbb{R}^3$ beschreibt $\langle \boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)} \rangle$ die Ebene $x_3 = 0$. Nehmen wir $\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}$ als weiteren Vektor hinzu, so erhalten wir keine neuen Vektoren, denn

$$a\boldsymbol{e}^{(1)} + b\boldsymbol{e}^{(2)} + c(\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}) = (a + c)\boldsymbol{e}^{(1)} + (b + c)\boldsymbol{e}^{(2)}.$$

Somit ist $\langle \boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)} \rangle = \langle \boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)} \rangle$. △

Beispiel 4.44. *Für die Ebenen aus Beispiel 4.39 gilt*

$$E_1 = \left\langle \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle, \qquad E_2 = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\rangle, \qquad E_1 \cap E_2 = \left\langle \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\rangle. \qquad \triangle$$

Beispiel 4.45. *Im Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.29 seien $\boldsymbol{f}^{(0)}$ und $\boldsymbol{f}^{(1)}$ die Funktionen mit $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ für alle $t \in \mathbb{R}$. Dann ist jede Funktion $\boldsymbol{f} \in \mathcal{F}$ von der Form $\boldsymbol{f} = a\boldsymbol{f}^{(0)} + b\boldsymbol{f}^{(1)}$, also*

$$\mathcal{F} = \langle \boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)} \rangle. \qquad \triangle$$

Beispiel 4.46. *Für einen Körper K bilden die* Monome *X^j mit $j \in \mathbb{N}$ ein Erzeugendensystem des Vektorraums $(K[X], +, \cdot)$ aus Beispiel 4.26. Denn jedes Polynom ist eine endliche Linearkombination dieser Monome, also*

$$K[X] = \left\langle \{X^j \mid j \in \mathbb{N}\} \right\rangle. \qquad \triangle$$

4.4 Basen

Wir interessieren uns jetzt für kleinstmögliche Erzeugendensysteme. Wir fragen also, welche Vektoren für die Erzeugung einer linearen Hülle überflüssig sind. So hatten wir in Beispiel 4.43 gesehen, dass wir $\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}$ zu den Erzeugern $\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}$ hinzufügen können, ohne die lineare Hülle $\langle \boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)} \rangle$ zu vergrößern – in diesem Sinne ist der zusätzliche Vektor $\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}$ für die Erzeugung von $\langle \boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)} \rangle$ überflüssig.

Wie können wir das formal erfassen? Bei der Definition einer Ebene in $\mathbb{R}^n$ (Definition 4.5) hatten wir für den zweiten Richtungsvektor $\boldsymbol{v}$ verlangt, dass $\boldsymbol{v} \neq \lambda \boldsymbol{u}$ für alle $\lambda \in \mathbb{R}$ gilt, damit wir nicht wieder eine Gerade erhalten. Für eine Gerade wäre der zweite Richtungsvektor auch wieder überflüssig. Wir verallgemeinern diese Bedingung jetzt von den zwei Richtungsvektoren $\boldsymbol{u}$ und $\boldsymbol{v}$ auf eine beliebige Menge von Vektoren.

Im gesamten Abschnitt sei K ein Körper und $(V, +, \cdot)$ ein Vektorraum über K.

Definition 4.47. Eine Menge $M \subseteq V$ heißt *linear unabhängig*, wenn für jedes $m \in \mathbb{N} \setminus \{0\}$ und alle paarweise verschiedenen Vektoren $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)} \in M$ gilt:

$$\forall \boldsymbol{\lambda} \in K^m \setminus \{\mathbf{0}\} \colon \sum_{j=1}^{m} \lambda_j \boldsymbol{v}^{(j)} \neq \mathbf{0}.$$

Andernfalls heißt M *linear abhängig*.

Beobachtung 4.48. Für $A \subseteq B \subseteq V$ gilt: Ist B linear unabhängig, so ist auch A linear unabhängig. Ist A linear abhängig, so ist auch B linear abhängig.

Beispiel 4.49. *Die leere Menge ist linear unabhängig, denn wir können keine Vektoren finden, die die Bedingung aus Definition 4.47 verletzen. Jede Menge, die den Nullvektor enthält, ist dagegen linear abhängig: Für $m = 1$, $\boldsymbol{v}^{(1)} = \mathbf{0}$ und $\boldsymbol{\lambda} = (1_K)$ ist die Unabhängigkeitsbedingung verletzt.* △

Beispiel 4.50. *Die Menge $\{\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}\} \subseteq \mathbb{R}^3$ aus Beispiel 4.43 ist linear abhängig, denn*

$$1\boldsymbol{e}^{(1)} + 1\boldsymbol{e}^{(2)} + (-1)(\boldsymbol{e}^{(1)} + \boldsymbol{e}^{(2)}) = \mathbf{0}.$$ △

Beispiel 4.51. *Die beiden Funktionen $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ aus Beispiel 4.45 bilden eine linear unabhängige Menge. Ist nämlich $a\boldsymbol{f}^{(0)} + b\boldsymbol{f}^{(1)}$ mit $a, b \in \mathbb{R}$ die Nullfunktion, so gilt $a + bt = 0$ für alle $t \in \mathbb{R}$. Auswerten bei $t = 0$ liefert $a = 0$, Auswerten bei $t = 1$ anschließend $b = 0$.* △

Wir zeigen nun, dass eine Menge genau dann linear abhängig ist, wenn einer ihrer Vektoren eine Linearkombination endlich vieler anderer Vektoren der Menge ist.

Satz 4.52. Eine Menge $M \subseteq V$ ist genau dann linear abhängig, wenn $\boldsymbol{v} \in M$ mit $\boldsymbol{v} \in \langle M \setminus \{\boldsymbol{v}\} \rangle$ existiert.

Beweis. Ist M linear abhängig, so gibt es paarweise verschiedene Vektoren $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)} \in M$ und Koeffizienten $\lambda_1, \ldots, \lambda_m \in K$ mit $\sum_{j=1}^{m} \lambda_j \boldsymbol{v}^{(j)} = \mathbf{0}$, wobei ein Koeffizient λ_k von null verschieden ist. Wir bringen die übrigen Summanden auf die rechte Seite und multiplizieren mit λ_k^{-1}:

$$\boldsymbol{v}^{(k)} = \sum_{j=1}^{k-1} \left(-\frac{\lambda_j}{\lambda_k}\right) \boldsymbol{v}^{(j)} + \sum_{j=k+1}^{m} \left(-\frac{\lambda_j}{\lambda_k}\right) \boldsymbol{v}^{(j)}.$$

Damit ist $\boldsymbol{v}^{(k)}$ eine Linearkombination von Vektoren aus $M \setminus \{\boldsymbol{v}^{(k)}\}$ und liegt nach Satz 4.42 in deren linearer Hülle.

Liegt umgekehrt $\boldsymbol{v} \in M$ in $\langle M \setminus \{\boldsymbol{v}\}\rangle$, so ist $\boldsymbol{v}$ nach Satz 4.42 eine Linearkombination endlich vieler anderer Vektoren aus M. Mehrfach auftretende Vektoren fassen wir durch Addieren ihrer Koeffizienten zusammen. Subtrahieren wir diese Linearkombination von $\boldsymbol{v}$, erhalten wir eine Darstellung des Nullvektors, in der der Koeffizient von $\boldsymbol{v}$ gleich $1_K \neq 0_K$ ist. Also ist M linear abhängig. □

Korollar 4.53 (Ergänzungskriterium)**.** Sei $L \subseteq V$ linear unabhängig und $\boldsymbol{w} \in V \setminus L$. Dann ist $L \cup \{\boldsymbol{w}\}$ genau dann linear unabhängig, wenn $\boldsymbol{w} \notin \langle L\rangle$ gilt.

Beweis. Ist $L \cup \{\boldsymbol{w}\}$ linear unabhängig, so gilt nach Satz 4.52$\boldsymbol{w} \notin \langle L\rangle$. Gilt umgekehrt $\boldsymbol{w} \notin \langle L\rangle$, so muss in jeder Linearkombination paarweise verschiedener Vektoren aus $L \cup \{\boldsymbol{w}\}$, die den Nullvektor ergibt, der Koeffizient von $\boldsymbol{w}$ null sein, sofern $\boldsymbol{w}$ darin vorkommt: Andernfalls könnten wir nach $\boldsymbol{w}$ auflösen und erhielten $\boldsymbol{w} \in \langle L\rangle$. Die übrigen Koeffizienten sind wegen der Unabhängigkeit von L ebenfalls null. Also ist $L \cup \{\boldsymbol{w}\}$ linear unabhängig. □

Eine linear unabhängige Menge enthält somit keinen Vektor, den wir durch die übrigen darstellen können. Erzeugt sie zugleich den ganzen Vektorraum, so haben wir genau die gesuchte, kleinstmögliche Auswahl von Vektoren.

Definition 4.54. Eine *Basis* von $(V, +, \cdot)$ ist eine linear unabhängige Menge $B \subseteq V$ mit $\langle B\rangle = V$. Ist B endlich, so heißt ein Tupel $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(|B|)})$ mit $\{\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(|B|)}\} = B$ eine *geordnete Basis*.

Insbesondere ist die leere Menge eine Basis des Nullvektorraums $(\{\mathbf{0}\}, +, \cdot)$, denn sie ist linear unabhängig und $\langle\emptyset\rangle = \{\mathbf{0}\}$.

Satz 4.55. Für $n \in \mathbb{N}$ ist $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)}) \in V^n$ genau dann eine geordnete Basis von $(V, +, \cdot)$, wenn zu jedem $\boldsymbol{v} \in V$ genau ein Tupel $(x_1, \ldots, x_n) \in K^n$ existiert mit

$$\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}.$$

Beweis. Für $n = 0$ sind beide Bedingungen genau dann erfüllt, wenn $V = \{\mathbf{0}\}$ gilt, denn die leere Linearkombination hat den Wert $\mathbf{0}$. Sei im Folgenden $n \geq 1$.

Sei $\mathfrak{B}$ eine geordnete Basis. Dann ist $B \coloneqq \{\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)}\}$ eine Basis und die Vektoren $\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)}$ sind paarweise verschieden. Wegen $\langle B\rangle = V$ besitzt nach Satz 4.42 jeder Vektor $\boldsymbol{v} \in V$ eine Darstellung als

Linearkombination von Vektoren in B. Nicht benötigten Basisvektoren können wir dabei den Koeffizienten 0_K zuordnen. Sind $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$ und $\boldsymbol{v} = \sum_{j=1}^{n} y_j \boldsymbol{b}^{(j)}$ zwei Darstellungen, so folgt durch Subtraktion

$$\sum_{j=1}^{n} (x_j - y_j) \boldsymbol{b}^{(j)} = \mathbf{0}.$$

Die lineare Unabhängigkeit liefert $x_j - y_j = 0_K$, also $x_j = y_j$ für alle $j \in \{1, \dots, n\}$.

Besitzt umgekehrt jeder Vektor genau eine Darstellung als Linearkombination, so gilt für $B \coloneqq \{\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(n)}\}$ zunächst $\langle B \rangle = V$. Die Vektoren $\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(n)}$ sind paarweise verschieden: Wäre $\boldsymbol{b}^{(i)} = \boldsymbol{b}^{(j)}$ für $i \neq j$, so könnten wir mit $x_i = 1_K$, $x_j = -1_K$ und $x_\ell = 0_K$ für alle $\ell \in \{1, \dots, n\} \setminus \{i, j\}$ eine zweite Darstellung des Nullvektors erhalten

$$\sum_{\ell=1}^{n} x_\ell \boldsymbol{b}^{(\ell)} = \boldsymbol{b}^{(i)} - \boldsymbol{b}^{(j)} = \mathbf{0} = \sum_{\ell=1}^{n} 0_K \boldsymbol{b}^{(\ell)}.$$

Zum Nachweis der linearen Unabhängigkeit sei $\boldsymbol{\lambda} \in K^n$ mit $\sum_{i=1}^{n} \lambda_i \boldsymbol{b}^{(i)} = \mathbf{0}$, wobei wir jede Linearkombination paarweise verschiedener Vektoren von B zu einer Linearkombination aller Vektoren mit zusätzlichen Koeffizienten 0_K erweitern können. Wegen der Eindeutigkeit der Darstellung $\mathbf{0} = \sum_{i=1}^{n} 0_K \boldsymbol{b}^{(i)}$ folgt $\boldsymbol{\lambda} = \mathbf{0}$. Also ist B linear unabhängig und damit eine Basis und $\mathfrak{B}$ eine geordnete Basis. □

Beobachtung 4.56. Eine Basis ist ein minimales Erzeugendensystem in dem Sinn, dass wir keinen Vektor weglassen können, ohne die Erzeugendeneigenschaft zu verlieren: Der weggelassene Vektor liegt nach Satz 4.52 nicht in der linearen Hülle der übrigen.

Beobachtung 4.57. Eine Basis B ist eine maximale linear unabhängige Teilmenge von V in dem Sinn, dass wir keinen Vektor aus $V \setminus B$ hinzufügen können, ohne die lineare Unabhängigkeit zu verlieren: Für jeden solchen Vektor $\boldsymbol{v}$ gilt $\boldsymbol{v} \in \langle B \rangle = V$, daher ist $B \cup \{\boldsymbol{v}\}$ nach Satz 4.52 linear abhängig.

Beispiel 4.58. *Für $n \in \mathbb{N} \setminus \{0\}$ ist die Menge $\{\boldsymbol{e}^{(1)}, \dots, \boldsymbol{e}^{(n)}\}$ eine Basis von $(\mathbb{R}^n, +, \cdot)$: Die Erzeugung kennen wir aus Beispiel 4.43, und aus $\sum_{j=1}^{n} x_j \boldsymbol{e}^{(j)} = \mathbf{0}$ folgt durch Vergleich der Komponenten $x_1 = \cdots = x_n = 0$. Diese Basis heißt* Standardbasis. *Die zugehörige geordnete Standardbasis ist $\mathfrak{E} \coloneqq (\boldsymbol{e}^{(1)}, \dots, \boldsymbol{e}^{(n)})$.* △

Beispiel 4.59. *Die beiden Erzeuger der Ebene E_1 aus Beispiel 4.44 bilden eine Basis des Untervektorraums $(E_1, +, \cdot)$, denn*

$$a\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ -a \\ b \end{pmatrix} = \mathbf{0}$$

impliziert $a = b = 0$. △

Beispiel 4.60. *Die Monome aus Beispiel 4.46 bilden eine Basis $B = \{X^j \mid j \in \mathbb{N}\}$ von $(K[X], +, \cdot)$. Nach Beispiel 4.46 erzeugt B den Vektorraum. Außerdem ist B linear unabhängig, denn ein Polynom ist genau dann das Nullpolynom, wenn alle seine Koeffizienten null sind (Definition 3.106). Diese Basis ist nicht endlich, aber jedes einzelne Polynom verwendet nur endlich viele Monome.* △

Nach Satz 4.55 sind die Koeffizienten bezüglich einer geordneten Basis eindeutig bestimmt. Dieses Tupel verwenden wir zur Beschreibung des Vektors.

Definition 4.61. Sei $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ eine geordnete Basis von $(V, +, \cdot)$. Für $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$ heißt

$$[\boldsymbol{v}]_{\mathfrak{B}} \coloneqq (x_1, \ldots, x_n)^\top \in K^n$$

der *Koordinatenvektor* von $\boldsymbol{v}$ bezüglich $\mathfrak{B}$.

Bezüglich der geordneten Standardbasis $\mathfrak{E}$ aus Beispiel 4.58 gilt $[\boldsymbol{x}]_{\mathfrak{E}} = (x_1, \ldots, x_n)^\top = \boldsymbol{x}$. Bei anderen Vektorräumen können die Vektoren hingegen ganz andere Objekte sein als ihre Koordinatentupel, wie das folgende Beispiel zeigt.

Beispiel 4.62. *Die Funktionen $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ erzeugen nach Beispiel 4.45 den Vektorraum $(\mathcal{F}, +, \cdot)$; ihre lineare Unabhängigkeit haben wir in Beispiel 4.51 gezeigt. Daher ist $\mathfrak{B} = (\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)})$ eine geordnete Basis, und für $\boldsymbol{f}(t) = a + bt$ mit $a, b \in \mathbb{R}$ gilt*

$$[\boldsymbol{f}]_{\mathfrak{B}} = \binom{a}{b}.$$

Der Vektor $\boldsymbol{f}$ ist eine Funktion, sein Koordinatenvektor ein Element von $\mathbb{R}^2$. Vertauschen wir die Basisvektoren, so erhalten wir $[\boldsymbol{f}]_{(\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(0)})} = \binom{b}{a}$. △

Beispiel 4.63. *Für $(\mathbb{C}, +, \cdot)$ hängt die Basis vom gewählten Skalarkörper ab (Beispiel 4.25): Über $\mathbb{R}$ ist $\{1, \mathrm{i}\}$ eine Basis, denn jede komplexe Zahl ist eindeutig von der Form $a + b\mathrm{i}$ mit $a, b \in \mathbb{R}$. Über $\mathbb{C}$ ist dagegen schon $\{1\}$ eine Basis, denn $z = z \cdot 1$ ist die eindeutige Darstellung mit einem komplexen Koeffizienten. Entsprechend gilt*

$$[a + b\mathrm{i}]_{(1,\mathrm{i})} = \binom{a}{b} \in \mathbb{R}^2, \qquad [z]_{(1)} = (z) \in \mathbb{C}^1.$$ △

4.5 Dimension

Eine geordnete Basis gibt jedem Vektor ein eindeutiges Koordinatentupel. Könnten wir durch die Wahl einer anderen Basis mit weniger Koordinaten auskommen? Wir zeigen, dass alle Basen eines Vektorraums mit endlichem Erzeugendensystem gleich viele Elemente haben. Diese Anzahl macht den Begriff der Dimension aus Abschnitt 4.1 unabhängig von der gewählten Darstellung.

Im gesamten Abschnitt sei K ein Körper und $(V, +, \cdot)$ ein Vektorraum über K. Wir nennen ihn *endlich erzeugt*, wenn es ein endliches Erzeugendensystem $E \subseteq V$ gibt. Zunächst vergleichen wir die Anzahl der Elemente eines solchen Erzeugendensystems mit der einer linear unabhängigen Menge.

Satz 4.64 (Austauschsatz)**.** Sei $E \subseteq V$ ein endliches Erzeugendensystem von $(V, +, \cdot)$. Dann ist jede linear unabhängige Menge $L \subseteq V$ endlich, und es gilt $|L| \leq |E|$.

Beweis. Angenommen, L enthielte mehr als $n \coloneqq |E|$ Elemente. Dann gäbe es eine linear unabhängige Teilmenge $L_0 \subseteq L$ mit $|L_0| = n + 1$. Für $n = 0$ wäre $V = \langle \emptyset \rangle = \{\mathbf{0}\}$ und damit $L_0 = \{\mathbf{0}\}$ nicht linear unabhängig; also ist $n \geq 1$. Unter allen Erzeugendensystemen $F \subseteq V$ mit $|F| = n$ wählen wir eines mit maximalem $|F \cap L_0|$. Das ist nach Satz 2.15 möglich, da $F = E$ eine zulässige Wahl ist und $|F \cap L_0| \in \{0, \ldots, n\}$ für alle F.

Wegen $|L_0| > |F|$ gibt es einen Vektor $\boldsymbol{v} \in L_0 \setminus F$. Schreiben wir $F = \{\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(n)}\}$, so gibt es wegen $\langle F \rangle = V$ Koeffizienten $\lambda_1, \dots, \lambda_n \in K$ mit

$$\boldsymbol{v} = \sum_{j=1}^{n} \lambda_j \boldsymbol{w}^{(j)}. \tag{4.2}$$

Da $(F \cap L_0) \cup \{\boldsymbol{v}\} \subseteq L_0$ als Teilmenge von L_0 nach Beobachtung 4.48 linear unabhängig ist, folgt aus Korollar 4.53, dass $\boldsymbol{v} \notin \langle F \cap L_0 \rangle$. Daher gilt$\lambda_k \neq 0_K$ für ein $\boldsymbol{w}^{(k)} \in F \setminus L_0$. Durch Auflösen von (4.2) nach $\boldsymbol{w}^{(k)}$ und Einsetzen können wir also $\boldsymbol{w}^{(k)}$ in jeder Linearkombination der Vektoren von F durch eine Linearkombination mit $\boldsymbol{v}$ ersetzen. Somit ist $F' \coloneqq (F \setminus \{\boldsymbol{w}^{(k)}\}) \cup \{\boldsymbol{v}\}$ wieder ein Erzeugendensystem mit $|F'| = n$ und $|F' \cap L_0| = |F \cap L_0| + 1$. Das ist ein Widerspruch zur Wahl von F. □

Nach Satz 4.64 sind insbesondere alle Basen eines Vektorraums mit endlichem Erzeugendensystem endlich. Für zwei solche Basen B und C liefert Satz 4.64 dann sowohl $|B| \leq |C|$ als auch $|C| \leq |B|$. Wir erhalten das folgende Korollar.

Korollar 4.65. Alle Basen eines endlich erzeugten Vektorraums sind endlich und haben dieselbe Anzahl von Elementen.

Mit Satz 4.64 können wir außerdem zeigen, dass sich aus jedem endlichen Erzeugendensystem eine Basis auswählen lässt, und dass jeder endlich erzeugte Vektorraum eine Basis besitzt.

Korollar 4.66 (Basisauswahl und Basisergänzung)**.** Sei $E \subseteq V$ ein endliches Erzeugendensystem und $L \subseteq V$ linear unabhängig. Dann gibt es eine Basis B von $(V, +, \cdot)$ mit

$$L \subseteq B \subseteq L \cup E.$$

Beweis. Nach Satz 4.64 ist L und damit auch $L \cup E$ endlich. Für jede linear unabhängige Menge B mit $L \subseteq B \subseteq L \cup E$ gilt $|B| \in \{|L|, \dots, |L \cup E|\} \subseteq \mathbb{N}$ und $B = L$ ist eine mögliche Wahl, daher können wir nach Satz 2.15 eine solche Menge B größter Kardinalität wählen. Es verbleibt zu zeigen, dass $\langle B \rangle = V$.

Wäre $\langle B \rangle \neq V$, so gäbe es einen Vektor $\boldsymbol{w} \in E \setminus \langle B \rangle$: Andernfalls wäre $E \subseteq \langle B \rangle$ und damit $V = \langle E \rangle \subseteq \langle B \rangle$. Dann ist $B \cup \{\boldsymbol{w}\}$ nach Korollar 4.53 linear unabhängig, im Widerspruch zur maximalen Kardinalität von B. □

Korollar 4.65 macht die folgende Definition unabhängig von der Wahl einer Basis.

Definition 4.67. Ein Vektorraum $(V, +, \cdot)$ über einem Körper K heißt *endlichdimensional*, wenn er eine endliche Basis B besitzt. Seine *Dimension* ist dann

$$\dim_K V \coloneqq |B|.$$

Ist der Körper festgelegt, schreiben wir auch $\dim V$. Ein Vektorraum ohne endliche Basis heißt *unendlichdimensional*.

Endlichdimensionalität und endliche Erzeugtheit sind nach Korollar 4.66 gleichbedeutend.

Beispiel 4.68. *Der Komponentenvergleich aus Beispiel 4.58 gilt über jedem Körper K: Die kanonischen Einheitsvektoren bilden für $n \in \mathbb{N} \setminus \{0\}$ eine Basis von $(K^n, +, \cdot)$. Daher ist*

$$\dim_K K^n = n.$$

△

Beispiel 4.69. *Für eine Gerade $G = \langle \boldsymbol{u} \rangle$ und eine Ebene $E = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$ durch den Ursprung in $\mathbb{R}^n$ gilt*

$$\dim_{\mathbb{R}} G = 1, \qquad \dim_{\mathbb{R}} E = 2.$$

Denn $\{\boldsymbol{u}\}$ beziehungsweise $\{\boldsymbol{u}, \boldsymbol{v}\}$ bilden Basen: Die Richtungsvektoren sind von null verschieden und bei einer Ebene keine Vielfachen voneinander, also linear unabhängig. △

Beispiel 4.70. *Für den Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.62 gilt $\dim_{\mathbb{R}} \mathcal{F} = 2$, denn $\{\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)}\}$ ist eine Basis. Die konstanten Funktionen bilden mit den eingeschränkten Operationen einen eindimensionalen Untervektorraum: Eine Basis ist $\{\boldsymbol{f}^{(0)}\}$.* △

Beispiel 4.71. *Die Basen aus Beispiel 4.63 liefern*

$$\dim_{\mathbb{R}} \mathbb{C} = 2, \qquad \dim_{\mathbb{C}} \mathbb{C} = 1.$$

Die Dimension hängt also auch vom Skalarkörper ab. △

Beispiel 4.72. *Der Vektorraum $(K[X], +, \cdot)$ ist unendlichdimensional. Besäße er ein endliches Erzeugendensystem aus n Vektoren, so dürfte nach Satz 4.64 jede linear unabhängige Menge höchstens n Elemente haben. Die $n + 1$ Monome $X^0, \ldots, X^n$ bilden aber nach Beispiel 4.60 eine linear unabhängige Menge. Also gibt es kein endliches Erzeugendensystem und damit auch keine endliche Basis.* △

Kennen wir die Dimension, genügt bei einer Menge der richtigen Größe bereits eine der beiden Eigenschaften, lineare Unabhängigkeit und Erzeugung, um Basis zu sein.

Satz 4.73. Sei $\dim V = n$ und $M \subseteq V$ mit $|M| = n$. Dann ist M genau dann eine Basis von $(V, +, \cdot)$, wenn M linear unabhängig ist, und genau dann, wenn $\langle M \rangle = V$ gilt.

Beweis. Übung. □

Satz 4.74. Jeder Untervektorraum $(U, +, \cdot)$ eines endlichdimensionalen Vektorraums $(V, +, \cdot)$ ist endlichdimensional, und es gilt

$$\dim U \leq \dim V.$$

Gleichheit gilt genau dann, wenn $U = V$ ist.

Beweis. Sei $n = \dim V$. Jede linear unabhängige Teilmenge von U hat nach Satz 4.64 höchstens n Elemente. Damit, und da die leere Menge linear unabhängig ist, können wir nach Satz 2.15 eine linear unabhängige Menge $L \subseteq U$ maximaler Größe wählen.

Wäre $\langle L \rangle \neq U$, so gäbe es einen Vektor $\boldsymbol{w} \in U \setminus \langle L \rangle$. Dann ist $L \cup \{\boldsymbol{w}\}$ nach Korollar 4.53 eine linear unabhängige Teilmenge von U mit mehr Elementen als L — ein Widerspruch zur maximalen Kardinalität von L. Also gilt $\langle L \rangle = U$, und L ist eine endliche Basis von $(U, +, \cdot)$ mit $\dim U = |L| \leq n$.

Ist $\dim U = n$, so ist die Basis L von U eine linear unabhängige Menge mit n Elementen und daher nach Satz 4.73 auch eine Basis von $(V, +, \cdot)$. Folglich gilt $U = \langle L \rangle = V$. Die umgekehrte Richtung ist unmittelbar. □

Für zwei Untervektorräume möchten wir nun die unabhängigen Richtungen zählen, die sie gemeinsam erzeugen. Dabei müssen wir die Richtungen im Schnitt nur einmal zählen. Dies führt zu einer Formel ähnlich zu Satz 2.27.

Satz 4.75 (Dimensionsformel für Untervektorräume)**.** Seien $(U, +, \cdot)$ und $(W, +, \cdot)$ Untervektorräume eines endlichdimensionalen Vektorraums $(V, +, \cdot)$. Für

$$U + W \coloneqq \{\boldsymbol{u} + \boldsymbol{w} \mid \boldsymbol{u} \in U,\ \boldsymbol{w} \in W\}$$

ist auch $(U + W, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$, und es gilt

$$\dim_K(U + W) = \dim_K U + \dim_K W - \dim_K(U \cap W).$$

Beweis. Jede Linearkombination von Vektoren aus $U \cup W$ lässt sich durch Zusammenfassen der Summanden als $\boldsymbol{u} + \boldsymbol{w}$ mit $\boldsymbol{u} \in U$ und $\boldsymbol{w} \in W$ schreiben. Umgekehrt ist jede solche Summe eine Linearkombination aus $U \cup W$. Damit gilt $U + W = \langle U \cup W \rangle$, also ist $U + W$ eine lineare Hülle und bildet nach Satz 4.38 einen Untervektorraum.

Nach Satz 4.74 sind die durch U, W und $U \cap W$ gebildeten Untervektorräume jeweils endlichdimensional. Wir wählen eine Basis B von $(U \cap W, +, \cdot)$ und ergänzen sie nach Korollar 4.66 zu Basen B_U von $(U, +, \cdot)$ und B_W von $(W, +, \cdot)$. Nach unserer anfänglichen Überlegung ist damit $\langle B_U \cup B_W \rangle = \langle U \cup W \rangle = U + W$, die Menge $B_U \cup B_W$ erzeugt also den Untervektorraum $(U + W, +, \cdot)$.

Zum Nachweis der Unabhängigkeit von $B_U \cup B_W$ betrachten wir eine Linearkombination, die $\mathbf{0}$ ergibt. Wir fassen die Beiträge aus B_U zu $\boldsymbol{u} \in U$ und die übrigen Beiträge aus $B_W \setminus B_U \subseteq B_W \setminus B$ zu $\boldsymbol{w} \in W$ zusammen, so dass $\boldsymbol{u} + \boldsymbol{w} = \mathbf{0}$. Aus $\boldsymbol{w} \in W$ und $\boldsymbol{w} = -\boldsymbol{u} \in U$ folgt $\boldsymbol{w} \in U \cap W$, daher besitzt $\boldsymbol{w}$ auch eine Darstellung allein durch $B \subseteq B_W$. In der Basis B_W haben wir somit zwei Darstellungen von $\boldsymbol{w}$: Eine verwendet nur B, die andere nur $B_W \setminus B$. Wegen der Eindeutigkeit nach Satz 4.55 sind alle Koeffizienten beider Darstellungen 0_K. Es bleibt $\boldsymbol{u} = \mathbf{0}$; die Unabhängigkeit von B_U liefert auch für die übrigen Koeffizienten 0_K.

Damit ist $B_U \cup B_W$ eine Basis von $(U + W, +, \cdot)$. Außerdem ist $B = B_U \cap B_W$: Nach Wahl der Basen ist $B \subseteq B_U \cap B_W$ und andersherum kann kein $\boldsymbol{v} \in (B_U \cap B_W) \setminus B$ existieren, da $\boldsymbol{v} \in (U \cap W) = \langle B \rangle$ von $B \subseteq B_U$ erzeugt wird und somit $B \cup \{\boldsymbol{v}\} \subseteq B_U$, also auch B_U linear abhängig wäre. Damit folgt

$$\begin{aligned} \dim(U + W) &= |B_U \cup B_W| \overset{2.27}{=} |B_U| + |B_W| - |B_U \cap B_W| \\ &= |B_U| + |B_W| - |B| = \dim U + \dim W - \dim(U \cap W). \end{aligned}$$ □

Für $U \cap W = \{\mathbf{0}\}$ addieren sich somit die Dimensionen, da der Schnitt die Dimension null hat.

Beispiel 4.76. *Für die Ebenen E_1 und E_2 aus Beispiel 4.39 kennen wir nach Beispiel 4.44 die Darstellungen*

$$E_1 = \left\langle \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle, \qquad E_2 = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\rangle, \qquad E_1 \cap E_2 = \left\langle \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right\rangle.$$

Die beiden Erzeuger der Ebenen sind jeweils linear unabhängig und der Erzeuger des Schnitts ist von null verschieden. Daher gilt $\dim E_1 = \dim E_2 = 2$ und $\dim(E_1 \cap E_2) = 1$. Die Dimensionsformel liefert

$$\dim(E_1 + E_2) = 2 + 2 - 1 = 3.$$

Nach Satz 4.74 folgt $E_1 + E_2 = \mathbb{R}^3$. Die beiden Ebenen erzeugen also gemeinsam den ganzen Raum. Mit anderen Worten: Jeder Vektor $\boldsymbol{v} \in \mathbb{R}^3$ lässt sich als die Summe eines Punkts auf E_1 und eines Punkts auf E_2 beschreiben. △

4.6 Skalarprodukte und Normen

In Abschnitt 4.1 haben wir Vektoren in $\mathbb{R}^n$ neben der Addition und der Multiplikation mit Skalaren auch mit dem Standardskalarprodukt und der euklidischen Länge ausgestattet. Mit dem Skalarprodukt konnten wir ausdrücken, wann zwei Vektoren orthogonal sind; mit der Länge konnten wir Abstände bestimmen.

Die Vektorraumaxiome allein legen weder ein Skalarprodukt noch einen Längenbegriff fest. Wir wollen Vektorräume jetzt um solche zusätzliche Struktur ergänzen und dafür weitere Eigenschaften fordern. So erhalten wir speziellere Strukturen, in denen wir auch für abstrakte Vektoren, etwa Funktionen, von Länge, Abstand und Orthogonalität sprechen können.

Im gesamten Abschnitt sei $(V, +, \cdot)$ ein Vektorraum über $\mathbb{R}$. Wir beginnen mit der Abstraktion des Skalarprodukts.

Definition 4.77. Eine Abbildung $(\cdot \mid \cdot)\colon V \times V \to \mathbb{R}$ heißt *Skalarprodukt* auf $(V, +, \cdot)$, falls sie für alle $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ und $\lambda, \mu \in \mathbb{R}$

(i) *symmetrisch* ist, also $(\boldsymbol{v} \mid \boldsymbol{w}) = (\boldsymbol{w} \mid \boldsymbol{v})$,

(ii) *bilinear* ist, also

$$(\lambda\boldsymbol{u} + \mu\boldsymbol{v} \mid \boldsymbol{w}) = \lambda(\boldsymbol{u} \mid \boldsymbol{w}) + \mu(\boldsymbol{v} \mid \boldsymbol{w}) \qquad \text{und} \qquad (\boldsymbol{u} \mid \lambda\boldsymbol{v} + \mu\boldsymbol{w}) = \lambda(\boldsymbol{u} \mid \boldsymbol{v}) + \mu(\boldsymbol{u} \mid \boldsymbol{w}),$$

(iii) *positiv definit* ist, also $(\boldsymbol{v} \mid \boldsymbol{v}) > 0$ für $\boldsymbol{v} \neq \boldsymbol{0}$.

Wegen der Symmetrie genügt es, eine der beiden Gleichungen für Bilinearität nachzuweisen; die andere folgt durch Vertauschen der Argumente.

Beobachtung 4.78. Für jedes Skalarprodukt gilt

$$(\boldsymbol{0} \mid \boldsymbol{v}) = (0\boldsymbol{0} \mid \boldsymbol{v}) = 0(\boldsymbol{0} \mid \boldsymbol{v}) = 0 = (\boldsymbol{v} \mid \boldsymbol{0}).$$

Damit ist $(\boldsymbol{v} \mid \boldsymbol{v}) \geq 0$ für alle $\boldsymbol{v} \in V$, mit Gleichheit genau für $\boldsymbol{v} = \boldsymbol{0}$.

Beispiel 4.79. *Das Standardskalarprodukt $(\boldsymbol{x} \mid \boldsymbol{y}) = \sum_{j=1}^{n} x_j y_j$ auf $(\mathbb{R}^n, +, \cdot)$ mit $n \in \mathbb{N} \setminus \{0\}$ aus Abschnitt 4.1 erfüllt nach Beobachtung 4.10 die drei Bedingungen aus Definition 4.77. Unsere Definition verallgemeinert also das Standardskalarprodukt.* △

Beispiel 4.80. *Auf $(\mathbb{R}^2, +, \cdot)$ können wir neben dem Standardskalarprodukt auch durch*

$$(\boldsymbol{x} \mid \boldsymbol{y})' := 2x_1y_1 + x_2y_2$$

ein Skalarprodukt definieren. Die Symmetrie folgt aus der Kommutativität der Multiplikation in $\mathbb{R}$, die Bilinearität durch Ausmultiplizieren: Es gilt

$$(\lambda\boldsymbol{x} + \mu\boldsymbol{y} \mid \boldsymbol{z})' = 2(\lambda x_1 + \mu y_1)z_1 + (\lambda x_2 + \mu y_2)z_2 = \lambda(\boldsymbol{x} \mid \boldsymbol{z})' + \mu(\boldsymbol{y} \mid \boldsymbol{z})'.$$

Für $\boldsymbol{x} \neq \boldsymbol{0}$ ist außerdem $(\boldsymbol{x} \mid \boldsymbol{x})' = 2x_1^2 + x_2^2 > 0$, da mindestens eine Komponente von null verschieden ist. △

Beispiel 4.81. *Der Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.62 besteht aus Funktionen $\boldsymbol{f}\colon \mathbb{R} \to \mathbb{R}$ mit $\boldsymbol{f}(t) = a+bt$ für $a, b \in \mathbb{R}$. Wir setzen für $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{F}$*

$$(\boldsymbol{f} \mid \boldsymbol{g}) \coloneqq \boldsymbol{f}(0)\boldsymbol{g}(0) + \boldsymbol{f}(1)\boldsymbol{g}(1).$$

Diese Abbildung ist symmetrisch, weil die Multiplikation in $\mathbb{R}$ kommutativ ist. Wegen der punktweisen Definition von Addition und Skalarmultiplikation gilt für $\lambda, \mu \in \mathbb{R}$ und $\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{F}$

$$\begin{aligned}(\lambda\boldsymbol{f} + \mu\boldsymbol{g} \mid \boldsymbol{h}) &= \sum_{j=0}^{1} (\lambda\boldsymbol{f}(j) + \mu\boldsymbol{g}(j))\boldsymbol{h}(j) \\ &= \lambda \sum_{j=0}^{1} \boldsymbol{f}(j)\boldsymbol{h}(j) + \mu \sum_{j=0}^{1} \boldsymbol{g}(j)\boldsymbol{h}(j) \\ &= \lambda(\boldsymbol{f} \mid \boldsymbol{h}) + \mu(\boldsymbol{g} \mid \boldsymbol{h}).\end{aligned}$$

Dies zeigt mit der Symmetrie die Bilinearität. Für $\boldsymbol{f}(t) = a + bt$ ist

$$(\boldsymbol{f} \mid \boldsymbol{f}) = a^2 + (a+b)^2 \geq 0.$$

Gleichheit gilt genau dann, wenn $a = 0$ und $a + b = 0$, also $a = b = 0$ gilt. In diesem Fall ist $\boldsymbol{f}$ die Nullfunktion. Damit ist die Abbildung positiv definit und ein Skalarprodukt. △

Wie in Abschnitt 4.1 können wir die Orthogonalität zweier Vektoren über ihr Skalarprodukt definieren.

Definition 4.82. Zwei Vektoren $\boldsymbol{v}, \boldsymbol{w} \in V$ heißen orthogonal bezüglich eines Skalarprodukts, wenn die Aussage

$$(\boldsymbol{v} \perp \boldsymbol{w}) \coloneqq ((\boldsymbol{v} \mid \boldsymbol{w}) = 0)$$

gilt. Insbesondere ist der Nullvektor zu jedem Vektor orthogonal.

Beispiel 4.83. *Das Standardskalarprodukt liefert eine andere Orthogonalitätsbedingung als das Skalarprodukt aus Beispiel 4.80: Für $\boldsymbol{u} = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ und $\boldsymbol{v} = \left(\begin{smallmatrix}1\\-1\end{smallmatrix}\right)$ gilt $(\boldsymbol{u} \mid \boldsymbol{v}) = 0$, aber $(\boldsymbol{u} \mid \boldsymbol{v})' = 1$. Der zugrunde liegende Vektorraum und damit seine Basen und seine Dimension bleiben dabei unverändert.* △

Beispiel 4.84. *Die Basisvektoren $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ aus Beispiel 4.62 erfüllen*

$$(\boldsymbol{f}^{(0)} \mid \boldsymbol{f}^{(1)}) = 1 \cdot 0 + 1 \cdot 1 = 1.$$

Die beiden linear unabhängigen Basisvektoren sind bezüglich des Skalarprodukts aus Beispiel 4.81 also nicht orthogonal. △

Wir folgern eine nützliche Ungleichung aus den Skalarproduktaxiomen.

Satz 4.85 (Cauchy–Schwarz-Ungleichung)**.** Sei $(\cdot \mid \cdot)$ ein Skalarprodukt auf $(V, +, \cdot)$. Für alle $\boldsymbol{v}, \boldsymbol{w} \in V$ gilt

$$(\boldsymbol{v} \mid \boldsymbol{w})^2 \leq (\boldsymbol{v} \mid \boldsymbol{v})(\boldsymbol{w} \mid \boldsymbol{w}).$$

Gleichheit gilt genau dann, wenn $\lambda \in \mathbb{R}$ existiert mit $\boldsymbol{v} = \lambda\boldsymbol{w}$ oder $\boldsymbol{w} = \lambda\boldsymbol{v}$.

Beweis. Für $\boldsymbol{w} = \mathbf{0}$ sind beide Seiten 0 (Beobachtung 4.78), und $\boldsymbol{w} = 0\boldsymbol{v}$ ist ein skalares Vielfaches von $\boldsymbol{v}$. Sei also $\boldsymbol{w} \neq \mathbf{0}$. Wegen positiver Definitheit ist $(\boldsymbol{w} \mid \boldsymbol{w}) > 0$ und wir können

$$\boldsymbol{r} \coloneqq \boldsymbol{v} - \frac{(\boldsymbol{v} \mid \boldsymbol{w})}{(\boldsymbol{w} \mid \boldsymbol{w})}\boldsymbol{w}$$

setzen. Mit Symmetrie und Bilinearität folgt

$$0 \leq (\boldsymbol{r} \mid \boldsymbol{r}) = (\boldsymbol{v} \mid \boldsymbol{v}) - 2\frac{(\boldsymbol{v} \mid \boldsymbol{w})^2}{(\boldsymbol{w} \mid \boldsymbol{w})} + \frac{(\boldsymbol{v} \mid \boldsymbol{w})^2}{(\boldsymbol{w} \mid \boldsymbol{w})} = (\boldsymbol{v} \mid \boldsymbol{v}) - \frac{(\boldsymbol{v} \mid \boldsymbol{w})^2}{(\boldsymbol{w} \mid \boldsymbol{w})}.$$

Multiplikation mit $(\boldsymbol{w} \mid \boldsymbol{w}) > 0$ liefert die Ungleichung. Gleichheit gilt genau für $(\boldsymbol{r} \mid \boldsymbol{r}) = 0$, also genau für $\boldsymbol{r} = \mathbf{0}$. Dann ist $\boldsymbol{v} = \lambda\boldsymbol{w}$ mit $\lambda = \frac{(\boldsymbol{v}|\boldsymbol{w})}{(\boldsymbol{w}|\boldsymbol{w})}$. Ist umgekehrt einer der beiden Vektoren ein skalares Vielfaches des anderen, so ergibt Einsetzen mit der Bilinearität unmittelbar die Gleichheit. □

Wir abstrahieren jetzt auch noch den Längenbegriff aus Abschnitt 4.1. Die folgende Definition umfasst nach den Sätzen 3.93 und 3.103 auch die Betragsfunktion auf $\mathbb{R}$ sowie auf $\mathbb{C}$, wenn wir $\mathbb{C}$ als reellen Vektorraum betrachten (Beispiel 4.25).

Definition 4.86. Eine Abbildung $\|\cdot\| \colon V \to \mathbb{R}_{\geq 0}$ heißt *Norm* auf $(V, +, \cdot)$, falls für alle $\boldsymbol{v}, \boldsymbol{w} \in V$ und $\lambda \in \mathbb{R}$ die folgenden Bedingungen gelten:

(i) Es gilt $\|\boldsymbol{v}\| = 0$ genau dann, wenn $\boldsymbol{v} = \mathbf{0}$ ist (*Definitheit*).

(ii) Es gilt $\|\lambda\boldsymbol{v}\| = |\lambda|\,\|\boldsymbol{v}\|$ (*absolute Homogenität*).

(iii) Es gilt $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ (Dreiecksungleichung).

Das Tupel $(V, +, \cdot, \|\cdot\|)$ heißt dann *normierter Vektorraum*.

In Abschnitt 4.1 haben wir die Länge über das Standardskalarprodukt definiert. Wir zeigen jetzt, dass wir auf diese Weise für jedes Skalarprodukt eine Norm erhalten können. In diesem Sinn induziert jeder Vektorraum mit Skalarprodukt auch einen normierten Raum. Umgekehrt benötigt eine Norm aber kein Skalarprodukt.

Satz 4.87. Jedes Skalarprodukt auf $(V, +, \cdot)$ definiert durch

$$\|\boldsymbol{v}\| \coloneqq \sqrt{(\boldsymbol{v} \mid \boldsymbol{v})}$$

eine Norm. Wir nennen sie die vom Skalarprodukt *induzierte Norm*.

Beweis. Es gilt $(\boldsymbol{v} \mid \boldsymbol{v}) \geq 0$, mit Gleichheit genau für $\boldsymbol{v} = \mathbf{0}$ (Beobachtung 4.78). Die Abbildung ist daher wohldefiniert und erfüllt die Definitheit. Aus der Bilinearität und Satz 3.93 folgt

$$\|\lambda\boldsymbol{v}\| = \sqrt{(\lambda\boldsymbol{v} \mid \lambda\boldsymbol{v})} = \sqrt{\lambda^2(\boldsymbol{v} \mid \boldsymbol{v})} = |\lambda|\,\|\boldsymbol{v}\|.$$

Für die Dreiecksungleichung multiplizieren wir aus, schätzen den gemischten Term ab und ziehen die Wurzel. Aus Cauchy–Schwarz (Satz 4.85) folgt zunächst

$$(\boldsymbol{v} \mid \boldsymbol{w}) \leq |(\boldsymbol{v} \mid \boldsymbol{w})| \leq \sqrt{(\boldsymbol{v} \mid \boldsymbol{v})(\boldsymbol{w} \mid \boldsymbol{w})} = \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|.$$

Damit ist

$$\begin{aligned}\|\boldsymbol{v}+\boldsymbol{w}\|^2 &= (\boldsymbol{v}+\boldsymbol{w} \mid \boldsymbol{v}+\boldsymbol{w}) = \|\boldsymbol{v}\|^2 + 2(\boldsymbol{v} \mid \boldsymbol{w}) + \|\boldsymbol{w}\|^2 \\ &\leq \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2 = (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2.\end{aligned}$$

Für nichtnegative reelle Zahlen a, b folgt aus $a^2 \leq b^2$ auch $a \leq b$: Wäre $a > b$, so wäre $a^2 - b^2 = (a-b)(a+b) > 0$. Mit $a = \|\boldsymbol{v}+\boldsymbol{w}\|$ und $b = \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ erhalten wir daher die Dreiecksungleichung. □

Die euklidische Länge aus Abschnitt 4.1 ist also eine Norm, die *euklidische Norm*. Der Satz des Pythagoras aus Abschnitt 4.1 überträgt sich auf abstrakte Vektorräume: Für orthogonale Vektoren und die vom Skalarprodukt induzierte Norm liefert Ausmultiplizieren (siehe Beweis von Satz 4.87) die folgende Identität.

Beobachtung 4.88. Sind $\boldsymbol{v}, \boldsymbol{w} \in V$ orthogonal, so gilt für die vom Skalarprodukt induzierte Norm

$$\|\boldsymbol{v}+\boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2.$$

Die Normaxiome lassen auch Längenbegriffe zu, die nicht von einem Skalarprodukt induziert werden.

Beispiel 4.89. *Auf* $(\mathbb{R}^n, +, \cdot)$ *mit* $n \in \mathbb{N} \setminus \{0\}$ *definieren wir die* Summennorm *und die* Maximumsnorm *durch*

$$\|\boldsymbol{x}\|_1 \coloneqq \sum_{j=1}^{n} |x_j| \qquad \textit{und} \qquad \|\boldsymbol{x}\|_\infty \coloneqq \max\{|x_j| \mid j \in \{1,\dots,n\}\}.$$

Beide Ausdrücke sind nichtnegativ und genau dann null, wenn alle Komponenten von $\boldsymbol{x}$ *null sind. Die absolute Homogenität folgt aus* $|\lambda x_j| = |\lambda|\,|x_j|$ *(Satz 3.93), indem wir den gemeinsamen nichtnegativen Faktor* $|\lambda|$ *aus der Summe bzw. dem Maximum ziehen. Die Dreiecksungleichung desselben Satzes liefert*

$$\|\boldsymbol{x}+\boldsymbol{y}\|_1 = \sum_{j=1}^{n} |x_j + y_j| \leq \sum_{j=1}^{n} (|x_j| + |y_j|) = \|\boldsymbol{x}\|_1 + \|\boldsymbol{y}\|_1.$$

Für die Maximumsnorm gilt für jedes $j \in \{1,\dots,n\}$

$$|x_j + y_j| \leq |x_j| + |y_j| \leq \|\boldsymbol{x}\|_\infty + \|\boldsymbol{y}\|_\infty.$$

Also erfüllt auch das Maximum der linken Seiten diese Schranke. Damit sind beide Abbildungen Normen.

Abbildung 4.2 zeigt für die drei Normen auf $\mathbb{R}^2$ *die jeweiligen* Einheitskreise*, also die Mengen aller Vektoren mit Norm* 1*. Insbesondere hat* $\boldsymbol{x} = \binom{1}{1}$ *die drei verschiedenen Normen*

$$\|\boldsymbol{x}\|_1 = 2, \qquad \|\boldsymbol{x}\|_\infty = 1, \qquad \|\boldsymbol{x}\| = \sqrt{2}.$$ △

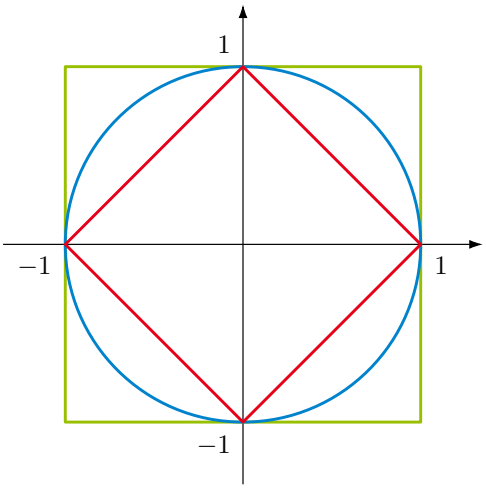


Abbildung 4.2: Einheitskreise der Summennorm (rot), der euklidischen Norm (blau) und der Maximumsnorm (grün) aus Beispiel 4.89.

Bemerkung 4.90. Für $n \geq 2$ werden die beiden Normen aus Beispiel 4.89 von keinem Skalarprodukt induziert. Denn jede induzierte Norm erfüllt durch Ausmultiplizieren wie im Beweis von Satz 4.87

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.$$

Für $\boldsymbol{x} = \boldsymbol{e}^{(1)}$ und $\boldsymbol{y} = \boldsymbol{e}^{(2)}$ ergibt die Summennorm links 8, die Maximumsnorm links 2, rechts dagegen jeweils 4.

Wie schon bei der euklidischen Länge erhalten wir aus einer Norm einen Abstand, indem wir die Differenz zweier Vektoren betrachten.

Definition 4.91. Für eine Norm auf $(V, +, \cdot)$ ist der *zugehörige Abstand*

$$d\colon V \times V \to \mathbb{R}_{\geq 0}, \qquad (\boldsymbol{v}, \boldsymbol{w}) \mapsto \|\boldsymbol{v}-\boldsymbol{w}\|.$$

Beobachtung 4.92. Für den zu einer Norm gehörigen Abstand und alle $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ gilt

$$\begin{aligned} d(\boldsymbol{v},\boldsymbol{w}) &= d(\boldsymbol{w},\boldsymbol{v}),\\ d(\boldsymbol{v},\boldsymbol{w}) = 0 &\Leftrightarrow \boldsymbol{v} = \boldsymbol{w},\\ d(\boldsymbol{u},\boldsymbol{w}) &\leq d(\boldsymbol{u},\boldsymbol{v}) + d(\boldsymbol{v},\boldsymbol{w}). \end{aligned}$$

Die letzte Ungleichung heißt ebenfalls *Dreiecksungleichung*.

Beweis. Aus der absoluten Homogenität folgt

$$d(\boldsymbol{v},\boldsymbol{w}) = \|\boldsymbol{v}-\boldsymbol{w}\| = \|-1(\boldsymbol{w}-\boldsymbol{v})\| = |-1|\,\|\boldsymbol{w}-\boldsymbol{v}\| = d(\boldsymbol{w},\boldsymbol{v}).$$

Die Definitheit der Norm liefert $d(\boldsymbol{v},\boldsymbol{w}) = 0$ genau für $\boldsymbol{v}-\boldsymbol{w} = \boldsymbol{0}$, also genau für $\boldsymbol{v} = \boldsymbol{w}$. Mit der Dreiecksungleichung der Norm erhalten wir schließlich

$$\begin{aligned} d(\boldsymbol{u},\boldsymbol{w}) &= \|\boldsymbol{u}-\boldsymbol{w}\| = \|(\boldsymbol{u}-\boldsymbol{v}) + (\boldsymbol{v}-\boldsymbol{w})\|\\ &\leq \|\boldsymbol{u}-\boldsymbol{v}\| + \|\boldsymbol{v}-\boldsymbol{w}\| = d(\boldsymbol{u},\boldsymbol{v}) + d(\boldsymbol{v},\boldsymbol{w}). \end{aligned}$$

□

Beispiel 4.93. *Wir betrachten auf dem Funktionenraum $(\mathcal{F}, +, \cdot)$ die vom Skalarprodukt aus Beispiel 4.81 induzierte Norm. Nach Definition 4.91 gilt für $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{F}$*

$$d(\boldsymbol{f}, \boldsymbol{g}) = \|\boldsymbol{f} - \boldsymbol{g}\| = \sqrt{(\boldsymbol{f}(0) - \boldsymbol{g}(0))^2 + (\boldsymbol{f}(1) - \boldsymbol{g}(1))^2}.$$

Der Abstand ergibt sich also aus den Unterschieden der Werte bei 0 und 1. Für die bekannten Basisvektoren $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ aus Beispiel 4.62 erhalten wir

$$d(\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)}) = \sqrt{(1-0)^2 + (1-1)^2} = 1. \qquad \triangle$$

4.7 Orthonormalbasen

In Abschnitt 4.1 haben wir die Länge eines Vektors als Wurzel der Summe seiner Koordinatenquadrate berechnet. Gilt diese Rechnung auch für die Koordinaten bezüglich einer beliebigen geordneten Basis aus Abschnitt 4.4? Bezüglich der geordneten Basis $\mathfrak{B} = \big((1,1)^\top, (1,0)^\top\big)$ hat etwa der Vektor $\boldsymbol{x} = (2,1)^\top$ die Koordinaten $(1,1)^\top$, denn $\boldsymbol{x} = 1 \cdot (1,1)^\top + 1 \cdot (1,0)^\top$. Seine Länge ist aber $\sqrt{5}$, während die Wurzel der Summe dieser Koordinatenquadrate $\sqrt{2}$ ergibt.

Die vertraute Längenberechnung beruht auf dem Satz von Pythagoras: Für senkrechte Richtungen addieren sich die Quadrate der Längen der Anteile (Beobachtung 4.88). Haben die Basisvektoren außerdem jeweils die Länge 1, so ist die Länge jedes Anteils gerade der Betrag seines Koeffizienten. Die Basisvektoren geben damit in jeder Richtung dieselbe Längeneinheit vor. Wir untersuchen nun Basen mit diesen beiden Eigenschaften und zeigen, wie wir sie aus einer gegebenen Basis gewinnen können.

Im gesamten Abschnitt sei $(V, +, \cdot)$ ein reeller Vektorraum mit einem Skalarprodukt $(\cdot \mid \cdot)$. Mit $\|\cdot\|$ bezeichnen wir die davon induzierte Norm (Satz 4.87).

Definition 4.94. Eine Menge $M \subseteq V$ heißt *orthogonal*, wenn $\boldsymbol{v} \perp \boldsymbol{w}$ für alle verschiedenen $\boldsymbol{v}, \boldsymbol{w} \in M$ gilt. Sie heißt *orthonormal*, wenn sie orthogonal ist und $\|\boldsymbol{v}\| = 1$ für alle $\boldsymbol{v} \in M$ gilt. Eine geordnete Basis heißt orthonormal, wenn ihre zugrunde liegende Menge orthonormal ist.

Beispiel 4.95. *Die Standardbasis aus Beispiel 4.58 ist eine orthonormale Basis bezüglich des Standardskalarprodukts: Für die kanonischen Einheitsvektoren gilt $(\boldsymbol{e}^{(i)} \mid \boldsymbol{e}^{(j)}) = 0$ für $i \neq j$ und $\|\boldsymbol{e}^{(i)}\| = 1$. Die Basis $\{\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)}\}$ des Funktionenraums $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.62 ist dagegen bezüglich des Auswertungsskalarprodukts aus Beispiel 4.81 nicht orthogonal, denn $(\boldsymbol{f}^{(0)} \mid \boldsymbol{f}^{(1)}) = 1$.* $\triangle$

Orthogonalität erleichtert nicht nur die Bestimmung von Koordinaten, sondern liefert bereits die lineare Unabhängigkeit, sofern der Nullvektor nicht vorkommt.

Satz 4.96. Jede orthogonale Menge $M \subseteq V \setminus \{\mathbf{0}\}$ ist linear unabhängig.

Beweis. Sei $m \in \mathbb{N} \setminus \{0\}$ und seien $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(m)} \in M$ paarweise verschieden und $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ mit $\sum_{j=1}^m \lambda_j \boldsymbol{v}^{(j)} = \mathbf{0}$. Für jedes $i \in \{1, \ldots, m\}$ folgt aus Bilinearität und Orthogonalität

$$0 = (\mathbf{0} \mid \boldsymbol{v}^{(i)}) = \left(\sum_{j=1}^m \lambda_j \boldsymbol{v}^{(j)} \,\middle|\, \boldsymbol{v}^{(i)} \right) = \sum_{j=1}^m \lambda_j (\boldsymbol{v}^{(j)} \mid \boldsymbol{v}^{(i)}) = \lambda_i (\boldsymbol{v}^{(i)} \mid \boldsymbol{v}^{(i)}).$$

Wegen $\boldsymbol{v}^{(i)} \neq \mathbf{0}$ ist $(\boldsymbol{v}^{(i)} \mid \boldsymbol{v}^{(i)}) > 0$ (positive Definitheit des Skalarprodukts), also $\lambda_i = 0$. Damit ist M linear unabhängig (Definition 4.47). □

Satz 4.97. Für jede geordnete, orthonormale Basis $\mathfrak{Q} = (\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(n)})$ von $(V, +, \cdot)$ mit $n \in \mathbb{N}$ und jeden Vektor $\boldsymbol{v} \in V$ gilt

$$\boldsymbol{v} = \sum_{j=1}^{n} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})\, \boldsymbol{q}^{(j)} \qquad \text{und} \qquad \|\boldsymbol{v}\|^2 = \sum_{j=1}^{n} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})^2.$$

Die Koordinaten von $\boldsymbol{v}$ bezüglich $\mathfrak{Q}$ sind insbesondere die Skalarprodukte mit den jeweiligen Basisvektoren.

Beweis. Für $n = 0$ gilt $V = \{\mathbf{0}\}$, und beide Aussagen folgen aus der Konvention für leere Summen (Definition 3.18). Sei also $n \geq 1$. Nach Satz 4.55 gibt es eindeutige Koeffizienten $x_1, \ldots, x_n \in \mathbb{R}$ mit $\boldsymbol{v} = \sum_{i=1}^{n} x_i \boldsymbol{q}^{(i)}$. Mit Bilinearität des Skalarprodukts und Orthogonalität der Basisvektoren folgt für $j \in \{1, \ldots, n\}$

$$(\boldsymbol{v} \mid \boldsymbol{q}^{(j)}) = \sum_{i=1}^{n} x_i (\boldsymbol{q}^{(i)} \mid \boldsymbol{q}^{(j)}) = x_j (\boldsymbol{q}^{(j)} \mid \boldsymbol{q}^{(j)}) = x_j$$

und damit

$$\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{q}^{(j)} = \sum_{j=1}^{n} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)}) \boldsymbol{q}^{(j)}.$$

Mit der Bilinearität folgt daraus weiter

$$\|\boldsymbol{v}\|^2 = (\boldsymbol{v} \mid \boldsymbol{v}) = \left(\boldsymbol{v} \,\middle|\, \sum_{j=1}^{n} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)}) \boldsymbol{q}^{(j)} \right) = \sum_{j=1}^{n} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})^2. \qquad \square$$

Bezüglich einer geordneten, orthonormalen Basis können wir die Koordinaten also durch Skalarprodukte bestimmen und die Norm wie bei der Standardbasis als die Wurzel der Summe der Quadrate der Koordinaten berechnen (siehe Definition 4.9).

Der folgende Satz liefert unmittelbar ein Verfahren, um eine gegebene Basis in eine orthonormale Basis umzurechnen. Dabei ergibt sich jeder zusätzliche Basisvektor aus dem Anteil eines gegebenen Basisvektors, der durch die bereits bestimmten Vektoren nicht erfasst wird.

Satz 4.98 (Gram–Schmidt-Verfahren)**.** Sei $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ eine geordnete Basis von $(V, +, \cdot)$ mit $n \in \mathbb{N}$ und $\mathfrak{Q} = (\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(n)})$ so, dass für alle $k \in \{1, \ldots, n\}$ gilt

$$\boldsymbol{q}^{(k)} = \frac{1}{\|\boldsymbol{u}^{(k)}\|} \boldsymbol{u}^{(k)} \qquad \text{für} \qquad \boldsymbol{u}^{(k)} := \boldsymbol{b}^{(k)} - \sum_{j=1}^{k-1} (\boldsymbol{b}^{(k)} \mid \boldsymbol{q}^{(j)})\, \boldsymbol{q}^{(j)}.$$

Dann ist $\boldsymbol{u}^{(k)} \neq \mathbf{0}$ für alle $k \in \{1, \ldots, n\}$ und $\mathfrak{Q}$ ist eine geordnete, orthonormale Basis von $(V, +, \cdot)$.

Beweis. Für $n = 0$ ist die leere Basis bereits orthonormal. Sei also $n \geq 1$. Wir zeigen per Induktion über $k \in \{1, \ldots, n\}$, dass $\boldsymbol{u}^{(k)} \neq \mathbf{0}$, dass $\{\boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(k)}\}$ orthonormal ist und dass $\langle \boldsymbol{q}^{(1)}, \ldots, \boldsymbol{q}^{(k)} \rangle = \langle \boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(k)} \rangle$. Für $k = n$ erhalten wir damit eine orthonormale Menge, die V erzeugt und nach Satz 4.96 linear unabhängig ist. Somit ist $\mathfrak{Q}$ eine geordnete Orthonormalbasis.

Für $k = 1$ ist die Aussage trivialerweise erfüllt, da $\boldsymbol{q}^{(1)} = \frac{1}{\|\boldsymbol{b}^{(1)}\|}\boldsymbol{b}^{(1)}$ mit $\boldsymbol{b}^{(1)} \neq \boldsymbol{0}$. Gelten die Behauptungen für ein $k \in \{1, \dots, n-1\}$, so ist $\boldsymbol{u}^{(k+1)} \neq \boldsymbol{0}$: Andernfalls läge $\boldsymbol{b}^{(k+1)}$ in $\langle \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)} \rangle = \langle \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(k)} \rangle$, im Widerspruch zur linearen Unabhängigkeit der Vektoren in $\mathfrak{B}$ (Satz 4.52). Wir dürfen also insbesondere durch $\|\boldsymbol{u}^{(k+1)}\| \neq 0$ teilen.

Nach der Induktionsvoraussetzung ist die Menge $\{\boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)}\}$ orthonormal. Für $i \in \{1, \dots, k\}$ gilt also

$$(\boldsymbol{u}^{(k+1)} \mid \boldsymbol{q}^{(i)}) = (\boldsymbol{b}^{(k+1)} \mid \boldsymbol{q}^{(i)}) - \sum_{j=1}^{k} (\boldsymbol{b}^{(k+1)} \mid \boldsymbol{q}^{(j)})(\boldsymbol{q}^{(j)} \mid \boldsymbol{q}^{(i)}) = (\boldsymbol{b}^{(k+1)} \mid \boldsymbol{q}^{(i)}) - (\boldsymbol{b}^{(k+1)} \mid \boldsymbol{q}^{(i)}) = 0.$$

Damit ist $\boldsymbol{q}^{(k+1)} \perp \boldsymbol{q}^{(i)}$ für alle $i \in \{1, \dots, k\}$ und insbesondere ist $\boldsymbol{q}^{(k+1)} \notin \{\boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)}\}$. Wegen der absoluten Homogenität der Norm gilt $\|\boldsymbol{q}^{(k+1)}\| = 1$. Schließlich ist

$$\boldsymbol{b}^{(k+1)} = \|\boldsymbol{u}^{(k+1)}\| \boldsymbol{q}^{(k+1)} + \sum_{j=1}^{k} (\boldsymbol{b}^{(k+1)} \mid \boldsymbol{q}^{(j)}) \boldsymbol{q}^{(j)} \in \langle \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k+1)} \rangle.$$

Nach Induktionsvoraussetzung ist $\langle \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(k)} \rangle = \langle \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)} \rangle$, also auch $\boldsymbol{b}^{(i)} \in \langle \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k+1)} \rangle$ für alle $i \in \{1, \dots, k\}$. Es folgt $\langle \boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(k+1)} \rangle \subseteq \langle \boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k+1)} \rangle$. Die linear unabhängige Menge $\{\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(k+1)}\}$ ist nach Definition 4.54 eine Basis ihrer linearen Hülle, die daher nach Definition 4.67 Dimension $k+1$ hat. Aus den $k+1$ erzeugenden Vektoren der rechten linearen Hülle lässt sich nach Korollar 4.66 mit $L = \emptyset$ eine Basis auswählen, sodass ihre Dimension höchstens $k+1$ ist. Zusammen mit der Inklusion folgt nach Satz 4.74 zunächst die Gleichheit der Dimensionen und damit die Gleichheit der linearen Hüllen. □

Insbesondere können wir Vektoren, die den Bedingungen aus Satz 4.98 genügen, der Reihe nach berechnen. Also haben wir konstruktiv die Existenz einer orthonormalen Basis im endlichdimensionalen Fall gezeigt.

Korollar 4.99. Jeder endlichdimensionale Vektorraum mit Skalarprodukt besitzt eine Orthonormalbasis.

Beispiel 4.100. *Wir betrachten die geordnete Basis $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)})$ von $(\mathbb{R}^2, +, \cdot)$ mit dem Standardskalarprodukt, wobei*

$$\boldsymbol{b}^{(1)} := \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \boldsymbol{b}^{(2)} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Wir können eine geordnete, orthonormale Basis $\mathfrak{Q} = (\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)})$ mit Hilfe des Gram-Schmidt-Verfahrens (Satz 4.98) bestimmen, indem wir die benötigten Vektoren der Reihe nach berechnen (siehe Abb. 4.3):

$$\begin{aligned} \boldsymbol{u}^{(1)} &= \boldsymbol{b}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \boldsymbol{q}^{(1)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \\ \boldsymbol{u}^{(2)} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, & \boldsymbol{q}^{(2)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{aligned}$$

Für den Punkt $\boldsymbol{x} = (2, 1)^\top$ gilt $\boldsymbol{x} = \boldsymbol{b}^{(1)} + \boldsymbol{b}^{(2)}$. Die Koordinaten bezüglich $\mathfrak{Q}$ berechnen wir mit den Skalarprodukten aus Satz 4.97 und erhalten insgesamt

$$[\boldsymbol{x}]_{\mathfrak{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad [\boldsymbol{x}]_{\mathfrak{Q}} = \begin{pmatrix} (\boldsymbol{x} \mid \boldsymbol{q}^{(1)}) \\ (\boldsymbol{x} \mid \boldsymbol{q}^{(2)}) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 3 \\ 1 \end{pmatrix}. \qquad \triangle$$

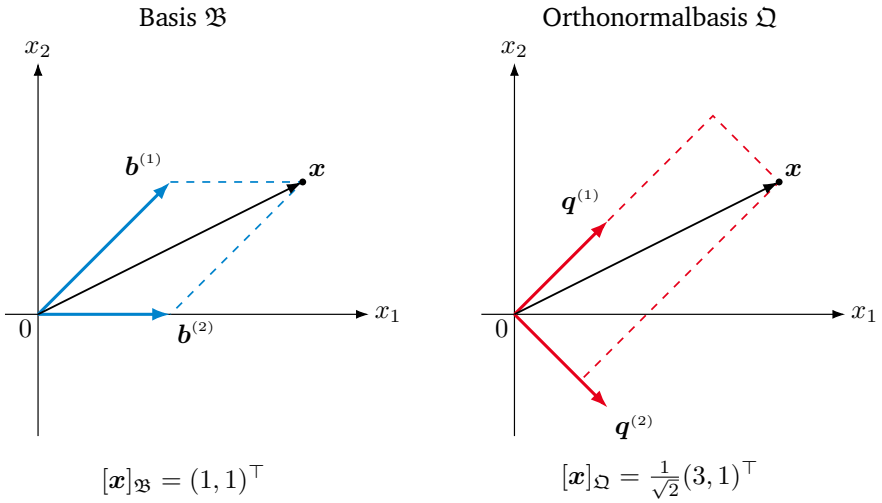


Abbildung 4.3: Die Basis $\mathfrak{B}$ (blau) und die durch Gram–Schmidt gewonnene Orthonormalbasis $\mathfrak{Q}$ (rot) aus Beispiel 4.100. Der Punkt $\boldsymbol{x} = (2,1)^\top$ ist derselbe; seine Koordinaten ändern sich mit der Basis. Gestrichelte Linien zeigen die Zerlegung in die beiden Basisrichtungen.

Beispiel 4.101. *Wir wenden Gram–Schmidt auf die geordnete Basis $(\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)})$ des Funktionenraums $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.62 an, mit dem Auswertungsskalarprodukt aus Beispiel 4.81. Wir behalten die Indizes 0 und 1 bei. Wegen $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ erhalten wir*

$$\|\boldsymbol{f}^{(0)}\| = \sqrt{2}, \qquad \boldsymbol{q}^{(0)}(t) = \frac{1}{\sqrt{2}},$$
$$(\boldsymbol{f}^{(1)} \mid \boldsymbol{q}^{(0)}) = \frac{1}{\sqrt{2}}, \qquad \boldsymbol{u}^{(1)}(t) = t - \frac{1}{2}.$$

Der Rest hat die Norm $\|\boldsymbol{u}^{(1)}\| = \sqrt{(-1/2)^2 + (1/2)^2} = 1/\sqrt{2}$, also ist $\boldsymbol{q}^{(1)}(t) = (2t-1)/\sqrt{2}$. Nach Satz 4.98 ist $\mathfrak{Q} = (\boldsymbol{q}^{(0)}, \boldsymbol{q}^{(1)})$ eine geordnete Orthonormalbasis. △

4.8 Projektionen

Beim Gram–Schmidt-Verfahren (Satz 4.98) haben wir von einem Vektor die Anteile in den bereits bestimmten orthonormalen Richtungen abgezogen. Der so erhaltene Basisvektor ist orthogonal zu den bisherigen Richtungen und erfasst den restlichen Anteil des Vektors, der außerhalb des bisher erzeugten Unterraums (also orthogonal dazu) liegt. Wir wollen jetzt systematischer untersuchen, inwiefern sich ein Vektorraum in zueinander orthogonale Unterräume zerlegen lässt.

Im gesamten Abschnitt sei $(V, +, \cdot)$ ein reeller Vektorraum mit einem Skalarprodukt $(\cdot \mid \cdot)$. Mit $\|\cdot\|$ bezeichnen wir die davon induzierte Norm.

Definition 4.102. Für eine Menge $U \subseteq V$ heißt

$$U^\perp := \{\boldsymbol{v} \in V \mid \forall \boldsymbol{u} \in U\colon (\boldsymbol{v} \mid \boldsymbol{u}) = 0\}$$

das *orthogonale Komplement* von U bezüglich des gewählten Skalarprodukts.

Tatsächlich bildet das orthogonale Komplement einen Untervektorraum.

Beobachtung 4.103. Für jede Menge $U \subseteq V$ ist $(U^\perp, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$. Ist $(U, +, \cdot)$ selbst ein Untervektorraum, so gilt $U \cap U^\perp = \{\mathbf{0}\}$.

Beweis. Nach Beobachtung 4.78 ist $\mathbf{0} \in U^\perp$. Für $\boldsymbol{v}, \boldsymbol{w} \in U^\perp$, $\lambda, \mu \in \mathbb{R}$ und jedes $\boldsymbol{u} \in U$ gilt

$$(\lambda \boldsymbol{v} + \mu \boldsymbol{w} \mid \boldsymbol{u}) = \lambda(\boldsymbol{v} \mid \boldsymbol{u}) + \mu(\boldsymbol{w} \mid \boldsymbol{u}) = 0.$$

Also ist $\lambda \boldsymbol{v} + \mu \boldsymbol{w} \in U^\perp$, und Satz 4.32 liefert die Untervektorraumeigenschaft. Ist auch $(U, +, \cdot)$ ein Untervektorraum, so enthält $U \cap U^\perp$ den Nullvektor. Für jeden Vektor $\boldsymbol{v} \in U \cap U^\perp$ gilt außerdem $(\boldsymbol{v} \mid \boldsymbol{v}) = 0$, also $\boldsymbol{v} = \mathbf{0}$. Daher ist $U \cap U^\perp = \{\mathbf{0}\}$. □

Wir zeigen jetzt, dass in einem endlichdimensionalen Vektorraum jeder Vektor eine eindeutige Zerlegung in die beiden gewünschten Anteile besitzt.

Satz 4.104. Sei $(U, +, \cdot)$ ein Untervektorraum des endlichdimensionalen Vektorraums $(V, +, \cdot)$. Jeder Vektor $\boldsymbol{v} \in V$ besitzt eine eindeutige Darstellung

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}, \qquad \boldsymbol{u} \in U, \quad \boldsymbol{w} \in U^\perp.$$

Für jede geordnete Orthonormalbasis $(\boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)})$ von $(U, +, \cdot)$ gilt

$$\boldsymbol{u} = \sum_{j=1}^{k} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})\, \boldsymbol{q}^{(j)}, \qquad \dim U^\perp = \dim V - \dim U.$$

Beweis. Nach Satz 4.74 ist $(U, +, \cdot)$ endlichdimensional. Mit Korollar 4.99 wählen wir eine geordnete Orthonormalbasis $(\boldsymbol{q}^{(1)}, \dots, \boldsymbol{q}^{(k)})$ von $(U, +, \cdot)$ und setzen

$$\boldsymbol{u} \coloneqq \sum_{j=1}^{k} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})\, \boldsymbol{q}^{(j)} \in U, \qquad \boldsymbol{w} \coloneqq \boldsymbol{v} - \boldsymbol{u}.$$

Für $i \in \{1, \dots, k\}$ gilt dann

$$(\boldsymbol{w} \mid \boldsymbol{q}^{(i)}) = (\boldsymbol{v} \mid \boldsymbol{q}^{(i)}) - (\boldsymbol{u} \mid \boldsymbol{q}^{(i)}) = (\boldsymbol{v} \mid \boldsymbol{q}^{(i)}) - \sum_{j=1}^{k} (\boldsymbol{v} \mid \boldsymbol{q}^{(j)})(\boldsymbol{q}^{(j)} \mid \boldsymbol{q}^{(i)}) = (\boldsymbol{v} \mid \boldsymbol{q}^{(i)}) - (\boldsymbol{v} \mid \boldsymbol{q}^{(i)}) = 0.$$

Wegen der Bilinearität ist $\boldsymbol{w}$ auch zu jeder Linearkombination der Basisvektoren orthogonal, also $\boldsymbol{w} \in U^\perp$.

Sind $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w} = \boldsymbol{u}' + \boldsymbol{w}'$ zwei solche Zerlegungen, so folgt $\boldsymbol{u} - \boldsymbol{u}' = \boldsymbol{w}' - \boldsymbol{w} \in U \cap U^\perp$. Nach Beobachtung 4.103 ist $U \cap U^\perp = \{\mathbf{0}\}$, also ist $\boldsymbol{u} = \boldsymbol{u}'$ und $\boldsymbol{w} = \boldsymbol{w}'$ und die Zerlegung eindeutig bestimmt.

Wegen $U + U^\perp = V$ ergibt die Summendimensionsformel aus Satz 4.75

$$\dim V = \dim U + \dim U^\perp - \dim\{\mathbf{0}\} = \dim U + \dim U^\perp.$$ □

Wir geben der Zuordnung, die jedem Vektor aus einem Vektorraum seinen nach Satz 4.104 eindeutigen Anteil aus einem Untervektorraum zuordnet, einen Namen.

Definition 4.105. Die *Orthogonalprojektion* eines endlichdimensionalen Vektorraums $(V, +, \cdot)$ auf einen Untervektorraum $(U, +, \cdot)$ ist die Abbildung $\pi_U \colon V \to V$ mit $\pi_U(\boldsymbol{v}) \coloneqq \boldsymbol{u}$, wobei $\boldsymbol{u} \in U$ der nach Satz 4.104 eindeutig bestimmte Anteil von $\boldsymbol{v}$ in der Zerlegung $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$ mit $\boldsymbol{w} \in U^\perp$ ist.

Beispiel 4.106. *Wir betrachten in $(\mathbb{R}^2, +, \cdot)$ mit dem Standardskalarprodukt die Gerade $G = \langle (1,1)^\top \rangle$ durch den Ursprung (siehe Abb. 4.4). Die Vektoren $(1,1)^\top$ und $(1,-1)^\top$ stehen nach Beispiel 4.12 senkrecht aufeinander. Mit der geordneten Orthonormalbasis aus Beispiel 4.100 gilt daher $G = \langle \boldsymbol{q}^{(1)} \rangle$ und $G^\perp = \langle \boldsymbol{q}^{(2)} \rangle$. Für $\boldsymbol{x} = (3,1)^\top$ erhalten wir nach Satz 4.104*

$$\pi_G(\boldsymbol{x}) = (\boldsymbol{x} \mid \boldsymbol{q}^{(1)})\, \boldsymbol{q}^{(1)} = \frac{3+1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \qquad \boldsymbol{x} - \pi_G(\boldsymbol{x}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

△

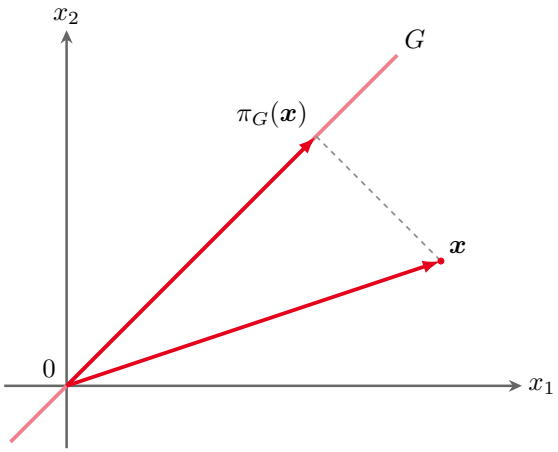


Abbildung 4.4: Orthogonalprojektion auf die Gerade G aus Beispiel 4.106: Der Vektor $\boldsymbol{x} = (3,1)^\top$ zerfällt in den Anteil $\pi_G(\boldsymbol{x}) = (2,2)^\top \in G$ und den senkrechten Rest $\boldsymbol{x} - \pi_G(\boldsymbol{x}) = (1,-1)^\top \in G^\perp$.

Beispiel 4.107. *Im Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.101 ist $\{\boldsymbol{q}^{(0)}\}$ mit $\boldsymbol{q}^{(0)}(t) = 1/\sqrt{2}$ eine Orthonormalbasis des Untervektorraums $(C, +, \cdot)$ der konstanten Funktionen. Nach Satz 4.104 gilt für $\boldsymbol{f} \in \mathcal{F}$*

$$\pi_C(\boldsymbol{f})(t) = (\boldsymbol{f} \mid \boldsymbol{q}^{(0)})\, \boldsymbol{q}^{(0)}(t) = \left(\frac{\boldsymbol{f}(0) + \boldsymbol{f}(1)}{\sqrt{2}} \right) \frac{1}{\sqrt{2}} = \frac{\boldsymbol{f}(0) + \boldsymbol{f}(1)}{2}.$$

Insbesondere wird $\boldsymbol{f}^{(1)}(t) = t$ auf die konstante Funktion mit Wert $1/2$ projiziert. Dies ist genau der Anteil, den wir beim Gram–Schmidt-Verfahren abgezogen haben. △

Beispiel 4.108. *Bei der Abstandsberechnung aus Satz 4.15 haben wir zu einem Punkt $\boldsymbol{x} \in \mathbb{R}^n$ einen nächstgelegenen Punkt in einer Hyperebene bestimmt (bezüglich des Standardskalarprodukts). Für eine Hyperebene $H = \{\boldsymbol{y} \in \mathbb{R}^n \mid (\boldsymbol{y} \mid \boldsymbol{a}) = 0\}$ durch den Ursprung mit $\boldsymbol{a} \neq \boldsymbol{0}$ ergibt die dortige Konstruktion*

$$\boldsymbol{p} = \boldsymbol{x} - \frac{(\boldsymbol{x} \mid \boldsymbol{a})}{\|\boldsymbol{a}\|^2} \boldsymbol{a} \in H, \qquad \boldsymbol{x} - \boldsymbol{p} = \frac{(\boldsymbol{x} \mid \boldsymbol{a})}{\|\boldsymbol{a}\|^2} \boldsymbol{a} \in H^\perp.$$

Damit ist $\boldsymbol{x} = \boldsymbol{p} + (\boldsymbol{x} - \boldsymbol{p})$ genau die eindeutige Zerlegung aus Satz 4.104. Nach Definition 4.105 gilt also $\boldsymbol{p} = \pi_H(\boldsymbol{x})$: Der dort bestimmte nächstgelegene Punkt ist die Orthogonalprojektion von $\boldsymbol{x}$ auf $(H, +, \cdot)$. △

4.9 Affine Räume

In Abschnitt 4.1 haben wir Geraden und Ebenen beschrieben, indem wir zu einem festen Stützvektor Vielfache von Richtungsvektoren addiert haben. Die Richtungsvektoren allein erzeugen durch ihre Linearkombinationen einen Untervektorraum; seine zugrunde liegende Menge ist ihre lineare Hülle (Definition 4.37 und Satz 4.42). Insbesondere bilden Geraden, Ebenen und Hyperebenen durch den Ursprung Untervektorräume (Beispiel 4.36).

Durch Addition eines Stützvektors verschieben wir diese Mengen. Die verschobene Menge enthält im Allgemeinen nicht mehr den Nullvektor und bildet dann keinen Vektorraum. Solche Verschiebungen erfassen wir nun mit dem Begriff des affinen Raums.

Im gesamten Abschnitt sei K ein Körper und $(V, +, \cdot)$ ein endlichdimensionaler Vektorraum über K.

Definition 4.109. Seien $(U, +, \cdot)$ ein Untervektorraum von $(V, +, \cdot)$ und $\boldsymbol{p} \in V$. Die Menge

$$\boldsymbol{p} + U \coloneqq \{\boldsymbol{p} + \boldsymbol{u} \mid \boldsymbol{u} \in U\}$$

heißt ein *affiner Raum* in $(V, +, \cdot)$. Wir nennen $\boldsymbol{p}$ einen *Stützvektor* und $(U, +, \cdot)$ einen *Richtungsvektorraum* dieses affinen Raums.

Für $\boldsymbol{p} = \boldsymbol{0}$ gilt $\boldsymbol{p} + U = U$. Durch das Verschieben muss die Menge aber weder unter Addition noch unter Skalarmultiplikation abgeschlossen bleiben.

Satz 4.110. Für jeden Punkt $\boldsymbol{q} \in A$ eines affinen Raums $A = \boldsymbol{p}+U$ in $(V, +, \cdot)$ gilt $A = \boldsymbol{q}+U$. Ist auch $A = \boldsymbol{r}+W$ mit $\boldsymbol{r} \in V$ und einem Untervektorraum $(W, +, \cdot)$ von $(V, +, \cdot)$, so gilt $U = W$. Der Richtungsvektorraum ist also eindeutig bestimmt.

Beweis. Sei $\boldsymbol{q} = \boldsymbol{p} + \boldsymbol{u}^{(0)}$ mit $\boldsymbol{u}^{(0)} \in U$. Wegen der Abgeschlossenheit unter Addition und Subtraktion ist $\{\boldsymbol{u}^{(0)} + \boldsymbol{u} \mid \boldsymbol{u} \in U\} = U$. Daher ist

$$\boldsymbol{q} + U = \{\boldsymbol{p} + \boldsymbol{u}^{(0)} + \boldsymbol{u} \mid \boldsymbol{u} \in U\} = \boldsymbol{p} + U = A.$$

Ist auch $A = \boldsymbol{r} + W$, so folgt wegen $\boldsymbol{p} \in A$ aus dem ersten Teil $A = \boldsymbol{p} + W$. Aus $\boldsymbol{p} + U = \boldsymbol{p} + W$ erhalten wir durch Subtraktion von $\boldsymbol{p}$ die Gleichheit $U = W$. □

Die Eindeutigkeit aus Satz 4.110 rechtfertigt die folgende Definition.

Definition 4.111. Die *Dimension* eines affinen Raums $A = \boldsymbol{p} + U$ ist

$$\dim_K A \coloneqq \dim_K U.$$

Beispiel 4.112. *Die Gerade durch zwei verschiedene Punkte $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ aus Satz 4.6 schreiben wir jetzt als*

$$G = \boldsymbol{p} + \langle \boldsymbol{q} - \boldsymbol{p} \rangle.$$

Ihr Richtungsvektorraum hat nach Beispiel 4.69 Dimension 1, also ist auch $\dim_{\mathbb{R}} G = 1$. Für drei nicht auf einer Geraden liegende Punkte $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}$ erhalten wir entsprechend

$$E = \boldsymbol{p} + \langle \boldsymbol{q} - \boldsymbol{p}, \boldsymbol{r} - \boldsymbol{p} \rangle, \qquad \dim_{\mathbb{R}} E = 2. \qquad \triangle$$

Beispiel 4.113. *Im Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.62 gilt für $c \in \mathbb{R}$*

$$A_c \coloneqq \{\boldsymbol{f} \in \mathcal{F} \mid \boldsymbol{f}(0) = c\} = c\boldsymbol{f}^{(0)} + \langle \boldsymbol{f}^{(1)} \rangle.$$

Denn für $\boldsymbol{f}(t) = a + bt$ ist $\boldsymbol{f}(0) = c$ genau dann, wenn $a = c$ gilt. Da $\{\boldsymbol{f}^{(1)}\}$ eine Basis des Richtungsvektorraums ist, hat A_c Dimension 1. Für $c \neq 0$ enthält dieser affine Raum die Nullfunktion nicht und bildet mit den punktweisen Operationen keinen Vektorraum. $\triangle$

Bei den Hyperebenen aus Definition 4.14 haben wir in Abschnitt 4.1 festgestellt, dass wir $n - 1$ Koordinaten frei wählen können. Mit dem orthogonalen Komplement aus Abschnitt 4.8 untermauern wir jetzt die Intuition, dass die Dimension einer Hyperebene genau dieser Anzahl an Freiheitsgraden entspricht.

Satz 4.114. Für $n \in \mathbb{N} \setminus \{0\}$ ist jede Hyperebene in $\mathbb{R}^n$ ein affiner Raum der Dimension $n - 1$.

Beweis. Seien $\boldsymbol{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ und $b \in \mathbb{R}$ mit $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid (\boldsymbol{x} \mid \boldsymbol{a}) = b\}$. Wir verwenden das Standardskalarprodukt und setzen

$$\boldsymbol{p} \coloneqq \frac{b}{\|\boldsymbol{a}\|^2}\boldsymbol{a}, \qquad U \coloneqq \langle \boldsymbol{a} \rangle^{\perp}.$$

Dann ist $(\boldsymbol{p} \mid \boldsymbol{a}) = b$. Wegen der Bilinearität ist ein Vektor genau dann zu $\boldsymbol{a}$ orthogonal, wenn er zu allen Vektoren aus $\langle \boldsymbol{a} \rangle$ orthogonal ist. Für $\boldsymbol{x} \in \mathbb{R}^n$ folgt daher

$$\boldsymbol{x} \in H \iff (\boldsymbol{x} - \boldsymbol{p} \mid \boldsymbol{a}) = 0 \iff \boldsymbol{x} - \boldsymbol{p} \in U \iff \boldsymbol{x} \in \boldsymbol{p} + U.$$

Also gilt $H = \boldsymbol{p} + U$. Nach Beobachtung 4.103 ist $(U, +, \cdot)$ ein Untervektorraum. Satz 4.104 liefert

$$\dim_{\mathbb{R}} U = n - \dim_{\mathbb{R}} \langle \boldsymbol{a} \rangle = n - 1,$$

da $\boldsymbol{a} \neq \mathbf{0}$ ist. Nach Definition 4.111 hat auch H Dimension $n - 1$. $\square$

5 Lineare Abbildungen

Analog zu den Gruppen-, Ring- und Körperhomomorphismen aus Kapitel 3 beschäftigen wir uns jetzt mit strukturerhaltenden Abbildungen zwischen Vektorräumen. Diese müssen also mit der Addition und Skalarmultiplikation verträglich sein. Wir werden sehen, dass sich solche Abbildungen auf einem endlichdimensionalen Vektorraum durch die Bilder einer Basis vollständig festlegen lassen. Deren Koordinaten führen uns zu Matrizen, mit denen wir Abbildungen beschreiben und rechnen können.

5.1 Lineare Abbildungen

Analog zu den Gruppen-, Ring- und Körperhomomorphismen (Definitionen 3.32, 3.55 und 3.71) definieren wir nun Vektorraumhomomorphismen, die auch lineare Abbildungen genannt werden.

Definition 5.1. Eine Abbildung $f\colon V \to W$ heißt *(Vektorraum-)Homomorphismus* oder *lineare Abbildung* zwischen den Vektorräumen $(V, +, \cdot)$ und $(W, \oplus, \odot)$ über demselben Körper K, falls für alle $\boldsymbol{v}, \boldsymbol{w} \in V$ und $\lambda \in K$ gilt, dass

$$f(\boldsymbol{v} + \boldsymbol{w}) = f(\boldsymbol{v}) \oplus f(\boldsymbol{w}) \qquad \text{und} \qquad f(\lambda \cdot \boldsymbol{v}) = \lambda \odot f(\boldsymbol{v}).$$

Ein bijektiver Vektorraumhomomorphismus heißt *(Vektorraum-)Isomorphismus*. Zwei Vektorräume heißen *isomorph*, wenn zwischen ihnen ein Vektorraumisomorphismus existiert. Ein Isomorphismus von V nach V heißt *(Vektorraum-)Automorphismus*.

Für eine lineare Abbildung $f\colon \mathbb{R} \to \mathbb{R}$ gilt insbesondere $f(x) = xf(1)$. Hier hat die Bezeichnung auch eine vertraute geometrische Bedeutung: Der Graph ist eine Gerade durch den Ursprung.

Im Folgenden schreiben wir die Verknüpfungen in beiden Vektorräumen wieder als $+$ und $\cdot$.

Beispiel 5.2. *Für einen Vektorraum $(V, +, \cdot)$ über K und $\alpha \in K$ ist die Streckung $s_\alpha\colon V \to V$, $\boldsymbol{v} \mapsto \alpha\boldsymbol{v}$ linear, für $\alpha \neq 0$ ist sie sogar ein Automorphismus.* △

Beispiel 5.3. *Für zwei Vektorräume $(V, +, \cdot)$ und $(W, +, \cdot)$ ist die Nullabbildung $f\colon V \to W$, $\boldsymbol{v} \mapsto \mathbf{0}_W$ linear:*

$$f(\boldsymbol{v} + \boldsymbol{v}') = \mathbf{0}_W = f(\boldsymbol{v}) + f(\boldsymbol{v}'), \qquad f(\lambda\boldsymbol{v}) = \mathbf{0}_W = \lambda f(\boldsymbol{v}).$$

Eine konstante Abbildung mit einem anderen Wert $\boldsymbol{w} \neq \mathbf{0}_W$ ist dagegen nicht linear: Für den Skalar 0 gilt $f(0\boldsymbol{v}) = \boldsymbol{w} \neq \mathbf{0}_W = 0f(\boldsymbol{v})$. △

Beispiel 5.4. *Im Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.29 hat jede Funktion die Form $\boldsymbol{f}(t) = a + bt$ mit eindeutig bestimmten $a, b \in \mathbb{R}$. Die Abbildung*

$$\kappa\colon \mathcal{F} \to \mathbb{R}^2, \qquad \boldsymbol{f} \mapsto (a, b)^\top$$

ordnet ihr den Koordinatenvektor bezüglich der bekannten Basis $(\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)})$ aus Beispiel 4.62 zu. Sie ist linear, weil bei Addition und Skalarmultiplikation der Funktionen die Koeffizienten komponentenweise addiert und skaliert werden. Außerdem gehört zu jedem $(a, b)^\top \in \mathbb{R}^2$ genau eine Funktion $t \mapsto a + bt$. Damit ist κ ein Isomorphismus zwischen $(\mathcal{F}, +, \cdot)$ und $(\mathbb{R}^2, +, \cdot)$. △

Beispiel 5.5. *Sei $(U, +, \cdot)$ ein Untervektorraum eines endlichdimensionalen reellen Vektorraums $(V, +, \cdot)$ mit Skalarprodukt. Die Orthogonalprojektion $\pi_U\colon V \to V$ aus Definition 4.105 ist linear. Nach Satz 4.104 besitzen $\boldsymbol{v}, \boldsymbol{v}' \in V$ eindeutige Zerlegungen*

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}, \qquad \boldsymbol{v}' = \boldsymbol{u}' + \boldsymbol{w}'$$

mit $\boldsymbol{u}, \boldsymbol{u}' \in U$ und $\boldsymbol{w}, \boldsymbol{w}' \in U^\perp$. Da $(U, +, \cdot)$ und $(U^\perp, +, \cdot)$ Untervektorräume sind, ist für $\alpha, \beta \in \mathbb{R}$ auch

$$\alpha\boldsymbol{v} + \beta\boldsymbol{v}' = (\alpha\boldsymbol{u} + \beta\boldsymbol{u}') + (\alpha\boldsymbol{w} + \beta\boldsymbol{w}')$$

eine Zerlegung in diese beiden Anteile. Wegen der Eindeutigkeit gilt daher

$$\pi_U(\alpha\boldsymbol{v} + \beta\boldsymbol{v}') = \alpha\boldsymbol{u} + \beta\boldsymbol{u}' = \alpha\pi_U(\boldsymbol{v}) + \beta\pi_U(\boldsymbol{v}').$$ △

Beispiel 5.6. *Eine Verschiebung $\tau_{\boldsymbol{a}}\colon V \to V$, $\boldsymbol{v} \mapsto \boldsymbol{v} + \boldsymbol{a}$ mit $\boldsymbol{a} \neq \boldsymbol{0}_V$ ist nicht linear: Für den Skalar 0 gilt $\tau_{\boldsymbol{a}}(0\boldsymbol{v}) = \boldsymbol{a} \neq \boldsymbol{0}_V = 0\tau_{\boldsymbol{a}}(\boldsymbol{v})$. Die Verschiebungen, mit denen wir in Abschnitt 4.9 affine Räume beschrieben haben, erhalten also im Allgemeinen die Vektorraumstruktur nicht.* △

Beispiel 5.7. *Wie bei der Dimension kommt es auch bei der Linearität darauf an, welchen Skalarkörper wir betrachten. Wir fassen $(\mathbb{C}, +, \cdot)$ wie in Beispiel 4.25 einmal als reellen und einmal als komplexen Vektorraum auf. Die Konjugation $c\colon \mathbb{C} \to \mathbb{C}$, $z \mapsto \overline{z}$ ist als Abbildung zwischen reellen Vektorräumen linear: Nach Satz 3.103 gilt für $z, w \in \mathbb{C}$ und $\lambda \in \mathbb{R}$*

$$\overline{z + w} = \overline{z} + \overline{w}, \qquad \overline{\lambda z} = \overline{\lambda}\,\overline{z} = \lambda\overline{z}.$$

Als Abbildung zwischen komplexen Vektorräumen ist sie aber nicht linear, denn

$$c(\mathrm{i} \cdot 1) = -\mathrm{i} \neq \mathrm{i} = \mathrm{i} \cdot c(1).$$

Dagegen ist für jedes feste $z \in \mathbb{C}$ die Multiplikation $m_z\colon \mathbb{C} \to \mathbb{C}$, $w \mapsto zw$ sowohl bei Betrachtung als Abbildung zwischen komplexen als auch zwischen reellen Vektorräumen linear. △

Die additive Bedingung aus Definition 5.1 verlangt, wie bereits bei Ring- und Körperhomomorphismen (Definitionen 3.55 und 3.71), einen Gruppenhomomorphismus; hinzu kommt die Verträglichkeit mit der Skalarmultiplikation. Die beiden Bedingungen zusammen bedeuten, dass Linearkombinationen zweier Vektoren erhalten bleiben:

$$f(\lambda\boldsymbol{v} + \mu\boldsymbol{w}) = \lambda f(\boldsymbol{v}) + \mu f(\boldsymbol{w}) \qquad \text{für alle } \boldsymbol{v}, \boldsymbol{w} \in V \text{ und } \lambda, \mu \in K.$$

Ein Vektorraumhomomorphismus erhält damit auch beliebige endliche Linearkombinationen, was die Bezeichnung als *lineare Abbildung* begründet. Die Regeln für den Nullvektor und additive Inverse kennen wir bereits von Gruppenhomomorphismen (Satz 3.38).

Satz 5.8. Für jede lineare Abbildung $f\colon V \to W$ zwischen Vektorräumen über demselben Körper K gilt

$$f(\mathbf{0}_V) = \mathbf{0}_W, \qquad f(-\boldsymbol{v}) = -f(\boldsymbol{v}) \qquad \text{für alle } \boldsymbol{v} \in V.$$

Außerdem gilt für alle $m \in \mathbb{N}$, $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(m)} \in V$ und $\lambda_1, \dots, \lambda_m \in K$

$$f\left(\sum_{j=1}^{m} \lambda_j \boldsymbol{v}^{(j)}\right) = \sum_{j=1}^{m} \lambda_j f(\boldsymbol{v}^{(j)}).$$

Umgekehrt ist jede Abbildung $f\colon V \to W$, die alle endlichen Linearkombinationen erhält, linear.

Beweis. Die ersten beiden Gleichheiten folgen aus Satz 3.38. Die Behauptung über Linearkombinationen zeigen wir durch Induktion über m. Für $m = 0$ ist sie wegen der Konvention für leere Summen (Definition 3.18) genau die Gleichheit $f(\mathbf{0}_V) = \mathbf{0}_W$. Gilt die Behauptung für m, so folgt

$$\begin{aligned} f\left(\sum_{j=1}^{m+1} \lambda_j \boldsymbol{v}^{(j)}\right) &= f\left(\sum_{j=1}^{m} \lambda_j \boldsymbol{v}^{(j)}\right) + f(\lambda_{m+1} \boldsymbol{v}^{(m+1)}) \\ &= \sum_{j=1}^{m} \lambda_j f(\boldsymbol{v}^{(j)}) + \lambda_{m+1} f(\boldsymbol{v}^{(m+1)}) = \sum_{j=1}^{m+1} \lambda_j f(\boldsymbol{v}^{(j)}). \end{aligned}$$

Umgekehrt erhalten wir aus der Erhaltung von Linearkombinationen mit $m = 1$ die Verträglichkeit mit der Skalarmultiplikation und mit $m = 2$, $\lambda_1 = \lambda_2 = 1$ die Verträglichkeit mit der Addition. □

Wie viele Bilder müssen wir kennen, um eine lineare Abbildung vollständig zu beschreiben? Besitzt der Ausgangsraum eine endliche Basis, so genügt es, die Bilder ihrer Vektoren zu kennen: Jeder weitere Vektor ist eine Linearkombination der Basisvektoren, deren Bild nach Satz 5.8 damit festliegt. Umgekehrt können wir die Basisbilder frei vorgeben.

Satz 5.9. Seien $(V, +, \cdot)$ und $(W, +, \cdot)$ Vektorräume über K, $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(n)})$ eine geordnete Basis von $(V, +, \cdot)$ mit $n \in \mathbb{N}$ und $\boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(n)} \in W$. Dann existiert genau eine lineare Abbildung $f\colon V \to W$ mit $f(\boldsymbol{b}^{(j)}) = \boldsymbol{w}^{(j)}$ für alle $j \in \{1, \dots, n\}$. Für diese gilt

$$f\left(\sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}\right) = \sum_{j=1}^{n} x_j \boldsymbol{w}^{(j)} \qquad \text{für alle } x_1, \dots, x_n \in K.$$

Beweis. Für $n = 0$ ist $V = \{\mathbf{0}_V\}$. Die Nullabbildung ist dann die einzige lineare Abbildung $V \to W$, und die angegebene Formel gilt wegen der Konvention für leere Summen. Sei im Folgenden $n \geq 1$. Nach Satz 4.55 besitzt jeder Vektor $\boldsymbol{v} \in V$ eine eindeutige Darstellung $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$. Damit ist

$$f\colon V \to W, \qquad \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)} \mapsto \sum_{j=1}^{n} x_j \boldsymbol{w}^{(j)}$$

wohldefiniert. Die Darstellung von $\boldsymbol{b}^{(j)}$ hat an Stelle j den Koeffizienten 1 und sonst nur Nullen, also gilt $f(\boldsymbol{b}^{(j)}) = \boldsymbol{w}^{(j)}$.

Für $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$, $\boldsymbol{v}' = \sum_{j=1}^{n} y_j \boldsymbol{b}^{(j)}$ und $\lambda \in K$ folgt aus den Vektorraumaxiomen

$$f(\boldsymbol{v} + \boldsymbol{v}') = f\left(\sum_{j=1}^{n} (x_j + y_j)\boldsymbol{b}^{(j)}\right) = \sum_{j=1}^{n} (x_j + y_j)\boldsymbol{w}^{(j)} = f(\boldsymbol{v}) + f(\boldsymbol{v}'),$$
$$f(\lambda \boldsymbol{v}) = f\left(\sum_{j=1}^{n} (\lambda x_j)\boldsymbol{b}^{(j)}\right) = \sum_{j=1}^{n} (\lambda x_j)\boldsymbol{w}^{(j)} = \lambda f(\boldsymbol{v}).$$

Also ist f linear.

Ist $g\colon V \to W$ eine weitere lineare Abbildung mit denselben Basisbildern, so gilt nach Satz 5.8 für jeden Vektor $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$

$$g(\boldsymbol{v}) = \sum_{j=1}^{n} x_j g(\boldsymbol{b}^{(j)}) = \sum_{j=1}^{n} x_j \boldsymbol{w}^{(j)} = f(\boldsymbol{v}).$$ □

Bemerkung 5.10. Die im vorigen Satz vorgegebenen Basisbilder müssen weder linear unabhängig noch paarweise verschieden noch von null verschieden sein. Betrachten wir etwa die Orthogonalprojektion π_G auf die Gerade $G = \langle (1,1)^\top \rangle$ aus Beispiel 4.106. Die beiden Standardbasisvektoren haben dasselbe Bild:

$$\pi_G\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \pi_G\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}.$$

Wegen der Linearität wird deshalb die Differenz der beiden Ausgangsvektoren auf null abgebildet:

$$\pi_G\left(\begin{pmatrix}1\\-1\end{pmatrix}\right) = \pi_G\left(\begin{pmatrix}1\\0\end{pmatrix}\right) - \pi_G\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \mathbf{0}.$$

Obwohl $(1,-1)^\top$ vom Nullvektor verschieden ist, haben beide dasselbe Bild. Eine lineare Abbildung ist also nicht notwendig injektiv.

Nachdem wir gesehen haben, wie sich lineare Abbildungen durch Basisbilder festlegen lassen, betrachten wir das Rechnen mit solchen Abbildungen. Wir können sie punktweise addieren und skalieren, aber auch verketten und im bijektiven Fall umkehren.

Beispiel 5.11. *Für zwei Vektorräume $(V,+,\cdot)$ und $(W,+,\cdot)$ über demselben Körper K bezeichnen wir die Menge ihrer linearen Abbildungen mit $\operatorname{Hom}_K(V,W)$. Mit punktweiser Addition und Skalarmultiplikation, also*

$$(f+g)(\boldsymbol{v}) \coloneqq f(\boldsymbol{v}) + g(\boldsymbol{v}), \qquad (\lambda f)(\boldsymbol{v}) \coloneqq \lambda f(\boldsymbol{v}),$$

ist $(\operatorname{Hom}_K(V,W),+,\cdot)$ selbst ein Vektorraum über K: Die Nullabbildung ist linear. Summen und skalare Vielfache linearer Abbildungen sind wieder linear. Die Linearität der Summe folgt für lineare Abbildungen $f, g\colon V \to W$, $\boldsymbol{v}, \boldsymbol{w} \in V$ und $\alpha, \beta \in K$ aus

$$\begin{aligned}(f+g)(\alpha\boldsymbol{v} + \beta\boldsymbol{w}) &= f(\alpha\boldsymbol{v} + \beta\boldsymbol{w}) + g(\alpha\boldsymbol{v} + \beta\boldsymbol{w})\\ &= \alpha f(\boldsymbol{v}) + \beta f(\boldsymbol{w}) + \alpha g(\boldsymbol{v}) + \beta g(\boldsymbol{w})\\ &= \alpha(f+g)(\boldsymbol{v}) + \beta(f+g)(\boldsymbol{w}).\end{aligned}$$

Für $\lambda \in K$ ist ebenso λf linear, denn

$$(\lambda f)(\alpha\boldsymbol{v} + \beta\boldsymbol{w}) = \lambda(\alpha f(\boldsymbol{v}) + \beta f(\boldsymbol{w})) = \alpha(\lambda f)(\boldsymbol{v}) + \beta(\lambda f)(\boldsymbol{w}).$$

Die Vektorraumaxiome gelten punktweise wie in Beispiel 4.28. △

Beispiel 5.12. *Die Projektion $\pi_G \colon \mathbb{R}^2 \to \mathbb{R}^2$ auf die Gerade $G = \langle (1,1)^\top \rangle$ ist nach Beispiel 5.5 linear. Nach Beispiel 5.11 ist daher auch die Abbildung*

$$s_G \coloneqq 2\pi_G - \mathrm{id}_{\mathbb{R}^2}$$

linear. Für die orthogonale Zerlegung $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$ mit $\boldsymbol{u} \in G$ und $\boldsymbol{w} \in G^\perp$ gilt

$$s_G(\boldsymbol{v}) = 2\pi_G(\boldsymbol{v}) - \boldsymbol{v} = 2\boldsymbol{u} - (\boldsymbol{u} + \boldsymbol{w}) = \boldsymbol{u} - \boldsymbol{w}.$$

Die Abbildung behält also den Anteil in G und kehrt das Vorzeichen des senkrechten Anteils um: Sie ist die Spiegelung an G. Zweimaliges Spiegeln liefert den ursprünglichen Vektor zurück; daher ist s_G sogar ein Automorphismus. △

Wir können lineare Abbildungen auch verketten und bijektive lineare Abbildungen umkehren. Wieder lassen sich die Ergebnisse für Gruppenhomomorphismen unmittelbar ergänzen.

Beobachtung 5.13. Sind $f \colon V \to W$ und $g \colon W \to X$ lineare Abbildungen zwischen Vektorräumen über demselben Körper, so ist auch $g \circ f \colon V \to X$ linear. Ist f ein Isomorphismus, so auch $f^{-1} \colon W \to V$.

Beweis. Die Verträglichkeit mit der Addition folgt aus den Beispielen 3.34 und 3.35. Für lineare Abbildungen $f \colon V \to W$ und $g \colon W \to X$ zwischen Vektorräumen über demselben Körper K gilt für alle $\lambda \in K$ und $\boldsymbol{v} \in V$ außerdem

$$(g \circ f)(\lambda \boldsymbol{v}) = g(\lambda f(\boldsymbol{v})) = \lambda (g \circ f)(\boldsymbol{v}).$$

Ist f ein Isomorphismus, so ist seine Umkehrabbildung bijektiv. Wir schreiben jedes $\boldsymbol{w} \in W$ als $\boldsymbol{w} = f(\boldsymbol{v})$ und erhalten

$$f^{-1}(\lambda \boldsymbol{w}) = f^{-1}(f(\lambda \boldsymbol{v})) = \lambda \boldsymbol{v} = \lambda f^{-1}(\boldsymbol{w}).$$ □

5.2 Kern, Bild und Rang

In Abschnitt 5.1 haben wir untersucht, wie sich lineare Abbildungen durch ihre Basisbilder beschreiben lassen. Nun betrachten wir die umgekehrte Richtung: Zu einem vorgegebenen Vektor $\boldsymbol{y}$ suchen wir alle Vektoren, die auf $\boldsymbol{y}$ abgebildet werden. Wir untersuchen, wann solche Vektoren existieren, wann sie eindeutig bestimmt sind und wie wir die Menge aller Vektoren mit Bild $\boldsymbol{y}$ beschreiben können.

Bei der Orthogonalprojektion π_G aus Bemerkung 5.10 haben die beiden Standardbasisvektoren dasselbe Bild. Ihre Differenz wird deshalb auf null abgebildet. Um zu verstehen, welche Vektoren dasselbe Bild haben, betrachten wir daher zunächst diejenigen Vektoren, die auf null abgebildet werden. Diese bilden den bereits von Gruppenhomomorphismen bekannten Kern (Definition 3.40). Für eine lineare Abbildung $f \colon V \to W$ ist dieser gegeben durch

$$\ker(f) = \{\boldsymbol{v} \in V \mid f(\boldsymbol{v}) = \boldsymbol{0}_W\}.$$

Das Bild $f(V)$ beschreibt dagegen, für welche Vektoren $\boldsymbol{y}$ überhaupt ein Urbild existiert. Wir zeigen zunächst, dass Kern und Bild Untervektorräume bilden.

Satz 5.14. Für eine lineare Abbildung $f\colon V \to W$ zwischen den Vektorräumen $(V, +, \cdot)$ und $(W, \oplus, \odot)$ bilden $\ker(f)$ und $f(V)$ Untervektorräume von $(V, +, \cdot)$ beziehungsweise $(W, \oplus, \odot)$. Ist $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ eine geordnete Basis von $(V, +, \cdot)$ mit $n \in \mathbb{N}$, so gilt

$$f(V) = \langle f(\boldsymbol{b}^{(1)}), \ldots, f(\boldsymbol{b}^{(n)}) \rangle.$$

Insbesondere bildet $f(V)$ dann einen endlichdimensionalen Vektorraum.

Beweis. Sei K der gemeinsame zugrunde liegende Körper. Da f insbesondere ein Gruppenhomomorphismus zwischen $(V, +)$ und $(W, \oplus)$ ist, folgen die Untergruppeneigenschaften aus den Sätzen 3.41 und 3.38 (iii). Nach Satz 4.32 bleibt die Abgeschlossenheit unter Skalarmultiplikation zu zeigen. Für $\lambda \in K$ und $\boldsymbol{v} \in \ker(f)$ gilt wegen der Linearität $f(\lambda \cdot \boldsymbol{v}) = \lambda \odot f(\boldsymbol{v}) = \mathbf{0}_W$, also $\lambda \cdot \boldsymbol{v} \in \ker(f)$. Ist $\boldsymbol{w} = f(\boldsymbol{v}) \in f(V)$, so ist ebenfalls wegen der Linearität $\lambda \odot \boldsymbol{w} = f(\lambda \cdot \boldsymbol{v}) \in f(V)$.

Nach Satz 4.55 lässt sich jeder Vektor $\boldsymbol{v} \in V$ als $\boldsymbol{v} = \sum_{j=1}^{n} x_j \cdot \boldsymbol{b}^{(j)}$ schreiben. Aus Satz 5.8 folgt daher

$$f(\boldsymbol{v}) = \sum_{j=1}^{n} x_j \odot f(\boldsymbol{b}^{(j)}).$$

Damit liegt jedes Bild in der angegebenen linearen Hülle. Umgekehrt gehört jede Linearkombination der Basisbilder zum Bild, denn dieses bildet einen Untervektorraum von $(W, \oplus, \odot)$. Die endliche Erzeugtheit liefert nach Korollar 4.66 die Endlichdimensionalität. □

Die Bilder der Basisvektoren erzeugen also das Bild. Dabei können verschiedene Basisvektoren dasselbe Bild haben, wie Bemerkung 5.10 zeigt. Die Dimension des Bildvektorraums gibt an, wie viele linear unabhängige Vektoren wir im Bild höchstens finden können.

Definition 5.15. Der *Rang* einer linearen Abbildung $f\colon V \to W$ von einem endlichdimensionalen Vektorraum V in einen beliebigen Vektorraum W ist

$$\operatorname{rang}(f) \coloneqq \dim f(V).$$

Beispiel 5.16. *Für die Orthogonalprojektion $\pi_G\colon \mathbb{R}^2 \to \mathbb{R}^2$ auf die Gerade $G = \langle (1,1)^\top \rangle$ aus Beispiel 4.106 gilt $\pi_G((x_1, x_2)^\top) = \frac{x_1+x_2}{2}(1,1)^\top$. Ein Vektor wird genau dann auf null abgebildet, wenn $x_1 + x_2 = 0$ ist. Also ist*

$$\ker(\pi_G) = \langle (1,-1)^\top \rangle \overset{4.106}{=} G^\perp.$$

Alle Bilder liegen in G, und jeder Vektor aus G ist sein eigenes Bild. Daher gilt $\pi_G(\mathbb{R}^2) = G$. Kern und Bild bilden somit jeweils einen eindimensionalen Vektorraum; insbesondere ist $\operatorname{rang}(\pi_G) = 1$. △

Bei der gerade betrachteten Projektion π_G haben die beiden Standardbasisvektoren dasselbe Bild (Bemerkung 5.10). Ihre Differenz wird auf null abgebildet und liegt somit in $\ker(\pi_G)$. Allgemein gilt: Zwei Vektoren haben genau dann dasselbe Bild, wenn ihre Differenz im Kern liegt.

Beobachtung 5.17. Für jede lineare Abbildung $f\colon V \to W$ und alle $\boldsymbol{v}, \boldsymbol{w} \in V$ gilt

$$f(\boldsymbol{v}) = f(\boldsymbol{w}) \iff \boldsymbol{v} - \boldsymbol{w} \in \ker(f).$$

Insbesondere ist f genau dann injektiv, wenn $\ker(f) = \{\mathbf{0}_V\}$ ist.

Beweis. Wegen der Linearität gilt

$$f(\boldsymbol{v}) = f(\boldsymbol{w}) \iff f(\boldsymbol{v}) - f(\boldsymbol{w}) = \mathbf{0}_W \iff f(\boldsymbol{v} - \boldsymbol{w}) = \mathbf{0}_W.$$

Das ist nach Definition gleichbedeutend mit $\boldsymbol{v} - \boldsymbol{w} \in \ker(f)$. Die Aussage zur Injektivität ist das bekannte Kernkriterium aus Satz 3.42 für die zugrunde liegenden additiven Gruppen. □

Beobachtung 5.17 beschreibt, wann zwei Vektoren dasselbe Bild haben. Halten wir nun einen Bildvektor $\boldsymbol{y}$ fest, können wir damit alle Vektoren bestimmen, die auf $\boldsymbol{y}$ abgebildet werden. Kennen wir einen solchen Vektor $\boldsymbol{v}^{(0)}$, so erhalten wir alle weiteren, indem wir die Vektoren aus dem Kern hinzuaddieren.

Satz 5.18. Sei $f\colon V \to W$ eine lineare Abbildung und $\boldsymbol{v}^{(0)} \in V$, $\boldsymbol{y} \in W$ mit $\boldsymbol{y} = f(\boldsymbol{v}^{(0)})$. Dann gilt

$$f^{-1}(\{\boldsymbol{y}\}) = \boldsymbol{v}^{(0)} + \ker(f) \coloneqq \{\boldsymbol{v}^{(0)} + \boldsymbol{u} \mid \boldsymbol{u} \in \ker(f)\}.$$

Ist $(V, +, \cdot)$ endlichdimensional, so bildet diese Menge einen affinen Raum mit Richtungsvektorraum $(\ker(f), +, \cdot)$.

Beweis. Für $\boldsymbol{v} \in V$ gilt nach Beobachtung 5.17

$$\boldsymbol{v} \in f^{-1}(\{\boldsymbol{y}\}) \iff f(\boldsymbol{v}) = f(\boldsymbol{v}^{(0)}) \iff \boldsymbol{v} - \boldsymbol{v}^{(0)} \in \ker(f).$$

Die letzte Bedingung bedeutet genau, dass $\boldsymbol{v} = \boldsymbol{v}^{(0)} + \boldsymbol{u}$ für ein $\boldsymbol{u} \in \ker(f)$ ist. Für einen endlichdimensionalen Ausgangsraum $(V, +, \cdot)$ ist diese Menge nach Definition 4.109 ein affiner Raum mit dem angegebenen Richtungsvektorraum. □

Beispiel 5.19. *Für die Projektion aus Beispiel 5.16 und $\boldsymbol{y} \in G$ ist $\pi_G(\boldsymbol{y}) = \boldsymbol{y}$. Nach Satz 5.18 gilt daher*

$$\pi_G^{-1}(\{\boldsymbol{y}\}) = \boldsymbol{y} + G^\perp = \boldsymbol{y} + \langle (1, -1)^\top \rangle.$$

Die Vektoren mit Bild $\boldsymbol{y}$ bilden also eine Gerade durch $\boldsymbol{y}$ senkrecht zu G. Für $\boldsymbol{y} \notin G$ ist das Urbild leer. △

Ist der Ausgangsraum endlichdimensional, so haben nach Satz 5.18 alle nichtleeren Urbilder dieselbe Dimension, denn ihr Richtungsvektorraum ist jeweils der Kernvektorraum. Wir zeigen nun, dass sich die Dimension des Ausgangsraums als Summe der Dimensionen von Kern und Bild ergibt.

Satz 5.20 (*Dimensionsformel*)**.** Für eine lineare Abbildung $f\colon V \to W$ von einem endlichdimensionalen Vektorraum V in einen beliebigen Vektorraum W gilt

$$\dim V = \dim \ker(f) + \operatorname{rang}(f).$$

Beweis. Sei K der gemeinsame zugrunde liegende Körper und $n \coloneqq \dim V$. Nach den Sätzen 5.14 und 4.74 besitzt $\ker(f)$ eine endliche Basis $\{\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(k)}\}$ mit $k \leq n$. Diese ergänzen wir nach Korollar 4.66 zu einer Basis $B = \{\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(n)}\}$ von $(V, +, \cdot)$. Es genügt zu zeigen, dass $\{f(\boldsymbol{b}^{(k+1)}), \dots, f(\boldsymbol{b}^{(n)})\}$ eine Basis des Bildraums $(f(V), +, \cdot)$ mit $n - k$ Elementen ist, da dann $\operatorname{rang}(f) = \dim f(V) = n - k$ gilt, was gemeinsam mit $\dim \ker(f) = k$ und $\dim V = n$ die Aussage zeigt.

Nach Satz 5.14 wird das Bild durch die Bilder aller Basisvektoren erzeugt. Per Definition des Kerns ist $f(\boldsymbol{b}^{(j)}) = \mathbf{0}_W$ für $1 \leq j \leq k$, daher folgt

$$f(V) = \langle f(\boldsymbol{b}^{(k+1)}), \ldots, f(\boldsymbol{b}^{(n)}) \rangle.$$

Für die lineare Unabhängigkeit von $\{f(\boldsymbol{b}^{(k+1)}), \ldots, f(\boldsymbol{b}^{(n)})\}$ seien $\lambda_{k+1}, \ldots, \lambda_n \in K$ mit

$$\sum_{j=k+1}^{n} \lambda_j f(\boldsymbol{b}^{(j)}) = \mathbf{0}_W. \tag{5.1}$$

Nach Satz 5.8 gilt

$$f\left(\sum_{j=k+1}^{n} \lambda_j \boldsymbol{b}^{(j)}\right) = \sum_{j=k+1}^{n} \lambda_j f(\boldsymbol{b}^{(j)}) = \mathbf{0}_W.$$

Also ist dann $\sum_{j=k+1}^{n} \lambda_j \boldsymbol{b}^{(j)} \in \ker(f)$. Da $\langle \boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(k)} \rangle = \ker(f)$, gibt es $\mu_1, \ldots, \mu_k \in K$ mit

$$\sum_{j=k+1}^{n} \lambda_j \boldsymbol{b}^{(j)} = \sum_{j=1}^{k} \mu_j \boldsymbol{b}^{(j)}, \qquad \text{also} \qquad \sum_{j=k+1}^{n} \lambda_j \boldsymbol{b}^{(j)} - \sum_{j=1}^{k} \mu_j \boldsymbol{b}^{(j)} = \mathbf{0}_V.$$

Die lineare Unabhängigkeit von B liefert, dass alle Koeffizienten λ_j und μ_j null sein müssen. Insbesondere müssen alle Koeffizienten in (5.1) null sein. Somit ist $\{f(\boldsymbol{b}^{(k+1)}), \ldots, f(\boldsymbol{b}^{(n)})\}$ linear unabhängig. Außerdem folgt $f(\boldsymbol{b}^{(j)}) \neq \mathbf{0}_W$ für alle $j \in \{k+1, \ldots, n\}$, da wir sonst in (5.1)$\lambda_j = 1$ und alle übrigen Koeffizienten gleich null wählen könnten. Genauso gilt $f(\boldsymbol{b}^{(i)}) \neq f(\boldsymbol{b}^{(j)})$ für alle verschiedenen $i, j \in \{k+1, \ldots, n\}$, da wir sonst in (5.1)$\lambda_i = 1$, $\lambda_j = -1$ und alle übrigen Koeffizienten gleich null wählen könnten. In beiden Fällen wäre die Gleichung erfüllt, obwohl nicht alle Koeffizienten null wären, im Widerspruch zur gerade gezeigten Aussage.

Somit ist $\{f(\boldsymbol{b}^{(k+1)}), \ldots, f(\boldsymbol{b}^{(n)})\}$ eine Basis des Bildraums mit $n - k$ Elementen. □

Damit kennen wir auch die Dimension der Menge aller Vektoren, die auf ein vorgegebenes $\boldsymbol{y} \in f(V)$ abgebildet werden.

Korollar 5.21. Für eine lineare Abbildung $f\colon V \to W$ von einem endlichdimensionalen Vektorraum V in einen beliebigen Vektorraum W hat jedes nichtleere Urbild $f^{-1}(\{\boldsymbol{y}\})$ die Dimension

$$\dim f^{-1}(\{\boldsymbol{y}\}) = \dim V - \operatorname{rang}(f).$$

Beweis. Nach Satz 5.18 hat das Urbild den Richtungsvektorraum $(\ker(f), +, \cdot)$. Die Behauptung folgt daher aus Definition 4.111 und Satz 5.20. □

Beispiel 5.22. *Wir betrachten die lineare Abbildung*

$$f\colon \mathbb{R}^3 \to \mathbb{R}^2, \qquad (x, y, z)^\top \mapsto (x + y, z)^\top.$$

Die Linearität folgt unmittelbar aus der komponentenweisen Addition und Skalarmultiplikation. Für den Kern suchen wir diejenigen Vektoren, die auf null abgebildet werden. Wie beim Schnitt einer Geraden mit einer

Ebene in Beispiel 4.7 übersetzen wir die Vektorgleichung in Gleichungen für die einzelnen Komponenten. Wir erhalten $x + y = 0$ *und* $z = 0$*. Wir können* $y = t$ *frei wählen und erhalten* $x = -t$ *sowie* $z = 0$*, also*

$$\ker(f) = \{(-t, t, 0)^\top \mid t \in \mathbb{R}\} = \langle (-1, 1, 0)^\top \rangle.$$

Zu jedem Vektor $(u, v)^\top \in \mathbb{R}^2$ *finden wir außerdem ein Urbild, denn* $f((u, 0, v)^\top) = (u, v)^\top$*. Somit gilt* $f(\mathbb{R}^3) = \mathbb{R}^2$ *und* $\operatorname{rang}(f) = 2$*. Die Dimensionsformel lautet hier* $3 = 1 + 2$*. Nach Satz 5.18 erhalten wir alle Vektoren mit Bild* $(u, v)^\top$ *durch*

$$f^{-1}(\{(u, v)^\top\}) = (u, 0, v)^\top + \langle (-1, 1, 0)^\top \rangle.$$

Dieses Urbild ist eine Gerade, hat also entsprechend Korollar 5.21 die Dimension $3 - 2 = 1$*.* △

Injektivität und Surjektivität sind im Allgemeinen verschiedene Eigenschaften. Für lineare Abbildungen zwischen endlichdimensionalen Vektorräumen gleicher Dimension genügt dagegen eine der beiden Eigenschaften, um bijektiv zu sein.

Satz 5.23. Eine lineare Abbildung $f\colon V \to W$ zwischen endlichdimensionalen Vektorräumen, für die $\dim V = \dim W$ gilt, ist genau dann injektiv, wenn sie surjektiv ist.

Beweis. Nach Beobachtung 5.17 ist f genau dann injektiv, wenn $\ker(f) = \{\mathbf{0}_V\}$ ist, also genau dann, wenn $\dim \ker(f) = 0$ gilt. Nach der Dimensionsformel (Satz 5.20) ist dies wegen $\dim V = \dim W$ äquivalent zu $\dim f(V) = \operatorname{rang}(f) = \dim W$. Nach Satz 4.74 gilt dies genau dann, wenn $f(V) = W$ ist, also genau dann, wenn f surjektiv ist. □

Bemerkung 5.24. Die Voraussetzung gleicher Dimension ist wesentlich: Die Abbildung aus Beispiel 5.22 ist surjektiv, aber wegen ihres nichttrivialen Kerns nicht injektiv. Umgekehrt ist die lineare Einbettung $\mathbb{R}^2 \to \mathbb{R}^3,\ (u, v)^\top \mapsto (u, 0, v)^\top$ injektiv, aber nicht surjektiv.

Ob zwei endlichdimensionale Vektorräume über einem festen Körper isomorph sind, können wir nun allein an ihren Dimensionen ablesen.

Korollar 5.25. Zwei endlichdimensionale Vektorräume über demselben Körper sind genau dann isomorph, wenn ihre Dimensionen übereinstimmen.

Beweis. Ist $f\colon V \to W$ ein Isomorphismus zwischen den Vektorräumen $(V, +, \cdot)$ und $(W, \oplus, \odot)$, so ist f nach Definition bijektiv. Wegen der Injektivität gilt $\ker(f) = \{\mathbf{0}_V\}$ (Beobachtung 5.17); wegen der Surjektivität ist $f(V) = W$. Aus der Dimensionsformel folgt $\dim V = \dim W$.

Sei umgekehrt $\dim V = \dim W = n$. Wir wählen geordnete Basen $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ von $(V, +, \cdot)$ und $\mathfrak{C} = (\boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(n)})$ von $(W, \oplus, \odot)$. Nach Satz 5.9 gibt es eine lineare Abbildung $f\colon V \to W$ mit $f(\boldsymbol{b}^{(j)}) = \boldsymbol{c}^{(j)}$ für alle $j \in \{1, \ldots, n\}$. Nach Satz 5.14 gilt $f(V) = \langle \boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(n)} \rangle$, außerdem ist $\langle \boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(n)} \rangle = W$, da $\mathfrak{C}$ eine geordnete Basis von $(W, \oplus, \odot)$ ist. Also ist $f(V) = W$ und somit f surjektiv. Satz 5.23 liefert die Bijektivität. □

Beispiel 5.26. *Der Funktionenraum* $(\mathcal{F}, +, \cdot)$ *aus Beispiel 4.70 und* $(\mathbb{R}^2, +, \cdot)$ *haben beide Dimension zwei und sind daher isomorph. Einen solchen Isomorphismus kennen wir bereits aus Beispiel 5.4: Er ordnet jeder Funktion* $\boldsymbol{f}(t) = a + bt$ *ihren Koordinatenvektor* $(a, b)^\top$ *bezüglich der dort verwendeten geordneten Basis zu.* △

5.3 Matrizen

In Abschnitt 5.2 haben wir die Struktur der Urbilder linearer Abbildungen untersucht. Nun entwickeln wir eine Darstellung, mit der wir solche Abbildungen durch Rechnungen mit ihren Koordinaten behandeln können. Ausgangspunkt ist Satz 5.9: Eine lineare Abbildung auf einem endlichdimensionalen Vektorraum ist durch ihre Basisbilder eindeutig bestimmt.

Beispiel 5.27. *Wir betrachten die lineare Abbildung*

$$f\colon \mathbb{R}^3 \to \mathbb{R}^2, \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ z \end{pmatrix}$$

aus Beispiel 5.22. Die Bilder der Standardbasisvektoren sind

$$f(\boldsymbol{e}^{(1)}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad f(\boldsymbol{e}^{(2)}) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad f(\boldsymbol{e}^{(3)}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ordnen wir sie nebeneinander als Spalten an, so erhalten wir das rechteckige Zahlenschema

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Um das Bild von $(x, y, z)^\top$ zu berechnen, bilden wir die Linearkombination dieser Spalten mit den Koordinaten x, y und z:

$$f\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x+y \\ z \end{pmatrix}.$$

Das Schema enthält somit alle Angaben, die wir zur Berechnung der Bilder benötigen. △

Allgemein gilt für einen Körper K, $m, n \in \mathbb{N}$ und eine lineare Abbildung $f\colon K^n \to K^m$ nach Satz 5.8

$$f(\boldsymbol{x}) = f\left(\sum_{j=1}^{n} x_j \boldsymbol{e}^{(j)}\right) = \sum_{j=1}^{n} x_j f(\boldsymbol{e}^{(j)}).$$

Wir fassen daher die Basisbilder zu einem Tupel zusammen, das wir durch nebeneinander angeordnete Spalten darstellen. Dazu führen wir eine Schreibweise für die Linearkombination dieser Spalten ein.

Definition 5.28. Eine $(m \times n)$-*Matrix* über einem Körper K mit $n, m \in \mathbb{N}$ ist ein Tupel $A = (\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(n)}) \in K^{m \times n} \coloneqq (K^m)^n$ von Vektoren $\boldsymbol{a}^{(j)} \in K^m$ für $j \in \{1, \ldots, n\}$. Den *Eintrag* an Position (i, j) bezeichnen wir für $1 \leq i \leq m$ und $1 \leq j \leq n$ mit $A_{ij} \coloneqq (\boldsymbol{a}^{(j)})_i$. Wir schreiben

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \ldots & A_{mn} \end{pmatrix}$$

und nennen $\boldsymbol{A_{\cdot j}} \coloneqq \boldsymbol{a}^{(j)}$ die j-te *Spalte*, $\boldsymbol{A_{i\cdot}} \coloneqq (A_{i1}, \ldots, A_{in})$ die i-te *Zeile* der Matrix.

Wir fassen nun die in Beispiel 5.27 verwendete Linearkombination der Spalten einer Matrix $A \in K^{m\times n}$ als Produkt auf. Im Folgenden sei K ein Körper.

Definition 5.29. Für $A \in K^{m\times n}$ und $\boldsymbol{x} = (x_1, \dots, x_n)^\top \in K^n$ mit $n, m \in \mathbb{N}$ definieren wir das *Produkt* $A\boldsymbol{x} \in K^m$ durch

$$A\boldsymbol{x} \coloneqq \sum_{j=1}^{n} x_j \boldsymbol{A_{\cdot j}}.$$

Die Definition beschreibt das Produkt als Linearkombination der Spalten. Für seine i-te Komponente ergibt sich

$$(A\boldsymbol{x})_i = \sum_{j=1}^{n} A_{ij} x_j \qquad \text{für } i \in \{1, \dots, m\}.$$

Wir multiplizieren also die Einträge der i-ten Zeile $\boldsymbol{A_{i\cdot}}$ mit den entsprechenden Komponenten von $\boldsymbol{x}$ und addieren die Ergebnisse. Beispielsweise lautet die Rechnung aus Beispiel 5.27 nun

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1\cdot x + 1\cdot y + 0\cdot z \\ 0\cdot x + 0\cdot y + 1\cdot z \end{pmatrix} = \begin{pmatrix} x + y \\ z \end{pmatrix}.$$

Bemerkung 5.30. Auch $m = 0$ oder $n = 0$ ist zugelassen. Die Matrix hat dann keine Einträge; ihr Format $m \times n$ bleibt angegeben. Dabei bildet $K^0 = \{()\}$ den Nullvektorraum mit leerer Standardbasis. Für $n = 0$ liefert die leere Summe (Definition 3.18) das Produkt $A\boldsymbol{x} = \mathbf{0} \in K^m$; für $m = 0$ ist $A\boldsymbol{x} = () \in K^0$.

Wir halten fest, dass sich jede lineare Abbildung zwischen solchen Koordinatenräumen durch eine Matrix beschreiben lässt.

Satz 5.31. Seien $n, m \in \mathbb{N}$. Jede Matrix $A \in K^{m\times n}$ definiert eine lineare Abbildung

$$f_A \colon K^n \to K^m, \qquad \boldsymbol{x} \mapsto A\boldsymbol{x}.$$

Umgekehrt gibt es zu jeder linearen Abbildung $f \colon K^n \to K^m$ genau eine Matrix $A \in K^{m\times n}$ mit $f = f_A$. Ihre Spalten sind die Bilder der Standardbasisvektoren:

$$A = \big(f(\boldsymbol{e}^{(1)}), \dots, f(\boldsymbol{e}^{(n)})\big).$$

Beweis. Nach Satz 5.9 gibt es genau eine lineare Abbildung, die für alle $1 \le j \le n$ den Standardbasisvektor $\boldsymbol{e}^{(j)}$ auf $\boldsymbol{A_{\cdot j}}$ abbildet. Für diese und jeden Vektor $\boldsymbol{x} = (x_1, \dots, x_n)^\top \in K^n$ gilt nach Satz 5.8

$$f\left(\sum_{j=1}^{n} x_j \boldsymbol{e}^{(j)}\right) = \sum_{j=1}^{n} x_j \boldsymbol{A_{\cdot j}} = A\boldsymbol{x}.$$

Sie ist also gerade f_A. Wählen wir umgekehrt die Bilder $f(\boldsymbol{e}^{(1)}), \dots, f(\boldsymbol{e}^{(n)})$ als Spalten von A, so ergibt die obige Rechnung $f = f_A$. Für jede Matrix mit $f = f_A$ muss ihre j-te Spalte gleich $A\boldsymbol{e}^{(j)} = f(\boldsymbol{e}^{(j)})$ sein. Damit ist auch die Matrix eindeutig bestimmt. □

Wir können somit die Begriffe des Kerns und Rangs aus Abschnitt 5.2 auf Matrizen übertragen. Das Bild der zugehörigen linearen Abbildung ist unmittelbar über die Spalten der Matrix beschrieben.

Definition 5.32. Für $A \in K^{m\times n}$ mit $n, m \in \mathbb{N}$ nennen wir $\ker A \coloneqq \ker(f_A)$ den *Kern* und $\operatorname{rang} A \coloneqq \operatorname{rang}(f_A)$ den *Rang* von A.

Beobachtung 5.33. Für eine Matrix $A \in K^{m\times n}$ gilt

$$\ker A = \{\boldsymbol{x} \in K^n \mid A\boldsymbol{x} = \boldsymbol{0}\}, \qquad f_A(K^n) = \langle \boldsymbol{A_{\cdot 1}}, \dots, \boldsymbol{A_{\cdot n}}\rangle.$$

Insbesondere ist $\operatorname{rang} A = \dim\langle \boldsymbol{A_{\cdot 1}}, \dots, \boldsymbol{A_{\cdot n}}\rangle$.

Beweis. Die Beschreibung des Kerns folgt aus dessen Definition. Da die Spalten die Bilder der Standardbasisvektoren sind, folgt die Beschreibung des Bilds aus Satz 5.14. Die Aussage über den Rang ergibt sich damit aus Definition 5.15. □

Beispiel 5.34. *Für die Matrix A aus Beispiel 5.27 stimmen die ersten beiden Spalten überein. Ihre Differenz ist also null, das heißt $A\begin{pmatrix}-1\\1\\0\end{pmatrix} = \boldsymbol{0}$. Die erste und die dritte Spalte bilden die Standardbasis von $(\mathbb{R}^2, +, \cdot)$, also gilt $\operatorname{rang} A = 2$. Damit erkennen wir den Kernvektor und den Rang aus Beispiel 5.22 an den Spalten wieder.* △

Auch die bereits betrachteten geometrischen Abbildungen lassen sich auf diese Weise durch ihre Basisbilder beschreiben.

Beispiel 5.35. *Wir betrachten die Orthogonalprojektion π_G auf die Gerade $G = \langle(1,1)^\top\rangle$ und die Spiegelung s_G an G aus den Beispielen 5.16 und 5.12. Für die Projektion gilt*

$$\pi_G(\boldsymbol{e}^{(1)}) = \begin{pmatrix}\frac{1}{2}\\\frac{1}{2}\end{pmatrix} = \pi_G(\boldsymbol{e}^{(2)}).$$

Die Spiegelung vertauscht die beiden Standardbasisvektoren, das heißt $s_G(\boldsymbol{e}^{(1)}) = \boldsymbol{e}^{(2)}$ und $s_G(\boldsymbol{e}^{(2)}) = \boldsymbol{e}^{(1)}$. Ordnen wir jeweils diese Bilder als Spalten an, so erhalten wir

$$P = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\\frac{1}{2} & \frac{1}{2}\end{pmatrix} \qquad \textit{und} \qquad R = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}.$$

Somit ist $\pi_G(\boldsymbol{x}) = P\boldsymbol{x}$ und $s_G(\boldsymbol{x}) = R\boldsymbol{x}$ für alle $\boldsymbol{x} \in \mathbb{R}^2$. △

Bisher lagen die Vektoren bereits als Spalten mit Einträgen aus dem Körper vor. In einem allgemeinen Vektorraum können die Vektoren aber beispielsweise Funktionen sein. Um auch dann eine Matrix verwenden zu können, greifen wir auf die Koordinaten bezüglich einer geordneten Basis zurück.

Beispiel 5.36. *Wir untersuchen, ob eine Funktion aus dem Funktionenraum $(\mathcal{F}, +, \cdot)$ aus Beispiel 4.29 durch ihre Werte bei 0 und 1 eindeutig bestimmt ist. Seine Elemente sind die Funktionen $\boldsymbol{f}\colon \mathbb{R} \to \mathbb{R}$ der Form $\boldsymbol{f}(t) = a + bt$ mit $a, b \in \mathbb{R}$. Dazu betrachten wir die Auswertung*

$$T\colon \mathcal{F} \to \mathbb{R}^2, \qquad \boldsymbol{f} \mapsto \begin{pmatrix}\boldsymbol{f}(0)\\\boldsymbol{f}(1)\end{pmatrix}.$$

Wegen der punktweisen Operationen gilt für $\boldsymbol{f}, \boldsymbol{g} \in \mathcal{F}$ und $\lambda \in \mathbb{R}$

$$T(\boldsymbol{f}+\boldsymbol{g}) = \begin{pmatrix} \boldsymbol{f}(0)+\boldsymbol{g}(0) \\ \boldsymbol{f}(1)+\boldsymbol{g}(1) \end{pmatrix} = T(\boldsymbol{f}) + T(\boldsymbol{g}), \qquad T(\lambda \boldsymbol{f}) = \lambda T(\boldsymbol{f}).$$

Damit ist T linear. Gemeint ist hier die Abbildung T zwischen den beiden Vektorräumen; die einzelne Funktion $\boldsymbol{f}\colon \mathbb{R} \to \mathbb{R}$ ist für $a \neq 0$ keine lineare Abbildung, da $\boldsymbol{f}(0) \neq 0$ ist.

Bezüglich der geordneten Basis $\mathfrak{B} = (\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)})$ mit $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ gilt nach Beispiel 4.62 die Darstellung $\boldsymbol{f} = a\boldsymbol{f}^{(0)} + b\boldsymbol{f}^{(1)}$ mit dem Koordinatenvektor $[\boldsymbol{f}]_{\mathfrak{B}} = (a,b)^\top$. Die Basisbilder sind

$$T(\boldsymbol{f}^{(0)}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad T(\boldsymbol{f}^{(1)}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Die Matrix mit diesen Spalten liefert daher

$$A[\boldsymbol{f}]_{\mathfrak{B}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ a+b \end{pmatrix} = T(\boldsymbol{f}).$$

Hier unterscheiden wir die Funktion $\boldsymbol{f}$, ihren Koordinatenvektor $(a,b)^\top$ und das Wertepaar $(a, a+b)^\top$. Die Matrix übersetzt die Koeffizienten der Funktion in ihre beiden Werte.

Für $T(\boldsymbol{f}) = \mathbf{0}$ muss $a = 0$ und $a + b = 0$ gelten, also auch $b = 0$. Somit besteht $\ker(T)$ nur aus der Nullfunktion. Nach Beobachtung 5.17 ist T injektiv. Beide Vektorräume haben Dimension zwei (Beispiel 4.70), also ist T nach Satz 5.23 sogar bijektiv. Jedes Wertepaar gehört daher zu genau einer Funktion aus $\mathcal{F}$. In Abschnitt 5.5 bestimmen wir die Umkehrabbildung und ihre Matrix ausdrücklich. △

Im Funktionenbeispiel war das Ergebnis der Auswertung bereits ein Vektor in $\mathbb{R}^2$. Bei der Multiplikation komplexer Zahlen sind dagegen sowohl die Eingabe als auch das Ergebnis komplexe Zahlen. Wir beschreiben nun beide durch ihre reellen Koordinaten, um auch diese Abbildung mit einer reellen Matrix auszuwerten.

Beispiel 5.37. *Für den reellen Vektorraum $(\mathbb{C}, +, \cdot)$ wählen wir die geordnete Basis $\mathfrak{B} = (1, \mathrm{i})$ aus Beispiel 4.63. Die Multiplikation mit einer festen komplexen Zahl $z = a + b\mathrm{i}$ und die Konjugation sind nach Beispiel 5.7 lineare Abbildungen zwischen reellen Vektorräumen. Schreiben wir $m_z(w) = zw$, so gilt nach Definition 3.95*

$$m_z(1) = a + b\mathrm{i}, \qquad m_z(\mathrm{i}) = -b + a\mathrm{i}.$$

Die Basisbilder sind hier komplexe Zahlen. Für eine reelle Matrix beschreiben wir sie durch ihre Koordinaten bezüglich $\mathfrak{B}$ und ordnen diese als Spalten an:

$$A = \big([m_z(1)]_{\mathfrak{B}}, [m_z(\mathrm{i})]_{\mathfrak{B}}\big) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Für $w = u + v\mathrm{i}$ gilt nach der Multiplikationsformel $m_z(w) = (au - bv) + (bu + av)\mathrm{i}$. Die Matrix liefert genau die Koordinaten dieses Bilds:

$$A[w]_{\mathfrak{B}} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au - bv \\ bu + av \end{pmatrix} = [m_z(w)]_{\mathfrak{B}}.$$

Die Konjugation $c\colon \mathbb{C} \to \mathbb{C}$, $w \mapsto \overline{w}$, bildet dagegen 1 auf 1 und i auf $-\mathrm{i}$ ab. Die Koordinaten dieser Basisbilder liefern entsprechend die Matrix

$$C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$ △

In den Beispielen 5.36 und 5.37 haben wir lineare Abbildungen durch Matrizen beschrieben, indem wir die Vektoren und ihre Bilder in Koordinaten dargestellt haben. Dieses Vorgehen übertragen wir nun auf beliebige endlichdimensionale Vektorräume. Dazu wählen wir im Ausgangs- und im Zielraum jeweils eine geordnete Basis und bilden die Matrix aus den Koordinatenvektoren der Basisbilder.

Definition 5.38. Sei $f\colon V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen $(V,+,\cdot)$ und $(W,+,\cdot)$ über einem Körper K. Für geordnete Basen $\mathfrak{B} = (\boldsymbol{b}^{(1)},\dots,\boldsymbol{b}^{(n)})$ von $(V,+,\cdot)$ und $\mathfrak{C} = (\boldsymbol{c}^{(1)},\dots,\boldsymbol{c}^{(m)})$ von $(W,+,\cdot)$ mit $m,n \in \mathbb{N}$ heißt

$$[f]_{\mathfrak{C}\leftarrow\mathfrak{B}} \coloneqq \left([f(\boldsymbol{b}^{(1)})]_{\mathfrak{C}},\dots,[f(\boldsymbol{b}^{(n)})]_{\mathfrak{C}}\right) \in K^{m\times n}$$

die *Darstellungsmatrix* von f bezüglich $\mathfrak{B}$ und $\mathfrak{C}$. Der Index $\mathfrak{C}\leftarrow\mathfrak{B}$ gibt an, dass die Matrix Koordinaten bezüglich $\mathfrak{B}$ in Koordinaten bezüglich $\mathfrak{C}$ übersetzt.

Beispiel 5.39. *Die Matrix zur Auswertung T aus Beispiel 5.36 schreiben wir mit der dort verwendeten Basis $\mathfrak{B}$ und der geordneten Standardbasis $\mathfrak{E}$ des Zielraums nun als*

$$[T]_{\mathfrak{E}\leftarrow\mathfrak{B}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Für die Multiplikation und die Konjugation aus Beispiel 5.37 verwenden wir dagegen in beiden Räumen dieselbe geordnete Basis $\mathfrak{B} = (1,\mathrm{i})$. Mit den dort berechneten Matrizen A und C gilt also

$$[m_z]_{\mathfrak{B}\leftarrow\mathfrak{B}} = A, \qquad [c]_{\mathfrak{B}\leftarrow\mathfrak{B}} = C. \qquad \triangle$$

In den Beispielen 5.36 und 5.37 übersetzt die jeweilige Matrix die Koordinaten eines Vektors in die Koordinaten seines Bilds. Wir zeigen nun, dass dies für jede lineare Abbildung zwischen endlichdimensionalen Vektorräumen gilt.

Satz 5.40. Sei $f\colon V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen $(V,+,\cdot)$ und $(W,+,\cdot)$ mit geordneten Basen $\mathfrak{B} = (\boldsymbol{b}^{(1)},\dots,\boldsymbol{b}^{(n)})$ beziehungsweise $\mathfrak{C} = (\boldsymbol{c}^{(1)},\dots,\boldsymbol{c}^{(m)})$. Für alle $\boldsymbol{v} \in V$ gilt

$$[f(\boldsymbol{v})]_{\mathfrak{C}} = [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}[\boldsymbol{v}]_{\mathfrak{B}}.$$

Beweis. Sei K der zugrunde liegende Körper und $A = [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$. Für $[\boldsymbol{v}]_{\mathfrak{B}} = (x_1,\dots,x_n)^\top \in K^n$ gilt nach Definition 4.61

$$\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}.$$

Die j-te Spalte von A ist der Koordinatenvektor von $f(\boldsymbol{b}^{(j)})$ bezüglich $\mathfrak{C}$, also gilt $f(\boldsymbol{b}^{(j)}) = \sum_{i=1}^{m} A_{ij}\boldsymbol{c}^{(i)}$. Mit der Linearität (Satz 5.8) erhalten wir

$$f(\boldsymbol{v}) = \sum_{j=1}^{n} x_j f(\boldsymbol{b}^{(j)}) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} A_{ij}\boldsymbol{c}^{(i)} = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j\right)\boldsymbol{c}^{(i)}.$$

Wegen der Eindeutigkeit der Basisdarstellung (Satz 4.55) ist somit die i-te Koordinate von $f(\boldsymbol{v})$ gleich $\sum_{j=1}^{n} A_{ij}x_j$. Dies ist gerade die i-te Komponente von $A[\boldsymbol{v}]_{\mathfrak{B}}$. □

Damit können wir auch die Urbildfrage aus Abschnitt 5.2 in Koordinaten ausdrücken. Für $\boldsymbol{v} \in V$, $\boldsymbol{y} \in W$ und $A = [f]_{\mathfrak{C} \leftarrow \mathfrak{B}}$ gilt

$$f(\boldsymbol{v}) = \boldsymbol{y} \quad \iff \quad A[\boldsymbol{v}]_{\mathfrak{B}} = [\boldsymbol{y}]_{\mathfrak{C}}.$$

Denn nach Satz 4.55 stimmen zwei Vektoren genau dann überein, wenn ihre Koordinatenvektoren bezüglich derselben Basis übereinstimmen. Wir können daher zunächst die Gleichung $A\boldsymbol{x} = [\boldsymbol{y}]_{\mathfrak{C}}$ für die unbekannten Koordinaten $\boldsymbol{x} = (x_1, \dots, x_n)^\top$ lösen und daraus die gesuchten Vektoren $\boldsymbol{v} = \sum_{j=1}^{n} x_j \boldsymbol{b}^{(j)}$ bilden.

Beispiel 5.41. *Wir wenden diese Übersetzung auf die Auswertung T aus Beispiel 5.36 an: Wir suchen eine Funktion $\boldsymbol{f}(t) = a + bt$ mit $\boldsymbol{f}(0) = 2$ und $\boldsymbol{f}(1) = 5$, also mit $T(\boldsymbol{f}) = \boldsymbol{y} = (2, 5)^\top$. Der gesuchte Vektor ist hier also die Funktion $\boldsymbol{f}$; ihre unbekannten Koordinaten sind $[\boldsymbol{f}]_{\mathfrak{B}} = (a, b)^\top$. Im Zielraum verwenden wir die Standardbasis $\mathfrak{E}$, also ist $[\boldsymbol{y}]_{\mathfrak{E}} = (2, 5)^\top$. Mit der Matrix $A = [T]_{\mathfrak{E} \leftarrow \mathfrak{B}}$ aus Beispiel 5.39 wird die obige Gleichung $A[\boldsymbol{f}]_{\mathfrak{B}} = [\boldsymbol{y}]_{\mathfrak{E}}$ damit zu*

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Die erste Zeile liefert $a = 2$ und die zweite $a + b = 5$, also $b = 3$. Die gesuchte Funktion ist damit $\boldsymbol{f}(t) = 2 + 3t$. △

5.4 Rechnen mit Matrizen

In Abschnitt 5.3 haben wir lineare Abbildungen durch Matrizen beschrieben. Nun übertragen wir Addition und Skalarmultiplikation aus Beispiel 5.11 sowie die Verkettung aus Beobachtung 5.13 auf ihre Matrizen. Insbesondere soll die Matrix einer Verkettung aus den Matrizen der beiden Abbildungen berechnet werden können.

Im Folgenden sei K ein Körper. Wir hatten bereits gesehen, dass eine lineare Abbildung eindeutig durch die Bilder einer Basis bestimmt ist (Satz 5.9). Bei Verwendung der Standardbasen stehen die Bilder der Basisvektoren in den Spalten der zugehörigen Matrix (Satz 5.31 bzw. 5.40). Es liegt daher nah, die Addition und Skalarmultiplikation auf die entsprechenden Operationen für die Spalten zurückzuführen.

Definition 5.42. Für $m, n \in \mathbb{N}$, Matrizen $A, B \in K^{m \times n}$ und $\lambda \in K$ definieren wir die *Summe* und das *skalare Vielfache* durch

$$A + B \coloneqq (\boldsymbol{A}_{\cdot \boldsymbol{1}} + \boldsymbol{B}_{\cdot \boldsymbol{1}}, \dots, \boldsymbol{A}_{\cdot \boldsymbol{n}} + \boldsymbol{B}_{\cdot \boldsymbol{n}}), \qquad \lambda A \coloneqq (\lambda \boldsymbol{A}_{\cdot \boldsymbol{1}}, \dots, \lambda \boldsymbol{A}_{\cdot \boldsymbol{n}}).$$

Für die Einträge gilt also $(A + B)_{ij} = A_{ij} + B_{ij}$ und $(\lambda A)_{ij} = \lambda A_{ij}$ für alle $1 \le i \le m$ und $1 \le j \le n$.

Beobachtung 5.43. Für $m, n \in \mathbb{N}$ bildet $(K^{m \times n}, +, \cdot)$ einen Vektorraum über K der Dimension mn. Die Menge aller Matrizen mit einem Eintrag 1 und allen anderen Einträgen 0 bildet eine Basis. Der Nullvektor ist die *Nullmatrix* $\mathbb{O}$, deren Einträge alle 0 sind.

Für die Addition und Skalarmultiplikation können wir somit einfach mit den Einträgen rechnen. Bei der Verkettung müssen wir beachten, dass das Ergebnis der ersten Abbildung die Eingabe der zweiten ist. Wir bestimmen die gesuchte Matrix zunächst an einem bekannten geometrischen Beispiel.

Beispiel 5.44. *Wir betrachten die Matrizen*

$$P = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

aus den Beispielen 5.35 und 5.37. Die Abbildung f_P ist die Orthogonalprojektion auf $G = \langle (1,1)^\top \rangle$. Die Abbildung f_C bildet $(x,y)^\top$ auf $(x,-y)^\top$ ab und ist somit die Spiegelung an der ersten Koordinatenachse. Sie beschreibt zugleich die komplexe Konjugation in reellen Koordinaten.

Wir spiegeln zuerst und projizieren anschließend auf G, betrachten also $f_P \circ f_C$. Da $f_C(\boldsymbol{e}^{(1)}) = \boldsymbol{e}^{(1)}$ und $f_C(\boldsymbol{e}^{(2)}) = -\boldsymbol{e}^{(2)}$ ist, gilt

$$(f_P \circ f_C)(\boldsymbol{e}^{(1)}) = f_P(\boldsymbol{e}^{(1)}) = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$
$$(f_P \circ f_C)(\boldsymbol{e}^{(2)}) = f_P(-\boldsymbol{e}^{(2)}) = -\frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Diese Basisbilder sind nach Satz 5.31 die Spalten der Matrix der Verkettung. Sie lautet daher

$$\frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}. \qquad \triangle$$

Die Identität verändert keinen Vektor. Ihre Matrix wird deshalb ebenfalls keine Wirkung haben.

Definition 5.45. Für $n \in \mathbb{N}$ heißt die Matrix der Identität id_{K^n} die *Einheitsmatrix* $\mathbb{I}_n \in K^{n\times n}$. Ihre Spalten sind die Standardbasisvektoren, also gilt

$$\mathbb{I}_n = (\boldsymbol{e}^{(1)}, \dots, \boldsymbol{e}^{(n)}), \qquad (\mathbb{I}_n)_{ij} = \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{falls } i \neq j. \end{cases}$$

Insbesondere ist $f_{\mathbb{I}_n} = \mathrm{id}_{K^n}$.

Allgemein seien $m, n, r \in \mathbb{N}$, $A \in K^{m\times n}$ und $B \in K^{n\times r}$. Bei der Verkettung $f_A \circ f_B \colon K^r \to K^m$ wird der Standardbasisvektor $\boldsymbol{e}^{(j)}$ zuerst auf die Spalte $\boldsymbol{B_{\cdot j}}$ von B und anschließend auf $A\boldsymbol{B_{\cdot j}}$ abgebildet. Die Spalten der gesuchten Matrix erhalten wir also, indem wir A auf die Spalten von B anwenden. Wir definieren ein Matrixprodukt, das die Matrix der Verkettung aus den Matrizen linearer Abbildungen berechnet.

Definition 5.46. Für $m, n, r \in \mathbb{N}$, $A \in K^{m\times n}$ und $B \in K^{n\times r}$ definieren wir das *Produkt* der Matrizen A und B durch

$$AB \coloneqq (A\boldsymbol{B_{\cdot 1}}, \dots, A\boldsymbol{B_{\cdot r}}) \in K^{m\times r}.$$

Für quadratische Matrizen $A \in K^{n\times n}$ definieren wir außerdem die *Matrixpotenzen* durch

$$A^0 \coloneqq \mathbb{I}_n, \qquad A^{k+1} \coloneqq AA^k \quad \text{für } k \in \mathbb{N}.$$

Die Zahl der Spalten von A muss mit der Zahl der Zeilen von B übereinstimmen, damit die einzelnen Produkte $A\boldsymbol{B_{\cdot j}}$ definiert sind. Für die Einträge erhalten wir aus dem Matrix-Vektor-Produkt

$$(AB)_{ij} = (A\boldsymbol{B_{\cdot j}})_i = \sum_{k=1}^{n} A_{ik}B_{kj} \qquad \text{für } i \in \{1,\dots,m\},\ j \in \{1,\dots,r\}.$$

Den Eintrag an Position (i, j) erhalten wir also, indem wir die entsprechenden Einträge der i-ten Zeile von A und der j-ten Spalte von B multiplizieren und die Produkte addieren. Für reelle Matrizen ist dies das Standardskalarprodukt (Definition 4.9) der als Spaltenvektor geschriebenen Zeile von A mit der Spalte von B. In diesem Sinne rechnen wir „Zeile mal Spalte“.

Beispiel 5.47. *Die Matrix der Verkettung aus Beispiel 5.44 schreiben wir nun als*

$$PC = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

Kehren wir die Reihenfolge um, so wird zunächst projiziert und dann gespiegelt. Beide Standardbasisvektoren haben unter f_P das Bild $\frac{1}{2}(1,1)^\top$, das f_C auf $\frac{1}{2}(1,-1)^\top$ abbildet. Somit ist

$$CP = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \neq PC.$$

Die Reihenfolge ist also entscheidend: Die zuerst angewandte Abbildung steht im Produkt rechts.

Wenden wir dieselbe Abbildung zweimal an, erhalten wir die Quadrate der Matrizen:

$$P^2 = \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = P,$$
$$C^2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbb{I}_2.$$

Das entspricht der geometrischen Wirkung: Nach der ersten Projektion liegt der Bildvektor bereits auf G; erneutes Projizieren verändert ihn nicht. Zweimaliges Spiegeln an derselben Achse bringt dagegen jeden Vektor an seine Ausgangsposition zurück. △

Wir halten den Zusammenhang zwischen den Matrixoperationen und den Abbildungsoperationen nun fest. Wie in Abschnitt 5.3 genügt dazu der Vergleich der Basisbilder.

Beobachtung 5.48. Seien $m, n, r \in \mathbb{N}$. Für $A, B \in K^{m\times n}$ und $\lambda \in K$ gilt

$$f_{A+B} = f_A + f_B, \qquad f_{\lambda A} = \lambda f_A.$$

Für $A \in K^{m\times n}$ und $B \in K^{n\times r}$ gilt

$$f_{AB} = f_A \circ f_B.$$

Beweis. Nach Beispiel 5.11 und Beobachtung 5.13 sind die Abbildungen auf den rechten Seiten linear. Auf einem Standardbasisvektor $\boldsymbol{e}^{(j)}$ gilt nach den Definitionen 5.42 und 5.46 jeweils

$$f_{A+B}(\boldsymbol{e}^{(j)}) = \boldsymbol{A_{\cdot j}} + \boldsymbol{B_{\cdot j}} = (f_A + f_B)(\boldsymbol{e}^{(j)}),$$
$$f_{\lambda A}(\boldsymbol{e}^{(j)}) = \lambda \boldsymbol{A_{\cdot j}} = (\lambda f_A)(\boldsymbol{e}^{(j)}),$$
$$f_{AB}(\boldsymbol{e}^{(j)}) = A\boldsymbol{B_{\cdot j}} = f_A(f_B(\boldsymbol{e}^{(j)})).$$

Somit stimmen die jeweiligen Basisbilder überein. Nach Satz 5.9 sind die Abbildungen daher gleich. □

Dasselbe gilt für Darstellungsmatrizen bezüglich beliebiger geordneter Basen. Bei der Verkettung muss die Zielbasis der ersten Abbildung mit der Ausgangsbasis der zweiten übereinstimmen, damit wir die Koordinaten direkt weiterverwenden können.

Satz 5.49. Seien $(V,+,\cdot)$, $(W,+,\cdot)$ und $(X,+,\cdot)$ endlichdimensionale Vektorräume über K mit geordneten Basen $\mathfrak{B}$, $\mathfrak{C}$ beziehungsweise $\mathfrak{D}$. Für lineare Abbildungen $f,h\colon V\to W$, $g\colon W\to X$ und $\lambda\in K$ gilt

$$[f+h]_{\mathfrak{C}\leftarrow\mathfrak{B}} = [f]_{\mathfrak{C}\leftarrow\mathfrak{B}} + [h]_{\mathfrak{C}\leftarrow\mathfrak{B}}, \qquad [\lambda f]_{\mathfrak{C}\leftarrow\mathfrak{B}} = \lambda [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$$

sowie

$$[g\circ f]_{\mathfrak{D}\leftarrow\mathfrak{B}} = [g]_{\mathfrak{D}\leftarrow\mathfrak{C}}\,[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}.$$

Beweis. Sei $\mathfrak{C}=(\boldsymbol{c}^{(1)},\dots,\boldsymbol{c}^{(m)})$ und seien $\boldsymbol{u}=\sum_{i=1}^m x_i\boldsymbol{c}^{(i)}$ und $\boldsymbol{w}=\sum_{i=1}^m y_i\boldsymbol{c}^{(i)}$ die Basisdarstellungen zweier Vektoren in W. Aus den Vektorraumaxiomen folgt

$$\boldsymbol{u}+\boldsymbol{w} = \sum_{i=1}^m (x_i+y_i)\boldsymbol{c}^{(i)}, \qquad \lambda\boldsymbol{u} = \sum_{i=1}^m (\lambda x_i)\boldsymbol{c}^{(i)}.$$

Wegen der Eindeutigkeit der Basisdarstellung (Satz 4.55) gilt daher

$$[\boldsymbol{u}+\boldsymbol{w}]_{\mathfrak{C}} = [\boldsymbol{u}]_{\mathfrak{C}} + [\boldsymbol{w}]_{\mathfrak{C}}, \qquad [\lambda\boldsymbol{u}]_{\mathfrak{C}} = \lambda[\boldsymbol{u}]_{\mathfrak{C}}.$$

Wenden wir dies auf $\boldsymbol{u}=f(\boldsymbol{b}^{(j)})$ und $\boldsymbol{w}=h(\boldsymbol{b}^{(j)})$ für jeden Basisvektor $\boldsymbol{b}^{(j)}$ von $\mathfrak{B}$ an, so erhalten wir nach Definition 5.38 genau die entsprechenden Spalten der Darstellungsmatrizen. Da Addition und Skalarmultiplikation von Matrizen spaltenweise definiert sind (Definition 5.42), folgen die ersten beiden Gleichheiten.

Setzen wir $F=[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$ und $G=[g]_{\mathfrak{D}\leftarrow\mathfrak{C}}$, so gilt für alle $\boldsymbol{v}\in V$ nach Satz 5.40

$$[g(f(\boldsymbol{v}))]_{\mathfrak{D}} = G[f(\boldsymbol{v})]_{\mathfrak{C}} = G\big(F[\boldsymbol{v}]_{\mathfrak{B}}\big) = (GF)[\boldsymbol{v}]_{\mathfrak{B}}.$$

Die letzte Gleichheit folgt aus Beobachtung 5.48. Setzen wir für $\boldsymbol{v}$ nacheinander die Vektoren der Basis $\mathfrak{B}$ ein, so ist $[\boldsymbol{v}]_{\mathfrak{B}}$ jeweils ein Standardbasisvektor. Die Spalten von GF sind daher genau die Koordinatenvektoren dieser Basisbilder unter $g\circ f$. Somit ist $GF=[g\circ f]_{\mathfrak{D}\leftarrow\mathfrak{B}}$. □

Die Rechenregeln für Matrizen gewinnen wir nun aus den bereits bekannten Regeln für die zugehörigen Abbildungen. Dass die Verkettung assoziativ ist und die Identität neutral wirkt, haben wir auch beim Aufbau der Permutationsgruppen in Beispiel 3.13 verwendet.

Korollar 5.50. Für Matrizen A,B,C über K gelten die folgenden Regeln, sofern die angegebenen Summen und Produkte definiert sind:

$$\begin{aligned}(AB)C &= A(BC),\\ A(B+C) &= AB+AC,\\ (A+B)C &= AC+BC.\end{aligned}$$

Außerdem gilt für $A\in K^{m\times n}$ mit $m,n\in\mathbb{N}$ die Gleichheit $\mathbb{I}_m A = A = A\mathbb{I}_n$ und für $A\in K^{m\times n}$, $B\in K^{n\times r}$ mit $r\in\mathbb{N}$ und $\lambda\in K$

$$(\lambda A)B = \lambda(AB) = A(\lambda B).$$

Beweis. Übung. □

Die quadratischen Matrizen eines festen Formats können wir sowohl addieren als auch miteinander multiplizieren. Die eben bewiesenen Regeln entsprechen genau den Ringaxiomen aus Definition 3.46. Analog zu Definition 3.52 nennen wir auch eine quadratische Matrix $A \neq \mathbb{O}$ einen Nullteiler, wenn es eine Matrix $B \neq \mathbb{O}$ desselben Formats mit $AB = \mathbb{O}$ oder $BA = \mathbb{O}$ gibt.

Beobachtung 5.51. Für $n \in \mathbb{N}$ bildet $(K^{n\times n}, +, \cdot)$ einen Ring mit Eins $\mathbb{I}_n$. Für $n \geq 2$ ist er nicht kommutativ und besitzt Nullteiler. Insbesondere bildet er für $n \geq 2$ keinen Körper.

Beweis. Nach Beobachtung 5.43 bildet $(K^{n\times n}, +)$ eine abelsche Gruppe. Die Assoziativität der Multiplikation, die Distributivität und die neutrale Wirkung von $\mathbb{I}_n$ folgen aus Korollar 5.50. Für $n \geq 2$ verwenden wir die Basismatrizen aus Beobachtung 5.43, wobei E_{ij} die Matrix mit einer 1 an Position (i, j) und sonst nur Nullen bezeichnet. Die Produktdefinition liefert

$$E_{11}E_{12} = E_{12} \neq \mathbb{O} = E_{12}E_{11}, \qquad E_{11}E_{22} = \mathbb{O}.$$

Die erste Gleichheit zeigt die fehlende Kommutativität. Da E_{11} und E_{22} beide von null verschieden sind, zeigt die zweite Gleichheit die Existenz von Nullteilern. □

Die fehlende Kommutativität haben wir über den reellen Zahlen bereits an Projektion und Spiegelung gesehen. Auch die Nullteiler lassen sich mit der Projektion veranschaulichen.

Beispiel 5.52. *Für die Matrix P aus Beispiel 5.44 gilt*

$$\mathbb{I}_2 - P = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Die Abbildung $f_{\mathbb{I}_2-P} = \mathrm{id}_{\mathbb{R}^2} - \pi_G$ liefert zu jedem Vektor seinen Anteil senkrecht zu G. Die anschließende Projektion auf G bildet diesen Anteil auf null ab. Deshalb gilt

$$P(\mathbb{I}_2 - P) = \mathbb{O}.$$

Beide Faktoren sind von null verschiedene Matrizen. Geometrisch entsteht das Nullprodukt also dadurch, dass die erste Abbildung nur Richtungen liefert, die die zweite auf null abbildet. △

Bei fest gewählten Basen können wir somit die Addition und Skalarmultiplikation linearer Abbildungen vollständig durch Rechnungen mit ihren Matrizen beschreiben. Dies verbindet den Abbildungsraum aus Beispiel 5.11 mit dem Matrizenraum.

Bemerkung 5.53. Seien $(V, +, \cdot)$ und $(W, +, \cdot)$ endlichdimensionale Vektorräume über K mit geordneten Basen $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ und $\mathfrak{C} = (\boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(m)})$ für $n, m \in \mathbb{N}$. Die Abbildung

$$\mathrm{Hom}_K(V, W) \to K^{m\times n}, \qquad f \mapsto [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$$

ist ein Vektorraumisomorphismus. Ihre Linearität folgt aus Satz 5.49. Jede Matrix legt durch ihre Spalten eindeutig die Koordinaten der Basisbilder fest, also nach Satz 4.55 auch die Basisbilder selbst. Nach Satz 5.9 gibt es dazu genau eine lineare Abbildung. Damit ist die Zuordnung auch bijektiv.

Bei der Matrixmultiplikation wirken Zeilen und Spalten zusammen. Zum Abschluss führen wir eine Operation ein, die ihre Rollen vertauscht. Sie wird uns bei orthogonalen und symmetrischen Matrizen in Abschnitt 6.5 wieder begegnen.

Definition 5.54. Für $m, n \in \mathbb{N}$ ist die *Transponierte* $A^\top \in K^{n\times m}$ einer Matrix $A \in K^{m\times n}$ gegeben durch

$$(A^\top)_{ji} \coloneqq A_{ij} \qquad \text{für } i \in \{1,\dots,m\},\ j \in \{1,\dots,n\}.$$

Beispiel 5.55. *Für die Matrix aus Beispiel 5.27 gilt*

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^\top = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Die erste Zeile wird zur ersten Spalte und die zweite Zeile zur zweiten Spalte. △

Definition 5.56. Für $m, n \in \mathbb{N}$ fassen wir bei Matrixprodukten einen Vektor $\boldsymbol{v} \in K^m$ auch als Matrix in $K^{m\times 1}$ mit einziger Spalte $\boldsymbol{v}$ auf. Matrizen aus $K^{1\times 1}$ und Vektoren aus K^1 identifizieren wir dabei mit ihrem einzigen Eintrag. Insbesondere erlaubt dies, das Standardskalarprodukt zweier Vektoren $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ als Matrixprodukt zu schreiben:

$$(\boldsymbol{u} \mid \boldsymbol{v}) = \boldsymbol{u}^\top \boldsymbol{v}.$$

Beispiel 5.57. *Für die Vektoren* $\boldsymbol{u} = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ *und* $\boldsymbol{v} = \left(\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right)$ *gilt*

$$\boldsymbol{u}^\top \boldsymbol{v} = \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = 1 \cdot 3 + 2 \cdot 4 = 11 = (\boldsymbol{u} \mid \boldsymbol{v}).$$

Das Produkt $\boldsymbol{u}\boldsymbol{v}^\top$ *ergibt dagegen eine* (2×2)*-Matrix:*

$$\boldsymbol{u}\boldsymbol{v}^\top = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 6 & 8 \end{pmatrix}.$$ △

Wir halten fest, wie sich das Transponieren mit Addition, Skalarmultiplikation und Matrixmultiplikation verträgt.

Beobachtung 5.58. Für $m, n, r \in \mathbb{N}$, $A, B \in K^{m\times n}$ und $\lambda \in K$ gilt

$$(A^\top)^\top = A, \qquad (A+B)^\top = A^\top + B^\top, \qquad (\lambda A)^\top = \lambda A^\top.$$

Für $A \in K^{m\times n}$ und $B \in K^{n\times r}$ gilt außerdem

$$(AB)^\top = B^\top A^\top.$$

Beweis. Übung. □

5.5 Invertierbarkeit

Nach Beobachtung 5.51 bilden die quadratischen Matrizen fester Zeilen-/Spaltenzahl einen Ring mit Eins. Im Allgemeinen besitzt darin nicht jede von null verschiedene Matrix eine multiplikative Inverse. Wir untersuchen nun, welche Matrizen invertierbar sind. Da die Matrixmultiplikation nach Beobachtung 5.48 der Verkettung entspricht, hängt dies damit zusammen, ob die zugehörige lineare Abbildung umkehrbar ist.

Im Folgenden sei K ein Körper.

Definition 5.59. Für $n \in \mathbb{N}$ heißt eine Matrix $A \in K^{n\times n}$ *invertierbar*, wenn eine Matrix $B \in K^{n\times n}$ mit

$$AB = BA = \mathbb{I}_n$$

existiert. Eine solche Matrix B heißt *Inverse* von A.

Wie bei inversen Elementen in Gruppen (Satz 3.9) ist diese Matrix eindeutig bestimmt.

Beobachtung 5.60. Für $n \in \mathbb{N}$ besitzt jede invertierbare Matrix $A \in K^{n\times n}$ genau eine Inverse. Wir bezeichnen sie mit A^{-1}.

Beweis. Sind B und C Inverse von A, so gilt nach Korollar 5.50

$$B = B\mathbb{I}_n = B(AC) = (BA)C = \mathbb{I}_n C = C. \qquad \square$$

Die beiden Gleichungen in Definition 5.59 entsprechen genau den Gleichungen für eine Umkehrabbildung aus Satz 1.76.

Satz 5.61. Für $n \in \mathbb{N}$ und $A \in K^{n\times n}$ ist A genau dann invertierbar, wenn $f_A \colon K^n \to K^n$ ein Automorphismus ist. In diesem Fall gilt

$$f_{A^{-1}} = (f_A)^{-1}.$$

Beweis. Ist A invertierbar, so liefert Beobachtung 5.48

$$\begin{aligned} f_A \circ f_{A^{-1}} &= f_{AA^{-1}} = f_{\mathbb{I}_n} = \mathrm{id}_{K^n}, \\ f_{A^{-1}} \circ f_A &= f_{A^{-1}A} = f_{\mathbb{I}_n} = \mathrm{id}_{K^n}. \end{aligned}$$

Nach Satz 1.76 ist f_A daher bijektiv und $f_{A^{-1}}$ seine Umkehrabbildung. Da f_A linear ist, ist sie ein Automorphismus.

Sei umgekehrt f_A ein Automorphismus. Nach Beobachtung 5.13 ist $(f_A)^{-1}$ linear und besitzt nach Satz 5.31 eine Matrix $B \in K^{n\times n}$ mit $f_B = (f_A)^{-1}$. Wegen

$$\begin{aligned} f_{AB} &= f_A \circ f_B = f_A \circ (f_A)^{-1} = \mathrm{id}_{K^n} = f_{\mathbb{I}_n}, \\ f_{BA} &= f_B \circ f_A = (f_A)^{-1} \circ f_A = \mathrm{id}_{K^n} = f_{\mathbb{I}_n} \end{aligned}$$

folgt aus der Eindeutigkeit der Matrix in Satz 5.31, dass $AB = BA = \mathbb{I}_n$ gilt. Somit ist A invertierbar und es gilt $B = A^{-1}$. $\square$

Beispiel 5.62. *Für die Spiegelung und die Projektion aus Beispiel 5.44 verwenden wir die Matrizen*

$$C = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad P = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Zweimaliges Spiegeln ergibt die Identität, also ist $C^{-1} = C$. Die Projektion bildet dagegen sowohl $(1,-1)^\top$ als auch den Nullvektor auf null ab (Beispiel 5.16). Damit ist f_P nicht injektiv und P nach Satz 5.61 nicht invertierbar. △

Wir übertragen nun die Aussage von Automorphismen des K^n auf Isomorphismen zwischen beliebigen endlichdimensionalen Vektorräumen. Auch bei Darstellungsmatrizen bezüglich beliebiger geordneter Basen entspricht die inverse Matrix der Umkehrabbildung. Dabei tauschen Ausgangs- und Zielbasis ihre Rollen.

Korollar 5.63. Seien $(V,+,\cdot)$ und $(W,+,\cdot)$ endlichdimensionale Vektorräume über K mit geordneten Basen $\mathfrak{B}$ beziehungsweise $\mathfrak{C}$. Ist $f\colon V \to W$ ein Isomorphismus, so ist $[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$ invertierbar und es gilt

$$[f^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{C}} = \left([f]_{\mathfrak{C}\leftarrow\mathfrak{B}}\right)^{-1}.$$

Beweis. Da f ein Isomorphismus ist, ist f bijektiv und besitzt nach Satz 1.76 eine Umkehrabbildung f^{-1}. Nach Beobachtung 5.13 ist diese ebenfalls linear. Nach Korollar 5.25 gilt $\dim V = \dim W = n$ für ein $n \in \mathbb{N}$; beide Darstellungsmatrizen sind daher quadratisch. Da die Identität jeden Basisvektor auf sich selbst abbildet, gilt $[\mathrm{id}_V]_{\mathfrak{B}\leftarrow\mathfrak{B}} = \mathbb{I}_n$ und $[\mathrm{id}_W]_{\mathfrak{C}\leftarrow\mathfrak{C}} = \mathbb{I}_n$. Aus Satz 5.49 folgt nun

$$\begin{aligned}[f^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{C}}\,[f]_{\mathfrak{C}\leftarrow\mathfrak{B}} &= [f^{-1}\circ f]_{\mathfrak{B}\leftarrow\mathfrak{B}} = \mathbb{I}_n,\\ [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}\,[f^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{C}} &= [f\circ f^{-1}]_{\mathfrak{C}\leftarrow\mathfrak{C}} = \mathbb{I}_n.\end{aligned}$$

Damit erfüllt $[f^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{C}}$ die beiden Gleichungen aus Definition 5.59 und ist die Inverse von $[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}$. □

Beispiel 5.64. *In Beispiel 5.41 haben wir aus $\boldsymbol{f}(0) = 2$ und $\boldsymbol{f}(1) = 5$ die Parameter $a = 2$ und $b = 3$ der Funktion $\boldsymbol{f}(t) = a + bt$ bestimmt. Wir betrachten diese Rechnung nun mithilfe der inversen Darstellungsmatrix. Dazu greifen wir die Abbildung*

$$T\colon \mathcal{F} \to \mathbb{R}^2, \qquad \boldsymbol{f} \mapsto \begin{pmatrix} \boldsymbol{f}(0) \\ \boldsymbol{f}(1) \end{pmatrix}$$

aus Beispiel 5.36 wieder auf. Dort haben wir bereits gezeigt, dass T ein Isomorphismus ist. Für $u, v \in \mathbb{R}$ suchen wir die Funktion $\boldsymbol{f}(t) = a + bt$ mit $T(\boldsymbol{f}) = (u,v)^\top$. Die beiden vorgegebenen Werte liefern $a = u$ und $a + b = v$, also $b = v - u$. Die Umkehrabbildung ist daher durch

$$\left(T^{-1}((u,v)^\top)\right)(t) = u + (v-u)t \qquad \textit{für } t \in \mathbb{R}$$

gegeben.

Bezüglich der Basis $\mathfrak{B} = (\boldsymbol{f}^{(0)}, \boldsymbol{f}^{(1)})$ mit $\boldsymbol{f}^{(0)}(t) = 1$ und $\boldsymbol{f}^{(1)}(t) = t$ hat diese Funktion den Koordinatenvektor $(u, v-u)^\top$. Mit der geordneten Standardbasis $\mathfrak{E}$ von $\mathbb{R}^2$ gilt somit

$$[T^{-1}((u,v)^\top)]_{\mathfrak{B}} = \begin{pmatrix} u \\ v-u \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

Die erste beziehungsweise zweite Spalte erhalten wir für die Standardbasisvektoren als Eingaben. Deshalb ist

$$[T^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{E}} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \left([T]_{\mathfrak{E}\leftarrow\mathfrak{B}}\right)^{-1},$$

wobei $[T]_{\mathfrak{E}\leftarrow\mathfrak{B}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ *aus Beispiel 5.39 bekannt ist. Für das Wertepaar aus Beispiel 5.41 erhalten wir konkret*

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Die inverse Matrix liefert also die Koeffizienten $a = 2$ *und* $b = 3$*; die Umkehrabbildung* T^{-1} *liefert die Funktion* $\boldsymbol{f}(t) = 2 + 3t$*.* △

Um die Invertierbarkeit einer Matrix festzustellen, betrachten wir Kern und Bild der zugehörigen linearen Abbildung: Sie ist genau dann injektiv, wenn ihr Kern nur den Nullvektor enthält, und genau dann surjektiv, wenn ihr Bild der gesamte Zielraum ist (Abschnitt 5.2). Bei quadratischen Matrizen genügt wegen der gleichen Dimension von Ausgangs- und Zielraum bereits eine dieser beiden Eigenschaften.

Satz 5.65. Für $n \in \mathbb{N}$ und $A \in K^{n\times n}$ sind die folgenden Aussagen äquivalent:

(i) A ist invertierbar.

(ii) $\ker A = \{\mathbf{0}\}$.

(iii) $\operatorname{rang} A = n$.

(iv) Das Tupel $(\boldsymbol{A_{\cdot 1}}, \ldots, \boldsymbol{A_{\cdot n}})$ ist eine geordnete Basis von $(K^n, +, \cdot)$.

Beweis. Nach Satz 5.61 ist A genau dann invertierbar, wenn f_A bijektiv ist. Wegen Satz 5.23 ist dies bereits zur Injektivität von f_A äquivalent, also nach Beobachtung 5.17 zu $\ker A = \{\mathbf{0}\}$. Ebenso ist es zur Surjektivität von f_A äquivalent. Diese gilt genau dann, wenn $f_A(K^n) = K^n$ ist, also nach Satz 4.74 genau dann, wenn $\dim f_A(K^n) = n$ gilt, also wenn $\operatorname{rang} A = \operatorname{rang}(f_A) = \dim f_A(K^n) = n$. Damit sind die Aussagen (i), (ii) und (iii) äquivalent.

Nach der Definition des Matrix-Vektor-Produkts gilt für $\boldsymbol{x} \in K^n$

$$f_A(\boldsymbol{x}) = A\boldsymbol{x} = \sum_{j=1}^{n} x_j \boldsymbol{A_{\cdot j}}.$$

Der Vektor $\boldsymbol{x}$ ist also genau dann ein Urbild von $\boldsymbol{y} \in K^n$, wenn $x_1, \ldots, x_n$ die Koeffizienten einer Darstellung von $\boldsymbol{y}$ als Linearkombination $\sum_{j=1}^{n} x_j \boldsymbol{A_{\cdot j}}$ sind. Somit ist f_A genau dann bijektiv, wenn jeder Vektor $\boldsymbol{y} \in K^n$ genau eine Darstellung als Linearkombination der Spalten von A besitzt. Nach Satz 4.55 ist dies genau dann der Fall, wenn $(\boldsymbol{A_{\cdot 1}}, \ldots, \boldsymbol{A_{\cdot n}})$ eine geordnete Basis von $(K^n, +, \cdot)$ ist. Damit sind auch die Aussagen (i) und (iv) äquivalent. □

Die gleiche Dimension von Ausgangs- und Zielraum hat noch eine weitere Folge: Für quadratische Matrizen genügt es, eine der beiden Inversengleichungen nachzuweisen.

Korollar 5.66. Für $n \in \mathbb{N}$ und $A, B \in K^{n\times n}$ folgt aus $AB = \mathbb{I}_n$ bereits $BA = \mathbb{I}_n$ und $B = A^{-1}$.

Beweis. Aus $AB = \mathbb{I}_n$ folgt für jeden Vektor $\boldsymbol{y} \in K^n$, dass $A(B\boldsymbol{y}) = \boldsymbol{y}$ gilt. Jeder Vektor besitzt also unter f_A ein Urbild. Damit ist f_A surjektiv, nach Satz 5.23 auch injektiv und nach Satz 5.61 die Matrix A invertierbar. Mit Korollar 5.50 erhalten wir

$$B = \mathbb{I}_n B = (A^{-1}A)B = A^{-1}(AB) = A^{-1}\mathbb{I}_n = A^{-1}.$$

Daraus folgt auch $BA = \mathbb{I}_n$. □

Beispiel 5.67. *Die Voraussetzung quadratischer Matrizen ist wesentlich. Für*

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix} \in \mathbb{R}^{1\times 2}, \qquad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{R}^{2\times 1}$$

gilt

$$AB = \mathbb{I}_1, \qquad BA = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \mathbb{I}_2.$$

Die Abbildung f_B bildet das einelementige Tupel $(x) \in \mathbb{R}^1$ auf $(x, 0)^\top \in \mathbb{R}^2$ ab; f_A liefert zu $(x, y)^\top$ das Tupel (x) der ersten Koordinate. So macht f_A die Einbettung rückgängig. Wenden wir die Abbildungen in umgekehrter Reihenfolge an, so wird dagegen $(x, y)^\top$ auf $(x, 0)^\top$ abgebildet. Die zweite Koordinate wird dabei nicht wiederhergestellt. △

Für die Inversen gelten die vertrauten Regeln für Umkehrabbildungen. Insbesondere kehrt sich beim Rückgängigmachen einer Verkettung die Reihenfolge um; dieselbe Regel kennen wir für inverse Gruppenprodukte aus Satz 3.16.

Satz 5.68. Für $n \in \mathbb{N}$ und invertierbare Matrizen $A, B \in K^{n\times n}$ sind auch AB, A^{-1} und $A^\top$ invertierbar und es gilt

$$(AB)^{-1} = B^{-1}A^{-1}, \qquad (A^{-1})^{-1} = A, \qquad (A^\top)^{-1} = (A^{-1})^\top.$$

Beweis. Übung. □

Die Menge der invertierbaren Matrizen ist also abgeschlossen bezüglich der Multiplikation und des Bildens von Inversen. Sie bildet sogar eine Gruppe.

Beobachtung 5.69. Für $n \in \mathbb{N}$ bildet die Menge

$$\mathrm{GL}_n(K) \coloneqq \{A \in K^{n\times n} \mid A \text{ ist invertierbar}\}$$

mit der Matrixmultiplikation eine Gruppe, die *allgemeine lineare Gruppe* $(\mathrm{GL}_n(K), \cdot)$ mit neutralem Element $\mathbb{I}_n$.

Beweis. Nach Satz 5.68 ist das Produkt zweier Elemente von $\mathrm{GL}_n(K)$ wieder ein Element dieser Menge. Die Multiplikation ist nach Korollar 5.50 assoziativ. Die Einheitsmatrix ist ihre eigene Inverse und liegt daher in $\mathrm{GL}_n(K)$; sie ist das neutrale Element. Jedes $A \in \mathrm{GL}_n(K)$ besitzt die Inverse A^{-1}, die nach Satz 5.68 ebenfalls in $\mathrm{GL}_n(K)$ liegt. Damit sind die Gruppenaxiome erfüllt. □

5.6 Basiswechsel

Die Darstellungsmatrix einer linearen Abbildung hängt von den gewählten Basen ab. Wir untersuchen nun, wie sie sich verändert, wenn wir andere Koordinaten verwenden. Dazu betrachten wir zunächst den Koordinatenwechsel eines einzelnen Vektors.

Beispiel 5.70. *Aus Beispiel 4.100 kennen wir die geordnete Basis $\mathfrak{B} = ((1,1)^\top, (1,0)^\top)$ von $\mathbb{R}^2$ und den Vektor $\boldsymbol{x} = (2,1)^\top$. Mit der geordneten Standardbasis $\mathfrak{E}$ gilt*

$$[\boldsymbol{x}]_{\mathfrak{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad [\boldsymbol{x}]_{\mathfrak{E}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Die Koordinaten $(r,s)^\top$ bezüglich $\mathfrak{B}$ beschreiben den Vektor

$$r \begin{pmatrix} 1 \\ 1 \end{pmatrix} + s \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} r+s \\ r \end{pmatrix}.$$

Umgekehrt gilt für jeden Vektor $\boldsymbol{y} = (u,v)^\top$ die Darstellung $\boldsymbol{y} = v(1,1)^\top + (u-v)(1,0)^\top$. Somit rechnen die Matrizen

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} [\boldsymbol{y}]_{\mathfrak{B}} = [\boldsymbol{y}]_{\mathfrak{E}}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} [\boldsymbol{y}]_{\mathfrak{E}} = [\boldsymbol{y}]_{\mathfrak{B}}$$

die Koordinaten ineinander um. Für unseren Vektor $\boldsymbol{x}$ erhalten wir insbesondere

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \qquad \triangle$$

Der Vektor bleibt bei dieser Umrechnung unverändert, seine Koordinaten ändern sich. Die beiden Matrizen beschreiben daher die Identität bezüglich verschiedener Ausgangs- und Zielbasen. Dies führt zu einer besonderen Darstellungsmatrix. Im Folgenden sei K ein Körper.

Definition 5.71. Sei $(V,+,\cdot)$ ein endlichdimensionaler Vektorraum über K mit geordneten Basen $\mathfrak{B}, \mathfrak{C}$. Die Matrix

$$S_{\mathfrak{C}\leftarrow\mathfrak{B}} \coloneqq [\mathrm{id}_V]_{\mathfrak{C}\leftarrow\mathfrak{B}}$$

heißt *Basiswechselmatrix* von $\mathfrak{B}$ nach $\mathfrak{C}$.

Beobachtung 5.72. Für geordnete Basen $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \dots, \boldsymbol{b}^{(n)})$ und $\mathfrak{C}$ eines endlichdimensionalen Vektorraums $(V,+,\cdot)$ über K und jeden Vektor $\boldsymbol{v} \in V$ gilt

$$S_{\mathfrak{C}\leftarrow\mathfrak{B}} = \big([\boldsymbol{b}^{(1)}]_{\mathfrak{C}}, \dots, [\boldsymbol{b}^{(n)}]_{\mathfrak{C}}\big), \qquad \text{und} \qquad [\boldsymbol{v}]_{\mathfrak{C}} = S_{\mathfrak{C}\leftarrow\mathfrak{B}} [\boldsymbol{v}]_{\mathfrak{B}}.$$

Beweis. Die Spalten der Darstellungsmatrix sind nach Definition 5.38 gegeben durch $[\mathrm{id}_V(\boldsymbol{b}^{(j)})]_{\mathfrak{C}} = [\boldsymbol{b}^{(j)}]_{\mathfrak{C}}$. Außerdem gilt

$$[\boldsymbol{v}]_{\mathfrak{C}} = [\mathrm{id}_V(\boldsymbol{v})]_{\mathfrak{C}} \overset{5.40}{=} [\mathrm{id}_V]_{\mathfrak{C}\leftarrow\mathfrak{B}} [\boldsymbol{v}]_{\mathfrak{B}} = S_{\mathfrak{C}\leftarrow\mathfrak{B}} [\boldsymbol{v}]_{\mathfrak{B}}. \qquad \square$$

Die Spalten der Basiswechselmatrix sind also die Koordinaten der alten Basisvektoren in der neuen Basis. Bei verschiedenen Basen ist die Darstellungsmatrix der Identität im Allgemeinen keine Einheitsmatrix, wie bereits Beispiel 5.70 zeigt. Ein Basiswechsel lässt sich jedoch stets rückgängig machen.

Korollar 5.73. Für zwei geordnete Basen $\mathfrak{B}, \mathfrak{C}$ eines endlichdimensionalen Vektorraums über K ist $S_{\mathfrak{C}\leftarrow\mathfrak{B}}$ invertierbar und es gilt

$$\left(S_{\mathfrak{C}\leftarrow\mathfrak{B}}\right)^{-1} = S_{\mathfrak{B}\leftarrow\mathfrak{C}}.$$

Beweis. Die Identität ist ein Automorphismus und ihre eigene Umkehrabbildung. Nach Korollar 5.63 ist ihre Darstellungsmatrix daher invertierbar, und wir erhalten

$$\left(S_{\mathfrak{C}\leftarrow\mathfrak{B}}\right)^{-1} = \left([\mathrm{id}_V]_{\mathfrak{C}\leftarrow\mathfrak{B}}\right)^{-1} \overset{5.63}{=} [\mathrm{id}_V^{-1}]_{\mathfrak{B}\leftarrow\mathfrak{C}} = [\mathrm{id}_V]_{\mathfrak{B}\leftarrow\mathfrak{C}} = S_{\mathfrak{B}\leftarrow\mathfrak{C}}. \qquad \square$$

Beispiel 5.74. *Die beiden Matrizen aus Beispiel 5.70 sind*

$$S_{\mathfrak{E}\leftarrow\mathfrak{B}} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \qquad S_{\mathfrak{B}\leftarrow\mathfrak{E}} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}.$$

Die erste rechnet Koordinaten bezüglich $\mathfrak{B}$ in Standardkoordinaten um, die zweite übernimmt den Rückweg. Nach Korollar 5.73 sind sie zueinander invers:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1\cdot 0 + 1\cdot 1 & 1\cdot 1 + 1\cdot(-1) \\ 1\cdot 0 + 0\cdot 1 & 1\cdot 1 + 0\cdot(-1) \end{pmatrix} = \mathbb{I}_2,$$
$$\begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0\cdot 1 + 1\cdot 1 & 0\cdot 1 + 1\cdot 0 \\ 1\cdot 1 + (-1)\cdot 1 & 1\cdot 1 + (-1)\cdot 0 \end{pmatrix} = \mathbb{I}_2. \qquad \triangle$$

Nun wechseln wir die Basen für eine lineare Abbildung $f\colon V \to W$. Wir kennen ihre Darstellungsmatrix bezüglich der alten Basen $\mathfrak{B}$ von V und $\mathfrak{C}$ von W und möchten stattdessen $\mathfrak{B}'$ und $\mathfrak{C}'$ verwenden. Dazu rechnen wir zunächst neue Eingabekoordinaten in alte um, wenden dann die bekannte Darstellungsmatrix an und wechseln zuletzt von alten zu neuen Ausgabekoordinaten. Für einen festen Vektor $\boldsymbol{v} \in V$ sind dies die drei Schritte

$$\begin{aligned} [\boldsymbol{v}]_{\mathfrak{B}} &= S_{\mathfrak{B}\leftarrow\mathfrak{B}'}[\boldsymbol{v}]_{\mathfrak{B}'}, \\ [f(\boldsymbol{v})]_{\mathfrak{C}} &= [f]_{\mathfrak{C}\leftarrow\mathfrak{B}}[\boldsymbol{v}]_{\mathfrak{B}}, \\ [f(\boldsymbol{v})]_{\mathfrak{C}'} &= S_{\mathfrak{C}'\leftarrow\mathfrak{C}}[f(\boldsymbol{v})]_{\mathfrak{C}}. \end{aligned}$$

Setzen wir die Gleichungen nacheinander ein, steht die zuerst angewandte Matrix im Produkt rechts.

Satz 5.75. Sei $f\colon V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen $(V, +, \cdot)$ und $(W, +, \cdot)$ über K. Für geordnete Basen $\mathfrak{B}, \mathfrak{B}'$ von V und $\mathfrak{C}, \mathfrak{C}'$ von W gilt

$$[f]_{\mathfrak{C}'\leftarrow\mathfrak{B}'} = S_{\mathfrak{C}'\leftarrow\mathfrak{C}}\,[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}\,S_{\mathfrak{B}\leftarrow\mathfrak{B}'}.$$

Beweis. Die drei Koordinatenschritte entsprechen der Verkettung $f = \mathrm{id}_W \circ f \circ \mathrm{id}_V$. Für id_V verwenden wir $\mathfrak{B}'$ als Ausgangs- und $\mathfrak{B}$ als Zielbasis; für id_W entsprechend $\mathfrak{C}$ und $\mathfrak{C}'$. Mit der Verkettungsregel erhalten wir

$$\begin{aligned} [f]_{\mathfrak{C}'\leftarrow\mathfrak{B}'} &= [\mathrm{id}_W \circ f \circ \mathrm{id}_V]_{\mathfrak{C}'\leftarrow\mathfrak{B}'} \\ &\overset{5.49}{=} [\mathrm{id}_W]_{\mathfrak{C}'\leftarrow\mathfrak{C}}\,[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}\,[\mathrm{id}_V]_{\mathfrak{B}\leftarrow\mathfrak{B}'} \\ &= S_{\mathfrak{C}'\leftarrow\mathfrak{C}}\,[f]_{\mathfrak{C}\leftarrow\mathfrak{B}}\,S_{\mathfrak{B}\leftarrow\mathfrak{B}'}. \end{aligned} \qquad \square$$

Wenn eine Abbildung einen Vektorraum in sich selbst abbildet, können wir für Eingabe und Ausgabe dieselbe Basis verwenden. Dann sind die beiden Koordinatenwechsel zueinander invers.

Korollar 5.76. Sei $f\colon V \to V$ eine lineare Abbildung eines endlichdimensionalen Vektorraums $(V, +, \cdot)$ über K mit geordneten Basen $\mathfrak{B}, \mathfrak{B}'$. Setzen wir $A \coloneqq [f]_{\mathfrak{B}\leftarrow\mathfrak{B}}$ und $S \coloneqq S_{\mathfrak{B}\leftarrow\mathfrak{B}'}$, so gilt

$$[f]_{\mathfrak{B}'\leftarrow\mathfrak{B}'} = S^{-1}AS.$$

Beweis. Es gilt

$$[f]_{\mathfrak{B}'\leftarrow\mathfrak{B}'} \overset{5.75}{=} S_{\mathfrak{B}'\leftarrow\mathfrak{B}}\,[f]_{\mathfrak{B}\leftarrow\mathfrak{B}}\,S_{\mathfrak{B}\leftarrow\mathfrak{B}'} \overset{5.73}{=} S^{-1}AS. \qquad \square$$

Dabei übersetzt S neue in alte Koordinaten; anschließend wenden wir A und zuletzt S^{-1} an. Für diesen Basiswechsel muss f nicht bijektiv sein. Wir nutzen ihn nun, um die Matrizen zweier bekannter geometrischer Abbildungen zu vereinfachen.

Beispiel 5.77. *Wir betrachten die Orthogonalprojektion π_G auf die Gerade $G = \langle(1,1)^\top\rangle$ und die Spiegelung s_G an G aus Beispiel 5.35. Bezüglich der geordneten Standardbasis $\mathfrak{E}$ haben sie die Matrizen*

$$P = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}, \qquad R = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

Wir wechseln zur geordneten Orthonormalbasis $\mathfrak{Q} = (\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)})$ aus Beispiel 4.100 mit

$$\boldsymbol{q}^{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \qquad \boldsymbol{q}^{(2)} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Der erste Basisvektor liegt auf G, der zweite steht senkrecht auf G. Daher gilt

$$\pi_G(\boldsymbol{q}^{(1)}) = \boldsymbol{q}^{(1)}, \qquad \pi_G(\boldsymbol{q}^{(2)}) = \boldsymbol{0}, \qquad s_G(\boldsymbol{q}^{(1)}) = \boldsymbol{q}^{(1)}, \qquad s_G(\boldsymbol{q}^{(2)}) = -\boldsymbol{q}^{(2)}.$$

Die Koordinaten dieser Basisbilder liefern unmittelbar

$$[\pi_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}, \qquad [s_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

In diesen Koordinaten behält die Projektion die erste Koordinate bei und setzt die zweite auf null; die Spiegelung kehrt das Vorzeichen der zweiten Koordinate um.

Wir überprüfen die beiden Darstellungsmatrizen mithilfe der Basiswechselmatrix. Ihre Spalten sind die Standardkoordinaten der neuen Basisvektoren, also

$$Q \coloneqq S_{\mathfrak{E}\leftarrow\mathfrak{Q}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}.$$

Hier ist $Q^2 = \mathbb{I}_2$, also $Q^{-1} = Q$. Die Gleichheit von Hin- und Rückwechsel ist eine Besonderheit dieser Basis; in Beispiel 5.74 waren die Matrizen verschieden. Nach Korollar 5.76 erhalten wir

$$[\pi_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}} = Q^{-1}PQ = \frac{1}{4}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

Für die Spiegelung erhalten wir entsprechend

$$[s_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}} = Q^{-1}RQ = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Wir projizieren und spiegeln nun den Punkt $\boldsymbol{x} = (2,1)^\top$ aus Beispiel 5.70. In Standardkoordinaten ergibt sich

$$[\pi_G(\boldsymbol{x})]_{\mathfrak{E}} = P[\boldsymbol{x}]_{\mathfrak{E}} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}3\\ 3\end{pmatrix},$$
$$[s_G(\boldsymbol{x})]_{\mathfrak{E}} = R[\boldsymbol{x}]_{\mathfrak{E}} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix} = \begin{pmatrix}1\\ 2\end{pmatrix}.$$

Für die Rechnung bezüglich $\mathfrak{Q}$ bestimmen wir zunächst die Koordinaten desselben Punkts:

$$[\boldsymbol{x}]_{\mathfrak{Q}} = Q^{-1}[\boldsymbol{x}]_{\mathfrak{E}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}3\\ 1\end{pmatrix}.$$

Die neuen Darstellungsmatrizen liefern damit

$$[\pi_G(\boldsymbol{x})]_{\mathfrak{Q}} = [\pi_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}}[\boldsymbol{x}]_{\mathfrak{Q}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}3\\ 1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}3\\ 0\end{pmatrix},$$
$$[s_G(\boldsymbol{x})]_{\mathfrak{Q}} = [s_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}}[\boldsymbol{x}]_{\mathfrak{Q}} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}3\\ 1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}3\\ -1\end{pmatrix}.$$

Zum Vergleich rechnen wir diese Ergebnisse mit Q in Standardkoordinaten zurück:

$$Q[\pi_G(\boldsymbol{x})]_{\mathfrak{Q}} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}3\\ 0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}3\\ 3\end{pmatrix} = [\pi_G(\boldsymbol{x})]_{\mathfrak{E}},$$
$$Q[s_G(\boldsymbol{x})]_{\mathfrak{Q}} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}3\\ -1\end{pmatrix} = \begin{pmatrix}1\\ 2\end{pmatrix} = [s_G(\boldsymbol{x})]_{\mathfrak{E}}.$$

Beide Rechnungen liefern also denselben Projektionspunkt $\pi_G(\boldsymbol{x}) = (\frac{3}{2}, \frac{3}{2})^\top$ und denselben Spiegelpunkt $s_G(\boldsymbol{x}) = (1,2)^\top$. Die Koordinaten der Ergebnisse unterscheiden sich, die Punkte selbst stimmen überein. △

In Beispiel 5.77 beschreiben die Matrizen P und $[\pi_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}}$ dieselbe Abbildung in unterschiedlichen Koordinaten und genauso R und $[s_G]_{\mathfrak{Q}\leftarrow\mathfrak{Q}}$. Wir geben dieser Beziehung zwischen quadratischen Matrizen einen Namen.

Definition 5.78. Für $n \in \mathbb{N}$ heißen zwei Matrizen $A, B \in K^{n\times n}$ *ähnlich*, wenn eine invertierbare Matrix $S \in K^{n\times n}$ existiert mit

$$B = S^{-1}AS.$$

Tatsächlich treten ähnliche Matrizen genau als Darstellungsmatrizen einer gemeinsamen Abbildung auf.

Satz 5.79. Zwei Matrizen $A, B \in K^{n\times n}$ sind genau dann ähnlich, wenn es einen n-dimensionalen Vektorraum $(V,+,\cdot)$ über K, eine lineare Abbildung $f\colon V \to V$ und geordnete Basen $\mathfrak{B}, \mathfrak{C}$ von V gibt mit

$$A = [f]_{\mathfrak{B}\leftarrow\mathfrak{B}}, \qquad B = [f]_{\mathfrak{C}\leftarrow\mathfrak{C}}.$$

Beweis. Sind A und B solche Darstellungsmatrizen, so folgt mit $S \coloneqq S_{\mathfrak{B}\leftarrow\mathfrak{C}}$ aus Korollar 5.76 die Gleichung $B = S^{-1}AS$. Die Wechselmatrix ist nach Korollar 5.73 invertierbar; damit sind A und B ähnlich.

Seien umgekehrt A und B ähnlich. Dann existiert eine invertierbare Matrix S mit $B = S^{-1}AS$. Wir wählen $V = K^n$ und $f = f_A$. Als $\mathfrak{B}$ wählen wir die geordnete Standardbasis von K^n und setzen $\mathfrak{C} \coloneqq (\boldsymbol{S_{\cdot 1}}, \ldots, \boldsymbol{S_{\cdot n}})$. Da S invertierbar ist, ist $\mathfrak{C}$ nach der Implikation $(i) \Rightarrow (iv)$ aus Satz 5.65 eine geordnete Basis von K^n. Die Spalten von S sind die Koordinatenvektoren der Basisvektoren aus $\mathfrak{C}$ bezüglich der Standardbasis $\mathfrak{B}$, nach Beobachtung 5.72 ist also $S = S_{\mathfrak{B}\leftarrow\mathfrak{C}}$. Damit folgt

$$[f_A]_{\mathfrak{B}\leftarrow\mathfrak{B}} = A, \qquad [f_A]_{\mathfrak{C}\leftarrow\mathfrak{C}} \overset{5.76}{=} S^{-1}AS = B. \qquad \square$$

Bei jeder der beiden Darstellungen wird im Ausgangs- und Zielraum dieselbe Basis verwendet. Die Abbildung f bleibt beim Basiswechsel unverändert. Die beiden Abbildungen f_A und f_B auf Koordinatenvektoren können dagegen verschieden sein: Sie beschreiben die Wirkung von f in unterschiedlichen Koordinaten.

Beobachtung 5.80. Für jedes $n \in \mathbb{N}$ ist Ähnlichkeit eine Äquivalenzrelation auf $K^{n\times n}$.

Beweis. Übung. $\square$

Ein Basiswechsel verändert die Darstellung einer Abbildung. Eigenschaften, die unabhängig von den Koordinaten sind, müssen dabei erhalten bleiben. Wir zeigen dies für den Rang und die Invertierbarkeit.

Satz 5.81. Sind $A, B \in K^{n\times n}$ ähnlich, so gilt $\operatorname{rang} A = \operatorname{rang} B$. Außerdem ist A genau dann invertierbar, wenn B invertierbar ist.

Beweis. Sei $B = S^{-1}AS$ mit invertierbarem S. Nach Beobachtung 5.48 gilt $f_B = f_{S^{-1}} \circ f_A \circ f_S$. Wegen der Surjektivität von f_S folgt

$$f_B(K^n) = f_{S^{-1}}\big(f_A(f_S(K^n))\big) = f_{S^{-1}}\big(f_A(K^n)\big).$$

Nach dieser Gleichheit ist die Abbildung

$$g\colon f_A(K^n) \to f_B(K^n), \qquad \boldsymbol{y} \mapsto f_{S^{-1}}(\boldsymbol{y}),$$

wohldefiniert (jeder Wert liegt im Zielraum) und surjektiv. Als Einschränkung der linearen Abbildung $f_{S^{-1}}$ ist auch g linear. Da S^{-1} invertierbar ist, ist $f_{S^{-1}}$ nach Satz 5.61 injektiv und damit auch seine Einschränkung g. Somit ist g ein Isomorphismus zwischen den beiden Bildräumen. Nach Korollar 5.25 stimmen deren Dimensionen überein, und nach Definition des Rangs erhalten wir

$$\operatorname{rang} A = \dim f_A(K^n) = \dim f_B(K^n) = \operatorname{rang} B.$$

Ist A invertierbar, so ist nach Satz 5.68 auch das Produkt $B = S^{-1}AS$ invertierbar. Die Umkehrung folgt ebenso aus $A = SBS^{-1}$. $\square$

5.7 Anwendungsbeispiel: Bildcodierung

Wir veranschaulichen nun das Grundprinzip der *Haar-Wavelet-Transformation*, die in der Bildcodierung verwendet wird.

Wir wollen ein Bild näherungsweise durch weniger Information beschreiben. Dazu nutzen wir, dass benachbarte Bildpunkte häufig ähnliche Helligkeitswerte haben. Ein Graustufenbild besteht aus Bildpunkten, den sogenannten Pixeln, denen jeweils ein Helligkeitswert zwischen null und eins zugeordnet ist: Null bedeutet schwarz, eins weiß, und die Werte dazwischen beschreiben verschiedene Grautöne. Ist das Bild w Pixel breit und h Pixel hoch, so enthält es $w \cdot h$ solche Werte. Lesen wir diese zeilenweise aus, erhalten wir einen Vektor in $[0,1]^{w \cdot h} \subseteq \mathbb{R}^{w \cdot h}$.

Für die Rechnung betrachten wir zunächst nur zwei benachbarte Pixel mit den Helligkeitswerten $a, b \in [0,1]$. Wir fassen sie zu einem Vektor $\boldsymbol{x} = (a,b)^\top$ zusammen. Statt beide Werte einzeln zu betrachten, verwenden wir ihren Mittelwert und ihre halbe Differenz:

$$m \coloneqq \frac{a+b}{2}, \qquad d \coloneqq \frac{a-b}{2}.$$

Der Mittelwert m beschreibt die gemeinsame Helligkeit der beiden Pixel, während d ihren Unterschied erfasst. Wegen

$$\begin{pmatrix} a \\ b \end{pmatrix} = m \begin{pmatrix} 1 \\ 1 \end{pmatrix} + d \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

sind $(m,d)^\top$ die Koordinaten von $\boldsymbol{x}$ bezüglich der geordneten Basis $\mathfrak{B} = ((1,1)^\top, (1,-1)^\top)$. Mit der geordneten Standardbasis $\mathfrak{E}$ gilt

$$T \coloneqq S_{\mathfrak{B} \leftarrow \mathfrak{E}} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad T^{-1} = S_{\mathfrak{E} \leftarrow \mathfrak{B}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Nach Beobachtung 5.72 liefert $T\boldsymbol{x}$ die neuen Koordinaten $(m,d)^\top$. Korollar 5.73 garantiert, dass der Basiswechsel umkehrbar ist und somit keine Information verliert. Aus m, d erhalten wir $a = m + d$ und $b = m - d$ zurück.

Bei ähnlichen Nachbarwerten ist $|d|$ klein. Wir ersetzen d durch null, falls $|d| < 0{,}02$, und behalten d andernfalls bei. Anschließend rechnen wir in die ursprünglichen Koordinaten zurück. Beispielsweise erhalten wir

$$\begin{pmatrix} 0{,}40 \\ 0{,}42 \end{pmatrix} \overset{T}{\longmapsto} \begin{pmatrix} 0{,}41 \\ -0{,}01 \end{pmatrix} \overset{D}{\longmapsto} \begin{pmatrix} 0{,}41 \\ 0 \end{pmatrix} \overset{T^{-1}}{\longmapsto} \begin{pmatrix} 0{,}41 \\ 0{,}41 \end{pmatrix} \qquad \text{mit} \qquad D \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

In den ursprünglichen Koordinaten würde das Nullsetzen eines Eintrags, also die Anwendung von D, einen Pixel schwarz färben. In den neuen Koordinaten verwerfen wir mit $d = -0{,}01$ nur den geringen Helligkeitsunterschied, während die gemeinsame Helligkeit $m = 0{,}41$ erhalten bleibt. Nach Korollar 5.76 können wir das Nullsetzen der Differenz in den ursprünglichen Koordinaten durch die Matrix

$$P = T^{-1} D T = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

beschreiben. Dies ist die Orthogonalprojektion aus Beispiel 5.35. Wir wechseln also mit T die Basis, setzen mit D die zweite Koordinate auf null und wechseln mit T^{-1} zurück. Der Informationsverlust lässt sich auch am Rang ablesen: Da P und D ähnlich sind, gilt

$$\operatorname{rang} P \overset{5.81}{=} \operatorname{rang} D = 1.$$

Von den beiden unabhängigen Koordinaten bleibt also nur eine erhalten; die ursprüngliche Differenz lässt sich aus dem projizierten Pixelpaar nicht zurückgewinnen.

Für ein ganzes Bild mit gerader Breite w teilen wir jede Zeile in Paare benachbarter Pixel auf und wenden denselben Koordinatenwechsel auf jedes Paar an. Die Matrizen für das ganze Bild lauten dann

$$T_{\text{Bild}} = \begin{pmatrix} T & & 0 \\ & \ddots & \\ 0 & & T \end{pmatrix}, \qquad D_{\text{Bild}} = \begin{pmatrix} D & & 0 \\ & \ddots & \\ 0 & & D \end{pmatrix}, \qquad P_{\text{Bild}} = \begin{pmatrix} P & & 0 \\ & \ddots & \\ 0 & & P \end{pmatrix},$$

wobei jeweils $\frac{w \cdot h}{2}$ Blöcke der Größe 2×2 auf der Diagonale stehen; alle Einträge außerhalb dieser Blöcke sind null. Dieser Basiswechsel ist umkehrbar, denn aus jedem Paar (m, d) können wir die ursprünglichen Werte $a = m + d$ und $b = m - d$ zurückgewinnen. Ersetzen wir in den neuen Koordinaten sämtliche Differenzen durch null, so erhalten wir nach dem Rückwechsel die Orthogonalprojektion mit der Matrix P_{Bild}. Sie ersetzt in jedem Pixelpaar beide Helligkeitswerte durch ihren Mittelwert. Beim oben beschriebenen Verfahren setzen wir nur die kleinen Differenzen auf null und erhalten so beim Rückwechsel ein angenähertes Bild. Bei vier Pixeln ergibt sich beispielsweise

$$\begin{pmatrix} 0{,}40 \\ 0{,}42 \\ 0{,}80 \\ 0{,}78 \end{pmatrix} \overset{T_{\text{Bild}}}{\longmapsto} \begin{pmatrix} 0{,}41 \\ -0{,}01 \\ 0{,}79 \\ 0{,}01 \end{pmatrix} \overset{D_{\text{Bild}}}{\longmapsto} \begin{pmatrix} 0{,}41 \\ 0 \\ 0{,}79 \\ 0 \end{pmatrix} \overset{T_{\text{Bild}}^{-1}}{\longmapsto} \begin{pmatrix} 0{,}41 \\ 0{,}41 \\ 0{,}79 \\ 0{,}79 \end{pmatrix}.$$

Der Helligkeitsunterschied zwischen den beiden Pixelpaaren bleibt erhalten; nur die kleinen Unterschiede innerhalb der Paare gehen verloren. Sind viele benachbarte Pixel ähnlich hell, entstehen dabei viele Nullkoordinaten. Wir können dann Speicherplatz sparen, indem wir nur die von null verschiedenen Einträge zusammen mit ihren Positionen speichern.

Abbildung 5.1 zeigt die Wirkung verschiedener Schwellwerte auf dasselbe Bild. Für die Darstellung verwenden wir das vollständige Motiv mit 64×64 Pixeln und vergrößern es, damit die Veränderungen sichtbar werden. Größere Schwellwerte führen zu mehr Nullkoordinaten und größerer Speicherersparnis, entfernen aber auch deutlichere Helligkeitsunterschiede.

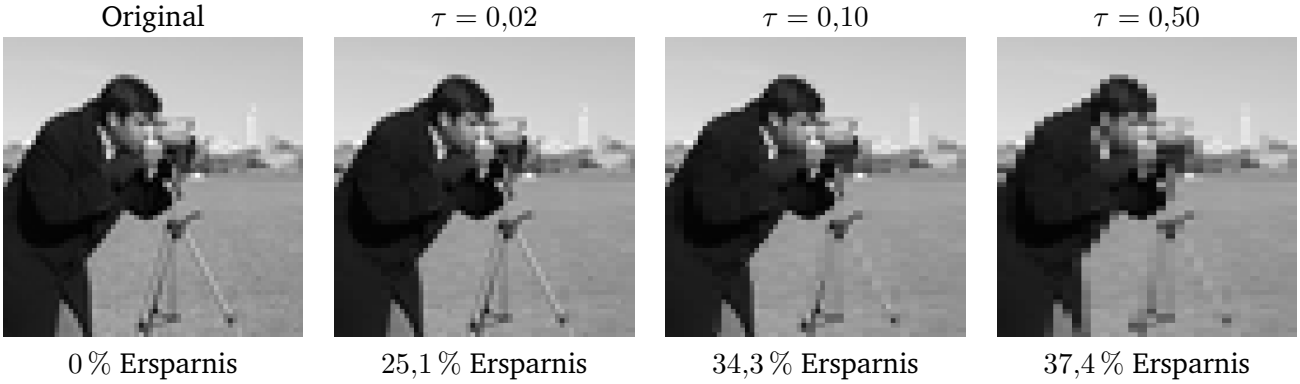


Abbildung 5.1: Bildnäherung durch das Nullsetzen von Differenzen mit $|d| < \tau$. Die Speicherersparnis gilt gegenüber unkomprimierten 8-Bit-Pixeln bei Speicherung der Mittelwerte, einer Positionsmaske und der verbleibenden Differenzen. Foto: Lav Varshney (scikit-image), CC0.

6 Lineare Gleichungssysteme und Eigenwerte

Lineare Gleichungssysteme sind in der Informatik von zentraler Bedeutung: Sie treten etwa in der Computergrafik bei der Berechnung von Schnittpunkten, in der medizinischen Bildverarbeitung bei der Rekonstruktion von Computertomographiebildern, in der technischen Informatik bei der Simulation elektrischer Schaltungen, in der Codierungstheorie bei der Wiederherstellung fehlender Daten, im wissenschaftlichen Rechnen bei der Berechnung von Kräften in Tragwerken und der Simulation von Wärmeverteilungen, im maschinellen Lernen bei der Anpassung linearer Vorhersagemodelle an Trainingsdaten, in der Robotik bei der Berechnung von Gelenkgeschwindigkeiten für eine gewünschte Greiferbewegung und bei vielen weiteren Aufgaben auf. Uns sind sie bereits mehrfach begegnet: beim Gerade-Ebene-Schnitt in Beispiel 4.7, bei der Bestimmung von Kernen und Urbildern in Beispiel 5.22 und bei der Rekonstruktion einer Funktion aus ihren Werten in Beispiel 5.64. Bisher haben wir diese Systeme durch passende Umformungen gelöst. Nun entwickeln wir ein allgemeines Verfahren, mit dem wir Lösbarkeit und Eindeutigkeit prüfen und sämtliche Lösungen bestimmen können.

Ein solches Lösungsverfahren hilft uns auch, wiederholte Anwendungen linearer Abbildungen besser zu verstehen und effizienter zu berechnen. Das ist beispielsweise wichtig, wenn wir die Entwicklung einer Population über viele Generationen, die Häufigkeit von Webseitenbesuchen bei zufälligen Klicks oder die Veränderung eines Bildes durch wiederholte Anwendung eines Filters untersuchen. Eine passende Basis kann diese Untersuchung wesentlich vereinfachen. Bei der Projektion und Spiegelung aus Beispiel 5.77 haben wir bereits Richtungen gefunden, in denen die Abbildung nur mit einem festen Faktor multipliziert. Bilden solche Richtungen eine Basis, können wir jede Koordinate einzeln betrachten und das Ergebnis nach vielen Schritten durch Potenzen der jeweiligen Faktoren berechnen. Im weiteren Verlauf des Kapitels untersuchen wir, wann eine solche Basis existiert und wie wir sie mithilfe linearer Gleichungssysteme bestimmen können.

6.1 Lineare Gleichungssysteme

Die Gleichung $A\boldsymbol{x} = \boldsymbol{b}$ tritt beim Schneiden von Geraden und Ebenen sowie beim Bestimmen von Kernen, Urbildern und Koordinaten auf: Wir suchen ein Urbild von $\boldsymbol{b}$ unter f_A, also Koeffizienten, mit denen sich $\boldsymbol{b}$ als Linearkombination der Spalten von A darstellen lässt. Bilden die Spalten in ihrer Reihenfolge eine geordnete Basis von K^m, so suchen wir gerade die Koordinaten von $\boldsymbol{b}$ bezüglich dieser Basis. Zunächst führen wir die Begriffe ein, mit denen wir solche Gleichungen und ihre Lösungen beschreiben.

Definition 6.1. Ein *lineares Gleichungssystem* über einem Körper K mit $m \in \mathbb{N}$ Gleichungen und $n \in \mathbb{N}$ Unbekannten $x_1, \dots, x_n$ hat die Form

$$\begin{aligned} \textstyle\sum_{j=1}^{n} A_{1j} x_j &= b_1, \\ \textstyle\sum_{j=1}^{n} A_{2j} x_j &= b_2, \\ &\vdots \\ \textstyle\sum_{j=1}^{n} A_{mj} x_j &= b_m, \end{aligned}$$

mit einer vorgegebenen *Koeffizientenmatrix* $A \in K^{m \times n}$ und *rechten Seite* $\boldsymbol{b} \in K^m$. Für $\boldsymbol{b} = \mathbf{0}$ heißt das System *homogen*, sonst *inhomogen*. Jeder Vektor $\boldsymbol{x} \in K^n$, der alle Gleichungen erfüllt, heißt *Lösung* und die Menge aller Lösungen heißt *Lösungsmenge* des Systems.
Wir schreiben das Gleichungssystem kurz als $A\boldsymbol{x} = \boldsymbol{b}$ und definieren die *erweiterte Koeffizientenmatrix*

$$(A \mid \boldsymbol{b}) \coloneqq \left(\begin{array}{ccc|c} A_{11} & \dots & A_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ A_{m1} & \dots & A_{mn} & b_m \end{array}\right) \in K^{m \times (n+1)}.$$

Beispiel 6.2. *Beim Gerade-Ebene-Schnitt aus Beispiel 4.7 haben wir das System*

$$\begin{aligned} s + t - r &= 0, \\ s + 2t - r &= 1, \\ 2t - r &= 0 \end{aligned} \qquad \textit{also} \qquad \underbrace{\begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -1 \\ 0 & 2 & -1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} s \\ t \\ r \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}_{\boldsymbol{b}}$$

gelöst. Dort haben wir $\boldsymbol{x} = (1,1,2)^\top$ als einzige Lösung bestimmt. Dieser Vektor enthält die Parameter der Geraden und der Ebene; Einsetzen in deren Darstellungen ergibt den Schnittpunkt $(2,3,3)^\top$. In der Sprache der linearen Abbildungen ist also $f_A^{-1}(\{\boldsymbol{b}\}) = \{(1,1,2)^\top\}$. △

Im Allgemeinen muss eine Lösung weder existieren noch eindeutig sein. Die Lösungsmenge ist ein Urbild unter f_A. Nach Satz 5.18 ist jedes nichtleere Urbild eine Verschiebung des Kerns.

Korollar 6.3. Ist $\boldsymbol{x}^{(0)} \in K^n$ mit $n \in \mathbb{N}$ Lösung eines linearen Gleichungssystems $A\boldsymbol{x} = \boldsymbol{b}$ über K, dann ist die Lösungsmenge des Gleichungssystems gegeben durch

$$\{\boldsymbol{x} \in K^n \mid A\boldsymbol{x} = \boldsymbol{b}\} = \boldsymbol{x}^{(0)} + \ker A.$$

Insbesondere ist $\ker A$ die Lösungsmenge des homogenen Systems $A\boldsymbol{x} = \mathbf{0}$ und jede nichtleere Lösungsmenge ein affiner Raum der Dimension $n - \operatorname{rang} A$.

Beweis. Die Lösungen sind genau diejenigen Vektoren $\boldsymbol{x} \in K^n$, für die $f_A(\boldsymbol{x}) = A\boldsymbol{x} = \boldsymbol{b}$ gilt. Die Lösungsmenge ist also das Urbild $f_A^{-1}(\{\boldsymbol{b}\})$. Nach Satz 5.18 ist $f_A^{-1}(\{\boldsymbol{b}\}) = \boldsymbol{x}^{(0)} + \ker(f_A) = \boldsymbol{x}^{(0)} + \ker A$ ein affiner Raum, da K^n endlichdimensional ist. Im homogenen Fall können wir $\boldsymbol{x}^{(0)} = \mathbf{0}$ wählen und erhalten als Lösungsmenge $\mathbf{0} + \ker A = \ker A$. Korollar 5.21 liefert $\dim f_A^{-1}(\{\boldsymbol{b}\}) = \dim K^n - \operatorname{rang}(f_A)$, also gilt die Dimensionsaussage wegen $\dim K^n = n$ und $\operatorname{rang}(f_A) = \operatorname{rang} A$. □

Für ein lösbares System genügt es somit, eine einzige Lösung $\boldsymbol{x}^{(0)}$ und den Kern von A zu bestimmen. Alle Lösungen erhalten wir, indem wir zu $\boldsymbol{x}^{(0)}$ jeweils einen Vektor aus $\ker A$ addieren. Ist $\{\boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(d)}\}$ eine Basis von $\ker A$, so hat jede Lösung nach Satz 4.55 genau eine Darstellung

$$\boldsymbol{x} = \boldsymbol{x}^{(0)} + t_1\boldsymbol{u}^{(1)} + \cdots + t_d\boldsymbol{u}^{(d)} \qquad \text{mit } t_1, \dots, t_d \in K.$$

Die Koeffizienten $t_1, \dots, t_d$ können wir unabhängig voneinander wählen; wir nennen sie *freie Parameter* der Lösungsdarstellung. Sie beschreiben die Freiheitsgrade der Lösungsmenge, wie zuvor die Parameter einer Geraden oder Ebene. Ihre Anzahl $d = n - \operatorname{rang} A$ stimmt mit der Dimension der Lösungsmenge überein.

Der Rang gibt auch Auskunft darüber, ob überhaupt eine Lösung existiert. Dafür muss $\boldsymbol{b}$ eine Linearkombination der Spalten von A sein. Das Anhängen von $\boldsymbol{b}$ als zusätzliche Spalte verändert dann die lineare Hülle und damit den Rang nicht. Andernfalls erhöht sich der Rang um eins. Die Lösbarkeit lässt sich daher durch Vergleich der Ränge von A und $(A \mid \boldsymbol{b})$ prüfen.

Satz 6.4. Sei $A \in K^{m\times n}$ und $\boldsymbol{b} \in K^m$. Die folgenden Aussagen sind äquivalent:

(i) Das System $A\boldsymbol{x} = \boldsymbol{b}$ ist lösbar.

(ii) Es gilt $\boldsymbol{b} \in \langle \boldsymbol{A}_{\cdot\boldsymbol{1}}, \dots, \boldsymbol{A}_{\cdot\boldsymbol{n}} \rangle$.

(iii) Es gilt $\operatorname{rang}(A \mid \boldsymbol{b}) = \operatorname{rang} A$.

Ein lösbares System $A\boldsymbol{x} = \boldsymbol{b}$ hat genau dann eine eindeutige Lösung, wenn $\operatorname{rang} A = n$ ist.

Beweis. Es ist $f_A(\boldsymbol{x}) = A\boldsymbol{x} = \boldsymbol{b}$ lösbar genau dann, wenn $\boldsymbol{b} \in f_A(K^n)$. Nach Beobachtung 5.33 ist dabei $f_A(K^n) = \langle \boldsymbol{A}_{\cdot\boldsymbol{1}}, \dots, \boldsymbol{A}_{\cdot\boldsymbol{n}} \rangle$, also sind (i) und (ii) äquivalent.

Setzen wir $U \coloneqq \langle \boldsymbol{A}_{\cdot\boldsymbol{1}}, \dots, \boldsymbol{A}_{\cdot\boldsymbol{n}} \rangle$ und $W \coloneqq \langle \boldsymbol{A}_{\cdot\boldsymbol{1}}, \dots, \boldsymbol{A}_{\cdot\boldsymbol{n}}, \boldsymbol{b} \rangle$, so gilt $U \subseteq W$ und $U = W$ genau dann, wenn $\boldsymbol{b} \in U$. Nach Satz 4.74 ist $U = W$ äquivalent zu $\dim U = \dim W$, also zu $\operatorname{rang} A = \operatorname{rang}(A \mid \boldsymbol{b})$. Also sind (ii) und (iii) äquivalent.

Ist das System lösbar, so ist seine Lösungsmenge nach Korollar 6.3 eine Verschiebung von $\ker A$. Sie enthält genau dann nur einen Vektor, wenn $\ker A = \{\boldsymbol{0}\}$ ist, also wenn $\dim \ker A = 0$ gilt. Nach der Dimensionsformel aus Satz 5.20 ist dies gleichbedeutend mit $\operatorname{rang} A = n$. □

Bei quadratischen Matrizen sind die Bedingungen für Existenz und Eindeutigkeit besonders eng miteinander verknüpft. Die bekannten Invertierbarkeitskriterien lassen sich unmittelbar als Aussagen über Gleichungssysteme lesen.

Korollar 6.5. Für $A \in K^{n\times n}$ sind die folgenden Aussagen äquivalent:

(i) A ist invertierbar.

(ii) Für jedes $\boldsymbol{b} \in K^n$ hat $A\boldsymbol{x} = \boldsymbol{b}$ genau eine Lösung.

(iii) Das homogene System $A\boldsymbol{x} = \boldsymbol{0}$ hat nur die Lösung $\boldsymbol{x} = \boldsymbol{0}$.

In diesem Fall ist $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ die eindeutige Lösung von $A\boldsymbol{x} = \boldsymbol{b}$.

Beweis. Nach Satz 5.61 ist A genau dann invertierbar, wenn f_A bijektiv ist. Dies bedeutet, dass für jedes $\boldsymbol{b} \in K^n$ genau ein $\boldsymbol{x} \in K^n$ mit $f_A(\boldsymbol{x}) = \boldsymbol{b}$, also mit $A\boldsymbol{x} = \boldsymbol{b}$, existiert. Damit sind (i) und (ii) äquivalent.

Nach Korollar 6.3 ist die Lösungsmenge des homogenen Systems $A\boldsymbol{x} = \boldsymbol{0}$ genau $\ker A$. Daher ist (iii) äquivalent zu $\ker A = \{\boldsymbol{0}\}$. Nach Satz 5.65 ist dies wiederum äquivalent zu (i).

Gilt (i), so erhalten wir für jede Lösung von $A\boldsymbol{x} = \boldsymbol{b}$

$$\boldsymbol{x} = \mathbb{I}_n \boldsymbol{x} = (A^{-1}A)\boldsymbol{x} = A^{-1}(A\boldsymbol{x}) = A^{-1}\boldsymbol{b}. \qquad \square$$

Das folgende Beispiel zeigt, wie Gleichungssysteme mit derselben Koeffizientenmatrix je nach rechter Seite keine oder mehrere Lösungen besitzen können.

Beispiel 6.6. *Wir betrachten über $\mathbb{R}$ das Gleichungssystem*

$$\begin{aligned} x_1 + x_2 + x_3 &= 2, \\ x_2 + x_3 + x_4 &= 1, \\ x_1 + 2x_2 + 2x_3 + x_4 &= 3 \end{aligned} \qquad \textit{mit} \qquad A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

Die dritte Gleichung ist die Summe der ersten beiden und liefert daher keine zusätzliche Bedingung. Durch Einsetzen sehen wir, dass $\boldsymbol{x}^{(0)} = (1, 1, 0, 0)^\top$ eine Lösung ist.

Um alle Lösungen zu bestimmen, berechnen wir den Kern von A. Nach Korollar 6.3 ist dies die Lösungsmenge des homogenen Systems $A\boldsymbol{x} = \boldsymbol{0}$. Dessen zweite Gleichung liefert $x_2 = -x_3 - x_4$ und nach Einsetzen in die erste Gleichung folgt $x_1 = x_4$. Die dritte Gleichung ist weiterhin die Summe der ersten beiden und stellt daher keine zusätzliche Bedingung. Für jede Wahl von x_3 und x_4 sind somit x_1 und x_2 eindeutig bestimmt, und alle drei Gleichungen sind erfüllt. Wir können also $x_3 = s$ und $x_4 = t$ mit beliebigen $s, t \in \mathbb{R}$ wählen und erhalten

$$\ker A = \left\{ \begin{pmatrix} t \\ -s-t \\ s \\ t \end{pmatrix} \,\middle|\, s, t \in \mathbb{R} \right\} = \left\langle \underbrace{\begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}}_{\boldsymbol{u}^{(1)}}, \underbrace{\begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix}}_{\boldsymbol{u}^{(2)}} \right\rangle.$$

Die beiden Vektoren sind linear unabhängig, denn aus $s\boldsymbol{u}^{(1)} + t\boldsymbol{u}^{(2)} = \boldsymbol{0}$ folgen an der dritten und vierten Komponente $s = t = 0$. Somit gilt $\dim \ker A = 2$ und nach der Dimensionsformel $\operatorname{rang} A = 4 - 2 = 2$. Korollar 6.3 liefert die gesamte Lösungsmenge

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix} \,\middle|\, s, t \in \mathbb{R} \right\}.$$

Sie ist eine affine Ebene in $\mathbb{R}^4$; ihre Dimension ist entsprechend dem Korollar $4 - \operatorname{rang} A = 2$.

Ersetzen wir die rechte Seite durch $\boldsymbol{b}' = (2, 1, 4)^\top$, so ist das System unlösbar: Aus den ersten beiden Gleichungen folgt, dass die linke Seite der dritten Gleichung den Wert 3 hat, dort wird aber der Wert 4 verlangt. Die rechte Seite $\boldsymbol{b}'$ liegt also nicht in der linearen Hülle der Spalten von A. Nach Satz 6.4 gilt folglich

$$\operatorname{rang}(A \mid \boldsymbol{b}) = \operatorname{rang} A = 2 < \operatorname{rang}(A \mid \boldsymbol{b}') = 3. \qquad \triangle$$

Beispiel 6.7. *Bei der Übertragung von Daten können einzelne Bits verloren gehen. Wir wollen sehen, wie zusätzliche Bits die Wiederherstellung der fehlenden Daten ermöglichen. Dazu rechnen wir über dem Körper* $\mathbb{F}_2 \coloneqq \mathbb{Z}/2\mathbb{Z}$*, dessen Elemente wir mit* 0 *und* 1 *bezeichnen. Die Addition entspricht der XOR-Verknüpfung aus Beispiel 3.14; insbesondere gilt* $1 + 1 = 0$*.*

Wir übertragen drei Datenbits $x_1, x_2, x_3 \in \mathbb{F}_2$ *und senden zusätzlich die beiden* Prüfbits

$$p_1 \coloneqq x_1 + x_2 \qquad \textit{und} \qquad p_2 \coloneqq x_2 + x_3.$$

Bei der Übertragung gehen x_1 *und* x_2 *verloren; erhalten bleiben* $x_3 = 1$ *sowie* $p_1 = 1$ *und* $p_2 = 0$*. Um die beiden fehlenden Bits wiederherzustellen, lösen wir das lineare Gleichungssystem*

$$\begin{aligned} x_1 + x_2 &= p_1 = 1, \\ x_2 + 1 &= p_2 = 0. \end{aligned}$$

Aus der zweiten Gleichung folgt $x_2 = -1 = 1$ *und damit aus der ersten* $x_1 = 1 - 1 = 0$*. Die ursprünglichen Datenbits waren also* $(0, 1, 1)$*. Die zusätzlichen Prüfbits ermöglichen hier die eindeutige Rekonstruktion.*

Geht außerdem p_2 *verloren, so bleibt für die fehlenden Datenbits nur die Gleichung* $x_1 + x_2 = 1$*. Setzen wir* $x_2 = t$*, so gilt* $x_1 = 1 - t = 1 + t$*. Die Lösungsmenge ist daher*

$$\left\{ \begin{pmatrix} 1+t \\ t \end{pmatrix} \;\middle|\; t \in \mathbb{F}_2 \right\} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Hier ist $\langle (1,1)^\top \rangle$ *der Kern der Koeffizientenmatrix* $\begin{pmatrix} 1 & 1 \end{pmatrix}$*, sodass unsere Darstellung der Lösungsmenge mit Korollar 6.3 übereinstimmt. Die Lösungsmenge hat einen freien Parameter und damit Dimension eins, enthält aber nur zwei Vektoren, da* t *nur die Werte* 0 *und* 1 *annehmen kann. Die verbleibenden Angaben reichen nicht aus, um zwischen den beiden möglichen Nachrichten* $(1, 0, 1)$ *und* $(0, 1, 1)$ *zu unterscheiden.* △

6.2 Das Gauß-Verfahren

Wir beschäftigen uns jetzt mit einem systematischen Verfahren, die Lösungsmenge eines linearen Gleichungssystems zu bestimmen. Nach Korollar 6.3 genügt es dazu, eine Lösung zu finden, zusammen mit einer Basis des Kerns. Nach Satz 6.4 können wir anhand des Rangs entscheiden, ob Lösungen existieren und ob die Lösung eindeutig ist.

Zum Einstieg betrachten wir über $\mathbb{R}$ das System $2x_2 = 4$ und $x_1 + x_2 = 3$. Wir vertauschen zunächst die beiden Gleichungen, teilen dann die zweite Gleichung durch 2 und ziehen sie anschließend von der ersten ab. In der erweiterten Koeffizientenmatrix lauten diese Schritte

$$\left(\begin{array}{cc|c} 0 & 2 & 4 \\ 1 & 1 & 3 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 2 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 1 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right).$$

Das umgeformte System entspricht dem System $x_1 = 1$, $x_2 = 2$, also können wir die Lösung direkt ablesen.

Wir halten jetzt die Operationen fest, die wir soeben verwendet haben: das Vertauschen, Skalieren und Kombinieren von Gleichungen. Wir werden sehen, dass diese drei Arten von Umformungen ausreichen, um jedes lineare Gleichungssystem über einem Körper zu lösen oder seine Unlösbarkeit festzustellen.

Definition 6.8. Die folgenden Operationen sind *elementare Zeilenumformungen* einer Matrix $M \in K^{m\times n}$:

(i) Vertauschen zweier Zeilen $\boldsymbol{M_{i\cdot}}$ und $\boldsymbol{M_{j\cdot}}$; wir schreiben $\boldsymbol{M_{i\cdot}} \leftrightarrow \boldsymbol{M_{j\cdot}}$.

(ii) Multiplizieren einer Zeile $\boldsymbol{M_{i\cdot}}$ mit einem Skalar $\lambda \in K \setminus \{0\}$; wir schreiben $\boldsymbol{M_{i\cdot}} \leftarrow \lambda \boldsymbol{M_{i\cdot}}$.

(iii) Addieren des λ-Fachen einer Zeile $\boldsymbol{M_{j\cdot}}$ zu einer anderen Zeile $\boldsymbol{M_{i\cdot}}$ mit $i \neq j$, wobei $\lambda \in K$ ist; wir schreiben $\boldsymbol{M_{i\cdot}} \leftarrow \boldsymbol{M_{i\cdot}} + \lambda \boldsymbol{M_{j\cdot}}$.

Zwei lineare Gleichungssysteme heißen *äquivalent*, wenn ihre erweiterten Koeffizientenmatrizen durch elementare Zeilenumformungen ineinander überführt werden können.

Die entscheidende Eigenschaft von elementaren Zeilenumformungen ist, dass sie die Lösungsmenge unverändert lassen: Jede Lösung vor der Umformung bleibt auch danach gültig und jede Lösung nach der Umformung muss auch vorher schon gültig gewesen sein.

Beobachtung 6.9. Äquivalente lineare Gleichungssysteme besitzen dieselbe Lösungsmenge.

Beispiel 6.10. *Wir betrachten über* $\mathbb{R}$ *das System*

$$\begin{aligned} 2x_3 + x_4 + x_5 &= 4, \\ 2x_1 + 2x_2 + x_3 + 3x_5 &= 3, \\ 4x_1 + 4x_2 + 4x_3 + 2x_4 + 6x_5 &= 12, \\ 2x_1 + 2x_2 + 5x_3 + 2x_4 + 5x_5 &= 11. \end{aligned}$$

Es hat die Koeffizientenmatrix A *und die rechte Seite* $\boldsymbol{b}$ *mit*

$$M \coloneqq (A \mid \boldsymbol{b}) = \left(\begin{array}{ccccc|c} 0 & 0 & 2 & 1 & 1 & 4 \\ 2 & 2 & 1 & 0 & 3 & 3 \\ 4 & 4 & 4 & 2 & 6 & 12 \\ 2 & 2 & 5 & 2 & 5 & 11 \end{array}\right).$$

Wir wollen zunächst x_1 *aus den unteren Gleichungen entfernen. Da sein Koeffizient in der ersten Gleichung null ist, vertauschen wir die ersten beiden Zeilen. Anschließend ziehen wir das Doppelte der ersten Zeile von der dritten und die erste Zeile von der vierten ab:*

$$\left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 4 & 4 & 4 & 2 & 6 & 12 \\ 2 & 2 & 5 & 2 & 5 & 11 \end{array}\right) \xrightarrow[\boldsymbol{M_{4\cdot}} \leftarrow \boldsymbol{M_{4\cdot}} - \boldsymbol{M_{1\cdot}}]{\boldsymbol{M_{3\cdot}} \leftarrow \boldsymbol{M_{3\cdot}} - 2\boldsymbol{M_{1\cdot}}} \left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 2 & 2 & 0 & 6 \\ 0 & 0 & 4 & 2 & 2 & 8 \end{array}\right).$$

Unterhalb der ersten Zeile sind damit auch alle Einträge der zweiten Spalte null. Wir gehen zur dritten Spalte weiter und verwenden die zweite Zeile, um x_3 *aus den darunterliegenden Gleichungen zu entfernen:*

$$\left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 2 & 2 & 0 & 6 \\ 0 & 0 & 4 & 2 & 2 & 8 \end{array}\right) \xrightarrow[\boldsymbol{M_{4\cdot}} \leftarrow \boldsymbol{M_{4\cdot}} - 2\boldsymbol{M_{2\cdot}}]{\boldsymbol{M_{3\cdot}} \leftarrow \boldsymbol{M_{3\cdot}} - \boldsymbol{M_{2\cdot}}} \left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Die letzte Zeile stellt keine Bedingung mehr. Wir beginnen daher mit der dritten Gleichung $x_4 - x_5 = 2$. Jede Lösung dieser Gleichung lässt sich zu einer Lösung des gesamten Systems ergänzen: Die zweite Gleichung $2x_3 + x_4 + x_5 = 4$ bestimmt zu den gewählten Werten von x_4 und x_5 eindeutig den Wert von x_3. Sie schränkt unsere Wahl von x_4 und x_5 also nicht weiter ein. Die erste Gleichung $2x_1 + 2x_2 + x_3 + 3x_5 = 3$ legt dann nur noch die Summe $x_1 + x_2$ fest. Wir können somit x_2 wählen und daraus x_1 bestimmen.

Um eine konkrete Lösung zu erhalten, wählen wir $x_5 = 0$. Aus der dritten Gleichung folgt $x_4 = 2$. Die zweite Gleichung wird damit zu $2x_3 + 2 = 4$, also ist $x_3 = 1$. Die erste Gleichung lautet nun $2x_1 + 2x_2 + 1 = 3$, also $x_1 + x_2 = 1$. Wählen wir zusätzlich $x_2 = 0$, so erhalten wir $x_1 = 1$. Nach Beobachtung 6.9 ist damit

$$\boldsymbol{x}^{(0)} = (1, 0, 1, 2, 0)^\top$$

eine Lösung des ursprünglichen Systems. △

Mit einigen weiteren Zeilenumformungen können wir die Lösung direkt ablesen.

Beispiel 6.11. *Wir formen die zuletzt erhaltene Matrix M aus Beispiel 6.10 weiter um:*

$$\left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow[\boldsymbol{M_2.} \leftarrow \frac{1}{2}\boldsymbol{M_2.}]{\boldsymbol{M_2.} \leftarrow \boldsymbol{M_2.} - \boldsymbol{M_3.}} \left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow[\boldsymbol{M_1.} \leftarrow \frac{1}{2}\boldsymbol{M_1.}]{\boldsymbol{M_1.} \leftarrow \boldsymbol{M_1.} - \boldsymbol{M_2.}} \left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Nun können wir die Lösung $\boldsymbol{x}^{(0)}$ ablesen, indem wir für jede (von $0 = 0$ verschiedene) Gleichung die erste Variable mit Koeffizient ungleich 0 auf den Wert der rechten Seite setzen und dann alle übrigen Variablen auf 0:

$$x_5^{(0)} = 0, \qquad x_4^{(0)} = 2 = 2 + x_5^{(0)}, \qquad x_3^{(0)} = 1 = 1 - x_5^{(0)}, \qquad x_2^{(0)} = 0, \qquad x_1^{(0)} = 1 = 1 - x_2^{(0)} - x_5^{(0)}. \quad \triangle$$

Wir führen Begriffe für die erhaltenen Matrixformen aus den Beispielen 6.10 und 6.11 ein und zeigen im Anschluss, wie wir jede Matrix systematisch mittels elementarer Zeilenumformungen in diese Formen bringen können.

Definition 6.12. Eine Matrix $M \in K^{m \times n}$ hat *Zeilenstufenform*, wenn es $r \in \{0, \dots, m\}$ und Spaltenindizes $p_1 < \cdots < p_r$ gibt, sodass:

(i) Alle Zeilen unterhalb der r-ten Zeile sind Nullzeilen, also

$$\forall i \in \{r+1, \dots, m\} \colon \boldsymbol{M_{i.}} = \mathbf{0}^\top.$$

(ii) In jeder Zeile $i \in \{1, \dots, r\}$ steht der erste von null verschiedene Eintrag in Spalte p_i, also

$$M_{ip_i} \neq 0, \qquad (M_{i1}, \dots, M_{i(p_i-1)}) = \mathbf{0}^\top.$$

Die Einträge M_{ip_i} heißen *Pivotelemente*, die Spalten $\boldsymbol{M_{\cdot p_i}}$ *Pivotspalten*. Die Zeilenstufenform heißt *reduziert*, wenn die Pivotelemente 1 sind und alle anderen Elemente ihrer Spalte 0, also

$$\forall i \in \{1, \dots, r\} \colon \boldsymbol{M_{\cdot p_i}} = \boldsymbol{e}^{(i)}.$$

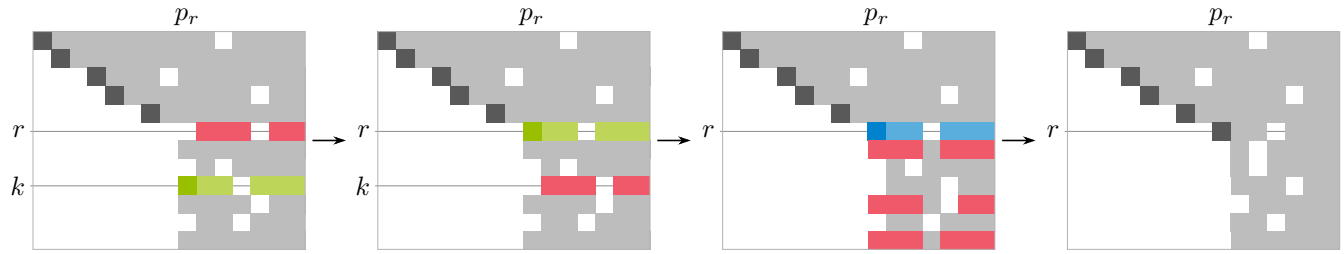


Abbildung 6.1: Eine Iteration der Vorwärtsphase, von links nach rechts: Pivot wählen, in die aktuelle Zeile r tauschen, Einträge unter dem Pivot beseitigen. Rote und grüne Einträge deuten Änderungen an.

Das *Gauß-Verfahren* fasst die bisherigen Rechnungen zu einem allgemeinen Verfahren zusammen. In der *Vorwärtsphase* bearbeiten wir die Zeilen von oben nach unten. In den noch nicht bearbeiteten Zeilen suchen wir die erste Spalte mit einem von null verschiedenen Eintrag. Einen solchen Eintrag holen wir durch Zeilentausch in die nächste Zeile r und beseitigen mit Vielfachen dieser Zeile die Einträge darunter (Abbildung 6.1). So entsteht Zeilenstufenform.

Für die reduzierte Zeilenstufenform folgt die *Rückwärtsphase*: Wir teilen jede Pivotzeile durch ihr Pivotelement und beseitigen mit Vielfachen dieser Zeile die Einträge darüber (Abbildung 6.2). Bereits bereinigte Pivotspalten bleiben dabei unverändert.

Wir formulieren das Gauß-Verfahren noch formal.

Algorithmus : Gauß-Verfahren

Eingabe: Matrix $M \in K^{m \times n}$ über einem Körper K
Ausgabe: M in reduzierter Zeilenstufenform

(Vorwärtsphase: Zeilenstufenform herstellen)
$r \leftarrow 0$ (Anzahl der bisherigen Pivots)
solange $\exists k \in \{r+1, \dots, m\} \colon \boldsymbol{M_{k\cdot}} \neq \mathbf{0}^\top$ **:** (Es gibt noch eine Nichtnullzeile)
  $r \leftarrow r + 1$
  $p_r \leftarrow \min\{j \in \{1, \dots, n\} \mid \exists k \in \{r, \dots, m\} \colon M_{kj} \neq 0\}$ (Pivotspalte suchen)
  wähle $k \in \{r, \dots, m\}$ mit $M_{kp_r} \neq 0$ (Pivotzeile wählen)
  falls $k \neq r$ **:** $\boldsymbol{M_{r\cdot}} \leftrightarrow \boldsymbol{M_{k\cdot}}$. (Pivot in Zeile r holen)
  für alle $k \in \{r+1, \dots, m\}$ **:**
    $\boldsymbol{M_{k\cdot}} \leftarrow \boldsymbol{M_{k\cdot}} - \frac{M_{kp_r}}{M_{rp_r}} \boldsymbol{M_{r\cdot}}$. (Eintrag unter dem Pivot beseitigen)

(Rückwärtsphase: Reduzierte Zeilenstufenform herstellen)
für alle $i \in \{r, r-1, \dots, 1\}$ **:**
  $\boldsymbol{M_{i\cdot}} \leftarrow \frac{1}{M_{ip_i}} \boldsymbol{M_{i\cdot}}$. (Pivot auf eins bringen)
  für alle $k \in \{1, \dots, i-1\}$ **:**
    $\boldsymbol{M_{k\cdot}} \leftarrow \boldsymbol{M_{k\cdot}} - M_{kp_i} \boldsymbol{M_{i\cdot}}$. (Eintrag über dem Pivot beseitigen)

Wir beobachten, dass sich das Verfahren wie gewünscht verhält.

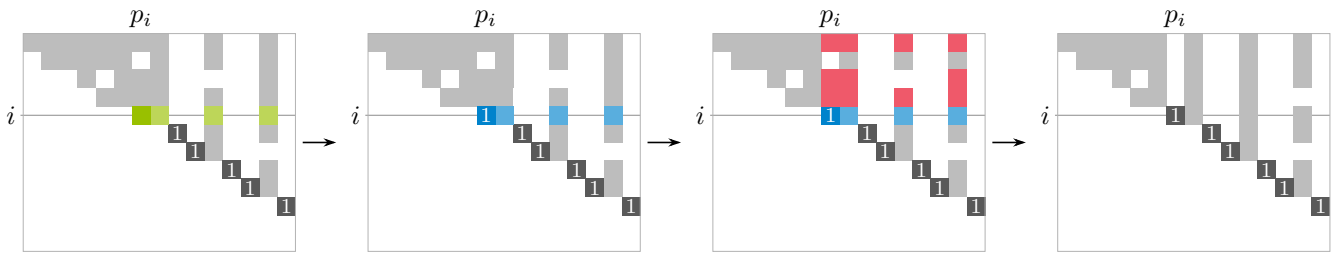


Abbildung 6.2: Eine Iteration der Rückwärtsphase, von links nach rechts: Pivot der aktuellen Zeile auf 1 bringen, dann Einträge über dem Pivot beseitigen. Rote und grüne Einträge deuten Änderungen an.

Satz 6.13. Das Gauß-Verfahren überführt jede Matrix $M \in K^{m \times n}$ durch elementare Zeilenumformungen zunächst in Zeilenstufenform und anschließend in reduzierte Zeilenstufenform.

Beweis. Für die Vorwärtsphase zeigen wir induktiv: Die bisherigen Pivots stehen treppenförmig, und alle noch unbearbeiteten Zeilen enthalten bis einschließlich der letzten Pivotspalte nur Nullen. Vor der ersten Iteration ist nichts zu zeigen. Im nächsten Schritt liegt das neue Pivotelement nach Induktionsvoraussetzung weiter rechts. Die gewählte Pivotzeile enthält links davon nur Nullen; beim Abziehen ihrer Vielfachen bleiben diese Nullen erhalten, und die Einträge unter dem neuen Pivot werden null. Die bereits bearbeiteten Zeilen bleiben unverändert. Damit gilt die Aussage auch nach dieser Iteration. Nach höchstens $\min\{m, n\}$ Iterationen endet die Phase; die verbleibenden Zeilen sind dann Nullzeilen. Es liegt also Zeilenstufenform vor.

Für die Rückwärtsphase zeigen wir induktiv: Die Zeilenstufenform bleibt erhalten, und jede bereits bearbeitete Pivotspalte enthält eine 1 an ihrer Pivotposition und sonst nur Nullen. Dies gilt vor der ersten Iteration. Im nächsten Schritt normieren wir das gewählte Pivotelement auf 1 und beseitigen die Einträge darüber. Da die Pivotzeile links vom Pivot nur Nullen enthält, bleiben die Stufenform und die anderen Pivots erhalten. Auch die bereits bearbeiteten Pivotspalten ändern sich nicht, denn die verwendete Zeile enthält dort nur Nullen. Damit gilt die Aussage nach jeder Iteration, unabhängig von der Bearbeitungsreihenfolge. Sind alle Pivotspalten bearbeitet, ist die Zeilenstufenform reduziert.

Alle Divisionen sind erlaubt, da Pivotelemente von null verschieden sind. Beide Phasen verwenden somit ausschließlich elementare Zeilenumformungen. □

Wenden wir das Gauß-Verfahren auf eine erweiterte Koeffizientenmatrix an, so erhalten wir nach Beobachtung 6.9 ein System mit derselben Lösungsmenge (ein Pivotelement in der letzten Spalte beschreibt eine nichterfüllbare Gleichung $0 = c$ mit $c \neq 0$).

Die Zeilenumformungen lassen sich dabei auch als Matrixmultiplikationen ausdrücken.

Beobachtung 6.14. Jede elementare Zeilenumformung einer Matrix M lässt sich ausdrücken als EM, wobei E eine der folgenden Matrizen ist (fehlende Einträge sind 1 auf der Diagonalen und sonst 0):

$$\begin{pmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} \begin{matrix} \\ i \\ j \\ \\ \\ \end{matrix} \qquad \begin{pmatrix} 1 & & & \\ & \lambda & & \\ & & & \\ & & & 1 \end{pmatrix} \begin{matrix} \\ i \\ \\ \\ \end{matrix} \qquad \begin{pmatrix} 1 & & & & \\ & 1 & \lambda & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix} \begin{matrix} \\ i \\ j \\ \\ \\ \end{matrix}$$

Jede solche Matrix E ist invertierbar und es gilt $\ker(EM) = \ker M$.

Beweis. Die Wirkung der Matrizen folgt direkt aus der Definition der Matrixmultiplikation. Die erste Matrix ist ihre eigene Inverse; die Inversen der beiden anderen entstehen durch Ersetzen von λ durch λ^{-1} beziehungsweise von λ durch $-\lambda$. Insbesondere gilt $EM\boldsymbol{x} = \mathbf{0} \iff M\boldsymbol{x} = E^{-1}\mathbf{0} = \mathbf{0}$ und damit $\ker(EM) = \ker M$. □

Eine Folge mehrerer elementarer Zeilenumformungen entspricht daher einer Multiplikation $T = FA = E^{(1)}E^{(2)}\cdots E^{(k)}A$, wobei F nach Satz 5.68 invertierbar ist. Dabei bleiben wegen $\ker T = \ker A$ alle linearen Abhängigkeiten zwischen den Spalten mit denselben Koeffizienten $\boldsymbol{t} \in K^n$ erhalten, denn

$$\mathbf{0} = T\boldsymbol{t} = t_1\boldsymbol{T_{\cdot 1}} + \cdots + t_n\boldsymbol{T_{\cdot n}} \quad \iff \quad \mathbf{0} = A\boldsymbol{t} = t_1\boldsymbol{A_{\cdot 1}} + \cdots + t_n\boldsymbol{A_{\cdot n}}.$$

Die von den Spalten erzeugten Untervektorräume müssen dagegen nicht gleich sein.

In Beispiel 6.11 haben wir eine Lösung aus der reduzierten Zeilenstufenform gewonnen, indem wir alle Einträge des Lösungsvektors, die zu keinem Pivotelement gehören, auf 0 gesetzt haben und die anderen auf die rechten Seiten der entsprechenden Pivotzeilen. Wir verallgemeinern dieses Vorgehen.

Beobachtung 6.15. Sei $(T \mid \boldsymbol{c})$ mit $T \in K^{m\times n}$ in reduzierter Zeilenstufenform, und seien $p_1 < \cdots < p_r$ die Indizes der Pivotspalten von T.

(i) Gibt es $i \in \{r+1, \ldots, m\}$ mit $c_i \neq 0$, so hat $T\boldsymbol{x} = \boldsymbol{c}$ keine Lösung.

(ii) Ist $c_i = 0$ für alle $i \in \{r+1, \ldots, m\}$, so ist eine Lösung $\boldsymbol{x}^{(0)} \in K^n$ von $T\boldsymbol{x} = \boldsymbol{c}$ gegeben durch

$$\boldsymbol{x}^{(0)} = \sum_{k=1}^{r} c_k\, \boldsymbol{e}^{(p_k)}.$$

Nach Korollar 6.3 ist die Lösungsmenge von $T\boldsymbol{x} = \boldsymbol{c}$ im lösbaren Fall gleich $\boldsymbol{x}^{(0)} + \ker T$. Um alle Lösungen anzugeben, benötigen wir also noch eine Basis von $\ker T$.

Beispiel 6.16. *Wir verwenden die reduzierte erweiterte Koeffizientenmatrix*

$$M = (T \mid \boldsymbol{c}) = \left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

aus Beispiel 6.11. Nach Beobachtung 6.14 gilt $\ker A = \ker T$. *Für einen Vektor* $\boldsymbol{x}$ *im Kern gilt* $T\boldsymbol{x} = \mathbf{0}$, *also*

$$x_1 + x_2 + x_5 = 0, \qquad x_3 + x_5 = 0, \qquad x_4 - x_5 = 0.$$

Wir können $x_2 = s$ *und* $x_5 = t$ *frei wählen mit* $s, t \in \mathbb{R}$; *die Gleichungen bestimmen dann* $x_1 = -s - t$, $x_3 = -t$ *und* $x_4 = t$. *Somit gilt*

$$\boldsymbol{x} = \begin{pmatrix} -s-t \\ s \\ -t \\ t \\ t \end{pmatrix} = s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{pmatrix}. \tag{6.1}$$

Mit der speziellen Lösung $\boldsymbol{x}^{(0)} = (1, 0, 1, 2, 0)^\top$ *aus Beispiel 6.10 erhalten wir nach Korollar 6.3 die Lösungsmenge als*

$$\boldsymbol{x}^{(0)} + \ker A = \boldsymbol{x}^{(0)} + \left\{ s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\}.$$

Die im Kern frei wählbaren Einträge x_2 und x_5 sind genau jene, die zu keiner der Pivotspalten $\boldsymbol{T_{\cdot p_1}} = \boldsymbol{T_{\cdot 1}}$, $\boldsymbol{T_{\cdot p_2}} = \boldsymbol{T_{\cdot 3}}$ und $\boldsymbol{T_{\cdot p_3}} = \boldsymbol{T_{\cdot 4}}$ gehören. Für jeden solchen Eintrag haben wir einen Parameter (s bzw. t) eingeführt. In (6.1) trägt der Parameter s zu seinem zugehörigen frei wählbaren Eintrag x_2 bei sowie zum Eintrag x_1 über den Koeffizienten T_{12}. Der Parameter t trägt zu seinem zugehörigen frei wählbaren Eintrag x_5 bei sowie zu den Einträgen x_1, x_3 und x_4 über die Koeffizienten T_{15}, T_{25} und T_{35}. Wir können also die erzeugenden Vektoren von $\ker A$ direkt von T ablesen, wobei wir Nullzeilen ignorieren:

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \boldsymbol{e}^{(2)} - \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{e}^{(5)} - \begin{pmatrix} 1 \\ 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{pmatrix}.$$

△

Wir verallgemeinern das Vorgehen aus Beispiel 6.16, um eine Basis des Kerns einer Matrix in reduzierter Zeilenstufenform zu gewinnen.

Satz 6.17. Sei $T \in K^{m \times n}$ in reduzierter Zeilenstufenform mit den Pivotspaltenindizes $p_1 < \cdots < p_r$. Dann ist eine Basis von $\ker T$ gegeben durch

$$\mathcal{B} = \left\{ \boldsymbol{e}^{(j)} - \sum_{k=1}^{r} T_{kj}\, \boldsymbol{e}^{(p_k)} \;\middle|\; j \in \{1, \ldots, n\} \setminus \{p_1, \ldots, p_r\} \right\}.$$

Beweis. Es ist $\boldsymbol{x} \in \ker T$ genau dann, wenn $T\boldsymbol{x} = \boldsymbol{0}$. Das ist äquivalent zu $\boldsymbol{T_{k\cdot}}\, \boldsymbol{x} = 0$ für alle $k \in \{1, \ldots, m\}$, also zu

$$\begin{aligned} \boldsymbol{x} = \sum_{j=1}^{n} x_j\, \boldsymbol{e}^{(j)} &= \sum_{j=1}^{n} x_j\, \boldsymbol{e}^{(j)} - \sum_{k=1}^{r} (\boldsymbol{T_{k\cdot}}\, \boldsymbol{x})\, \boldsymbol{e}^{(p_k)} \\ &= \sum_{j=1}^{n} x_j\, \boldsymbol{e}^{(j)} - \sum_{k=1}^{r} \left(\sum_{j=1}^{n} T_{kj} x_j \right) \boldsymbol{e}^{(p_k)} \\ &= \sum_{j=1}^{n} x_j \left(\boldsymbol{e}^{(j)} - \sum_{k=1}^{r} T_{kj}\, \boldsymbol{e}^{(p_k)} \right). \end{aligned} \tag{6.2}$$

Da T in reduzierter Zeilenstufenform ist, ist die rechte Seite eine Linearkombination von Vektoren aus $\mathcal{B}$: Für $j = p_i$ ist die Pivotspalte $\boldsymbol{T_{\cdot j}} = \boldsymbol{e}^{(i)}$, also verschwindet der Summand

$$\boldsymbol{e}^{(j)} - \sum_{k=1}^{r} T_{kj}\, \boldsymbol{e}^{(p_k)} = \boldsymbol{e}^{(j)} - \sum_{k=1}^{r} e_k^{(i)}\, \boldsymbol{e}^{(p_k)} = \boldsymbol{e}^{(j)} - \boldsymbol{e}^{(p_i)} = \boldsymbol{0}.$$

Also ist $\ker T \subseteq \langle \mathcal{B} \rangle$. Umgekehrt ist $\mathcal{B} \subseteq \ker T$ und damit $\langle \mathcal{B} \rangle \subseteq \ker T$, da wir jeden Vektor aus $\mathcal{B}$ durch eine triviale Linearkombination der Form (6.2) mit einem Summanden schreiben können. Also ist $\ker T = \langle \mathcal{B} \rangle$.

Da nur der Vektor $\boldsymbol{e}^{(j)} - \sum_{k=1}^{r} T_{kj}\, \boldsymbol{e}^{(p_k)} \in \mathcal{B}$ einen j-ten Eintrag ungleich 0 hat, ist $\mathcal{B}$ linear unabhängig und somit eine Basis von $\ker T$. □

Da Zeilenumformungen die Lösungsmenge und den Kern erhalten (Beobachtungen 6.9 und 6.14), können wir Beobachtung 6.15 und Satz 6.17 auf das ursprüngliche System übertragen.

Korollar 6.18. Entsteht die reduzierte Zeilenstufenform $(T \mid \boldsymbol{c})$ aus $(A \mid \boldsymbol{b})$ durch elementare Zeilenumformungen, so ist $A\boldsymbol{x} = \boldsymbol{b}$ genau dann lösbar, wenn keine Zeile $(0, \dots, 0 \mid c)$ mit $c \neq 0$ auftritt. Die Lösungsmenge ist dann $\boldsymbol{x}^{(0)} + \langle \mathcal{B} \rangle$, mit dem Vektor $\boldsymbol{x}^{(0)}$ aus Beobachtung 6.15 und der Basis $\mathcal{B}$ aus Satz 6.17.

Das Gauß-Verfahren funktioniert über beliebigen Körpern.

Beispiel 6.19. *Im Prüfbit-Beispiel 6.7 hatten wir die Gleichungen $x_1 + x_2 = 1$ und $x_2 = 1$ über $\mathbb{F}_2$. In Matrixschreibweise ergibt sich, mit $1 + 1 = 0$,*

$$M \coloneqq \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 1 \end{array}\right) \xrightarrow{\boldsymbol{M_1.} \leftarrow \boldsymbol{M_1.} + \boldsymbol{M_2.}} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right).$$

Wir können die Lösung $\boldsymbol{x} = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ also direkt ablesen. △

Über $\mathbb{F}_2$ ist jedes von null verschiedene Pivotelement automatisch 1. Das nächste Beispiel zeigt auch die Pivotnormierung über einem endlichen Körper.

Beispiel 6.20. *Wir rechnen über $\mathbb{F}_5 \coloneqq \mathbb{Z}/5\mathbb{Z}$, das nach Satz 3.69 ein Körper ist. Wir bezeichnen die Restklassen durch ihre Repräsentanten $0, 1, 2, 3, 4$ und lösen das System mit der erweiterten Matrix*

$$M \coloneqq (A \mid \boldsymbol{b}) = \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 1 & 3 & 4 & 2 \\ 3 & 4 & 0 & 3 \end{array}\right).$$

Wegen $2 \cdot 3 = 1$ in $\mathbb{F}_5$ ist $2^{-1} = 3$. Nach dem Gauß-Verfahren formen wir wie folgt um:

$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 1 & 3 & 4 & 2 \\ 3 & 4 & 0 & 3 \end{array}\right) \xrightarrow{\boldsymbol{M_1.} \leftarrow 3\boldsymbol{M_1.}} \left(\begin{array}{ccc|c} 1 & 3 & 3 & 3 \\ 1 & 3 & 4 & 2 \\ 3 & 4 & 0 & 3 \end{array}\right) \xrightarrow[\boldsymbol{M_3.} \leftarrow \boldsymbol{M_3.} - 3\boldsymbol{M_1.}]{\boldsymbol{M_2.} \leftarrow \boldsymbol{M_2.} - \boldsymbol{M_1.}} \left(\begin{array}{ccc|c} 1 & 3 & 3 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 1 & 4 \end{array}\right)$$

$$\xrightarrow{\boldsymbol{M_3.} \leftarrow \boldsymbol{M_3.} - \boldsymbol{M_2.}} \left(\begin{array}{ccc|c} 1 & 3 & 3 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\boldsymbol{M_1.} \leftarrow \boldsymbol{M_1.} - 3\boldsymbol{M_2.}} \left(\begin{array}{ccc|c} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Also ist x_2 frei wählbar und wir können $\boldsymbol{x}^{(0)} = \left(\begin{smallmatrix} 1 \\ 0 \\ 4 \end{smallmatrix}\right)$ von der rechten Seite ablesen. Der Erzeuger des Kerns ergibt sich aus der zweiten Spalte zu $\boldsymbol{e}^{(2)} - \left(\begin{smallmatrix} 3 \\ 0 \\ 0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 2 \\ 1 \\ 0 \end{smallmatrix}\right)$. Damit ist die Lösungsmenge

$$\boldsymbol{x}^{(0)} + \ker A = \left\{ \left(\begin{smallmatrix} 1 \\ 0 \\ 4 \end{smallmatrix}\right) + t\left(\begin{smallmatrix} 2 \\ 1 \\ 0 \end{smallmatrix}\right) \;\middle|\; t \in \mathbb{F}_5 \right\} = \left\{ \left(\begin{smallmatrix} 1 \\ 0 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 3 \\ 1 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 2 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 \\ 3 \\ 4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 4 \\ 4 \\ 4 \end{smallmatrix}\right) \right\}.$$ △

Nach Satz 6.17 und Korollar 6.18 gehört zu jedem frei wählbaren Eintrag ein Basisvektor des Kerns. Die Dimension des Kerns und damit der affinen Lösungsmenge (sofern sie nicht leer ist) entspricht also der Anzahl der Freiheitsgrade, wie wir bereits in Abschnitt 6.1 überlegt hatten. Die Dimensionsformel aus Satz 5.20 liefert mit Beobachtung 5.33 außerdem einen Zusammenhang mit dem Rang der zugrundeliegenden Matrix. Konkret gibt es bei n Spalten und r Pivotelementen $n - r$ frei wählbare Einträge, also $\dim \ker A = n - r$ und nach der Dimensionsformel $\operatorname{rang} A = n - (n - r) = r$.

Beobachtung 6.21. Hat eine durch Zeilenumformungen gewonnene Zeilenstufenform von $A \in K^{m\times n}$ genau r Pivotelemente, so gilt

$$\dim \ker A = n - r, \qquad \operatorname{rang} A = r.$$

Nach Satz 6.4 ist das System $A\boldsymbol{x} = \boldsymbol{b}$ genau dann lösbar, wenn A und $(A \mid \boldsymbol{b})$ denselben Rang haben. Beobachtung 6.21 setzt keine reduzierte Zeilenstufenform voraus; wir können beide Ränge daher bereits durch Abzählen der Pivots in der gewöhnlichen Zeilenstufenform bestimmen. Die Lösbarkeit entspricht genau der Bedingung, dass kein Pivot in der rechten Seite liegt, also keine Zeile $(0, \dots, 0 \mid c)$ mit $c \neq 0$ auftritt.

Im lösbaren Fall entstehen nach Korollar 6.3 alle Lösungen aus einer speziellen Lösung $\boldsymbol{x}^{(0)}$, indem wir eine beliebige Linearkombination einer Kernbasis addieren. Nach Beobachtung 6.21 hat eine solche Basis genau $n - r$ Vektoren. Die $n - r$ Koeffizienten der Linearkombination entsprechen den Freiheitsgraden der Lösung. Insbesondere ist die Lösung genau dann eindeutig, wenn jede Koeffizientenspalte ein Pivotelement enthält.

Beispiel 6.22. *Die erweiterte Matrix*

$$\left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

aus Beispiel 6.10 hat in ihrer Stufenform drei Pivots. Nach Beobachtung 6.21 ist ihr Rang 3*. Die Lösungsmenge des ursprünglichen Systems ist also eine affine Ebene der Dimension* $5 - 3 = 2$ *in* $\mathbb{R}^5$*. Ersetzen wir nur die rechte Seite* $\boldsymbol{b} = (4, 3, 12, 11)^\top$ *durch* $\boldsymbol{b}' = (4, 3, 12, 12)^\top$*, so führen dieselben Zeilenumformungen zur Matrix*

$$\left(\begin{array}{ccccc|c} 2 & 2 & 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right).$$

Die letzte Zeile fordert $0 = 1$*. Das System ist daher unlösbar; die Koeffizientenmatrix hat weiterhin Rang* 3*, die erweiterte Matrix nun aber Rang* 4*.* △

Auch über einem endlichen Körper können wir die Dimension auf diese Weise ablesen.

Beispiel 6.23. *Die reduzierte Stufenform*

$$\left(\begin{array}{ccc|c} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

der Matrix $(A \mid \boldsymbol{b})$ *aus Beispiel 6.20 hat* $r = 2$ *Pivotelemente und* A *hat* $n = 3$ *Spalten. Nach Beobachtung 6.21 gilt daher*

$$\operatorname{rang} A = 2, \qquad \dim \ker A = 3 - 2 = 1.$$

Die dort gefundene Kernbasis besteht entsprechend aus einem Vektor und die fünfelementige Lösungsmenge hat Dimension 1. △

Wir können mit derselben Rechnung auch eine Bildbasis bestimmen: Die Pivotspalten der Zeilenstufenform sind wegen ihrer versetzten Pivotpositionen linear unabhängig. Da Zeilenumformungen die linearen Abhängigkeiten zwischen den Spalten erhalten (Beobachtung 6.14), gilt dies auch für die entsprechenden Spalten der ursprünglichen Matrix. Ihre Anzahl stimmt nach Beobachtung 6.21 mit der Dimension $\dim f_A(K^n) = \operatorname{rang} A$ des Bildes überein, sodass sie nach Satz 4.73 eine Bildbasis bilden.

Beobachtung 6.24. Sei $A \in K^{m\times n}$, und seien $p_1 < \cdots < p_r$ die Indizes der Pivotspalten einer durch Zeilenumformungen gewonnenen Zeilenstufenform von A. Dann ist eine Basis von $f_A(K^n)$ gegeben durch

$$\{\boldsymbol{A_{\cdot p_1}}, \ldots, \boldsymbol{A_{\cdot p_r}}\}.$$

Beispiel 6.25. *Die Indizes der Pivotspalten der Zeilenstufenform*

$$\begin{pmatrix} 2 & 2 & 1 & 0 & 3 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \textit{von} \quad A = \begin{pmatrix} 0 & 0 & 2 & 1 & 1 \\ 2 & 2 & 1 & 0 & 3 \\ 4 & 4 & 4 & 2 & 6 \\ 2 & 2 & 5 & 2 & 5 \end{pmatrix}$$

aus Beispiel 6.10 sind 1, 3 *und* 4. *Nach Beobachtung 6.24 bilden daher die Originalspalten*

$$\begin{pmatrix} 0 \\ 2 \\ 4 \\ 2 \end{pmatrix}, \qquad \begin{pmatrix} 2 \\ 1 \\ 4 \\ 5 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$$

eine Basis von $f_A(\mathbb{R}^5)$. *Achtung: Die entsprechenden Spalten der Zeilenstufenform oder der reduzierten Zeilenstufenform aus Beispiel 6.11 bilden hingegen* keine *Basis von* $f_A(\mathbb{R}^5)$. △

Das Gauß-Verfahren erlaubt es uns auch, Koordinatendarstellungen und Basiswechsel systematisch zu berechnen. Ist $\mathfrak{C} = (\boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(n)})$ eine geordnete Basis von K^n, so bestimmen wir die Koordinaten $\boldsymbol{x} \coloneqq [\boldsymbol{v}]_{\mathfrak{C}}$ eines Vektors $\boldsymbol{v}$ gemäß Definition 4.61 durch Lösen von

$$\begin{pmatrix} \boldsymbol{c}^{(1)} & \ldots & \boldsymbol{c}^{(n)} \end{pmatrix} \boldsymbol{x} = \boldsymbol{v}.$$

Für eine weitere geordnete Basis $\mathfrak{B} = (\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)})$ erhalten wir die Basiswechselmatrix $S_{\mathfrak{C}\leftarrow\mathfrak{B}}$ gemäß Beobachtung 5.72, indem wir dieses System für alle Vektoren $\boldsymbol{v} \in \{\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(n)}\}$ lösen. Dabei können wir jeweils dieselben elementaren Zeilenumformungen verwenden, daher können wir das System simultan für alle rechten Seiten $\boldsymbol{v}$ lösen.

Beispiel 6.26. *Wir greifen die Basis* $\mathfrak{B} = (\binom{1}{1}, \binom{1}{0})$ *und die Standardbasis* $\mathfrak{E} = (\binom{1}{0}, \binom{0}{1})$ *aus Beispiel 5.70 wieder auf. Wir verwenden die Vektoren von* $\mathfrak{B}$ *als Spalten der Koeffizientenmatrix* A *und lösen* $A\boldsymbol{x} = \boldsymbol{e}^{(1)}$ *und* $A\boldsymbol{x} = \boldsymbol{e}^{(2)}$ *gleichzeitig:*

$$M \coloneqq (A \mid \mathbb{I}_2) = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\boldsymbol{M_2.\leftarrow M_2.-M_1.}} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & -1 & -1 & 1 \end{array}\right) \xrightarrow[\boldsymbol{M_1.\leftarrow M_1.-M_2.}]{\boldsymbol{M_2.\leftarrow -M_2.}} \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & -1 \end{array}\right).$$

Rechts stehen nun die $\mathfrak{B}$*-Koordinaten der beiden Standardbasisvektoren. Nach Beobachtung 5.72 erhalten wir damit die bereits bekannte Basiswechselmatrix*

$$S_{\mathfrak{B}\leftarrow\mathfrak{E}} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}.$$ △

Dasselbe Prinzip führt zur Berechnung einer Inversen. Zu $A \in K^{n\times n}$ suchen wir eine Matrix B mit $AB = \mathbb{I}_n$. Die j-te Spalte von B muss also das System $A\boldsymbol{x} = \boldsymbol{e}^{(j)}$ lösen. Wir schreiben deshalb alle n rechten Seiten zur erweiterten Matrix $(A \mid \mathbb{I}_n)$ zusammen. Erhalten wir durch Zeilenumformungen $(\mathbb{I}_n \mid B)$, so stehen rechts die Lösungen dieser Systeme (Beobachtung 6.9). Damit gilt $AB = \mathbb{I}_n$ und nach Korollar 5.66 ist A invertierbar mit $B = A^{-1}$.

Ist umgekehrt $A \in K^{n\times n}$ invertierbar, so hat A nach Satz 5.65 Rang n. Nach Beobachtung 6.21 hat ihre Zeilenstufenform daher in jeder Spalte ein Pivotelement. Die reduzierte Zeilenstufenform ist somit $\mathbb{I}_n$. Wir können die linke Hälfte also durch das Gauß-Verfahren auf die gewünschte Form bringen und führen dabei alle Umformungen gleichzeitig an der rechten Hälfte aus.

Korollar 6.27. Eine Matrix $A \in K^{n\times n}$ ist genau dann invertierbar, wenn sich $(A \mid \mathbb{I}_n)$ durch elementare Zeilenumformungen in eine Matrix der Form $(\mathbb{I}_n \mid B)$ überführen lässt. In diesem Fall gilt $B = A^{-1}$.

Beispiel 6.28. *Wir bestimmen die Inverse der reellen Matrix*

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix},$$

indem wir das Gauß-Verfahren auf die erweiterte Matrix $M := (B \mid \mathbb{I}_3)$ anwenden:

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{\boldsymbol{M_2.} \leftarrow \boldsymbol{M_2.} - 2\boldsymbol{M_1.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{\boldsymbol{M_3.} \leftarrow \boldsymbol{M_3.} - \boldsymbol{M_2.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right)$$

$$\xrightarrow{\boldsymbol{M_2.} \leftarrow \boldsymbol{M_2.} - \boldsymbol{M_3.}} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -4 & 2 & -1 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right)$$

$$\xrightarrow{\boldsymbol{M_1.} \leftarrow \boldsymbol{M_1.} - \boldsymbol{M_2.}} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 5 & -2 & 1 \\ 0 & 1 & 0 & -4 & 2 & -1 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right).$$

Nach Korollar 6.27 ist B somit invertierbar mit

$$B^{-1} = \begin{pmatrix} 5 & -2 & 1 \\ -4 & 2 & -1 \\ 2 & -1 & 1 \end{pmatrix}.$$

Zur Kontrolle berechnen wir

$$BB^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 5 & -2 & 1 \\ -4 & 2 & -1 \\ 2 & -1 & 1 \end{pmatrix} = \mathbb{I}_3. \qquad \triangle$$

Bemerkung 6.29. Die Dimension der linearen Hülle der Zeilen einer Matrix A heißt ihr *Zeilenrang*. Er stimmt mit dem Rang von A, also ihrem *Spaltenrang*, überein. Denn elementare Zeilenumformungen erhalten die lineare Hülle der Zeilen. In Zeilenstufenform sind die Nichtnullzeilen wegen ihrer versetzten Pivotpositionen linear unabhängig und bilden somit eine Basis dieser Hülle. Ihre Anzahl ist nach Beobachtung 6.21 gerade $\operatorname{rang} A$. Insbesondere gilt $\operatorname{rang} A^\top = \operatorname{rang} A$.

6.3 Determinanten

Mit dem Gauß-Verfahren können wir den Rang einer Matrix und damit die Dimension des Bildes der zugehörigen linearen Abbildung bestimmen (Beobachtung 6.21). Für einen Vektorraumautomorphismus $\mathbb{R}^n \to \mathbb{R}^n$ ist diese Dimension stets n (Satz 5.65). Wie stark er Flächen und Volumina verändert, erkennen wir am Rang jedoch nicht. Dafür untersuchen wir, wie er das Einheitsquadrat beziehungsweise den Einheitswürfel verformt.

Die Spalten einer reellen $n \times n$-Matrix geben die Bilder der Standardbasisvektoren an. Im $\mathbb{R}^2$ sind sie daher die Kanten des Parallelogramms, auf das die Abbildung das Einheitsquadrat abbildet. Im $\mathbb{R}^3$ entsteht entsprechend ein Spat. Wir suchen eine Größe, an der wir den Flächeninhalt des Parallelogramms beziehungsweise das Volumen des Spats ablesen können.

Beispiel 6.30. *Wir betrachten die Wirkung der Matrizen*

$$S = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad P = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

auf das Einheitsquadrat $\mathcal{Q} \coloneqq [0,1]^2$ (siehe Abb. 6.3). Die Spalten der Matrizen sind genau die Bilder von $\boldsymbol{e}^{(1)}$ und $\boldsymbol{e}^{(2)}$ unter den entsprechenden Abbildungen. Da jeder Punkt in $\mathcal{Q}$ eine Linearkombination von $\boldsymbol{e}^{(1)}$ und $\boldsymbol{e}^{(2)}$ ist, können wir die Wirkungen direkt an den Matrizen ablesen: Die Matrix S bildet $\mathcal{Q}$ auf ein Rechteck mit Seitenlängen 3 und 2 ab; sein Flächeninhalt ist $3 \cdot 2 = 6$. Die Matrix H schert $\mathcal{Q}$ zu einem Parallelogramm gleicher Grundseite und Höhe 1; sein Flächeninhalt ist $1 \cdot 1 = 1$. Die Matrix R spiegelt $\mathcal{Q}$ (siehe Beispiel 5.35); sein Flächeninhalt bleibt 1, aber die Reihenfolge der Basisvektoren ändert sich. Die Matrix P projiziert $\mathcal{Q}$ auf die Gerade $\langle (1,1)^\top \rangle$ (siehe Beispiel 5.35), wobei der Flächeninhalt 0 wird. △

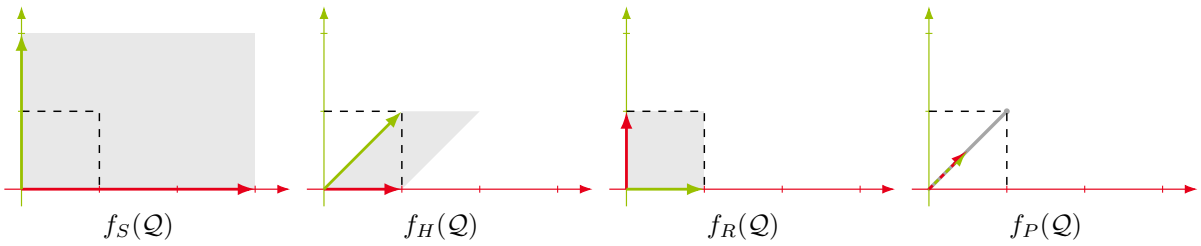


Abbildung 6.3: Die Bilder des Einheitsquadrats unter den Abbildungen aus Beispiel 6.30.

Wir hatten in den Beobachtungen 4.20 und 4.21 bereits überlegt, dass der Flächeninhalt des durch $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ aufgespannten Parallelogramms $|u_1 v_2 - u_2 v_1|$ beträgt und das Volumen des durch $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$ aufgespannten Spats $|(\boldsymbol{u} \times \boldsymbol{v} \mid \boldsymbol{w})|$. Wir können also bereits Fläche und Volumen des Bildes des Einheitsquadrats $[0,1]^2$ bzw. des Einheitswürfels $[0,1]^3$ unter einer linearen Abbildung berechnen, indem wir für $\boldsymbol{u}$, $\boldsymbol{v}$ und ggf. $\boldsymbol{w}$ die Spalten der entsprechenden Matrix einsetzen. Wir wollen das jetzt auf beliebige Dimensionen und zugrunde liegende Körper verallgemeinern.

Analog zu den natürlichen Zahlen (Abschnitt 2.1), den reellen Zahlen (Abschnitt 3.6) und dem Skalarprodukt (Abschnitt 4.6) gehen wir wieder axiomatisch vor, indem wir Eigenschaften festlegen, die wir uns für ein verallgemeinertes Volumen (bzw. eine verallgemeinerte Fläche) wünschen.

Wir versehen das Volumen zunächst mit einem Vorzeichen, damit wir die Wirkung der Spiegelung R in Beispiel 6.30 auf das Einheitsquadrat am Vorzeichenwechsel bei der Fläche seines Bildes erkennen können.

Wir gehen dazu von den Formeln $\mathrm{fl}(\boldsymbol{u}, \boldsymbol{v}) \coloneqq u_1 v_2 - u_2 v_1$ und $\mathrm{vol}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \coloneqq (\boldsymbol{u} \times \boldsymbol{v} \mid \boldsymbol{w})$, also ohne den Betrag zu nehmen, aus.

Eine zentrale Eigenschaft der Abbildungen fl und vol ist, dass sie *multilinear* sind, also linear, wenn wir alle bis auf ein beliebiges Argument festhalten: Bei Streckung einer Spalte um einen Faktor λ ändert sich der Wert der Abbildung um denselben Faktor λ, also

$$\mathrm{fl}(\lambda \boldsymbol{u}, \boldsymbol{v}) = \mathrm{fl}(\boldsymbol{u}, \lambda \boldsymbol{v}) = \lambda \,\mathrm{fl}(\boldsymbol{u}, \boldsymbol{v}) \qquad \text{bzw.} \qquad \mathrm{vol}(\lambda \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \mathrm{vol}(\boldsymbol{u}, \lambda \boldsymbol{v}, \boldsymbol{w}) = \mathrm{vol}(\boldsymbol{u}, \boldsymbol{v}, \lambda \boldsymbol{w}) = \lambda \,\mathrm{vol}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w});$$

insbesondere kehrt die Spiegelung einer Spalte, also Streckung um Faktor $\lambda = -1$, das Vorzeichen um. Andererseits addieren sich die (vorzeichenbehafteten) Höhen bei festgehaltener Grundseite bzw. Grundfläche, wenn wir Spalten addieren, also beispielsweise

$$\mathrm{fl}(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{v'}) = \mathrm{fl}(\boldsymbol{u}, \boldsymbol{v}) + \mathrm{fl}(\boldsymbol{u}, \boldsymbol{v'}) \qquad \text{bzw.} \qquad \mathrm{vol}(\boldsymbol{u}, \boldsymbol{v} + \boldsymbol{v'}, \boldsymbol{w}) = \mathrm{vol}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) + \mathrm{vol}(\boldsymbol{u}, \boldsymbol{v'}, \boldsymbol{w}).$$

Spannen die Spalten nicht den ganzen Raum auf, so sind Fläche und Volumen 0; das gilt insbesondere, wenn zwei Spalten identisch sind. Das Einheitsquadrat hat Flächeninhalt 1, der Einheitswürfel Volumen 1.

Diese Eigenschaften genügen tatsächlich, um das verallgemeinerte Volumen eindeutig festzulegen.

Definition 6.31. Sei K ein Körper und $n \in \mathbb{N}$. Eine Abbildung $\det : K^{n \times n} \to K$ heißt *Determinante*, wenn

(i) sie *multilinear* in den Spalten ist, also für jede Spalte, alle $\boldsymbol{u}, \boldsymbol{v} \in K^n$ und $\lambda, \mu \in K$ gilt:

$$\det((\cdots \ \lambda \boldsymbol{u} + \mu \boldsymbol{v} \ \cdots)) = \lambda \det((\cdots \ \boldsymbol{u} \ \cdots)) + \mu \det((\cdots \ \boldsymbol{v} \ \cdots)),$$

(ii) sie *alternierend* ist, also $\det(A) = 0$ für alle $A \in K^{n \times n}$ gilt, für die $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ und $\boldsymbol{A}_{\cdot \boldsymbol{i}} = \boldsymbol{A}_{\cdot \boldsymbol{j}}$ existieren.

(iii) sie *normiert* ist, also $\det(\mathbb{I}_n) = 1$ gilt.

Wir setzen ohne Beweis voraus, dass es immer eine Abbildung gibt, die die Bedingungen einer Determinante erfüllt. Um das zu beweisen, müssen eine solche Abbildung konstruiert und ihre Eigenschaften nachgewiesen werden.

Satz 6.32. Für jeden Körper K und jedes $n \in \mathbb{N}$ gibt es eine Determinante $\det_{K,n} : K^{n \times n} \to K$.

Wir bezeichnen im Folgenden der Einfachheit halber die Determinante aus Satz 6.32 jeweils mit $\det \coloneqq \det_{K,n}$, wenn K und n aus dem Kontext ersichtlich sind. Wir werden später sehen (Korollar 6.37), dass die Determinante für festes K und n eindeutig bestimmt ist.

Zunächst halten wir fest, wie sich die Determinante bei Spaltenumformungen, analog zu den Zeilenumformungen aus Definition 6.8, ändert.

Beobachtung 6.33. Seien $A \in K^{n \times n}$ und $\lambda \in K$. Dann gilt

(i) $\forall i \in \{1, \ldots, n\}\colon \det((\cdots \ \lambda \boldsymbol{A}_{\cdot \boldsymbol{i}} \ \cdots)) = \lambda \det(A)$ (Spaltenskalierung).

(ii) $\forall i, j \in \{1, \ldots, n\}, i \neq j\colon \det((\cdots \ \boldsymbol{A}_{\cdot \boldsymbol{i}} + \lambda \boldsymbol{A}_{\cdot \boldsymbol{j}} \ \cdots)) = \det(A)$ (Spaltenaddition).

(iii) $\forall i, j \in \{1, \ldots, n\}, i < j\colon \det((\cdots \ \boldsymbol{A}_{\cdot \boldsymbol{j}} \ \cdots \ \boldsymbol{A}_{\cdot \boldsymbol{i}} \ \cdots)) = -\det(A)$ (Spaltenvertauschung).

Beweis. (i) folgt aus der Multilinearität. Für (ii) gilt

$$\det((\cdots\ \boldsymbol{A_{\cdot i}} + \lambda \boldsymbol{A_{\cdot j}}\ \cdots)) \overset{6.31\,(i)}{=} \det(A) + \lambda \det((\cdots\ \boldsymbol{A_{\cdot j}}\ \cdots)) \overset{6.31\,(ii)}{=} \det(A).$$

Für (iii) gilt

$$\begin{aligned}
0 &\overset{6.31\,(ii)}{=} \det((\cdots\ \boldsymbol{A_{\cdot i}} + \boldsymbol{A_{\cdot j}}\ \cdots\ \boldsymbol{A_{\cdot i}} + \boldsymbol{A_{\cdot j}}\ \cdots)) \\
&\overset{6.31\,(i)}{=} \det((\cdots\ \boldsymbol{A_{\cdot i}}\ \cdots\ \boldsymbol{A_{\cdot i}}\ \cdots)) + \det((\cdots\ \boldsymbol{A_{\cdot i}}\ \cdots\ \boldsymbol{A_{\cdot j}}\ \cdots)) \\
&\qquad + \det((\cdots\ \boldsymbol{A_{\cdot j}}\ \cdots\ \boldsymbol{A_{\cdot i}}\ \cdots)) + \det((\cdots\ \boldsymbol{A_{\cdot j}}\ \cdots\ \boldsymbol{A_{\cdot j}}\ \cdots)) \\
&\overset{6.31\,(ii)}{=} \det((\cdots\ \boldsymbol{A_{\cdot i}}\ \cdots\ \boldsymbol{A_{\cdot j}}\ \cdots)) + \det((\cdots\ \boldsymbol{A_{\cdot j}}\ \cdots\ \boldsymbol{A_{\cdot i}}\ \cdots)).
\end{aligned}$$

□

Diese Regeln erklären, wie sich die Determinante bei Spaltenumformungen ändert. Das Gauß-Verfahren verwendet allerdings Zeilenumformungen, deren Einfluss auf die Determinante wir noch nicht verstehen. Wir behelfen uns, indem wir das Verfahren auf die transponierte Matrix $A^\top$ anwenden und A in jedem Schritt über $A = (A^\top)^\top$ mitverändern. Jede Zeilenumformung von $A^\top$ entspricht dann einer Spaltenumformung von A. Nach der Vorwärtsphase des Gauß-Verfahrens ist $A^\top$ in Zeilenstufenform, also A in transponierter Zeilenstufenform, und wir können die Determinante wie folgt ablesen.

Satz 6.34. Sei $T \in K^{n\times n}$ mit $n \in \mathbb{N}$. Ist $T^\top$ in Zeilenstufenform, so gilt

$$\det(T) = \prod_{i=1}^{n} T_{ii}.$$

Beweis. Für $n = 0$ ist das leere Produkt gleich 1, in Übereinstimmung mit der Normierung $\det(\mathbb{I}_0) = 1$. Für $n = 1$ gilt nach Multilinearität und Normierung $\det(T) = T_{11}\det(\mathbb{I}_1) = T_{11}$. Sei nun $n \geq 2$. Da $T^\top$ in Zeilenstufenform mit Pivotspalte $p_j \geq j$ für jede Nichtnullzeile ist (Definition 6.12), hat die j-te Spalte von T die Form $\boldsymbol{T_{\cdot j}} = \sum_{i=j}^{n} T_{ij}\,\boldsymbol{e}^{(i)}$. Damit gilt

$$\begin{aligned}
\det(T) &\overset{6.31\,(i)}{=} T_{nn}\det((\boldsymbol{T_{\cdot 1}}\ \cdots\ \boldsymbol{T_{\cdot (n-1)}}\ \boldsymbol{e}^{(n)})) \\
&= T_{nn}\det((\cdots\ (T_{n-1,n-1}\boldsymbol{e}^{(n-1)} + T_{n,n-1}\boldsymbol{e}^{(n)})\ \boldsymbol{e}^{(n)})) \\
&\overset{6.31\,(i)}{=} T_{nn}\big(T_{n-1,n-1}\det((\cdots\ \boldsymbol{e}^{(n-1)}\ \boldsymbol{e}^{(n)})) + T_{n,n-1}\det((\cdots\ \boldsymbol{e}^{(n)}\ \boldsymbol{e}^{(n)}))\big) \\
&\overset{6.31\,(ii)}{=} T_{nn}T_{n-1,n-1}\det((\cdots\ \boldsymbol{e}^{(n-1)}\ \boldsymbol{e}^{(n)})) \\
&\ \ \vdots \\
&= \left(\prod_{i=1}^{n} T_{ii}\right)\det((\boldsymbol{e}^{(1)}\ \cdots\ \boldsymbol{e}^{(n)})) = \left(\prod_{i=1}^{n} T_{ii}\right)\det(\mathbb{I}_n) \\
&\overset{6.31\,(iii)}{=} \prod_{i=1}^{n} T_{ii}.
\end{aligned}$$

□

Die Vorwärtsphase des Gauß-Verfahrens verwendet ausschließlich Zeilentausch und Addition von Zeilenvielfachen. Nach Beobachtung 6.33 müssen wir nur die Vertauschungen mitzählen, um die Veränderung der Determinante während der Vorwärtsphase zu bestimmen. Zusammen mit Satz 6.34 haben wir somit ein Berechnungsverfahren für die Determinante.

Satz 6.35. Sei $A \in K^{n\times n}$ mit $n \in \mathbb{N}$. Ist $s \in \mathbb{N}$ die Anzahl der Zeilenvertauschungen, die die Vorwärtsphase des Gauß-Verfahrens benötigt, um $A^\top$ in die Zeilenstufenform T zu überführen, so gilt

$$\det(A) = (-1)^s \prod_{i=1}^{n} T_{ii}. \tag{6.3}$$

Beispiel 6.36. *Wir greifen die Kantenvektoren $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ des Spats aus Beispiel 4.22 auf und schreiben sie als Spalten einer Matrix:*

$$A = (\boldsymbol{u}\ \boldsymbol{v}\ \boldsymbol{w}) = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 0 \\ 0 & 2 & 1 \end{pmatrix}.$$

Zur Berechnung von $\det(A)$ nach Satz 6.35 wenden wir die Vorwärtsphase auf $A^\top$ an. Zunächst ziehen wir die erste Zeile von der zweiten und das Doppelte der ersten Zeile von der dritten ab. Anschließend addieren wir das Doppelte der neuen zweiten Zeile zur dritten. Wir bezeichnen mit M jeweils die aktuelle Matrix:

$$A^\top = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 2 \\ 2 & 0 & 1 \end{pmatrix} \xrightarrow[\boldsymbol{M_3.} \leftarrow \boldsymbol{M_3.} - 2\boldsymbol{M_1.}]{\boldsymbol{M_2.} \leftarrow \boldsymbol{M_2.} - \boldsymbol{M_1.}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -2 & 1 \end{pmatrix}$$
$$\xrightarrow{\boldsymbol{M_3.} \leftarrow \boldsymbol{M_3.} + 2\boldsymbol{M_2.}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 5 \end{pmatrix} = T.$$

Es wurde keine Zeile vertauscht, also ist $s = 0$. Aus den Diagonaleinträgen von T folgt

$$\det(A) = (-1)^0 \cdot 1 \cdot 1 \cdot 5 = 5.$$

Der Betrag stimmt mit dem in Beispiel 4.22 berechneten Spatvolumen überein. △

Nach Satz 6.13 können wir die Vorwärtsphase für jede Matrix $A^\top$ durchführen. Die Gleichheit (6.3) gilt für jede Wahl einer Determinante $\det$; ihre rechte Seite wird allein durch diese Gauß-Rechnung bestimmt und legt daher $\det(A)$ und damit die Abbildung $\det$ eindeutig fest. Für $n = 0$ ist $\det(\mathbb{I}_0) = 1$ bereits durch die Normierung festgelegt. Zusammen mit der Existenz aus Satz 6.32 erhalten wir folgendes Korollar.

Korollar 6.37. Für jeden Körper K und jedes $n \in \mathbb{N}$ gibt es genau eine Determinante $\det : K^{n\times n} \to K$.

Das Diagonalprodukt zeigt auch, wann die Determinante verschwindet. Damit erhalten wir eine weitere der äquivalenten Bedingungen aus Satz 5.65.

Satz 6.38. Eine Matrix $A \in K^{n\times n}$ ist genau dann invertierbar, wenn $\det(A) \neq 0$ gilt.

Beweis. Sei T die Zeilenstufenform von $A^\top$ aus Satz 6.35. Nach (6.3) ist $\det(A) \neq 0$ genau dann, wenn alle Diagonaleinträge von T von null verschieden sind (die von null verschiedenen Elemente bilden nach Definition 3.64 eine multiplikative Gruppe). In Zeilenstufenform ist dies genau dann der Fall, wenn jede Zeile ein Pivotelement enthält. Nach Beobachtung 6.21 bedeutet dies $\operatorname{rang} A^\top = n$ (es ist $r = n$), nach Satz 5.65 ist das äquivalent zur Invertierbarkeit von $A^\top$. Diese ist nach Satz 5.68 äquivalent zur Invertierbarkeit von A. □

Wenden wir zwei lineare Abbildungen hintereinander an, so erwarten wir, dass sich ihre Flächen- beziehungsweise Volumenfaktoren multiplizieren. Algebraisch entspricht dies der folgenden Produktregel.

Satz 6.39. Für $A, B \in K^{n\times n}$ gilt
$$\det(AB) = \det(A)\,\det(B).$$

Beweis. Ist $\det(A) = 0$, so ist A nach Satz 6.38 nicht invertierbar. Nach den Sätzen 5.61 und 5.23 ist f_A dann nicht surjektiv. Also ist auch $f_{AB} = f_A \circ f_B$ nicht surjektiv. Somit ist AB nicht invertierbar und nach Satz 6.38 gilt $\det(AB) = 0 = \det(A)\,\det(B)$.

Sei also $\lambda := \det(A) \neq 0$. Wir betrachten die Abbildung
$$d_A : K^{n\times n} \to K, \qquad B \mapsto \lambda^{-1}\det(AB).$$
Nach Definition 5.46 ist die j-te Spalte von AB gleich $A\boldsymbol{B}_{\cdot j}$. Da die Multiplikation von A mit einer Spalte nach Satz 5.31 linear und die Determinante multilinear ist, ist auch d_A multilinear. Gleiche Spalten von B ergeben gleiche Spalten von AB, also $\det(AB) = 0$; deshalb ist d_A auch alternierend. Wegen $d_A(\mathbb{I}_n) = \lambda^{-1}\det(A) = 1$ ist d_A eine Determinante nach Definition 6.31. Wegen der Eindeutigkeit aus Korollar 6.37 gilt $d_A(B) = \det(B)$ und damit
$$\det(AB) = \lambda d_A(B) = \lambda\det(B) = \det(A)\,\det(B). \qquad \square$$

Das Transponieren vertauscht die Rollen von Zeilen und Spalten. Wir zeigen nun, dass sich dabei die Determinante nicht ändert.

Satz 6.40. Für $A \in K^{n\times n}$ gilt $\det(A^\top) = \det(A)$.

Beweis. Ist A nicht invertierbar, so ist nach Satz 5.68 auch $A^\top$ nicht invertierbar. Mit Satz 6.38 folgt daher $\det(A^\top) = 0 = \det(A)$.

Sei nun A invertierbar. Nach Satz 5.65 ist $\operatorname{rang} A = n$, also liefert das Gauß-Verfahren nach Beobachtung 6.21 eine reduzierte Zeilenstufenform mit n Pivotelementen, die jeweils 1 sind. Die reduzierte Zeilenstufenform ist also $\mathbb{I}_n$. Nach Beobachtung 6.14 kann jede elementare Zeilenumformung durch eine inverse Zeilenumformung rückgängig gemacht werden und jede solche Umformung entspricht einer Multiplikation von links mit einer bestimmten Matrix. Also ist
$$A = E^{(\ell)} \cdots E^{(1)}\mathbb{I}_n = E^{(\ell)} \cdots E^{(1)},$$
wobei die Matrizen $E^{(1)}, \ldots, E^{(\ell)}$ der Form aus Beobachtung 6.14 genügen. Für diese Matrizen gilt $\det(E^\top) = \det(E)$: Für Zeilentausch und Zeilenskalierung ist sogar $E^\top = E$; die Matrix E für Zeilenaddition und ihre Transponierte $E^\top$ entstehen jeweils durch eine Spaltenaddition aus $\mathbb{I}_n$, daher gilt für sie $\det(E) = \det(\mathbb{I}_n) = \det(E^\top)$ nach Beobachtung 6.33.

Damit ist
$$\begin{aligned}
\det(A^\top) &\overset{5.58}{=} \det\big((E^{(1)})^\top \cdots (E^{(\ell)})^\top\big)\\
&\overset{6.39}{=} \det((E^{(1)})^\top) \cdots \det((E^{(\ell)})^\top)\\
&= \det(E^{(1)}) \cdots \det(E^{(\ell)})\\
&= \det(E^{(\ell)}) \cdots \det(E^{(1)})\\
&\overset{6.39}{=} \det(E^{(\ell)} \cdots E^{(1)})\\
&= \det(A). \qquad \square
\end{aligned}$$

Nach Satz 6.40 gilt $\det(A^\top) = \det(A)$. Wenden wir Satz 6.35 auf $A^\top$ statt auf A an, so arbeitet die Vorwärtsphase auf $(A^\top)^\top = A$. Damit können wir die Determinante unmittelbar durch Zeilenumformungen berechnen.

Korollar 6.41. Überführt die Vorwärtsphase des Gauß-Verfahrens $A \in K^{n\times n}$ durch s Zeilenvertauschungen und Addition von Zeilenvielfachen in die Zeilenstufenform T, so gilt

$$\det(A) = (-1)^s \prod_{i=1}^{n} T_{ii}.$$

Die oben eingeführten Abbildungen fl und vol sind multilinear, alternierend und normiert, erfüllen also die Eigenschaften aus Definition 6.31. Wegen der Eindeutigkeit aus Korollar 6.37 sind ihre Werte daher die Determinanten der Matrizen mit den jeweiligen Vektoren als Spalten. Ihr Betrag liefert nach den Beobachtungen 4.20 und 4.21 den Flächeninhalt beziehungsweise das Volumen. Für die Flächenformel rechnen wir das noch einmal nach.

Beispiel 6.42. *Für eine Matrix aus $\mathbb{R}^{2\times 2}$ ergibt sich*

$$\begin{aligned}\det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) &= \det\left(\begin{pmatrix} a\boldsymbol{e}^{(1)} + c\boldsymbol{e}^{(2)} & b\boldsymbol{e}^{(1)} + d\boldsymbol{e}^{(2)} \end{pmatrix}\right) \\ &= ab\det((\boldsymbol{e}^{(1)}\ \boldsymbol{e}^{(1)})) + ad\det((\boldsymbol{e}^{(1)}\ \boldsymbol{e}^{(2)})) \\ &\quad + cb\det((\boldsymbol{e}^{(2)}\ \boldsymbol{e}^{(1)})) + cd\det((\boldsymbol{e}^{(2)}\ \boldsymbol{e}^{(2)})) \\ &= ad\det((\boldsymbol{e}^{(1)}\ \boldsymbol{e}^{(2)})) + cb\det((\boldsymbol{e}^{(2)}\ \boldsymbol{e}^{(1)})) = ad - bc.\end{aligned}$$

Für Spalten $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$ ist $|\det((\boldsymbol{u}\ \boldsymbol{v}))|$ somit nach Beobachtung 4.20 der Flächeninhalt des aufgespannten Parallelogramms.

Für die Matrizen aus Beispiel 6.30 erhalten wir $\det(S) = 6$, $\det(H) = 1$, $\det(R) = -1$ und $\det(P) = 0$. Die Beträge stimmen mit den dort bestimmten Flächeninhalten überein. Das negative Vorzeichen bei der Spiegelung hält zusätzlich die Umkehrung der Reihenfolge der Kanten fest. △

Beispiel 6.43. *Wir greifen die Koeffizientenmatrix aus dem Eingangsbeispiel von Abschnitt 6.2 wieder auf. Durch Vertauschen der beiden Zeilen erhalten wir bereits eine Zeilenstufenform:*

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

Daher ist $\det(A) = -1 \cdot 1 \cdot 2 = -2$.

Für die Matrix B aus Beispiel 6.28 haben wir die ersten beiden Schritte bereits ausgeführt:

$$B = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dabei wurde von der zweiten Zeile das Doppelte der ersten und danach von der dritten Zeile die neue zweite abgezogen. Beide Schritte erhalten die Determinante. Somit ist $\det(B) = 1 \cdot 1 \cdot 1 = 1$. Die weiteren Schritte zur Inversenberechnung benötigen wir hier nicht. △

Wir können die Determinante über beliebigen Körpern berechnen. Das folgende Beispiel gibt eine Anwendung aus der Informatik.

Beispiel 6.44. *Wie in Beispiel 6.7 wollen wir verlorene Datenbits aus erhaltenen Prüfbits wiederherstellen. Für fünf Datenbits $\boldsymbol{x} = (x_1, \dots, x_5)^\top \in \mathbb{F}_2^5$ speichern wir zusätzlich die fünf Prüfbits*

$$\boldsymbol{p} \coloneqq A\boldsymbol{x}, \qquad A \coloneqq \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Die ersten vier Prüfbits sind Summen benachbarter Datenbits, das fünfte ist die Summe aller fünf Datenbits. Bestimmt $\boldsymbol{p}$ die Daten $\boldsymbol{x}$ eindeutig, auch wenn alle fünf Datenbits verloren gehen?

Wir wenden die Vorwärtsphase des Gauß-Verfahrens an. Dabei bezeichnet M jeweils die aktuelle Matrix. Wegen $1 + 1 = 0$ erhalten wir

$$A \xrightarrow{\boldsymbol{M_1.} \leftrightarrow \boldsymbol{M_2.}} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \xrightarrow[\boldsymbol{M_5.} \leftarrow \boldsymbol{M_5.} + \boldsymbol{M_3.}]{\boldsymbol{M_5.} \leftarrow \boldsymbol{M_5.} + \boldsymbol{M_1.}} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = T.$$

Nach Korollar 6.41 ist $\det(A) = (-1) \cdot 1^5 = 1$, denn in $\mathbb{F}_2$ gilt $-1 = 1$; eine Zeilenvertauschung ändert die Determinante hier also nicht. Nach Satz 6.38 ist A invertierbar. Damit besitzt $A\boldsymbol{x} = \boldsymbol{p}$ nach Korollar 6.5 für jedes $\boldsymbol{p}$ genau eine Lösung: Die Rekonstruktion gelingt eindeutig.

Warum hilft gerade das fünfte Prüfbit? Drehen wir alle Datenbits um, so bleiben die vier Summen benachbarter Bits gleich. Die Summe aller fünf Bits ändert sich dagegen um 1. Würden wir stattdessen $p_5 = x_1 + x_5$ wählen, wäre $p_5 = p_1 + p_2 + p_3 + p_4$ schon durch die ersten vier Prüfbits bestimmt. Addieren wir dann die ersten vier Zeilen zur fünften, entsteht eine Nullzeile; die Determinante ist 0. Tatsächlich liefern nun etwa $(0, 0, 0, 0, 0)^\top$ und $(1, 1, 1, 1, 1)^\top$ dieselben Prüfbits. △

Bemerkung 6.45. Die Determinante ist in jeder einzelnen Spalte linear. Skalieren wir dagegen die ganze Matrix $A \in K^{n \times n}$, so skalieren wir alle n Spalten und erhalten für $\lambda \in K$

$$\det(\lambda A) = \lambda^n \det(A).$$

Insbesondere ist $\det(\lambda \mathbb{I}_n) = \lambda^n$. Die Linearität in einer Spalte ist also von der Linearität als Funktion der ganzen Matrix zu unterscheiden.

Für invertierbares A liefert die Produktregel $1 = \det(AA^{-1}) = \det(A)\det(A^{-1})$. Die Determinante der Inversen ist daher die multiplikative Inverse von $\det(A)$. Für beliebiges $A \in K^{n \times n}$ und invertierbares $S \in K^{n \times n}$ gilt außerdem

$$\det(S^{-1}AS) = \det(S^{-1})\det(A)\det(S) = \det(S)^{-1}\det(A)\det(S) = \det(A).$$

Wir halten beide Folgerungen fest.

Korollar 6.46. Für invertierbares $A \in K^{n\times n}$ gilt $\det(A^{-1}) = \det(A)^{-1}$. Ähnliche Matrizen haben dieselbe Determinante.

Nach Korollar 5.76 sind die Darstellungsmatrizen einer linearen Abbildung eines Vektorraums in sich selbst bezüglich verschiedener Basen ähnlich. Neben Rang und Invertierbarkeit (Satz 5.81) bleibt somit auch die Determinante beim Basiswechsel erhalten. Für reelle Abbildungen ist insbesondere der Flächen- beziehungsweise Volumenfaktor unabhängig von der gewählten Basis.

6.4 Eigenwerte, Eigenvektoren und Eigenräume

Die Determinante beschreibt die Verformung von Flächen und Volumina durch einen Vektorraumautomorphismus mit einer Zahl. Um die Wirkung einer linearen Abbildung genauer zu erfassen, wollen wir jetzt Richtungen untersuchen, die unter der Abbildung erhalten bleiben. Finden wir eine Basis aus Vektoren in solchen Richtungen, können wir die Abbildung besonders einfach beschreiben: Wir zerlegen einen Vektor bezüglich dieser Basis und müssen seine Bestandteile nur mit den jeweiligen Faktoren multiplizieren.

In diesem Abschnitt bezeichnet K einen Körper und n eine natürliche Zahl.

Definition 6.47. Ein Vektor $\boldsymbol{v} \in K^n \setminus \{\mathbf{0}\}$ ist ein *Eigenvektor* von $A \in K^{n\times n}$ zu dem *Eigenwert* $\lambda \in K$, wenn

$$A\boldsymbol{v} = \lambda\boldsymbol{v}.$$

Beispiel 6.48. *Für die Skalierung $S = \left(\begin{smallmatrix} 3 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ aus Beispiel 6.30 gilt*

$$S\boldsymbol{e}^{(1)} = S\begin{pmatrix}1\\0\end{pmatrix} = 3\begin{pmatrix}1\\0\end{pmatrix}, \qquad S\begin{pmatrix}0\\1\end{pmatrix} = S\boldsymbol{e}^{(2)} = 2\begin{pmatrix}0\\1\end{pmatrix},$$

also sind $\boldsymbol{e}^{(1)}$, $\boldsymbol{e}^{(2)}$ Eigenvektoren zu den Eigenwerten 3 beziehungsweise 2. Zerlegen wir einen Vektor $\boldsymbol{v} \in \mathbb{R}^2$ in diese beiden Richtungen, so folgt

$$S\boldsymbol{v} = S\left(v_1\boldsymbol{e}^{(1)} + v_2\boldsymbol{e}^{(2)}\right) = 3v_1\boldsymbol{e}^{(1)} + 2v_2\boldsymbol{e}^{(2)}.$$

Die beiden Bestandteile werden also unabhängig voneinander mit den jeweiligen Faktoren multipliziert. Bei einer gleichmäßigen Streckung mit der Matrix $\lambda\mathbb{I}_n$ ist sogar jeder von null verschiedene Vektor ein Eigenvektor zum Eigenwert λ. △

Beispiel 6.49. *Für die Spiegelung und die Projektion aus Beispiel 5.77 mit den Matrizen*

$$R = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \qquad P = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$$

kennen wir bereits die Wirkung auf die Vektoren $\boldsymbol{u} = (1,1)^\top$ und $\boldsymbol{v} = (1,-1)^\top$:

$$R\boldsymbol{u} = \boldsymbol{u}, \qquad R\boldsymbol{v} = -\boldsymbol{v}, \qquad P\boldsymbol{u} = \boldsymbol{u}, \qquad P\boldsymbol{v} = \mathbf{0} = 0\boldsymbol{v}.$$

Für R und P ist $\boldsymbol{u}$ jeweils Eigenvektor zum Eigenwert 1 und $\boldsymbol{v}$ ist Eigenvektor von R zum Eigenwert -1. Der Bildvektor $R\boldsymbol{v}$ zeigt also in die entgegengesetzte Richtung zu $\boldsymbol{v}$, liegt aber weiterhin auf derselben Geraden. Für P

ist $\boldsymbol{v}$ ein Eigenvektor zum Eigenwert 0: Sein Bild ist der Nullvektor, obwohl $\boldsymbol{v}$ selbst von null verschieden ist. Der Nullvektor wird in der Definition als Eigenvektor ausgeschlossen, da $A\mathbf{0} = \lambda\mathbf{0}$ für jedes λ gilt.

Die beiden Vektoren $\boldsymbol{u}$ und $\boldsymbol{v}$ bilden eine Basis von $(\mathbb{R}^2, +, \cdot)$; nach Normierung erhalten wir die Orthonormalbasis aus Beispiel 5.77. Schreiben wir einen Vektor als $a\boldsymbol{u} + b\boldsymbol{v}$, so können wir die Wirkung auf die Bestandteile also wieder trennen:

$$R(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{u} - b\boldsymbol{v}, \qquad P(a\boldsymbol{u} + b\boldsymbol{v}) = a\boldsymbol{u}.$$

△

Eine lineare Abbildung muss allerdings keine Richtungen erhalten.

Beispiel 6.50. *Die Matrix*

$$D = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

beschreibt eine Vierteldrehung der Ebene. Sie hat keine reellen Eigenwerte und Eigenvektoren. △

Um die Wirkung einer linearen Abbildung in Bestandteile zu zerlegen, fassen wir die Eigenvektoren zusammen, auf die die Matrix mit demselben Faktor wirkt.

Definition 6.51. Der *Eigenraum* von $A \in K^{n\times n}$ zu $\lambda \in K$ ist

$$E_\lambda(A) \coloneqq \{\boldsymbol{v} \in K^n \mid A\boldsymbol{v} = \lambda\boldsymbol{v}\}.$$

Wir beziehen dabei den Nullvektor in den Eigenraum mit ein, damit wir einen Untervektorraum erhalten.

Satz 6.52. Für $A \in K^{n\times n}$ und $\lambda \in K$ gilt

$$E_\lambda(A) = \ker(A - \lambda\mathbb{I}_n).$$

Insbesondere ist $E_\lambda(A)$ ein Untervektorraum von $(K^n, +, \cdot)$.

Beweis. Per Definition gehört ein Vektor $\boldsymbol{v}$ genau dann zu $E_\lambda(A)$, wenn $A\boldsymbol{v} = \lambda\boldsymbol{v}$ gilt. Wegen

$$A\boldsymbol{v} = \lambda\boldsymbol{v} \quad\iff\quad (A - \lambda\mathbb{I}_n)\boldsymbol{v} = \mathbf{0} \quad\iff\quad \boldsymbol{v} \in \ker(A - \lambda\mathbb{I}_n)$$

stimmen die Mengen überein. Als Kern der zu $A - \lambda\mathbb{I}_n$ gehörenden linearen Abbildung ist $E_\lambda(A)$ nach Satz 5.14 ein Untervektorraum. □

Eine Basis des Eigenraums beschreibt somit alle Eigenvektoren zum jeweiligen Eigenwert: Es sind genau die von null verschiedenen Linearkombinationen der Basisvektoren. Bevor wir diese Basis berechnen können, müssen wir jedoch wissen, welche Faktoren λ überhaupt Eigenwerte sind. Wir gehen deshalb in zwei Schritten vor: Zuerst bestimmen wir die Eigenwerte, danach die zugehörigen Eigenräume.

Ein Skalar λ ist genau dann ein Eigenwert, wenn $E_\lambda(A) \neq \{\mathbf{0}\}$ ist. Nach Satz 6.52 suchen wir also diejenigen λ, für die das homogene System $(A - \lambda\mathbb{I}_n)\boldsymbol{v} = \mathbf{0}$ eine von null verschiedene Lösung besitzt. Damit gilt

$$\begin{aligned} \lambda \text{ ist Eigenwert von } A &\iff \ker(A - \lambda\mathbb{I}_n) \neq \{\mathbf{0}\} \\ &\overset{5.65}{\iff} (A - \lambda\mathbb{I}_n) \text{ ist nicht invertierbar} \\ &\overset{6.38}{\iff} \det(A - \lambda\mathbb{I}_n) = 0. \end{aligned}$$

Wir suchen also die Nullstellen eines Ausdrucks, den wir als Polynom auffassen können: Jede Spalte von $A - t\mathbb{I}_n$ hat die Form $\boldsymbol{A_{\cdot j}} - t\boldsymbol{e}^{(j)}$, also können wir nacheinander die Multilinearität bezüglich der verschiedenen Spalten anwenden, um den Faktor t jeweils aus der Determinante zu ziehen. Konkreter ist

$$\begin{aligned} \det(A - t\mathbb{I}_n) &= \det\big((\boldsymbol{A_{\cdot 1}} - t\boldsymbol{e}^{(1)} \;\; \boldsymbol{A_{\cdot 2}} - t\boldsymbol{e}^{(2)} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &= \det\big((\boldsymbol{A_{\cdot 1}} \;\; \boldsymbol{A_{\cdot 2}} - t\boldsymbol{e}^{(2)} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &\quad - t\det\big((\boldsymbol{e}^{(1)} \;\; \boldsymbol{A_{\cdot 2}} - t\boldsymbol{e}^{(2)} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &= \det\big((\boldsymbol{A_{\cdot 1}} \;\; \boldsymbol{A_{\cdot 2}} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &\quad - t\det\big((\boldsymbol{A_{\cdot 1}} \;\; \boldsymbol{e}^{(2)} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &\quad - t\det\big((\boldsymbol{e}^{(1)} \;\; \boldsymbol{A_{\cdot 2}} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &\quad + t^2\det\big((\boldsymbol{e}^{(1)} \;\; \boldsymbol{e}^{(2)} \;\; \cdots \;\; \boldsymbol{A_{\cdot n}} - t\boldsymbol{e}^{(n)})\big) \\ &= \cdots \\ &= \det(A) + c_1 t + \cdots + c_{n-1}t^{n-1} + (-t)^n \det(\mathbb{I}_n) \end{aligned}$$

ein Polynom vom Grad n.

Definition 6.53. Das charakteristische Polynom einer Matrix $A \in K^{n\times n}$ ist

$$\chi_A(t) \coloneqq \det(A - t\mathbb{I}_n).$$

Damit können wir die Eigenwerte über die folgende Bedingung charakterisieren.

Satz 6.54. Ein Skalar $\lambda \in K$ ist genau dann ein Eigenwert von $A \in K^{n\times n}$, wenn

$$\chi_A(\lambda) = 0.$$

Zur Berechnung bestimmen wir also zunächst die Nullstellen des charakteristischen Polynoms in K. Für jeden so gefundenen Eigenwert λ lösen wir anschließend das homogene System $(A - \lambda\mathbb{I}_n)\boldsymbol{v} = \mathbf{0}$ und geben eine Basis seines Lösungsraums an. Dieser ist nach Satz 6.52 der gesuchte Eigenraum. Eine Basis bestimmen wir mit dem Gauß-Verfahren wie in Satz 6.17; die von null verschiedenen Linearkombinationen dieser Basisvektoren sind die Eigenvektoren.

Beispiel 6.55. *Für die Matrix* $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right) \in \mathbb{R}^{2\times 2}$ *erhalten wir mit der Formel aus Beispiel 6.42*

$$\chi_A(t) = \det\left(\begin{pmatrix} 2-t & 1 \\ 1 & 2-t \end{pmatrix}\right) = (2-t)^2 - 1 = (3-t)(1-t).$$

Nach Satz 6.54 sind die Eigenwerte also 3 *und* 1*. Für den Eigenwert* 3 *gilt*

$$(A - 3\mathbb{I}_2)\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \iff \begin{cases} -x + y = 0, \\ x - y = 0, \end{cases} \iff y = x.$$

Die zweite Gleichung ist das Negative der ersten und kann durch Addition der ersten beseitigt werden. Wir können y *frei wählen und erhalten* $E_3(A) = \langle (1,1)^\top \rangle$*. Für den Eigenwert* 1 *folgt analog*

$$(A - \mathbb{I}_2)\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \iff \begin{cases} x + y = 0, \\ x + y = 0, \end{cases} \iff y = -x,$$

also $E_1(A) = \langle (1,-1)^\top \rangle$*. Beide Eigenräume haben Dimension* 1*; ihre Eigenvektoren sind jeweils die von null verschiedenen Vielfachen des angegebenen Basisvektors.* △

Bei einem Basiswechsel wird eine Matrix A durch eine ähnliche Matrix $S^{-1}AS$ ersetzt (Korollar 5.76). Wir zeigen nun, dass dabei das charakteristische Polynom und damit nach Satz 6.54 auch die Eigenwerte erhalten bleiben.

Satz 6.56. Ähnliche Matrizen haben dasselbe charakteristische Polynom.

Beweis. Wir verwenden ohne Beweis, dass die Determinantenregeln aus Abschnitt 6.3 auch beim formalen Ausmultiplizieren von Polynomen gelten.[1] Für $A \in K^{n\times n}$ und eine invertierbare Matrix $S \in K^{n\times n}$ erhalten wir

$$\begin{aligned} \chi_{S^{-1}AS}(t) &= \det(S^{-1}AS - t\mathbb{I}_n) = \det(S^{-1}AS - tS^{-1}S) \\ &= \det(S^{-1}AS - tS^{-1}\mathbb{I}_n S) = \det\big(S^{-1}(A - t\mathbb{I}_n)S\big) \\ &\overset{6.39}{=} \det(S^{-1})\,\chi_A(t)\,\det(S) \overset{6.46}{=} \chi_A(t). \end{aligned}$$

□

Ob das charakteristische Polynom überhaupt Nullstellen besitzt, ob es also Eigenwerte gibt, hängt wesentlich vom zugrunde liegenden Körper ab.

Beispiel 6.57. *Für die Vierteldrehung* D *aus Beispiel 6.50 erhalten wir*

$$\chi_D(t) = \det\left(\begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix}\right) = t^2 + 1.$$

Das Polynom hat keine reelle Nullstelle; dies bestätigt unsere geometrische Feststellung, dass D *keine reellen Eigenwerte besitzt. Fassen wir dieselbe Matrix dagegen als Matrix über* $\mathbb{C}$ *auf, so gilt* $t^2 + 1 = (t - \mathrm{i})(t + \mathrm{i})$*. Über* $\mathbb{C}$ *hat* D *daher die Eigenwerte* i *und* $-\mathrm{i}$*.* △

Über $\mathbb{C}$ sichert uns der Fundamentalsatz der Algebra (Satz 3.109) die Existenz einer Nullstelle jedes Polynoms von Grad $n > 0$, insbesondere also des charakteristischen Polynoms. Daraus folgt die allgemeine Existenz von Eigenwerten.

[1] Die Determinantenregeln übertragen sich auf die zugehörigen Polynomfunktionen. Dass deren Gleichheit erhalten bleibt, genügt hier aber nicht, da unterschiedliche Polynome dieselbe Polynomfunktion haben können.

Korollar 6.58. Jede Matrix $A \in \mathbb{C}^{n\times n}$ mit $n > 0$ besitzt einen Eigenwert.

Das charakteristische Polynom bestimmt somit die Eigenwerte. Die Dimensionen der zugehörigen Eigenräume sind durch das charakteristische Polynom allein jedoch nicht bestimmt.

Beispiel 6.59. *Die Scherungsmatrix* $H = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ *aus Beispiel 6.30 und die Einheitsmatrix* $\mathbb{I}_2$ *über* $\mathbb{R}$ *haben dasselbe charakteristische Polynom:*

$$\chi_H(t) = \det\left(\begin{pmatrix} 1-t & 1 \\ 0 & 1-t \end{pmatrix}\right) = (1-t)^2 = \chi_{\mathbb{I}_2}(t).$$

Beide haben also nur den Eigenwert 1*. Für die Scherung gilt*

$$(H - \mathbb{I}_2)\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \quad \iff \quad y = 0,$$

also $E_1(H) = \langle (1,0)^\top \rangle$*. Dagegen ist* $E_1(\mathbb{I}_2) = \mathbb{R}^2$*. Trotz gleicher charakteristischer Polynome haben die Eigenräume somit unterschiedliche Dimensionen, nämlich* 1 *und* 2*.* △

Bemerkung 6.60. Eigenwerte, -vektoren und -räume einer linearen Abbildung $f\colon V \to V$ können wir analog definieren: Ein Vektor $\boldsymbol{v} \in V \setminus \{\mathbf{0}\}$ heißt Eigenvektor zum Eigenwert λ, wenn $f(\boldsymbol{v}) = \lambda\boldsymbol{v}$ gilt, und der zugehörige Eigenraum ist $E_\lambda(f) \coloneqq \{\boldsymbol{v} \in V \mid f(\boldsymbol{v}) = \lambda\boldsymbol{v}\}$. Bildet V einen endlichdimensionalen Vektorraum, können wir zur Berechnung eine geordnete Basis $\mathfrak{B}$ wählen und mit der Darstellungsmatrix $A = [f]_{\mathfrak{B}\leftarrow\mathfrak{B}}$ arbeiten, da $f(\boldsymbol{v}) = \lambda\boldsymbol{v} \iff A[\boldsymbol{v}]_{\mathfrak{B}} = \lambda[\boldsymbol{v}]_{\mathfrak{B}}$ (Satz 5.40) und $\boldsymbol{v} \neq \mathbf{0} \Leftrightarrow [\boldsymbol{v}]_{\mathfrak{B}} \neq \mathbf{0}$ gilt. Nach Korollar 5.76 sind die Darstellungsmatrizen bezüglich verschiedener geordneter Basen ähnlich, so dass wir wegen Satz 6.56 das charakteristische Polynom von f unabhängig von der gewählten Basis über $\chi_f \coloneqq \chi_A$ definieren können.

Diagonalisierung

Wir können jetzt Eigenwerte und Eigenräume berechnen. Wann finden wir darin genügend Eigenvektoren, um jeden Vektor in solche Richtungen zu zerlegen? Bei der Scherung aus Beispiel 6.59 reicht der einzige, eindimensionale Eigenraum dafür nicht aus. Bei der Matrix aus Beispiel 6.55 gelingt es dagegen.

Beispiel 6.61. *Für* $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ *haben wir in Beispiel 6.55 die Eigenvektoren* $\boldsymbol{u} \coloneqq (1,1)^\top$ *und* $\boldsymbol{v} \coloneqq (1,-1)^\top$ *zu den Eigenwerten* 3 *beziehungsweise* 1 *bestimmt. Sie bilden eine Basis von* $(\mathbb{R}^2, +, \cdot)$*. Stellen wir sie als Spalten einer Matrix zusammen, erhalten wir*

$$S \coloneqq \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad S^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

und damit

$$S^{-1}AS = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 3 & -1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Bezüglich der geordneten Basis $\mathfrak{C} = (\boldsymbol{u}, \boldsymbol{v})$ *wird* f_A *nach Korollar 5.76 also durch die rechte Matrix dargestellt. Für* $\boldsymbol{x} = c_1\boldsymbol{u} + c_2\boldsymbol{v}$ *bedeutet dies*

$$A\boldsymbol{x} = 3c_1\boldsymbol{u} + c_2\boldsymbol{v}.$$

Wir multiplizieren die beiden Bestandteile einzeln mit den Eigenwerten. In diesen Koordinaten beeinflussen sie sich nicht gegenseitig. △

Eine solche Basis vereinfacht die Beschreibung der Abbildung. Wir geben ihr einen Namen.

Definition 6.62. Eine Basis von $(K^n, +, \cdot)$ aus Eigenvektoren von $A \in K^{n\times n}$ heißt *Eigenbasis* von A. Besitzt A eine Eigenbasis, so heißt A *diagonalisierbar*.

Eine quadratische Matrix, deren Einträge außerhalb der Diagonale null sind, nennen wir *Diagonalmatrix* und schreiben

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) \coloneqq \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Wie im Beispiel liefern die Vektoren einer Eigenbasis eine Basiswechselmatrix, bezüglich derer A durch eine Diagonalmatrix dargestellt wird.

Satz 6.63. Seien $A, S \in K^{n\times n}$, wobei S invertierbar ist, und $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ mit $\lambda_1, \ldots, \lambda_n \in K$. Dann gilt $S^{-1}AS = D$ genau dann, wenn jede Spalte $\boldsymbol{S_{\cdot j}}$ ein Eigenvektor von A zum Eigenwert λ_j ist. Insbesondere ist A genau dann diagonalisierbar, wenn A zu einer Diagonalmatrix ähnlich ist.

Beweis. Die Spalten von S bilden nach Satz 5.65 eine Basis und sind daher von null verschieden. Die Aussage ergibt sich damit aus

$$S^{-1}AS = D \quad\Longleftrightarrow\quad AS = SD \quad\Longleftrightarrow\quad \forall j \in \{1, \ldots, n\}\colon A\boldsymbol{S_{\cdot j}} = \lambda_j \boldsymbol{S_{\cdot j}}.$$

Insbesondere folgt in Vorwärtsrichtung aus der Ähnlichkeit von A zu der Diagonalmatrix D, dass die Spalten von S eine Eigenbasis bilden, A also diagonalisierbar ist. Ist umgekehrt A diagonalisierbar, dann existiert per Definition eine Eigenbasis, deren Vektoren nach Satz 5.65 die Spalten einer invertierbaren Matrix S bilden, mit den zugehörigen Eigenwerten $\lambda_1, \ldots, \lambda_n$. Aus der Rückwärtsrichtung der obigen Äquivalenz folgt, dass A ähnlich zu $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ ist. □

Es bleibt die Frage, wann wir eine Eigenbasis finden können. Wir zeigen, dass die Eigenvektoren zu verschiedenen Eigenwerten gute Kandidaten für Basisvektoren sind.

Satz 6.64. Sind $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(r)}$ Eigenvektoren von $A \in K^{n\times n}$ zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$, dann sind $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(r)}$ linear unabhängig.

Beweis. Wir führen eine Induktion über r. Für $r = 1$ gilt die Aussage, da $\mathbf{0}$ kein Eigenvektor sein kann. Sei nun $r > 1$ und seien $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(r-1)}$ linear unabhängig. Wir haben zu zeigen, dass

$$\sum_{j=1}^{r} a_j \boldsymbol{v}^{(j)} = \mathbf{0}$$

mit $\boldsymbol{a} \in K^r$ impliziert, dass $\boldsymbol{a} = \mathbf{0}$.

Um den letzten Summanden zu beseitigen, multiplizieren wir mit $A - \lambda_r \mathbb{I}_n$. Wegen $A\boldsymbol{v}^{(j)} = \lambda_j \boldsymbol{v}^{(j)}$ erhalten wir

$$\mathbf{0} = (A - \lambda_r \mathbb{I}_n) \sum_{j=1}^{r} a_j \boldsymbol{v}^{(j)} = \sum_{j=1}^{r-1} a_j (\lambda_j - \lambda_r) \boldsymbol{v}^{(j)}.$$

Nach Induktionsannahme sind die ersten $r-1$ Eigenvektoren linear unabhängig. Also gilt $a_j(\lambda_j - \lambda_r) = 0$ und wegen $\lambda_j \neq \lambda_r$ auch $a_j = 0$ für $j < r$. In der ursprünglichen Gleichung bleibt $a_r \boldsymbol{v}^{(r)} = \mathbf{0}$, woraus wegen $\boldsymbol{v}^{(r)} \neq \mathbf{0}$ auch $a_r = 0$ folgt. □

Insbesondere kann eine Matrix in $K^{n\times n}$ nach Satz 4.64 höchstens n Eigenwerte haben, wegen $\dim K^n = n$ (Beispiel 4.68). Ein Eigenraum kann mehrere linear unabhängige Eigenvektoren enthalten. Um eine Eigenbasis zusammenzusetzen, wählen wir daher in jedem Eigenraum eine Basis. Das gelingt genau dann, wenn wir so auf $\dim K^n = n$ Vektoren kommen.

Satz 6.65. Eine Matrix $A \in K^{n\times n}$ mit den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$ ist genau dann diagonalisierbar, wenn

$$\sum_{j=1}^{r} \dim E_{\lambda_j}(A) = n.$$

In diesem Fall ergibt die Vereinigung von Basen der Eigenräume eine Eigenbasis. Insbesondere ist A diagonalisierbar, wenn sie n verschiedene Eigenwerte besitzt.

Beweis. Wir wählen für jedes $i \in \{1, \ldots, r\}$ eine Basis $B_i = \{\boldsymbol{v}^{(i,1)}, \ldots, \boldsymbol{v}^{(i,d_i)}\}$ von $E_{\lambda_i}(A)$ mit $d_i := \dim E_{\lambda_i}(A)$ und setzen $B := \bigcup_{i=1}^{r} B_i$. Da jeder Eigenvektor nur zu einem Eigenwert gehören kann, sind die B_i paarweise disjunkt. Um die lineare Unabhängigkeit von B zu zeigen, seien $a_{ij} \in K$ für alle $i \in \{1, \ldots, r\}$ und $j \in \{1, \ldots, d_i\}$ mit

$$\sum_{i=1}^{r} \boldsymbol{w}^{(i)} := \sum_{i=1}^{r} \sum_{j=1}^{d_i} a_{ij} \boldsymbol{v}^{(i,j)} = \mathbf{0},$$

wobei $\boldsymbol{w}^{(i)} \in E_{\lambda_i}(A)$, da jeder dieser Vektoren eine Linearkombination von Vektoren des jeweiligen Untervektorraums ist (Satz 6.52). Die von null verschiedenen $\boldsymbol{w}^{(i)}$ sind Eigenvektoren zu verschiedenen Eigenwerten und deshalb nach Satz 6.64 linear unabhängig. Damit die Summe null ist, muss daher $\boldsymbol{w}^{(i)} = \mathbf{0}$ für alle $i \in \{1, \ldots, r\}$ gelten. Da jede Basis B_i selbst linear unabhängig ist, folgt daraus $a_{ij} = 0$ für alle $i \in \{1, \ldots, r\}$ und $j \in \{1, \ldots, d_i\}$. Also ist B linear unabhängig.

Nach Satz 4.64 gilt somit

$$\sum_{i=1}^{r} \dim E_{\lambda_i}(A) = \sum_{i=1}^{r} |B_i| = |B| \leq n.$$

Ist die Summe gleich n, insbesondere also wenn A genau $r = n$ verschiedene Eigenwerte besitzt, so ist B nach Satz 4.73 eine Basis von $(K^n, +, \cdot)$ und damit eine Eigenbasis. Also ist A diagonalisierbar.

Ist umgekehrt A diagonalisierbar, so besitzt A nach Definition 6.62 eine Eigenbasis. Jeder Vektor dieser Eigenbasis liegt im zugehörigen Eigenraum $E_{\lambda_i}(A) = \langle B_i \rangle \subseteq \langle B \rangle$. Folglich kann jedes $\boldsymbol{v} \in K^n$ als Linearkombination von Vektoren aus $\langle B \rangle$ dargestellt werden, also $\boldsymbol{v} \in \langle B \rangle$; damit erzeugt B ganz K^n. Zusammen mit der bereits gezeigten linearen Unabhängigkeit ist B also eine Basis. Daher gilt $\sum_{i=1}^{r} \dim E_{\lambda_i}(A) = |B| = n$. □

Damit können wir auch die bisherigen Beispiele einordnen.

Beispiel 6.66. *Die Scherungsmatrix H aus Beispiel 6.59 ist nicht diagonalisierbar: Ihr einziger Eigenraum $E_1(H)$ hat Dimension 1, während wir für eine Basis von $(\mathbb{R}^2, +, \cdot)$ zwei Vektoren benötigen. Auch die Vierteldrehung aus Beispiel 6.57 ist über $\mathbb{R}$ nicht diagonalisierbar, da sie keine reellen Eigenwerte hat. Über $\mathbb{C}$ besitzt sie dagegen die beiden verschiedenen Eigenwerte i und $-\mathrm{i}$ und ist deshalb diagonalisierbar.* △

Auch wenn eine Matrix weniger als n verschiedene Eigenwerte hat, können die Eigenräume zusammen genügend unabhängige Vektoren für eine Eigenbasis liefern. Wir prüfen dies an einem größeren Beispiel.

Beispiel 6.67. *Wir betrachten die Matrix $A \in \mathbb{R}^{5\times 5}$ und die Vektoren $\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(5)}$, die wir als Spalten von S zusammenfassen:*

$$A \coloneqq \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \qquad S \coloneqq \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Um zu prüfen, ob die vorgegebenen Vektoren Eigenvektoren sind, berechnen wir ihre Bilder gleichzeitig durch

$$AS = \begin{pmatrix} 3 & 0 & 1 & 0 & 0 \\ 3 & 0 & -1 & 0 & 0 \\ 0 & 3 & 0 & 1 & 0 \\ 0 & 3 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 3\boldsymbol{v}^{(1)} & 3\boldsymbol{v}^{(2)} & \boldsymbol{v}^{(3)} & \boldsymbol{v}^{(4)} & 2\boldsymbol{v}^{(5)} \end{pmatrix}.$$

Die ersten beiden Spalten sind also Eigenvektoren zum Eigenwert 3, die nächsten beiden zum Eigenwert 1 und die letzte zum Eigenwert 2.

Die Vektoren $\boldsymbol{v}^{(1)}$ und $\boldsymbol{v}^{(2)}$ sind linear unabhängig: In $a\boldsymbol{v}^{(1)} + b\boldsymbol{v}^{(2)} = \mathbf{0}$ liefert die erste Koordinate $a = 0$ und die dritte $b = 0$. Dasselbe Argument gilt für $\boldsymbol{v}^{(3)}$ und $\boldsymbol{v}^{(4)}$. Wegen $\boldsymbol{v}^{(5)} \neq \mathbf{0}$ gilt somit

$$\dim E_3(A) \geq 2, \qquad \dim E_1(A) \geq 2, \qquad \dim E_2(A) \geq 1.$$

Wie im Beweis von Satz 6.65 gezeigt, ist die Summe der Dimensionen aller Eigenräume höchstens 5. Die drei gefundenen Eigenräume erreichen diese Schranke bereits. Also gilt

$$\dim E_3(A) = 2, \qquad \dim E_1(A) = 2, \qquad \dim E_2(A) = 1,$$

und es kann keine weiteren Eigenwerte geben, da jeder weitere Eigenraum mindestens Dimension 1 hätte. Nach Satz 6.65 ist A daher diagonalisierbar, obwohl sie nur drei verschiedene Eigenwerte besitzt.

Die beiden ersten Spalten bilden eine Basis von $E_3(A)$, die nächsten beiden eine Basis von $E_1(A)$ und die letzte eine Basis von $E_2(A)$. Ihre Vereinigung ist nach Satz 6.65 eine Eigenbasis. Damit ist S nach Satz 5.65 invertierbar und nach Satz 6.63 gilt

$$S^{-1}AS = \operatorname{diag}(3, 3, 1, 1, 2).$$

Wie sich die Wirkung von A in dieser Eigenbasis zerlegen lässt, sehen wir etwa am Vektor

$$\boldsymbol{x} \coloneqq (3, 1, 2, 0, 1)^\top = 2\boldsymbol{v}^{(1)} + \boldsymbol{v}^{(2)} + \boldsymbol{v}^{(3)} + \boldsymbol{v}^{(4)} + \boldsymbol{v}^{(5)}.$$

Die ersten beiden Basisvektoren werden mit 3 multipliziert, die nächsten beiden bleiben unverändert und der letzte wird mit 2 multipliziert. Daher erhalten wir

$$A\boldsymbol{x} = 6\boldsymbol{v}^{(1)} + 3\boldsymbol{v}^{(2)} + \boldsymbol{v}^{(3)} + \boldsymbol{v}^{(4)} + 2\boldsymbol{v}^{(5)} = (7, 5, 4, 2, 2)^\top.$$

In der Eigenbasis berechnen wir die Wirkung von A also, indem wir jede Koordinate mit dem zugehörigen Eigenwert multiplizieren. △

Besonders nützlich ist die Diagonalisierung, wenn wir eine Matrix wiederholt auf einen Vektor anwenden: In einer Eigenbasis wird jeder Bestandteil bei jedem Schritt mit demselben Eigenwert multipliziert. Für $k \in \mathbb{N}$ aufeinanderfolgende Anwendungen verwenden wir die Matrixpotenz A^k aus Definition 5.46.

Bei einer Diagonalmatrix können wir die Potenzen direkt berechnen: Beim Multiplizieren werden nur die jeweiligen Diagonaleinträge miteinander multipliziert. Durch den Basiswechsel genügt diese einfache Rechnung auch für jede diagonalisierbare Matrix.

Satz 6.68. Ist $A = SDS^{-1}$ mit invertierbarem $S \in K^{n\times n}$ und $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, so gilt für alle $k \in \mathbb{N}$

$$A^k = SD^kS^{-1} \qquad \text{und} \qquad D^k = \operatorname{diag}(\lambda_1^k, \dots, \lambda_n^k).$$

Beweis. Die Aussage über D^k folgt direkt aus der Definition des Matrixprodukts durch Induktion über k. Auch die erste Gleichheit zeigen wir durch Induktion. Für $k = 0$ gilt $SD^0S^{-1} = S\mathbb{I}_nS^{-1} = \mathbb{I}_n = A^0$. Im Induktionsschritt erhalten wir

$$A^{k+1} = AA^k = SDS^{-1}SD^kS^{-1} = SD^{k+1}S^{-1}.$$ □

Anwendungsbeispiel: Fibonacci-Zahlen

Die Fibonacci-Zahlen entstehen, indem wir jeweils die beiden vorherigen Zahlen addieren. Wir wollen eine Formel finden, mit der sich jede dieser Zahlen direkt aus ihrem Index berechnen lässt.

Beispiel 6.69. *Die* Fibonacci-Zahlen *sind rekursiv definiert durch*

$$f_0 \coloneqq 0, \qquad f_1 \coloneqq 1, \qquad f_{k+2} \coloneqq f_{k+1} + f_k \quad \textit{für } k \in \mathbb{N}.$$

Die ersten Werte sind $0, 1, 1, 2, 3, 5, 8, 13, \dots$. Um die nächste Zahl zu berechnen, benötigen wir jeweils die beiden letzten Werte. Wir fassen sie deshalb in einem Vektor zusammen und setzen

$$\boldsymbol{x}^{(k)} \coloneqq \begin{pmatrix} f_{k+1} \\ f_k \end{pmatrix}, \qquad A \coloneqq \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Dann gilt

$$\boldsymbol{x}^{(k+1)} = \begin{pmatrix} f_{k+1} + f_k \\ f_{k+1} \end{pmatrix} = A\boldsymbol{x}^{(k)}, \qquad \boldsymbol{x}^{(0)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Per Induktion folgt sofort $\boldsymbol{x}^{(k)} = A^k\boldsymbol{x}^{(0)}$. Wir bestimmen die Potenzen mithilfe einer Eigenbasis und Satz 6.68.

Das charakteristische Polynom von A ist

$$\chi_A(t) = \det\left(\begin{pmatrix} 1-t & 1 \\ 1 & -t \end{pmatrix}\right) = (1-t)(-t) - 1 = t^2 - t - 1.$$

Nach Satz 6.54 sind die beiden Nullstellen

$$\lambda_+ \coloneqq \frac{1+\sqrt{5}}{2}, \qquad \lambda_- \coloneqq \frac{1-\sqrt{5}}{2}$$

die Eigenwerte. Für beide gilt $\lambda^2 = \lambda + 1$ und daher

$$A\begin{pmatrix}\lambda\\1\end{pmatrix} = \begin{pmatrix}\lambda+1\\\lambda\end{pmatrix} = \lambda\begin{pmatrix}\lambda\\1\end{pmatrix}.$$

Somit sind $\boldsymbol{u} \coloneqq (\lambda_+, 1)^\top$ und $\boldsymbol{v} \coloneqq (\lambda_-, 1)^\top$ Eigenvektoren zu λ_+ beziehungsweise λ_-. Da die Eigenwerte verschieden sind, sind $\boldsymbol{u}$ und $\boldsymbol{v}$ nach Satz 6.64 linear unabhängig und bilden eine Basis von $(\mathbb{R}^2, +, \cdot)$. Mit

$$S \coloneqq \begin{pmatrix}\lambda_+ & \lambda_-\\ 1 & 1\end{pmatrix}, \qquad D \coloneqq \operatorname{diag}(\lambda_+, \lambda_-)$$

gilt also nach Satz 6.63 $A = SDS^{-1}$.

Wegen $\lambda_+ - \lambda_- = \sqrt{5}$ lässt sich der Anfangsvektor besonders einfach in die beiden Eigenrichtungen zerlegen:

$$\boldsymbol{x}^{(0)} = \begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix}\lambda_+ - \lambda_-\\ 1-1\end{pmatrix} = \frac{1}{\sqrt{5}}\boldsymbol{u} - \frac{1}{\sqrt{5}}\boldsymbol{v}.$$

Die Koeffizienten $1/\sqrt{5}$ und $-1/\sqrt{5}$ sind seine Koordinaten in der Eigenbasis. Nach Satz 6.68 werden sie bei k Anwendungen von A mit λ_+^k beziehungsweise λ_-^k multipliziert. Wir erhalten

$$\boldsymbol{x}^{(k)} = A^k\boldsymbol{x}^{(0)} = \frac{\lambda_+^k}{\sqrt{5}}\boldsymbol{u} - \frac{\lambda_-^k}{\sqrt{5}}\boldsymbol{v} = \frac{1}{\sqrt{5}}\begin{pmatrix}\lambda_+^{k+1} - \lambda_-^{k+1}\\ \lambda_+^k - \lambda_-^k\end{pmatrix}.$$

An der zweiten Koordinate lesen wir die gesuchte Formel ab:

$$f_k = \frac{\lambda_+^k - \lambda_-^k}{\sqrt{5}} = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k\right) \qquad \text{für } k \in \mathbb{N}.$$

Die Eigenbasis macht aus der Rekursion zwei getrennte Multiplikationen. Dadurch können wir f_k direkt aus k bestimmen, ohne zuvor alle früheren Fibonacci-Zahlen berechnen zu müssen. Obwohl in der Formel irrationale Zahlen auftreten, ist ihr Wert für jedes $k \in \mathbb{N}$ eine natürliche Zahl. △

6.5 Symmetrische Matrizen

Eine Eigenbasis zerlegt die Wirkung einer Matrix in einzelne Multiplikationen. Können wir die Eigenbasis zusätzlich orthonormal wählen? Dann vereinfacht der Basiswechsel die Matrix und erhält zugleich Längen und Skalarprodukte. Wir werden sehen, dass für symmetrische Matrizen eine solche Eigenbasis stets möglich ist.

Um Aussagen über Längen, Orthogonalität und Vorzeichen machen zu können, arbeiten wir in diesem Abschnitt über dem angeordneten Körper $(\mathbb{R}, +, \cdot, \leq)$ und verwenden auf $\mathbb{R}^n$ das Standardskalarprodukt, wobei im Folgenden $n \in \mathbb{N}$ gilt.

Definition 6.70. Eine Matrix $A \in \mathbb{R}^{n\times n}$ heißt *symmetrisch*, wenn $A^\top = A$ gilt.

Die Vektoren einer geordneten Basis bilden die Spalten ihrer Basiswechselmatrix Q zur Standardbasis (Definition 5.71). Ihre gegenseitigen Skalarprodukte erkennen wir an einem Matrixprodukt: Der Eintrag an Position (i, j) von $Q^\top Q$ ist das Skalarprodukt der i-ten und der j-ten Spalte von Q. Bei einer Orthonormalbasis entsteht daher die Einheitsmatrix. In dem Fall heißt die zugehörige Matrix orthogonal, obwohl *orthonormal* eine deskriptivere Benennung wäre.

Definition 6.71. Eine Matrix $Q \in \mathbb{R}^{n\times n}$ heißt *orthogonal*, wenn $Q^\top Q = \mathbb{I}_n$ gilt, also wenn ihre Spalten eine Orthonormalbasis von $(\mathbb{R}^n, +, \cdot)$ bilden.

Symmetrie und Orthogonalität sind verschiedene Eigenschaften, wie die bekannten geometrischen Abbildungen zeigen.

Beispiel 6.72. *Die Matrix $D = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ der Vierteldrehung aus Beispiel 6.50 ist orthogonal, denn $D^\top D = \mathbb{I}_2$, aber nicht symmetrisch, da $D^\top \neq D$. Für die Projektionsmatrix $P = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ aus Beispiel 5.47 gilt dagegen $P^\top = P$ und $P^\top P = P^2 = P \neq \mathbb{I}_2$. Sie ist also symmetrisch, aber nicht orthogonal. Die Spiegelungsmatrix $C = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ aus dem gleichen Beispiel besitzt beide Eigenschaften, denn $C^\top = C$ und $C^2 = \mathbb{I}_2$.* △

Bei einem Basiswechsel zu einer Orthonormalbasis können wir die neuen Koordinaten durch Transponieren der Basiswechselmatrix berechnen. Zugleich bleiben Skalarprodukte und Längen erhalten. Auch Flächen und Volumina bleiben unverändert: Der Betrag der Determinante, den wir in Abschnitt 6.3 als Flächen- beziehungsweise Volumenfaktor kennengelernt haben, ist hier 1. Wir halten diese Vorteile in einem Satz fest.

Satz 6.73. Jede orthogonale Matrix $Q \in \mathbb{R}^{n\times n}$ ist invertierbar mit $Q^{-1} = Q^\top$ und erfüllt $\det(Q) \in \{-1, 1\}$. Für alle $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ gilt

$$(Q\boldsymbol{x} \mid Q\boldsymbol{y}) = (\boldsymbol{x} \mid \boldsymbol{y}) \qquad \text{und} \qquad \|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|.$$

Beweis. Nach Definition 6.71 gilt $Q^\top Q = \mathbb{I}_n$. Nach Korollar 5.66 ist Q daher invertierbar mit $Q^{-1} = Q^\top$. Für $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ folgt außerdem

$$(Q\boldsymbol{x} \mid Q\boldsymbol{y}) \overset{5.56}{=} (Q\boldsymbol{x})^\top Q\boldsymbol{y} \overset{5.58}{=} \boldsymbol{x}^\top Q^\top Q\boldsymbol{y} = \boldsymbol{x}^\top Q^{-1} Q\boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y} = (\boldsymbol{x} \mid \boldsymbol{y}).$$

Wir erhalten insbesondere $\|Q\boldsymbol{x}\|^2 = (Q\boldsymbol{x} \mid Q\boldsymbol{x}) = (\boldsymbol{x} \mid \boldsymbol{x}) = \|\boldsymbol{x}\|^2$. Da Normen nichtnegativ sind, folgt $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$. Für die Determinante gilt $\det(Q) \in \{-1, 1\}$, denn

$$1 = \det(\mathbb{I}_n) = \det(Q^\top Q) \overset{6.39}{=} \det(Q^\top)\det(Q) \overset{6.40}{=} \det(Q)^2.$$ □

Die Inversenformel aus Satz 6.73 vereinfacht die Koordinatenberechnung bezüglich einer Orthonormalbasis: Sind die Spalten von Q die Vektoren einer geordneten Orthonormalbasis $\mathfrak{Q}$ und bezeichnet $\mathfrak{E}$ die geordnete Standardbasis, so gilt nach Beobachtung 5.72 sowohl $Q = S_{\mathfrak{E}\leftarrow\mathfrak{Q}}$ als auch $\boldsymbol{x} = [\boldsymbol{x}]_{\mathfrak{E}} = Q[\boldsymbol{x}]_{\mathfrak{Q}}$. Damit erhalten wir die Koordinaten $[\boldsymbol{x}]_{\mathfrak{Q}} = Q^{-1}\boldsymbol{x} = Q^\top \boldsymbol{x}$. Dies sind genau die Skalarprodukte mit den Basisvektoren, wie wir sie bereits aus Satz 4.97 kennen.

Nachdem wir uns von den Vorteilen einer orthonormalen Eigenbasis überzeugt haben, fragen wir uns jetzt, wie wir eine solche finden können. Wir zeigen zunächst, dass Eigenvektoren zu verschiedenen Eigenwerten bereits orthogonal sind. Die Existenz einer Orthonormalbasis aus Eigenvektoren werden wir anschließend ohne Beweis übernehmen.

Beobachtung 6.74. Sei $A \in \mathbb{R}^{n\times n}$ symmetrisch. Sind $\boldsymbol{u}$ und $\boldsymbol{v}$ Eigenvektoren von A zu verschiedenen Eigenwerten, so gilt $\boldsymbol{u} \perp \boldsymbol{v}$.

Beweis. Seien $A\boldsymbol{u} = \lambda\boldsymbol{u}$ und $A\boldsymbol{v} = \mu\boldsymbol{v}$ mit $\lambda \neq \mu$. Wegen $A^\top = A$ erhalten wir

$$\lambda(\boldsymbol{u} \mid \boldsymbol{v}) = (A\boldsymbol{u} \mid \boldsymbol{v}) = (A\boldsymbol{u})^\top \boldsymbol{v} \overset{5.58}{=} \boldsymbol{u}^\top A^\top \boldsymbol{v} = \boldsymbol{u}^\top A\boldsymbol{v} = (\boldsymbol{u} \mid A\boldsymbol{v}) = \mu(\boldsymbol{u} \mid \boldsymbol{v}).$$

Also gilt $(\lambda - \mu)(\boldsymbol{u} \mid \boldsymbol{v}) = 0$ und wegen $\lambda \neq \mu$ somit $(\boldsymbol{u} \mid \boldsymbol{v}) = 0$. □

Damit ist noch nicht gezeigt, dass genügend reelle Eigenvektoren für eine Basis existieren. Dies sichert der folgende Satz, den wir ohne Beweis verwenden.

Satz 6.75 (Spektralsatz)**.** Jede symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ besitzt eine Orthonormalbasis aus Eigenvektoren. Insbesondere gibt es eine orthogonale Matrix $Q \in \mathbb{R}^{n\times n}$ und eine Diagonalmatrix $D \in \mathbb{R}^{n\times n}$ mit

$$Q^\top AQ = D \qquad \text{und} \qquad A = QDQ^\top.$$

Eine Eigenbasis liefert bereits nach Satz 6.63 eine Zerlegung in einzelne Multiplikationen. Für symmetrische Matrizen garantiert der Spektralsatz, dass eine solche Basis existiert und sogar orthonormal gewählt werden kann. In diesen Koordinaten bleiben zugleich die gewohnten Formeln für Längen und Skalarprodukte gültig. Wir können dadurch sowohl die Wirkung der Matrix als auch geometrische Fragen untersuchen, indem wir die einzelnen Eigenrichtungen getrennt betrachten.

Zur Berechnung einer orthonormalen Eigenbasis einer symmetrischen Matrix bestimmen wir wie in Abschnitt 6.4 zunächst die Eigenwerte und Basen der Eigenräume. In jedem Eigenraum wenden wir auf seine Basis das Gram–Schmidt-Verfahren aus Satz 4.98 an. Bei einem eindimensionalen Eigenraum genügt es, den gewählten Eigenvektor durch seine Norm zu teilen. Vektoren aus verschiedenen Eigenräumen sind nach Beobachtung 6.74 dann zueinander orthogonal. Nach Satz 6.65 ergeben die so gefundenen Vektoren gemeinsam eine Orthonormalbasis aus Eigenvektoren.

Beispiel 6.76. *Wir greifen die Matrix $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ aus Beispiel 6.61 wieder auf. Sie ist symmetrisch. Ihre Eigenvektoren $(1,1)^\top$ und $(1,-1)^\top$ zu den Eigenwerten 3 und 1 sind orthogonal und haben jeweils Länge $\sqrt{2}$. Wir müssen sie also nur normieren. Damit ist*

$$Q \coloneqq \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

orthogonal, und nach Satz 6.63 gilt

$$Q^\top AQ = D \coloneqq \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}.$$

Die neuen Koordinaten $(u,v)^\top$ eines Vektors $(x,y)^\top$ sind

$$\begin{pmatrix} u \\ v \end{pmatrix} = Q^\top \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} x+y \\ x-y \end{pmatrix}.$$

In diesen Koordinaten multipliziert A die erste Komponente mit 3 und lässt die zweite unverändert. Der Koordinatenwechsel erhält dabei Längen und Skalarprodukte. Abbildung 6.4 zeigt die Zerlegung $A = QDQ^\top$ am Einheitsquadrat: $Q^\top$ übersetzt in Eigenbasiskoordinaten, D streckt die erste Koordinate um den Faktor 3, und Q übersetzt zurück. Die beiden Basiswechsel erhalten Längen und Flächen; nur der mittlere Schritt verändert sie. △

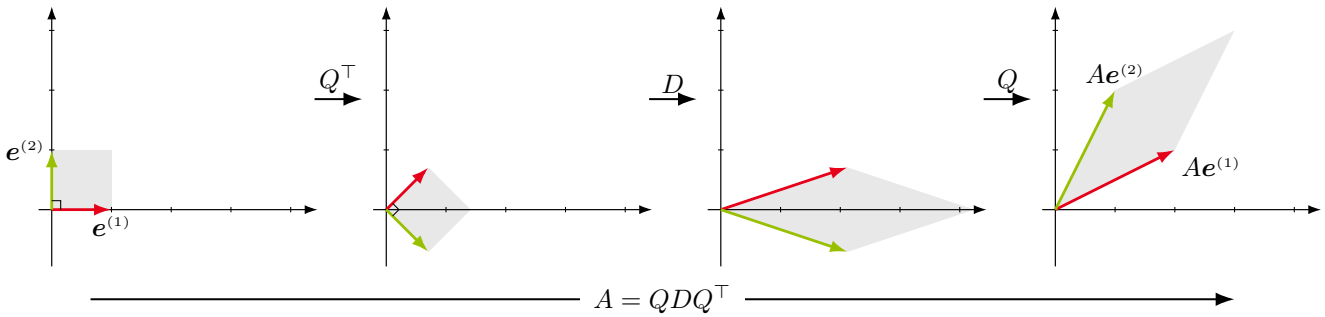


Abbildung 6.4: Die Wirkung von A aus Beispiel 6.76 auf das Einheitsquadrat als Folge von $Q^\top$, D und Q.

Wir führen nun das gesamte Verfahren an einer größeren Matrix durch, bei der wir innerhalb eines Eigenraums auch orthogonalisieren müssen.

Beispiel 6.77. *Wir wollen die symmetrische Matrix*

$$A := \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

orthogonal diagonalisieren. Dazu bestimmen wir zunächst die Eigenwerte, dann Basen der Eigenräume und daraus eine Orthonormalbasis aus Eigenvektoren.

Eigenwerte. *Wir addieren die zweite und dritte Spalte von $A - t\mathbb{I}_3$ zur ersten und klammern dort $4 - t$ aus (Beobachtung 6.33). Danach ziehen wir die erste Zeile von den beiden anderen ab. Mit Korollar 6.41 ergibt sich*

$$\begin{aligned} \chi_A(t) &= \det\left(\begin{pmatrix} 2-t & 1 & 1 \\ 1 & 2-t & 1 \\ 1 & 1 & 2-t \end{pmatrix}\right) = (4-t)\det\left(\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2-t & 1 \\ 1 & 1 & 2-t \end{pmatrix}\right) \\ &= (4-t)\det\left(\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1-t & 0 \\ 0 & 0 & 1-t \end{pmatrix}\right) = (4-t)(1-t)^2. \end{aligned}$$

Die Eigenwerte sind nach Satz 6.54 also 4 und 1.

Eigenräume. *Nach Satz 6.52 lösen wir für jeden Eigenwert λ das homogene System $(A - \lambda\mathbb{I}_3)\boldsymbol{x} = \mathbf{0}$. Für $\lambda = 4$ lautet es*

$$\begin{cases} -2x_1 + x_2 + x_3 = 0, \\ x_1 - 2x_2 + x_3 = 0, \\ x_1 + x_2 - 2x_3 = 0. \end{cases}$$

Die Differenz der ersten beiden Gleichungen liefert $3(x_2 - x_1) = 0$, also $x_2 = x_1$. Die erste Gleichung ergibt dann $x_3 = x_1$. Umgekehrt erfüllt jede Wahl $x_1 = x_2 = x_3$ alle drei Gleichungen. Daher gilt

$$E_4(A) = \langle \boldsymbol{b}^{(1)} \rangle \qquad \textit{mit} \qquad \boldsymbol{b}^{(1)} \coloneqq (1,1,1)^\top.$$

Für $\lambda = 1$ lauten alle drei Gleichungen $x_1 + x_2 + x_3 = 0$. Wir können x_2 und x_3 frei wählen; dann ist $x_1 = -x_2 - x_3$ und jeder Lösungsvektor hat die Form (vgl. Satz 6.17)

$$\boldsymbol{x} = \begin{pmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

Somit gilt

$$E_1(A) = \langle \boldsymbol{b}^{(2)}, \boldsymbol{b}^{(3)} \rangle \qquad \textit{mit} \qquad \boldsymbol{b}^{(2)} \coloneqq (-1,1,0)^\top, \quad \boldsymbol{b}^{(3)} \coloneqq (-1,0,1)^\top.$$

Orthonormalbasis. *Im eindimensionalen Eigenraum $E_4(A)$ genügt das Normieren:*

$$\boldsymbol{q}^{(1)} \coloneqq \frac{\boldsymbol{b}^{(1)}}{\|\boldsymbol{b}^{(1)}\|} = \frac{1}{\sqrt{3}}(1,1,1)^\top.$$

Die Basisvektoren von $E_1(A)$ sind dagegen nicht orthogonal, denn $(\boldsymbol{b}^{(2)} \mid \boldsymbol{b}^{(3)}) = 1 \neq 0$. Wir wenden das Gram–Schmidt-Verfahren aus Satz 4.98 an. Zunächst normieren wir $\boldsymbol{b}^{(2)}$ und ziehen dann von $\boldsymbol{b}^{(3)}$ seinen Anteil in dieser Richtung ab:

$$\begin{aligned} \boldsymbol{q}^{(2)} &\coloneqq \frac{1}{\sqrt{2}}(-1,1,0)^\top, \\ \boldsymbol{u}^{(3)} &\coloneqq \boldsymbol{b}^{(3)} - (\boldsymbol{b}^{(3)} \mid \boldsymbol{q}^{(2)})\, \boldsymbol{q}^{(2)} = (-1,0,1)^\top - \frac{1}{2}(-1,1,0)^\top = \frac{1}{2}(-1,-1,2)^\top. \end{aligned}$$

Wegen $\|\boldsymbol{u}^{(3)}\| = \sqrt{6}/2$ erhalten wir

$$\boldsymbol{q}^{(3)} \coloneqq \frac{\boldsymbol{u}^{(3)}}{\|\boldsymbol{u}^{(3)}\|} = \frac{1}{\sqrt{6}}(-1,-1,2)^\top.$$

Der Vektor $\boldsymbol{u}^{(3)}$ liegt als Linearkombination von $\boldsymbol{b}^{(2)}$ und $\boldsymbol{b}^{(3)}$ weiterhin in $E_1(A)$. Die Vektoren $\boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}$ bilden eine Orthonormalbasis dieses Eigenraums. Nach Beobachtung 6.74 stehen beide senkrecht auf $\boldsymbol{q}^{(1)}$. Zusammen bilden die drei Vektoren somit eine Orthonormalbasis aus Eigenvektoren.

Basiswechsel. *Wir setzen die gefundenen Vektoren als Spalten zusammen:*

$$Q \coloneqq (\boldsymbol{q}^{(1)}\ \boldsymbol{q}^{(2)}\ \boldsymbol{q}^{(3)}) = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}.$$

Die Matrix Q ist orthogonal. Da $A\boldsymbol{q}^{(1)} = 4\boldsymbol{q}^{(1)}$, $A\boldsymbol{q}^{(2)} = \boldsymbol{q}^{(2)}$ und $A\boldsymbol{q}^{(3)} = \boldsymbol{q}^{(3)}$, gilt

$$AQ = (4\boldsymbol{q}^{(1)}\ \boldsymbol{q}^{(2)}\ \boldsymbol{q}^{(3)}) = QD \qquad \textit{mit} \qquad D \coloneqq \begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Multiplikation mit $Q^\top = Q^{-1}$ liefert $Q^\top A Q = D$. In den neuen Koordinaten vervierfacht A also die erste Komponente und lässt die beiden anderen unverändert.

Längen und Volumina. *Wir prüfen noch einmal konkret, dass der Basiswechsel die Geometrie erhält. Für* $\boldsymbol{x} = (1,2,3)^\top$ *lauten die neuen Koordinaten*

$$\boldsymbol{y} = Q^\top \boldsymbol{x} = \begin{pmatrix} 2\sqrt{3} \\ 1/\sqrt{2} \\ 3/\sqrt{6} \end{pmatrix}, \qquad \text{und es gilt} \qquad \|\boldsymbol{y}\|^2 = 12 + \frac{1}{2} + \frac{9}{6} = 14 = \|\boldsymbol{x}\|^2.$$

Die Länge bleibt also unverändert. Für die Determinante ziehen wir die drei Normierungsfaktoren aus den Spalten und bringen die verbleibende Matrix durch Zeilenadditionen in Zeilenstufenform:

$$\det(Q) = \frac{1}{\sqrt{3}\sqrt{2}\sqrt{6}} \det\left(\begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & 0 & 2 \end{pmatrix}\right) = \frac{1}{6} \det\left(\begin{pmatrix} 1 & -1 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\right) = 1.$$

Der von $\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}$ *aufgespannte Spat hat daher wie der Einheitswürfel das Volumen* 1*; auch der Rückwechsel mit* Q *erhält dieses Volumen. Die Abbildung mit Matrix* A *verändert das Volumen hingegen um den Faktor* $\det(A) = \det(D) = 4$*.* △

Quadratische Formen

Eine wichtige Anwendung des Spektralsatzes ist die Untersuchung quadratischer Ausdrücke: Durch einen orthogonalen Basiswechsel können wir sie als Summen von Quadraten schreiben, deren Koeffizienten die Eigenwerte sind. So erkennen wir insbesondere, ob ein solcher Ausdruck für alle von null verschiedenen Vektoren positiv ist. Diese Frage tritt häufig bei Optimierungsproblemen auf, etwa beim Auswerten ungenauer Messungen oder beim Trainieren von KI-Modellen: Dort wollen wir beurteilen, ob gefundene Parameterwerte den Fehler tatsächlich minimieren.

Beispiel 6.78. *Zwei aneinander anschließende Strecken werden einzeln mit jeweils* 1 *und zusammen mit* 3 *Längeneinheiten gemessen. Wir behandeln die drei Messungen als gleich zuverlässig und suchen Längen* $a, b \in \mathbb{R}$*, für die die Fehlerquadratsumme*

$$F(a,b) \coloneqq (a-1)^2 + (b-1)^2 + (a+b-3)^2$$

möglichst klein ist. Die drei Abweichungen können nicht zugleich null sein, da die Einzelmessungen nicht zur Gesamtmessung passen.

Ein naheliegender Kandidat entsteht, wenn wir die beiden Einzelmessungen um jeweils $1/3$ *erhöhen und die Gesamtmessung um* $1/3$ *verringern: Dann stimmen die korrigierten Werte mit* $a = b = 4/3$ *überein. Ob dies tatsächlich die beste Wahl ist, prüfen wir durch Einsetzen von* $a = 4/3 + x$ *und* $b = 4/3 + y$*. Es gilt*

$$\begin{aligned} F\left(\frac{4}{3}+x, \frac{4}{3}+y\right) &= \left(\frac{1}{3}+x\right)^2 + \left(\frac{1}{3}+y\right)^2 + \left(-\frac{1}{3}+x+y\right)^2 \\ &= \frac{1}{3} + 2x^2 + 2xy + 2y^2. \end{aligned}$$

Der Kandidat liefert also den Wert $F(4/3, 4/3) = 1/3$*. Ist* $2x^2 + 2xy + 2y^2 > 0$ *für alle* $(x,y) \neq (0,0)$*, so liefert jede andere Wahl von* a, b *einen größeren Fehler.*

Hier hilft uns die Koordinatenwahl aus Beispiel 6.76: Für $u = (x+y)/\sqrt{2}$ und $v = (x-y)/\sqrt{2}$ gilt umgekehrt $x = (u+v)/\sqrt{2}$ und $y = (u-v)/\sqrt{2}$. Einsetzen liefert

$$\begin{aligned} 2x^2 + 2xy + 2y^2 &= (u+v)^2 + (u+v)(u-v) + (u-v)^2 \\ &= 3u^2 + v^2. \end{aligned}$$

Der gemischte Term ist verschwunden. Weil der Koordinatenwechsel invertierbar ist, gilt $(x,y) \neq (0,0)$ genau dann, wenn $(u,v) \neq (0,0)$ gilt. Dann ist mindestens eines der beiden Quadrate positiv, also $3u^2 + v^2 > 0$. Damit ist $a = b = 4/3$ die eindeutig bestimmte Wahl mit kleinster Fehlerquadratsumme. △

Den quadratischen Ausdruck aus dem Beispiel können wir mithilfe der Matrix A aus Beispiel 6.76 schreiben als

$$2x^2 + 2xy + 2y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Die Faktoren 3 und 1 vor den Quadraten in den neuen Koordinaten, also im Ausdruck $3u^2 + v^2$, sind gerade ihre Eigenwerte. Wir untersuchen nun allgemein, wie das Vorzeichen solcher Ausdrücke mit den Eigenwerten zusammenhängt.

Definition 6.79. Die zu einer symmetrischen Matrix $A \in \mathbb{R}^{n\times n}$ gehörige quadratische Form mit Variablen $\boldsymbol{x} \in \mathbb{R}^n$ ist

$$\boldsymbol{x}^\top A \boldsymbol{x}.$$

Beim Ausmultiplizieren treten für $i < j$ die beiden Terme $A_{ij}x_ix_j$ und $A_{ji}x_jx_i$ auf. Wegen der Symmetrie ergeben sie zusammen $2A_{ij}x_ix_j$. Daher gilt

$$\boldsymbol{x}^\top A \boldsymbol{x} = \sum_{i=1}^{n} A_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} A_{ij}x_ix_j.$$

Umgekehrt können wir jeden Ausdruck aus Vielfachen von x_i^2 und x_ix_j mit $i < j$ als quadratische Form schreiben: Die Koeffizienten der Quadrate kommen auf die Diagonale; den Koeffizienten von x_ix_j tragen wir jeweils zur Hälfte an den Positionen (i,j) und (j,i) ein. So gehört zum Ausdruck $3x^2 + 4xy + 5y^2$ die symmetrische Matrix $\left(\begin{smallmatrix} 3 & 2 \\ 2 & 5 \end{smallmatrix}\right)$.

Wie in Beispiel 6.78 wollen wir wissen, ob eine quadratische Form für jede von null verschiedene Änderung positiv ist. Entsprechend können wir nach negativen oder nach nichtnegativen beziehungsweise nichtpositiven Werten fragen. Da $\mathbf{0}^\top A \mathbf{0} = 0$ gilt, nehmen wir bei einer strikten Ungleichung den Nullvektor aus.

Definition 6.80. Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ heißt

(i) *positiv definit* (bzw. *negativ definit*), wenn $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ (bzw. $\boldsymbol{x}^\top A \boldsymbol{x} < 0$) für alle $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ gilt;

(ii) *positiv semidefinit* (bzw. *negativ semidefinit*), wenn $\boldsymbol{x}^\top A \boldsymbol{x} \geq 0$ (bzw. $\boldsymbol{x}^\top A \boldsymbol{x} \leq 0$) für alle $\boldsymbol{x} \in \mathbb{R}^n$ gilt;

(iii) *indefinit*, wenn sie weder positiv semidefinit noch negativ semidefinit ist.

Jede positiv (bzw. negativ) definite Matrix ist insbesondere positiv (bzw. negativ) semidefinit.

Beispiel 6.81. *Für die symmetrischen Matrizen*

$$A_1 \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A_2 \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad A_3 \coloneqq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

und $\boldsymbol{x} = (x, y)^\top$ *erhalten wir* $\boldsymbol{x}^\top A_1 \boldsymbol{x} = x^2 + y^2$, $\boldsymbol{x}^\top A_2 \boldsymbol{x} = x^2$ *und* $\boldsymbol{x}^\top A_3 \boldsymbol{x} = x^2 - y^2$. *Somit ist* A_1 *positiv definit. Die Matrix* A_2 *ist positiv semidefinit, aber nicht positiv definit, denn für* $\boldsymbol{x} = (0, 1)^\top \neq \mathbf{0}$ *gilt* $\boldsymbol{x}^\top A_2 \boldsymbol{x} = 0$. *Die Matrix* A_3 *ist indefinit, da* $(1, 0)^\top$ *den Wert* 1 *und* $(0, 1)^\top$ *den Wert* -1 *liefert.* △

Bei Ausdrücken mit gemischten Termen ist das Vorzeichen weniger leicht abzulesen. Wie in Beispiel 6.78 liefert die orthogonale Diagonalisierung eine Summe einzelner Quadrate. Deren Koeffizienten sind die Eigenwerte.

Satz 6.82. Eine symmetrische Matrix ist

(i) genau dann positiv definit (bzw. negativ definit), wenn alle ihre Eigenwerte positiv (bzw. negativ) sind;

(ii) genau dann positiv semidefinit (bzw. negativ semidefinit), wenn alle ihre Eigenwerte nichtnegativ (bzw. nichtpositiv) sind;

(iii) genau dann indefinit, wenn sie sowohl einen positiven als auch einen negativen Eigenwert besitzt.

Beweis. Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch. Nach Satz 6.75 gibt es eine orthogonale Matrix Q mit $Q^\top A Q = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Für $\boldsymbol{x} \in \mathbb{R}^n$ und $\boldsymbol{y} \coloneqq Q^{-1}\boldsymbol{x}$, also $\boldsymbol{x} = Q\boldsymbol{y}$, erhalten wir

$$\boldsymbol{x}^\top A \boldsymbol{x} = (Q\boldsymbol{y})^\top A (Q\boldsymbol{y}) \overset{5.58}{=} \boldsymbol{y}^\top Q^\top A Q \boldsymbol{y} = \boldsymbol{y}^\top D \boldsymbol{y} = \sum_{j=1}^{n} \lambda_j y_j^2. \tag{6.4}$$

Da die lineare Abbildung f_Q invertierbar ist, ist $f_Q(\mathbb{R}^n) = \mathbb{R}^n$, also $\{\boldsymbol{x}^\top A \boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^n\} = \{\sum_{j=1}^n \lambda_j y_j^2 \mid \boldsymbol{y} \in \mathbb{R}^n\}$ und $Q\boldsymbol{y} = \mathbf{0}$ gilt genau für $\boldsymbol{y} = \mathbf{0}$. Da alle y_j^2 nichtnegativ sind und für $\boldsymbol{y} \neq \mathbf{0}$ mindestens eines positiv ist, sind die Bedingungen an die Eigenwerte in (i) und (ii) hinreichend. Sie sind auch notwendig, denn für $\boldsymbol{y} = \boldsymbol{e}^{(j)}$ hat die Summe den Wert λ_j. Schließlich bedeutet „indefinit“ nach Definition weder positiv noch negativ semidefinit; nach (ii) ist dies genau dann der Fall, wenn ein positiver und ein negativer Eigenwert vorkommen. □

Die positive Definitheit der Matrix aus Beispiel 6.78 lässt sich somit an den beiden positiven Eigenwerten 3 und 1 ablesen. Damit genügt die Untersuchung endlich vieler Eigenwerte, um das Vorzeichenverhalten für alle Vektoren zu kennen.

Beispiel 6.83. *Beim Trainieren eines KI-Modells verändern wir dessen Parameter, um den Fehler seiner Vorhersagen zu verkleinern. Als Fehlermaß können wir wie in Beispiel 6.78 eine Summe quadrierter Abweichungen zwischen vorhergesagten und vorgegebenen Werten verwenden. Diese Summe ist nichtnegativ; ihre Änderung kann jedoch positiv oder negativ sein. Haben wir bereits eine optimale Einstellung gefunden, wenn das einzelne Verändern jedes Parameters den Fehler erhöht? Zur Veranschaulichung betrachten wir ein vorgegebenes, vereinfachtes Modell: Kleine Änderungen dreier Parameter gegenüber der aktuellen Einstellung fassen wir in* $\boldsymbol{x} = (x_1, x_2, x_3)^\top$ *zusammen. Die Änderung dieser Fehlerquadratsumme modellieren wir für kleine Parameteränderungen durch*

$$\Delta F = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3 = \boldsymbol{x}^\top B \boldsymbol{x}, \qquad B \coloneqq \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

Ein negativer Wert von ΔF bedeutet also, dass der Fehler kleiner wird. Verändern wir nur einen der drei Parameter um $t \neq 0$, so ist dagegen jeweils $\Delta F = t^2 > 0$. Jede dieser einzelnen Änderungen verschlechtert somit die Vorhersagen in unserem Modell.

Um auch gemeinsame Änderungen zu beurteilen, verwenden wir die Matrizen

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$$

aus Beispiel 6.77. Wegen $B = 3\mathbb{I}_3 - A$ gilt

$$Q^\top B Q = 3Q^\top Q - Q^\top A Q = 3\mathbb{I}_3 - \operatorname{diag}(4,1,1) = \operatorname{diag}(-1,2,2).$$

Die Spalten von Q bilden daher auch eine Eigenbasis von B, nun zu den Eigenwerten $-1, 2, 2$. Nach Satz 6.82 ist B indefinit: Es gibt sowohl Änderungen, die den Fehler erhöhen, als auch solche, die ihn verringern.

Die Eigenbasis zeigt uns auch, welche Änderungen helfen. Mit $\boldsymbol{y} = Q^\top \boldsymbol{x}$ wird die Fehleränderung nach (6.4) zu

$$\Delta F = -y_1^2 + 2y_2^2 + 2y_3^2.$$

Der negative Eigenwert gehört zur ersten Spalte $\boldsymbol{q}^{(1)} = (1,1,1)^\top/\sqrt{3}$ von Q. Verändern wir alle drei Parameter um dasselbe kleine $t \neq 0$, so gilt $\boldsymbol{x} = t(1,1,1)^\top = \sqrt{3}t\boldsymbol{q}^{(1)}$ und damit $\boldsymbol{y} = (\sqrt{3}t, 0, 0)^\top$. Wir erhalten $\Delta F = -3t^2 < 0$: Die gemeinsame Änderung senkt den Fehler, obwohl jede einzelne Änderung ihn erhöht.

Wäre B positiv definit, so würde jede von null verschiedene Änderung den Fehler in diesem Modell erhöhen; die aktuelle Einstellung wäre damit im Gültigkeitsbereich des Modells eindeutig optimal. Die hier festgestellte Indefinitheit widerlegt dies, und der Eigenvektor zum negativen Eigenwert liefert eine konkrete Verbesserungsrichtung. △

7 Wachstumsverhalten und Grenzwerte

Wir untersuchen nun, wie sich Zahlenwerte entwickeln, wenn eine Größe zunimmt oder sich ein Vorgang immer weiter wiederholt. Dabei interessiert uns, wie schnell die Werte wachsen und ob sie sich einem festen Wert annähern. In der Informatik begegnen uns diese Fragen etwa in der *Algorithmik* (Wachstum des Zeit- und Speicheraufwands mit der Datenmenge), der *Numerik* (Abnahme von Näherungsfehlern), dem *maschinellen Lernen* (Entwicklung des Fehlers beim Training), den *verteilten Systemen* (Ausgleich der Arbeitslast durch wiederholtes Verschieben zwischen Rechnern), der *Computergrafik* (Annäherung an Formen durch feinere Unterteilungen), der *Simulation* (Verhalten bei feineren Zeit- und Raumschritten) und der *Datenanalyse* (Stabilisierung von Schätzwerten bei mehr Beobachtungen).

Damit wenden wir uns der *Analysis* zu, dem Teilgebiet der Mathematik, das Veränderungen und Grenzwerte untersucht. Bisher haben wir vor allem mit Gleichungen und algebraischen Strukturen gearbeitet, insbesondere mit Vektorräumen, linearen Abbildungen und linearen Gleichungssystemen als Gegenständen der *Linearen Algebra*. In der Analysis rücken dagegen Fragen der Nähe und Annäherung in den Mittelpunkt: Wie verändern sich Werte, wie klein können ihre Abstände werden und welchem Wert nähern sie sich? Dabei arbeiten wir mit reellen Zahlen.

7.1 Reelle Folgen

Wie können wir eine Quadratwurzel $\sqrt{q}$ einer positiven rationalen Zahl $q \in \mathbb{Q}$, $q > 0$ mit den Grundrechenarten berechnen? Wir wissen bereits, dass die Quadratwurzel über den reellen Zahlen existiert (Satz 3.87), im Allgemeinen aber keine rationale Zahl sein muss (Satz 1.18). Um bei dem konkreten Rechenverfahren mit den endlich darstellbaren rationalen Zahlen arbeiten zu können, suchen wir eine Vorschrift, die nur rationale Zahlen verwendet und uns Näherungen mit jeder gewünschten Genauigkeit liefert.

Ein Rechteck mit Seitenlängen $x > 0$ und $y = q/x$ hat Fläche $x \cdot y = q$. Wären beide Seiten gleich lang, hätten sie die gesuchte Länge $\sqrt{q}$, ansonsten ist $x > \sqrt{q}$ und $y < \sqrt{q}$ oder umgekehrt. Wir können uns der Lösung nähern, indem wir x auf den Mittelwert $x' = \frac{x+y}{2}$ setzen und y auf $y' = q/x'$ und diesen Schritt wiederholen.

Wenn wir mit dem Wert $x_0 = q$ für x beginnen, erhalten wir so die Werte x_0, x_1, … mit

$$x_{n+1} \coloneqq \frac{1}{2}\left(x_n + \frac{q}{x_n}\right) > 0 \qquad \text{für } n \in \mathbb{N}.$$

Für $q = 2$ ergeben die ersten Schritte

$$x_0 = 2, \qquad x_1 = \frac{3}{2} = 1{,}5, \qquad x_2 = \frac{17}{12} \approx 1{,}416667, \qquad x_3 = \frac{577}{408} \approx 1{,}414216.$$

Dass die Werte nahe bei $\sqrt{2}$ liegen, beweist noch nicht, dass die Vorschrift zuverlässig funktioniert. Können wir durch weitere Schritte jeden vorgegebenen Fehler unterschreiten? Um diese Frage zu beantworten, beschreiben wir zuerst die entstehenden Werte als Folge.

Definition 7.1. Eine *(reelle) Folge* ist eine Abbildung $a\colon \mathbb{N} \to \mathbb{R}$. Ihren Wert schreiben wir als $a_n \coloneqq a(n)$ und nennen ihn das *n-te Folgenglied*. Die ganze Folge bezeichnen wir mit $(a_n)_{n\in\mathbb{N}}$.

Da $0 \in \mathbb{N}$ gilt, beginnt die Indizierung bei a_0. Eine Folge ist von ihrer Wertemenge zu unterscheiden: Reihenfolge und Wiederholungen der Glieder gehören zur Folge.

Beispiel 7.2. *Eine Folge a wird durch eine Abbildungsvorschrift $n \mapsto a_n$ beschrieben. Wir können also zum Beispiel die Folgen $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ angeben als*

$$\begin{aligned} a_n &\coloneqq \frac{1}{n+1} && (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots), \\ b_n &\coloneqq (-1)^n && (1, -1, 1, -1, \dots), \\ c_n &\coloneqq 2^{-n} && (1, 1/2, 1/4, 1/8, \dots), \\ d_n &\coloneqq \frac{n}{n+1} && (0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \dots). \end{aligned}$$

△

Beispiel 7.3. *Wir können Folgen auch durch eine rekursive Vorschrift angeben. Die Folge $(c_n)_{n\in\mathbb{N}}$ aus Beispiel 7.2 können wir alternativ (Beweis durch Induktion) schreiben als*

$$c_0 \coloneqq 1, \qquad c_{n+1} \coloneqq \frac{c_n}{2} \quad \text{für } n \in \mathbb{N}.$$

Die Folge $(f_n)_{n\in\mathbb{N}}$ der Fibonacci-Zahlen aus Beispiel 6.69 hatten wir dort zunächst rekursiv beschrieben und dann eine explizite Formel für f_n abgeleitet. △

Die folgenden Begriffe knüpfen an die bereits eingeführte Ordnung der reellen Zahlen an.

Definition 7.4. Eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ heißt

(i) *nach oben (bzw. unten) beschränkt*, wenn $\exists s \in \mathbb{R}\colon \forall n \in \mathbb{N}\colon a_n \leq s$ (bzw. $\exists m \in \mathbb{R}\colon \forall n \in \mathbb{N}\colon m \leq a_n$);

(ii) *beschränkt*, wenn sie sowohl nach oben als auch nach unten beschränkt ist, also $\exists c \geq 0\colon \forall n \in \mathbb{N}\colon |a_n| \leq c$;

(iii) *monoton wachsend*, wenn $\forall n \in \mathbb{N}\colon a_n \leq a_{n+1}$, und *monoton fallend*, wenn $\forall n \in \mathbb{N}\colon a_n \geq a_{n+1}$;

(iv) *streng monoton wachsend/fallend*, wenn sie monoton wachsend/fallend ist und $\forall n \in \mathbb{N}\colon a_n \neq a_{n+1}$.

Beispiel 7.5. *Die Folge $(a_n)_{n\in\mathbb{N}}$ aus Beispiel 7.2 ist streng monoton fallend und durch $0 < a_n \leq 1$ beschränkt. Die Folge $(b_n)_{n\in\mathbb{N}}$ ist wegen $|b_n| = 1$ beschränkt, aber weder monoton wachsend noch monoton fallend. Die Folge $(c_n)_{n\in\mathbb{N}}$ ist wegen $0 < c_n \leq 1$ beschränkt und wegen $c_{n+1} = c_n/2 < c_n$ streng monoton fallend. Für $(d_n)_{n\in\mathbb{N}}$ gilt*

$$0 \leq d_n < 1, \qquad d_{n+1} - d_n = \frac{n+1}{n+2} - \frac{n}{n+1} = \frac{1}{(n+1)(n+2)} > 0.$$

Sie ist also beschränkt und streng monoton wachsend. △

Für die Frage nach beliebig kleinen Fehlern benötigen wir zunächst eine Folgerung aus der Vollständigkeit der reellen Zahlen: Natürliche Zahlen können jede reelle Schranke überschreiten.

Satz 7.6. Die Menge $\mathbb{N}$ ist in $\mathbb{R}$ nicht nach oben beschränkt. Insbesondere gibt es zu jedem $\varepsilon > 0$ einen Index $n_0 \in \mathbb{N}$ mit $1/(n_0 + 1) < \varepsilon$.

Beweis. Wäre $\mathbb{N}$ nach oben beschränkt, so besäße sie nach Definition 3.80 ein Supremum $s \in \mathbb{R}$. Da $s - 1 < s$ keine obere Schranke sein kann, gäbe es ein $n \in \mathbb{N}$ mit $s - 1 < n$. Dann wäre aber auch $n + 1 \in \mathbb{N}$ und $s < n + 1$, im Widerspruch dazu, dass s eine obere Schranke ist. Zu $\varepsilon > 0$ können wir somit $n_0 > 1/\varepsilon$ wählen; daraus folgt $1/(n_0 + 1) < \varepsilon$. □

Wir können nun die Forderung an unsere Näherungen genau ausdrücken. Eine vorgegebene Fehlertoleranz $\varepsilon > 0$ bestimmt ein offenes Intervall $(a - \varepsilon, a + \varepsilon)$ um den angestrebten Wert a. Ab einem passenden Index sollen alle Folgenglieder darin liegen.

Definition 7.7. Eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ *konvergiert gegen* $a \in \mathbb{R}$, wenn

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

Dann heißt a *Grenzwert* von $(a_n)_{n\in\mathbb{N}}$ und die Folge heißt *konvergent*. Ist $a = 0$, heißt sie *Nullfolge*.
Eine Folge *divergiert* und heißt *divergent*, wenn sie nicht konvergent ist.

In Definition 7.7 darf der Index n_0 von ε abhängen. Je kleiner die erlaubte Abweichung, desto weiter müssen wir möglicherweise in der Folge gehen. Es genügt nicht, dass nur einzelne, beliebig spät auftretende Glieder nahe bei a liegen: Alle Glieder ab n_0 müssen die Abschätzung erfüllen. Ein Grenzwert muss selbst kein Folgenglied sein.

Beispiel 7.8. *Jede konstante Folge $a_n = c \in \mathbb{R}$ konvergiert gegen c, denn $|a_n - c| = 0 < \varepsilon$ gilt für jedes $\varepsilon > 0$ und jeden Index. Die Folge $a_n = (-1)^n/(n + 1)$ ist eine Nullfolge: Zu $\varepsilon > 0$ wählen wir nach Satz 7.6 einen Index n_0 mit $1/(n_0 + 1) < \varepsilon$. Dann gilt für alle $n \geq n_0$*

$$|a_n - 0| = \frac{1}{n+1} \leq \frac{1}{n_0+1} < \varepsilon.$$

Konvergenz verlangt also keine Monotonie. △

Wir halten Eigenschaften konvergenter Folgen fest.

Satz 7.9. Für reelle Folgen gilt:

(i) Jede konvergente Folge hat genau einen Grenzwert.

(ii) Jede konvergente Folge ist beschränkt.

(iii) Gibt es zu zwei Folgen $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ einen Index $n_0 \in \mathbb{N}$ mit $a_n = b_n$ für alle $n \geq n_0$, so konvergiert die eine genau dann gegen $c \in \mathbb{R}$, wenn die andere gegen c konvergiert.

Beweis. Zu (i): Per Definition hat jede konvergente Folge mindestens einen Grenzwert. Angenommen, $(a_n)_{n\in\mathbb{N}}$ konvergiert gegen zwei verschiedene Zahlen $a \in \mathbb{R}$ und $b \in \mathbb{R}$. Für $\varepsilon := |a-b|/3 > 0$ gibt es dann per Definition Indizes n_a und n_b, ab denen $|a_n - a| < \varepsilon$ beziehungsweise $|a_n - b| < \varepsilon$ gilt. Für $n \geq \max\{n_a, n_b\}$ folgt mit der Dreiecksungleichung (Satz 3.93)

$$|a-b| \leq |a-a_n| + |a_n - b| < 2\varepsilon = \frac{2}{3}|a-b|,$$

ein Widerspruch zu $a \neq b$.

Zu (ii): Sei $(a_n)_{n\in\mathbb{N}}$ eine konvergente Folge mit Grenzwert a und sei $\varepsilon := 1$. Per Definition des Grenzwerts gibt es ein n_0, sodass $|a_n - a| < 1$ für alle $n \geq n_0$ gilt. Damit ist mit der Dreiecksungleichung $|a_n| = |a_n - 0| \leq |a_n - a| + |a - 0| < 1 + |a|$ für alle $n \geq n_0$. Es folgt, dass $(a_n)_{n\in\mathbb{N}}$ beschränkt ist durch $|a_n| \leq c$ für alle $n \in \mathbb{N}$, wobei

$$c := \max\big(\{1 + |a|\} \cup \{|a_n| \mid n \in \{0, \dots, n_0 - 1\}\}\big).$$

Zu (iii): Sei $a_n = b_n$ für alle $n \geq n_0$, und die Folge $(a_n)_{n\in\mathbb{N}}$ konvergiere gegen c. Zu $\varepsilon > 0$ wählen wir n_1, sodass $|a_n - c| < \varepsilon$ für alle $n \geq n_1$ gilt. Dann ist für $n \geq \max\{n_0, n_1\}$ auch $|b_n - c| = |a_n - c| < \varepsilon$. Somit konvergiert auch $(b_n)_{n\in\mathbb{N}}$ gegen c. Die umgekehrte Richtung folgt durch Vertauschen der Folgen. □

Beispiel 7.10. *Beschränktheit allein reicht nicht für Konvergenz: Die Folge $a_n = (-1)^n$ ist beschränkt, aber divergent. Gäbe es einen Grenzwert a, so gäbe es zu $\varepsilon = 1/2$ einen Index $n_0 \in \mathbb{N}$ mit $|a_n - a| < 1/2$ für alle $n \geq n_0$. Für $n \geq n_0$ ergäbe sich dann der Widerspruch*

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

Nach Satz 7.9 (i) und (iii) ist der Grenzwert eindeutig und unabhängig von endlich vielen Anfangsgliedern. Wir können die Grenzwertnotation daher auch für Ausdrücke verwenden, die erst ab einem bestimmten Index definiert sind.

Definition 7.11. Seien $n_0 \in \mathbb{N}$ und $f\colon \{n \in \mathbb{N} \mid n \geq n_0\} \to \mathbb{R}$. Existiert eine gegen $a \in \mathbb{R}$ konvergente reelle Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_n = f(n)$ für alle $n \geq n_0$, so definieren wir

$$\lim_{n\to\infty} f(n) := a.$$

Beispielsweise gilt

$$\lim_{n\to\infty} \frac{1}{n} = 0.$$

Zu $\varepsilon > 0$ wählen wir nach Satz 7.6 einen Index $n_0 > 1/\varepsilon$. Für alle $n \geq n_0$ gilt dann $|1/n - 0| = 1/n \leq 1/n_0 < \varepsilon$. Einen Wert bei $n = 0$ müssen wir für diese Grenzwertschreibweise nicht festlegen.

Wir wollen nun begründen, dass das Näherungsverfahren aus unserem Einstieg tatsächlich konvergiert. Dafür genügt eine Kombination der bereits bekannten Eigenschaften Beschränktheit und Monotonie.

Satz 7.12 (Monotoniekriterium)**.** Jede monoton wachsende (bzw. fallende), nach oben (bzw. unten) beschränkte reelle Folge $(a_n)_{n\in\mathbb{N}}$ konvergiert gegen $\sup\{a_n \mid n \in \mathbb{N}\}$ (bzw. $\inf\{a_n \mid n \in \mathbb{N}\}$).

Beweis. Wir zeigen die Aussage für den Fall, dass $(a_n)_{n\in\mathbb{N}}$ monoton wachsend und nach oben beschränkt ist, der andere Fall ist analog. Nach dem Vollständigkeitsaxiom aus Definition 3.80 existiert $a \coloneqq \sup\{a_n \mid n \in \mathbb{N}\}$. Zu $\varepsilon > 0$ ist $a - \varepsilon$ keine obere Schranke der Wertemenge. Also gibt es einen Index n_0 mit $a - \varepsilon < a_{n_0}$. Für alle $n \geq n_0$ gilt wegen der Monotonie

$$a - \varepsilon < a_{n_0} \leq a_n \leq a.$$

Damit ist $|a_n - a| < \varepsilon$ für alle $n \geq n_0$, also $\lim_{n\to\infty} a_n = a$. □

Beispiel 7.13. *Die Folge $(x_n)_{n\in\mathbb{N}}$ für die Näherung der Quadratwurzel von $q > 0$ war*

$$x_0 = q, \qquad x_{n+1} = \frac{1}{2}\left(x_n + \frac{q}{x_n}\right).$$

Für jedes $n \in \mathbb{N}$ gilt

$$x_{n+1}^2 - q = \frac{1}{4}\left(x_n + \frac{q}{x_n}\right)^2 - q = \frac{(x_n^2 - q)^2}{4x_n^2} \geq 0.$$

Wegen $x_n > 0$ folgt $x_n \geq \sqrt{q}$ für alle $n \geq 1$. Für diese Indizes ist außerdem

$$x_n - x_{n+1} = \frac{x_n^2 - q}{2x_n} \geq 0.$$

Ersetzen wir das erste Glied x_0 durch x_1, ist die Folge monoton fallend und nach unten beschränkt. Nach Satz 7.12 konvergiert sie gegen eine Zahl $a \geq \sqrt{q} > 0$. Nach Satz 7.9 (iii) konvergiert auch die ursprüngliche Folge $(x_n)_{n\in\mathbb{N}}$ gegen a. △

Wir wissen nun, dass die im Einstieg definierte Folge konvergiert. Wie bestimmen wir ihren Grenzwert aus der Rekursion? Die folgenden Regeln erlauben uns, Grenzwerte aus bereits bekannten Grenzwerten zu berechnen.

Satz 7.14. Seien $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ reelle Folgen mit $\lim_{n\to\infty} a_n = a$ und $\lim_{n\to\infty} b_n = b$, und sei $\lambda \in \mathbb{R}$. Dann gilt

$$\lim_{n\to\infty}(a_n + b_n) = a + b, \qquad \lim_{n\to\infty}(\lambda a_n) = \lambda a,$$
$$\lim_{n\to\infty}(a_n b_n) = ab, \qquad \lim_{n\to\infty}|a_n| = |a|.$$

Ist zusätzlich $b \neq 0$ und $b_n \neq 0$ für alle $n \in \mathbb{N}$, so gilt

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

Beweis. Für die Summenregel wählen wir zu $\varepsilon > 0$ einen gemeinsamen Index n_0, ab dem $|a_n - a| < \varepsilon/2$ und $|b_n - b| < \varepsilon/2$ gelten. Dann ist

$$|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \varepsilon.$$

Für die Multiplikation mit $\lambda \neq 0$ genügt entsprechend die Abschätzung $|a_n - a| < \varepsilon/|\lambda|$; für $\lambda = 0$ ist die Aussage unmittelbar klar.

Für die Produktregel verwenden wir, dass $(a_n)_{n\in\mathbb{N}}$ nach Satz 7.9 beschränkt ist. Wir wählen $s \geq 1$ mit $|a_n| \leq s$ für alle n. Zu $\varepsilon > 0$ wählen wir einen gemeinsamen Index n_0, ab dem

$$|b_n - b| < \frac{\varepsilon}{2s} \qquad \text{und} \qquad |a_n - a| < \frac{\varepsilon}{2(|b|+1)}$$

gelten. Für $n \geq n_0$ folgt

$$|a_n b_n - ab| = |a_n(b_n - b) + b(a_n - a)| \leq s|b_n - b| + |b|\,|a_n - a| < \varepsilon.$$

Aus der Dreiecksungleichung folgen $|a_n| \leq |a_n - a| + |a|$ und $|a| \leq |a_n - a| + |a_n|$. Zusammen ergibt dies

$$\big||a_n| - |a|\big| \leq |a_n - a|.$$

Jede Abschätzung $|a_n - a| < \varepsilon$ liefert daher dieselbe Fehlerschranke für die Beträge.

Sei nun $b \neq 0$. Es gibt einen Index $n_0 \in \mathbb{N}$, sodass für alle $n \geq n_0$ die Abschätzung $|b_n - b| < |b|/2$ gilt und damit

$$|b_n| \geq |b| - |b_n - b| > \frac{|b|}{2} > 0.$$

Für diese Indizes erhalten wir

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

Zu jedem $\varepsilon > 0$ gibt es wegen $\lim_{n\to\infty} b_n = b$ einen Index $n_1 \in \mathbb{N}$ mit $|b_n - b| < \varepsilon|b|^2/2$ für alle $n \geq n_1$. Für $n \geq \max\{n_0, n_1\}$ ist der letzte Ausdruck daher kleiner als ε. Also konvergieren die Kehrwerte gegen $1/b$. Die Quotientenregel folgt nun aus der Produktregel. □

Beispiel 7.15. *Wir verwenden den Grenzwert $\lim_{n\to\infty} \frac{1}{n} = 0$. Die beiden Faktoren im folgenden Produkt konvergieren nach der Summen- und Faktorregel gegen 2 beziehungsweise 3. Mit der Produktregel folgt*

$$\lim_{n\to\infty} \left(2 + \frac{1}{n}\right)\left(3 - \frac{1}{n}\right) = 2 \cdot 3 = 6. \qquad \triangle$$

Beispiel 7.16. *Die Betragsregel lässt sich nicht umkehren: Für $a_n = (-1)^n$ gilt $\lim_{n\to\infty} |a_n| = 1$, aber die Folge $(a_n)_{n\in\mathbb{N}}$ ist divergent (Beispiel 7.10).* $\triangle$

Beispiel 7.17. *Welchen Grenzwert hat die Folge mit den Gliedern $\frac{3n^2-n+2}{2n^2+1}$? Zähler und Nenner wachsen beide unbeschränkt, sodass wir die Quotientenregel nicht unmittelbar anwenden können. Für $n \geq 1$ teilen wir Zähler und Nenner durch n^2 und erhalten*

$$\frac{3n^2 - n + 2}{2n^2 + 1} = \frac{3 - 1/n + 2/n^2}{2 + 1/n^2}.$$

Aus $\lim_{n\to\infty} 1/n = 0$ folgt mit der Produktregel $\lim_{n\to\infty} 1/n^2 = 0$. Der Nenner $2 + 1/n^2$ ist für alle $n \geq 1$ positiv und hat den Grenzwert $2 \neq 0$. Mit den Grenzwertregeln folgt daher

$$\lim_{n\to\infty} \frac{3n^2 - n + 2}{2n^2 + 1} = \lim_{n\to\infty} \frac{3 - 1/n + 2/n^2}{2 + 1/n^2} = \frac{3}{2}. \qquad \triangle$$

Beispiel 7.18. *Die Bedingung $b \neq 0$ in der Quotientenregel ist wesentlich. Obwohl $\lim_{n\to\infty} 1/n = \lim_{n\to\infty} 1/n^2 = 0$ gilt, können Quotienten dieser Ausdrücke unterschiedliches Grenzverhalten haben:*

$$\lim_{n\to\infty} \frac{1/n}{1/n} = 1, \qquad \lim_{n\to\infty} \frac{1/n^2}{1/n} = \lim_{n\to\infty} \frac{1}{n} = 0.$$

Dagegen ist $\frac{1/n}{1/n^2} = n$ für $n \geq 1$ unbeschränkt und besitzt deshalb nach Satz 7.9 (ii) keinen reellen Grenzwert. Aus den Grenzwerten von Zähler und Nenner allein können wir in diesem Fall also nicht auf das Grenzverhalten des Quotienten schließen. △

Mit den Rechenregeln können wir nun auch feststellen, dass unser Näherungsverfahren für die Quadratwurzel funktioniert.

Beispiel 7.19. *Wir kehren zur Folge aus Beispiel 7.13 zurück. Dort haben wir ihre Konvergenz gegen eine Zahl $a \geq \sqrt{q} > 0$ bewiesen. Nun bestimmen wir den Grenzwert mithilfe der Rekursion. Aus $\lim_{n\to\infty} x_n = a$ folgt auch $\lim_{n\to\infty} x_{n+1} = a$, also liefern die Grenzwertregeln aus Satz 7.14 mit $a > 0$, dass*

$$a = \frac{1}{2}\left(a + \frac{q}{a}\right), \qquad \text{also} \qquad a^2 = q.$$

Wegen $a > 0$ ist nach der Eindeutigkeit der positiven Quadratwurzel $a = \sqrt{q}$ (Satz 3.87). Das Näherungsverfahren erreicht somit jede vorgegebene Genauigkeit. △

Der folgende Satz ist nützlich, da er es erlaubt, Grenzwerte schwieriger zu behandelnder Folgen auf Grenzwerte einfacherer Folgen zurückzuführen.

Satz 7.20. Für reelle Folgen $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ und $(c_n)_{n\in\mathbb{N}}$ mit $\lim_{n\to\infty} a_n = a$ und $\lim_{n\to\infty} b_n = b$ sowie einen Index $n_0 \in \mathbb{N}$ gilt:

(i) Ist $a_n \leq b_n$ für alle $n \geq n_0$, so folgt $a \leq b$.

(ii) Ist $a_n \geq b_n$ für alle $n \geq n_0$, so folgt $a \geq b$.

(iii) Ist $a_n \leq c_n \leq b_n$ für alle $n \geq n_0$ und ist $a = b$, so ist $\lim_{n\to\infty} c_n = a = b$.

Beweis. Zu (i): Wäre $a > b$, so könnten wir $\varepsilon := (a-b)/3 > 0$ wählen. Wegen der beiden Konvergenzannahmen gibt es einen Index $n_1 \in \mathbb{N}$, ab dem $|a_n - a| < \varepsilon$ und $|b_n - b| < \varepsilon$ gelten. Für alle $n \geq \max\{n_0, n_1\}$ wären dann gleichzeitig

$$a_n > a - \varepsilon > b + \varepsilon > b_n \qquad \text{und} \qquad a_n \leq b_n$$

erfüllt, ein Widerspruch.

Zu (ii): Dies folgt aus (i) durch Vertauschen der beiden Folgen.

Zu (iii): Wegen $a = b$ konvergieren beide äußeren Folgen gegen a. Zu $\varepsilon > 0$ wählen wir einen gemeinsamen Index, ab dem die vorausgesetzten Ungleichungen sowie $|a_n - a| < \varepsilon$ und $|b_n - a| < \varepsilon$ gelten. Dann ist

$$a - \varepsilon < a_n \leq c_n \leq b_n < a + \varepsilon,$$

also $|c_n - a| < \varepsilon$. □

Bemerkung 7.21. Strikte Ungleichungen können beim Grenzübergang verloren gehen: Obwohl $1/(n+1) > 0$ für alle n gilt, ist der Grenzwert gleich null.

Beispiel 7.22. *Für $n \geq 1$ gilt*

$$0 \leq \frac{n}{n^2+1} \leq \frac{1}{n}.$$

Wegen $\lim_{n\to\infty} 0 = 0 = \lim_{n\to\infty} 1/n$ folgt nach Satz 7.20 (iii) auch $\lim_{n\to\infty} \frac{n}{n^2+1} = 0$. △

Wir haben in Satz 7.9 (ii) bereits gezeigt, dass jede unbeschränkte Folge divergiert. Besonders interessant sind dabei Folgen, die langfristig wachsen oder fallen.

Definition 7.23. Eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ heißt *bestimmt divergent*

(i) *gegen ∞*, wenn $\forall s > 0 \colon \exists n_0 \in \mathbb{N} \colon \forall n \geq n_0 \colon a_n > s$;

(ii) *gegen $-\infty$*, wenn $\forall s > 0 \colon \exists n_0 \in \mathbb{N} \colon \forall n \geq n_0 \colon a_n < -s$.

Für $n_0 \in \mathbb{N}$ und $f \colon \{n \in \mathbb{N} \mid n \geq n_0\} \to \mathbb{R}$ definieren wir analog zu Definition 7.11 die Schreibweise $\lim_{n\to\infty} f(n) = \infty$ (bzw. $-\infty$), wenn eine gegen ∞ (bzw. $-\infty$) bestimmt divergente reelle Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_n = f(n)$ für alle $n \geq n_0$ existiert.

Beispiel 7.24. *Es gilt $\lim_{n\to\infty} n = \infty$: Zu $s > 0$ wählen wir nach Satz 7.6 einen natürlichen Index $n_0 > s$; dann gilt $n > s$ für alle $n \geq n_0$. Entsprechend gilt $\lim_{n\to\infty}(-n) = -\infty$. Die Folge $(-1)^n n$ ist dagegen zwar unbeschränkt, aber weder bestimmt divergent gegen ∞ noch gegen $-\infty$. Die beschränkte Folge $(-1)^n$ ist divergent, aber nicht bestimmt divergent.* △

Wir untersuchen jetzt geometrische Folgen, bei denen sich aufeinander folgende Glieder um einen festen Faktor unterscheiden. Sie treten in der Mathematik, Informatik und in der Natur immer wieder auf.

Satz 7.25. Für eine *geometrische Folge* $(q^n)_{n\in\mathbb{N}}$ mit $q \in \mathbb{R}$ gilt:

(i) Für $|q| < 1$ ist $\lim_{n\to\infty} q^n = 0$.

(ii) Für $q = 1$ ist $\lim_{n\to\infty} q^n = 1$.

(iii) Für $q > 1$ ist $\lim_{n\to\infty} q^n = \infty$.

(iv) Für $q \leq -1$ ist die Folge divergent, aber nicht bestimmt divergent.

Beweis. Wir verwenden zunächst die Abschätzung

$$(1+c)^n \geq 1 + nc \qquad \text{für } c > 0 \text{ und } n \in \mathbb{N}. \tag{7.1}$$

Sie folgt durch vollständige Induktion: Für $n = 0$ gilt Gleichheit, und aus der Induktionsannahme für n folgt

$$(1+c)^{n+1} \overset{\text{IV}}{\geq} (1+nc)(1+c) = 1 + (n+1)c + nc^2 \geq 1 + (n+1)c.$$

Zu (i): Für $q = 0$ ist die Aussage unmittelbar klar. Sei $0 < |q| < 1$. Mit $c \coloneqq 1/|q| - 1 > 0$ gilt

$$0 \le |q^n| = \frac{1}{(1+c)^n} \overset{(7.1)}{\le} \frac{1}{1+nc}.$$

Zu $\varepsilon > 0$ wählen wir $n_0 > 1/(c\varepsilon)$. Dann ist für alle $n \ge n_0$ nach (7.1) auch $|q^n| \le 1/(1+nc) < \varepsilon$. Somit gilt $\lim_{n\to\infty} q^n = 0$.

Zu (ii): Dies ist die konstante Folge 1.

Zu (iii): Für $c \coloneqq q - 1 > 0$ gilt nach (7.1) auch $q^n \ge 1 + nc$. Zu jeder Schranke $s > 0$ wählen wir $n_0 > s/c$; dann ist $q^n > s$ für alle $n \ge n_0$.

Zu (iv): Für $q = -1$ haben wir die Divergenz bereits in Beispiel 7.10 gezeigt. Für $q < -1$ wachsen die Beträge $|q|^n$ nach (iii) über jede Schranke; die Folge ist daher unbeschränkt und nicht konvergent. Für gerade Indizes sind die Glieder positiv, für ungerade negativ. In beiden Fällen $q \le -1$ schließen diese beliebig spät auftretenden Vorzeichen eine bestimmte Divergenz in eine der beiden Richtungen aus. □

Grenzwerte helfen uns auch zu verstehen, was geschieht, wenn wir eine lineare Abbildung immer wieder anwenden. Wir verbinden dazu die geometrischen Folgen mit den Matrixpotenzen aus Satz 6.68.

Beispiel 7.26. *Zwei Rechner sollen ihre Arbeitslast ausgleichen. In jeder Runde gibt jeder Rechner ein Drittel seiner aktuellen Last an den anderen ab. Wir modellieren die Last als beliebig teilbare Arbeitsmenge; während des Ausgleichs kommen keine Aufgaben hinzu und es werden keine abgearbeitet. Bezeichnen u_n und v_n die nichtnegativen Lasten nach n Runden, so gilt*

$$\boldsymbol{x}^{(n+1)} = A\boldsymbol{x}^{(n)}, \qquad \boldsymbol{x}^{(n)} \coloneqq \begin{pmatrix} u_n \\ v_n \end{pmatrix}, \qquad A \coloneqq \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Die anfängliche Verteilung $(9,0)^\top$ wird beispielsweise zu $(6,3)^\top$, dann zu $(5,4)^\top$. Wir betrachten die beiden Lastfolgen $(u_n)_{n\in\mathbb{N}}$ und $(v_n)_{n\in\mathbb{N}}$ und fragen uns: Konvergieren sie gegen denselben Wert?

Nach n Runden gilt $\boldsymbol{x}^{(n)} = A^n\boldsymbol{x}^{(0)}$. Die Matrix A ist ein Drittel der Matrix aus Beispiel 6.76. Die Eigenvektoren $(1,1)^\top$ und $(1,-1)^\top$ bleiben also erhalten und haben die Eigenwerte 1 und 1/3. Mit

$$m \coloneqq \frac{u_0 + v_0}{2}, \qquad d \coloneqq \frac{u_0 - v_0}{2}$$

zerlegen wir den Anfangsvektor als $\boldsymbol{x}^{(0)} = m(1,1)^\top + d(1,-1)^\top$. Nach Satz 6.68 erhalten wir

$$\boldsymbol{x}^{(n)} = m\begin{pmatrix}1\\1\end{pmatrix} + d\left(\frac{1}{3}\right)^n\begin{pmatrix}1\\-1\end{pmatrix}, \qquad \text{also} \qquad u_n = m + d\left(\frac{1}{3}\right)^n, \quad v_n = m - d\left(\frac{1}{3}\right)^n.$$

Wegen $\lim_{n\to\infty}(1/3)^n = 0$ folgt mit den Grenzwertregeln

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} v_n = m = \frac{u_0 + v_0}{2}.$$

Beide Lastfolgen konvergieren also gegen die Hälfte der Gesamtlast. Der Unterschied beträgt nach n Runden $|u_n - v_n| = |u_0 - v_0|\,3^{-n}$ und wird damit beliebig klein. Die Eigenrichtung $(1,1)^\top$ beschreibt den erhaltenen gemeinsamen Anteil, die Eigenrichtung $(1,-1)^\top$ den abklingenden Unterschied. Wir haben so aus einer Formel für die einzelnen Runden eine Aussage über das langfristige Verhalten gewonnen. △

Teilfolgen

Die Divergenz einer Folge sagt uns noch wenig über ihr langfristiges Verhalten: Sie kann unbegrenzt wachsen oder fallen, zwischen zwei Werten wechseln oder sich mehreren verschiedenen Werten immer wieder annähern. Teilfolgen helfen, diese Verhaltensweisen genauer zu unterscheiden.

Definition 7.27. Sei $(a_n)_{n\in\mathbb{N}}$ eine reelle Folge. Zu einer streng wachsenden Folge $(n_k)_{k\in\mathbb{N}}$ natürlicher Zahlen heißt $(a_{n_k})_{k\in\mathbb{N}}$ eine *Teilfolge* von $(a_n)_{n\in\mathbb{N}}$. Eine Zahl $a \in \mathbb{R}$ heißt *Häufungswert* von $(a_n)_{n\in\mathbb{N}}$, wenn eine Teilfolge von $(a_n)_{n\in\mathbb{N}}$ gegen a konvergiert.

Die Teilfolge wird durch k indiziert; ihr k-tes Glied ist a_{n_k}. Ihren Grenzwert schreiben wir daher als $\lim_{k\to\infty} a_{n_k}$, sofern sie konvergiert.

Beispiel 7.28. *Für $a_n = (-1)^n$ sind $(a_{2k})_{k\in\mathbb{N}}$ und $(a_{2k+1})_{k\in\mathbb{N}}$ Teilfolgen von $(a_n)_{n\in\mathbb{N}}$. Die erste ist konstant 1, die zweite konstant -1. Damit sind 1 und -1 Häufungswerte, obwohl die ursprüngliche Folge divergent ist.* △

Wir können Häufungswerte natürlich in Bezug zur Gesamtfolge ausdrücken.

Beobachtung 7.29. Für eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ gilt:

(i) Konvergiert $(a_n)_{n\in\mathbb{N}}$ gegen a, so konvergiert jede Teilfolge gegen den einzigen Häufungswert a.

(ii) Eine Zahl $a \in \mathbb{R}$ ist genau dann Häufungswert, wenn

$$\forall \varepsilon > 0 \colon \forall j_0 \in \mathbb{N} \colon \exists n \geq j_0 \colon |a_n - a| < \varepsilon.$$

Eine beschränkte Folge muss nicht konvergieren. Können wir aus ihr aber stets eine Teilfolge auswählen, die sich einem festen Wert annähert?

Satz 7.30 (*Bolzano–Weierstraß*)**.** Jede beschränkte reelle Folge besitzt eine konvergente Teilfolge und damit mindestens einen Häufungswert.

Beweis. Sei $(a_n)_{n\in\mathbb{N}}$ beschränkt. Wir wählen $c > 0$ so, dass alle Glieder im Intervall $I_0 \coloneqq [-c, c]$ liegen. Wir halbieren dieses Intervall. Mindestens eine der beiden abgeschlossenen Hälften enthält a_n für unendlich viele Indizes n, denn sonst könnte auch ihre Vereinigung nach Satz 2.14 nur Glieder zu endlich vielen Indizes enthalten. Diese Hälfte nennen wir I_1. Wiederholen wir das Verfahren, erhalten wir ineinander enthaltene Intervalle $I_k = [l_k, r_k]$, die jeweils Glieder zu unendlich vielen Indizes enthalten, mit

$$r_k - l_k = \frac{2c}{2^k}.$$

Die Folge $(l_k)_{k\in\mathbb{N}}$ ist monoton wachsend und nach oben durch c beschränkt. Nach Satz 7.12 besitzt sie einen Grenzwert a. Wegen $\lim_{k\to\infty} 2^{-k} = 0$ und $r_k = l_k + 2c\,2^{-k}$ gilt nach den Grenzwertregeln aus Satz 7.14 auch $\lim_{k\to\infty} r_k = a$.

Wir wählen nun n_0 mit $a_{n_0} \in I_0$ und danach für jedes $k \geq 1$ einen Index $n_k > n_{k-1}$ mit $a_{n_k} \in I_k$. Dies ist möglich, weil I_k Glieder zu unendlich vielen Indizes enthält. Aus

$$l_k \leq a_{n_k} \leq r_k$$

folgt mit Satz 7.20 (iii), dass $\lim_{k\to\infty} a_{n_k} = a$. Damit ist die gesuchte Teilfolge gefunden. □

Die ursprüngliche Folge muss dabei nicht konvergieren, wie $(-1)^n$ zeigt. Wenn eine beschränkte Folge jedoch nur einen einzigen Häufungswert besitzt, bestimmt dieser das Verhalten der ganzen Folge.

Korollar 7.31. Eine beschränkte reelle Folge konvergiert genau dann gegen $a \in \mathbb{R}$, wenn a ihr einziger Häufungswert ist.

Beweis. Für eine konvergente Folge folgt die Aussage aus Beobachtung 7.29 (i). Sei umgekehrt $(a_n)_{n\in\mathbb{N}}$ beschränkt und a ihr einziger Häufungswert. Angenommen, $(a_n)_{n\in\mathbb{N}}$ konvergiert nicht gegen a. Dann gibt es ein $\varepsilon > 0$, sodass nach jedem Index ein Glied mit $|a_n - a| \geq \varepsilon$ auftritt. Wir können deshalb eine Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$ mit $|a_{n_k} - a| \geq \varepsilon$ für alle k wählen. Sie ist ebenfalls beschränkt und besitzt nach Satz 7.30 eine konvergente Teilfolge. Diese ist auch eine Teilfolge von $(a_n)_{n\in\mathbb{N}}$; ihren Grenzwert nennen wir b. Nach der Betragsregel aus Satz 7.14 und Satz 7.20 (ii) gilt $|b-a| \geq \varepsilon$, also $b \neq a$. Damit wäre b ein weiterer Häufungswert, ein Widerspruch zur Wahl von $(a_n)_{n\in\mathbb{N}}$. □

Beispiel 7.32. *Die Beschränktheit ist in diesem Korollar wesentlich. Für die Folge mit*

$$a_{2k} = 0, \qquad a_{2k+1} = k + 1 \qquad \text{für } k \in \mathbb{N}$$

ist 0 *ein Häufungswert, denn* $\lim_{k\to\infty} a_{2k} = 0$. *Sei nun* $a \in \mathbb{R} \setminus \{0\}$. *Wir setzen* $\varepsilon := |a|/2 > 0$ *und wählen* $k_0 \in \mathbb{N}$ *mit* $k_0 > |a| + \varepsilon$. *Für alle* $k \geq k_0$ *gilt*

$$|a_{2k} - a| = |a| > \varepsilon, \qquad |a_{2k+1} - a| = |k + 1 - a| \geq k + 1 - |a| > \varepsilon.$$

Damit ist $|a_n - a| \geq \varepsilon$ *für alle* $n \geq 2k_0$. *Nach Beobachtung 7.29 (ii) ist* a *daher kein Häufungswert. Die Folge besitzt somit genau einen Häufungswert, ist aber wegen* $a_{2k+1} = k + 1$ *unbeschränkt und deshalb nach Satz 7.9 (ii) divergent.* △

Teilfolgen liefern auch ein Konvergenzkriterium, das keinen vermuteten Grenzwert voraussetzt. Wenn wir einer festen Zahl beliebig nahe kommen, müssen die späten Glieder auch untereinander beliebig nahe liegen.

Definition 7.33. Eine reelle Folge $(a_n)_{n\in\mathbb{N}}$ heißt *Cauchy-Folge*, wenn

$$\forall \varepsilon > 0 \colon \exists n_0 \in \mathbb{N} \colon \forall n, m \geq n_0 \colon |a_n - a_m| < \varepsilon.$$

Satz 7.34 (*Cauchy-Kriterium*)**.** Eine reelle Folge konvergiert genau dann, wenn sie eine Cauchy-Folge ist.

Beweis. Sei zunächst $(a_n)_{n\in\mathbb{N}}$ eine Folge mit Grenzwert $\lim_{n\to\infty} a_n = a$. Zu $\varepsilon > 0$ wählen wir n_0 mit $|a_n - a| < \varepsilon/2$ für alle $n \geq n_0$. Dann ist $(a_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge, denn für alle $n, m \geq n_0$ gilt

$$|a_n - a_m| \overset{3.93}{\leq} |a_n - a| + |a - a_m| < \varepsilon.$$

Sei umgekehrt $(a_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge. Dann gibt es einen Index $n_0 \in \mathbb{N} \setminus \{0\}$ mit $|a_n - a_m| < 1$ für alle $n, m \geq n_0$, also insbesondere $|a_n| < |a_{n_0}| + 1$ für alle $n \geq n_0$. Damit ist $(a_n)_{n\in\mathbb{N}}$ beschränkt durch $|a_n| \leq c$ für alle $n \in \mathbb{N}$, wobei

$$c \coloneqq \max\{|a_0|, \ldots, |a_{n_0-1}|, |a_{n_0}| + 1\}.$$

Nach Satz 7.30 besitzt die Folge daher eine Teilfolge $(a_{n_k})_{k\in\mathbb{N}}$, die gegen eine reelle Zahl a konvergiert. Für beliebiges $\varepsilon > 0$ können wir, da $(a_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist, jeweils einen Index $\ell_0 \in \mathbb{N}$ mit $|a_n - a_m| < \varepsilon/2$ für alle $n, m \geq \ell_0$ wählen. Außerdem können wir, wegen $\lim_{k\to\infty} a_{n_k} = a$, ein k wählen mit $n_k \geq \ell_0$ und $|a_{n_k} - a| < \varepsilon/2$. Damit konvergiert $(a_n)_{n\in\mathbb{N}}$, denn für jedes $n \geq \ell_0$ gilt

$$|a_n - a| \overset{3.93}{\leq} |a_n - a_{n_k}| + |a_{n_k} - a| < \varepsilon. \qquad \square$$

Im Beweis führt die gegenseitige Nähe der Glieder zunächst zur Beschränktheit. Bolzano–Weierstraß liefert einen Grenzwert für eine Teilfolge; die Cauchy-Eigenschaft zwingt dann auch alle hinreichend späten Glieder in seine Nähe. Die Vollständigkeit der reellen Zahlen ist dabei über das Monotoniekriterium in den Beweis von Bolzano–Weierstraß eingegangen.

7.2 Reihen

Bisher haben wir untersucht, wann eine Folge von Werten konvergiert. Nun betrachten wir aufeinanderfolgende Änderungen eines Wertes, wie sie etwa bei der schrittweisen Verbesserung einer Näherung auftreten. Uns interessiert, wann der durch diese Änderungen entstehende Gesamtwert konvergiert.

Anschaulich muss die Folge der Änderungen eine Nullfolge sein, damit sich der Gesamtwert stabilisieren kann. Wir werden aber sehen, dass das nicht ausreicht: Beispielsweise wächst die Summe$1 + 1/2 + 1/3 + \ldots$ beim Hinzufügen weiterer Summanden unbeschränkt, obwohl die Änderungen immer kleiner werden. Wir untersuchen jetzt Bedingungen, die die Konvergenz einer solchen Summe garantieren.

Definition 7.35. Eine *(reelle) Reihe* ist eine Folge $(s_m)_{m\in\mathbb{N}}$ von *Partialsummen*

$$s_m \coloneqq \sum_{n=0}^{m} a_n \qquad \text{für } m \in \mathbb{N},$$

wobei $(a_n)_{n\in\mathbb{N}}$ eine reelle Folge von *Summanden* bezeichnet. Konvergiert die Reihe oder divergiert sie bestimmt, ist ihr *Wert* definiert als

$$\sum_{n=0}^{\infty} a_n \coloneqq \lim_{m\to\infty} s_m = \lim_{m\to\infty} \sum_{n=0}^{m} a_n.$$

Das Symbol $\sum_{n=0}^{\infty} a_n$ bezeichnet je nach Zusammenhang die Reihe selbst oder, falls sie konvergiert oder bestimmt divergiert, ihren Grenzwert. Beginnt die Summation bei $n_0 \in \mathbb{N}$, schreiben wir entsprechend $\sum_{n=n_0}^{\infty} a_n$. Dafür setzen wir die Summanden vor n_0 gleich null; die Partialsummen $s_m = \sum_{n=n_0}^{m} a_n$ sind für $m < n_0$ leere Summen. Die Summanden können auch durch einen Ausdruck angegeben werden: $\sum_{n=n_0}^{\infty} f(n)$ bezeichnet die Reihe mit den Summanden $a_n = f(n)$ für $n \geq n_0$.

Beispiel 7.36. *Wir betrachten die* geometrische Reihe $\sum_{n=0}^{\infty} q^n$ *mit* $q \in \mathbb{R} \setminus \{1\}$. *Für* $m \in \mathbb{N}$ *gilt*

$$(1-q)\sum_{n=0}^{m} q^n = \sum_{n=0}^{m} q^n - \sum_{n=0}^{m} q^{n+1} = \sum_{n=0}^{m} q^n - \sum_{n=1}^{m+1} q^n = 1 - q^{m+1}.$$

Da $1 - q \neq 0$, *können wir durch* $1 - q$ *teilen und erhalten die endliche Summenformel*

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

Ist $|q| < 1$, *so gilt nach Satz 7.25* $\lim_{m\to\infty} q^{m+1} = 0$ *und die geometrische Reihe ist konvergent mit Wert*

$$\sum_{n=0}^{\infty} q^n = \lim_{m\to\infty} \sum_{n=0}^{m} q^n = \lim_{m\to\infty} \frac{1-q^{m+1}}{1-q} \overset{7.14}{=} \frac{1-0}{1-q} = \frac{1}{1-q}.$$

Für $q > 1$ *gilt dagegen*

$$\sum_{n=0}^{m} q^n \geq \sum_{n=0}^{m} 1 = m+1 \qquad \textit{für alle } m \in \mathbb{N}.$$

Die Partialsummen wachsen daher über jede Schranke; die geometrische Reihe divergiert bestimmt gegen ∞. △

Beispiel 7.37. *Für die* harmonische Reihe *gilt* $\sum_{n=1}^{\infty} 1/n = \infty$. *Ihre Partialsummen* $h_m := \sum_{n=1}^{m} 1/n$ *sind nach*

$$h_{2^k} = 1 + \sum_{j=1}^{k} \sum_{n=2^{j-1}+1}^{2^j} \frac{1}{n} \geq 1 + \sum_{j=1}^{k} \sum_{n=2^{j-1}+1}^{2^j} \frac{1}{2^j} = 1 + \sum_{j=1}^{k} 2^{j-1}\frac{1}{2^j} = 1 + \frac{k}{2} \qquad \textit{für } k \in \mathbb{N}$$

unbeschränkt und können nach Satz 7.9 (ii) nicht konvergieren. Da sie monoton wachsen, gilt sogar $\lim_{m\to\infty} h_m = \infty$. △

Beispiel 7.38. *Für* $m \geq 1$ *gilt*

$$\sum_{n=1}^{m} \frac{1}{n(n+1)} = \sum_{n=1}^{m} \left(\frac{1}{n} - \frac{1}{n+1}\right) = \frac{1}{1} - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} - \frac{1}{4} + \cdots - \frac{1}{m+1} = 1 - \frac{1}{m+1},$$

also

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{m\to\infty} \sum_{n=1}^{m} \frac{1}{n(n+1)} = \lim_{m\to\infty} \left(1 - \frac{1}{m+1}\right) \overset{7.14}{=} 1.$$

△

Definition 7.39. Eine Reihe $\sum_{n=0}^{\infty} a_n$ heißt *absolut konvergent*, wenn die Reihe $\sum_{n=0}^{\infty} |a_n|$ konvergiert.

Beispiel 7.40. *Die Reihe* $\sum_{n=0}^{\infty}(-1/2)^n$ *ist absolut konvergent, denn nach Beispiel 7.36 gilt*

$$\sum_{n=0}^{\infty} \left|\left(-\frac{1}{2}\right)^n\right| = \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = 2.$$

△

Konvergenzkriterien für Reihen

Wir leiten Kriterien her, mit deren Hilfe wir die Konvergenz oder Divergenz einer Reihe nachweisen können.

Es ist anschaulich klar, dass eine Reihe nur konvergieren kann, wenn ihre Summanden eine Nullfolge bilden. Umgekehrt genügt es für die Konvergenz nicht, dass die Summanden gegen 0 konvergieren, wie wir bei der harmonischen Reihe (Beispiel 7.37) gesehen haben.

Beobachtung 7.41. Für jede konvergente Reihe $\sum_{n=0}^{\infty} a_n$ gilt $\lim_{n\to\infty} a_n = 0$.

Beweis. Konvergiert eine Reihe $\sum_{n=0}^{\infty} a_n$ mit Partialsummen $s_m = \sum_{n=0}^{m} a_n$ gegen s, so gilt $a_n = s_n - s_{n-1}$ für $n \geq 1$. Mit den Grenzwertregeln aus Satz 7.14 folgt

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = s - s = 0. \qquad \square$$

Beobachtung 7.42. Seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ konvergente Reihen und $\lambda, \mu \in \mathbb{R}$. Dann konvergiert auch $\sum_{n=0}^{\infty}(\lambda a_n + \mu b_n)$, und es gilt

$$\sum_{n=0}^{\infty}(\lambda a_n + \mu b_n) = \lambda \sum_{n=0}^{\infty} a_n + \mu \sum_{n=0}^{\infty} b_n.$$

Beweis. Wir bezeichnen die Partialsummen der beiden Reihen mit $s_m := \sum_{n=0}^{m} a_n$ und $t_m := \sum_{n=0}^{m} b_n$ für $m \in \mathbb{N}$. Nach Voraussetzung konvergieren $(s_m)_{m\in\mathbb{N}}$ und $(t_m)_{m\in\mathbb{N}}$ gegen die Werte $s := \sum_{n=0}^{\infty} a_n$ und $t := \sum_{n=0}^{\infty} b_n$.

Für die Partialsummen der Reihe mit Summanden $\lambda a_n + \mu b_n$ gilt

$$u_m := \sum_{n=0}^{m}(\lambda a_n + \mu b_n) = \lambda \sum_{n=0}^{m} a_n + \mu \sum_{n=0}^{m} b_n = \lambda s_m + \mu t_m.$$

Wir wenden nun zuerst die Summenregel und dann die Faktorregel für Grenzwerte an:

$$\lim_{m\to\infty} u_m = \lim_{m\to\infty}(\lambda s_m + \mu t_m) \overset{7.14}{=} \lim_{m\to\infty}(\lambda s_m) + \lim_{m\to\infty}(\mu t_m) \overset{7.14}{=} \lambda \lim_{m\to\infty} s_m + \mu \lim_{m\to\infty} t_m = \lambda s + \mu t.$$

Damit konvergiert die Partialsummenfolge $(u_m)_{m\in\mathbb{N}}$, also auch die zugehörige Reihe. Ihr Wert ist

$$\sum_{n=0}^{\infty}(\lambda a_n + \mu b_n) = \lambda s + \mu t = \lambda \sum_{n=0}^{\infty} a_n + \mu \sum_{n=0}^{\infty} b_n. \qquad \square$$

Das Ändern, Hinzufügen oder Weglassen endlich vieler Anfangssummanden ändert die Konvergenz einer Reihe nicht. Bei Folgen haben wir in Satz 7.9 (iii) gesehen, dass Änderungen endlich vieler Glieder sogar den Grenzwert unverändert lassen. Bei Reihen unterscheiden sich dagegen die hinreichend späten Partialsummen um eine feste Konstante, gegebenenfalls nach einer festen Indexverschiebung.

Beobachtung 7.43. Für jedes $n_0 \in \mathbb{N}$ konvergiert $\sum_{n=0}^{\infty} a_n$ genau dann, wenn $\sum_{n=n_0}^{\infty} a_n$ konvergiert.

Um die Konvergenz einer Reihe ohne Kenntnis ihres Wertes zu prüfen, wenden wir nun das Cauchy-Kriterium für Folgen aus Satz 7.34 auf ihre Partialsummen an.

Satz 7.44 (Cauchy-Kriterium)**.** Eine Reihe $\sum_{n=0}^{\infty} a_n$ konvergiert genau dann, wenn

$$\forall \varepsilon > 0 \colon \exists n_0 \in \mathbb{N} \colon \forall n, m \in \mathbb{N}, m \geq n \geq n_0 \colon \left| \sum_{k=n}^{m} a_k \right| < \varepsilon.$$

Beweis. Sei $s_m = \sum_{k=0}^{m} a_k$. Nach Satz 7.34 konvergiert die Reihe genau dann, wenn $(s_m)_{m\in\mathbb{N}}$ eine Cauchy-Folge ist.

Ist $(s_m)_{m\in\mathbb{N}}$ Cauchy-Folge, dann existiert per Definition für alle $\varepsilon > 0$ ein $j_0 \in \mathbb{N}$ mit

$$\left| \sum_{k=n}^{m} a_k \right| = |s_m - s_{n-1}| < \varepsilon \qquad \text{für } m \geq n-1 \geq j_0,$$

also ist die Bedingung des Satzes für $n_0 = j_0 + 1$ erfüllt.

Gilt umgekehrt diese Bedingung für ein beliebiges $\varepsilon > 0$ und sind $m > \ell \geq n_0$, so ist $|s_m - s_\ell| = |\sum_{k=\ell+1}^{m} a_k| < \varepsilon$. Für $m = \ell$ ist die Differenz null; für $m < \ell$ vertauschen wir die Indizes. Damit ist die Partialsummenfolge eine Cauchy-Folge. □

Mit dem Cauchy-Kriterium können wir nun zeigen, dass die absolute Konvergenz eine hinreichende Bedingung für die Konvergenz einer Reihe ist.

Satz 7.45. Jede absolut konvergente Reihe konvergiert.

Beweis. Sei $\sum_{n=0}^{\infty} |a_n|$ konvergent und $\varepsilon > 0$. Nach Satz 7.44 gibt es $n_0 \in \mathbb{N}$ mit $\sum_{k=n}^{m} |a_k| < \varepsilon$ für alle $m \geq n \geq n_0$. Durch wiederholtes Anwenden der Dreiecksungleichung aus Satz 3.93 folgt

$$\left| \sum_{k=n}^{m} a_k \right| \leq \sum_{k=n}^{m} |a_k| < \varepsilon \qquad \text{für alle } m \geq n \geq n_0.$$

Nochmals Satz 7.44 liefert die Konvergenz von $\sum_{n=0}^{\infty} a_n$. □

Für Reihen mit nichtnegativen Summanden können wir Konvergenz oder Divergenz auch durch den Vergleich mit einer bereits bekannten Reihe nachweisen.

Satz 7.46. Seien $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ reelle Folgen mit $0 \leq a_n \leq b_n$ für alle $n \geq n_0$, wobei $n_0 \in \mathbb{N}$. Dann gilt:

(i) Konvergiert $\sum_{n=0}^{\infty} b_n$, so konvergiert auch $\sum_{n=0}^{\infty} a_n$ (*Majorantenkriterium*).

(ii) Divergiert $\sum_{n=0}^{\infty} a_n$, so divergiert auch $\sum_{n=0}^{\infty} b_n$ (*Minorantenkriterium*).

Beweis. Zu (i): Nach Beobachtung 7.43 können wir die endlich vielen Summanden vor n_0 weglassen. Für alle $m \geq n_0$ ist

$$0 \leq \sum_{n=n_0}^{m} a_n \leq \sum_{n=n_0}^{m} b_n.$$

Die rechten Partialsummen sind wegen ihrer Konvergenz beschränkt. Damit sind auch die linken Partialsummen beschränkt. Wegen der nichtnegativen Summanden wachsen sie monoton und konvergieren daher nach Satz 7.12. Teil (ii) ist die Kontraposition von Teil (i). □

Beispiel 7.47. *Die Reihe $\sum_{n=1}^{\infty} 1/n^2$ lässt sich nicht unmittelbar so aufteilen wie in Beispiel 7.38. Für $n \geq 1$ gilt wegen $n+1 \leq 2n$ aber*

$$0 \leq \frac{1}{n^2} \leq \frac{2}{n(n+1)}.$$

Die Reihe $\sum_{n=1}^{\infty} \frac{2}{n(n+1)}$ konvergiert nach Beispiel 7.38 und Beobachtung 7.42. Satz 7.46 (i) liefert somit die Konvergenz von $\sum_{n=1}^{\infty} 1/n^2$, ohne ihren Reihenwert bestimmen zu müssen. △

Beispiel 7.48. *Die Reihe $\sum_{n=1}^{\infty} 1/\sqrt{n}$ divergiert: Wegen $\sqrt{n} \leq n$ ist $1/\sqrt{n} \geq 1/n$, sodass Beispiel 7.37 und Satz 7.46 (ii) anwendbar sind.* △

Bei nichtnegativen Summanden können die Partialsummen nur wachsen. Vorzeichenwechsel ermöglichen dagegen Ausgleich: Positive und negative Beiträge können sich gegenseitig aufheben. Die harmonische Reihe erhält dadurch ein anderes Konvergenzverhalten.

Satz 7.49 (Leibniz-Kriterium)**.** Sei $(b_n)_{n\in\mathbb{N}}$ eine monoton fallende Folge mit $b_n \geq 0$ für alle $n \in \mathbb{N}$ und $\lim_{n\to\infty} b_n = 0$. Dann konvergiert die Reihe

$$\sum_{n=0}^{\infty} (-1)^n b_n.$$

Beweis. Bezeichnen wir die Partialsummen mit $s_m = \sum_{n=0}^{m} (-1)^n b_n$. Die Partialsummen mit geraden Indizes fallen, die mit ungeraden wachsen, denn

$$s_{2k+2} - s_{2k} = -b_{2k+1} + b_{2k+2} \leq 0, \qquad s_{2k+3} - s_{2k+1} = b_{2k+2} - b_{2k+3} \geq 0, \qquad \text{für alle } k \in \mathbb{N}.$$

Zudem ist $s_{2k+1} = s_{2k} - b_{2k+1} \leq s_{2k}$ und damit

$$s_1 \leq s_{2k+1} \leq s_{2k} \leq s_0 \qquad \text{für alle } k \in \mathbb{N}.$$

Nach Satz 7.12 konvergieren $(s_{2k})_{k\in\mathbb{N}}$ und $(s_{2k+1})_{k\in\mathbb{N}}$ also. Da $\lim_{n\to\infty} b_n = 0$, konvergiert auch die Teilfolge $(b_{2k+1})_{k\in\mathbb{N}}$ gegen 0 (Beobachtung 7.29 (i)). Wegen $s_{2k} - s_{2k+1} = b_{2k+1}$ ist nach den Grenzwertregeln aus Satz 7.14 auch die Differenz der beiden Grenzwerte 0; sie stimmen also überein. Wir nennen den gemeinsamen Grenzwert s. Für jedes $\varepsilon > 0$ liegen alle hinreichend späten geraden und ungeraden Partialsummen im Abstand kleiner als ε von s; also konvergiert die gesamte Partialsummenfolge gegen s. □

Beispiel 7.50. *Die Folge $b_n = 1/(n+1)$ ist monoton fallend und nach der Abschätzung in Beispiel 7.8 eine Nullfolge. Daher liefert Satz 7.49 die Konvergenz der* alternierenden harmonischen Reihe

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots.$$

Allerdings konvergiert die Reihe nicht absolut, wie wir in Beispiel 7.37 gesehen haben. △

Wir haben uns in Beispiel 7.36 überlegt, dass die geometrische Reihe $\sum_{n=0}^{\infty} q^n$ für $0 \le q < 1$ konvergiert und für $q > 1$ divergiert. Die folgenden beiden Kriterien verallgemeinern diese Erkenntnis. Zunächst müssen sich die Beträge aufeinanderfolgender Summanden nicht genau um einen festen Faktor unterscheiden: Es genügt, dass ihre Quotienten gegen $q < 1$ bzw. $q > 1$ konvergieren.

Satz 7.51 (Quotientenkriterium)**.** Sei $(a_n)_{n\in\mathbb{N}}$ eine reelle Folge mit $a_n \neq 0$ für alle $n \ge n_1$, wobei $n_1 \in \mathbb{N}$ ist. Außerdem gelte $\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = q \in \mathbb{R}$. Ist $q < 1$, so konvergiert $\sum_{n=0}^{\infty} a_n$ absolut. Ist $q > 1$, so divergiert die Reihe.

Beweis. Wegen $|a_{n+1}/a_n| \ge 0$ ist $q \ge 0$ nach Satz 7.20. Sei zunächst $q < 1$. Wir wählen r mit $q < r < 1$ und $\varepsilon := r - q > 0$. Nach der Grenzwertdefinition gibt es $n_0 \ge n_1$ mit $\big||a_{n+1}/a_n| - q\big| < \varepsilon = r - q$ für alle $n \ge n_0$. Also ist $|a_{n+1}/a_n| = |a_{n+1}|/|a_n| < r$ und damit $|a_{n+1}| < r\,|a_n|$. Wiederholtes Anwenden liefert

$$|a_{n_0+k}| < r|a_{n_0+k-1}| < \dots < |a_{n_0}|r^k \qquad \text{für alle } k \in \mathbb{N}\setminus\{0\}.$$

Die geometrische Reihe $\sum_{k=0}^{\infty} |a_{n_0}|r^k$ konvergiert nach Beispiel 7.36 und Beobachtung 7.42. Das Majorantenkriterium (Satz 7.46 (i)) zeigt wegen $0 \le |a_{n_0+k}| \le |a_{n_0}|r^k$ für alle $k \in \mathbb{N}$, dass $\sum_{k=0}^{\infty} |a_{n_0+k}|$ und damit $\sum_{n=n_0}^{\infty} |a_n|$ konvergiert. Das Hinzufügen der endlich vielen Summanden mit $n < n_0$ ändert nach Beobachtung 7.43 die Konvergenz nicht. Also konvergiert auch $\sum_{n=0}^{\infty} |a_n|$, das heißt, $\sum_{n=0}^{\infty} a_n$ konvergiert absolut.

Sei nun $q > 1$. Wir wählen r mit $1 < r < q$ und $\varepsilon := q - r > 0$. Nach der Grenzwertdefinition gibt es $n_0 \ge n_1$ mit $\big||a_{n+1}/a_n| - q\big| < \varepsilon = q - r$ für alle $n \ge n_0$. Also ist $|a_{n+1}/a_n| = |a_{n+1}|/|a_n| > r$ und damit $|a_{n+1}| > r\,|a_n|$. Wiederholtes Anwenden liefert

$$|a_{n_0+k}| > r|a_{n_0+k-1}| > \dots > |a_{n_0}|r^k \ge |a_{n_0}| > 0 \qquad \text{für alle } k \in \mathbb{N}\setminus\{0\}.$$

Die Beträge aller Summanden ab n_0 sind also mindestens so groß wie die feste positive Zahl $|a_{n_0}|$. Damit ist $(a_n)_{n\in\mathbb{N}}$ keine Nullfolge, und nach Beobachtung 7.41 divergiert die Reihe $\sum_{n=0}^{\infty} a_n$. □

Beispiel 7.52. *Für $a_n = 1/n!$ gilt $|a_{n+1}/a_n| = 1/(n+1)$. Wegen $\lim_{n\to\infty} |a_{n+1}/a_n| = 0$ (Beispiel 7.8) konvergiert $\sum_{n=0}^{\infty} 1/n!$ nach Satz 7.51. Ebenso konvergiert $\sum_{n=1}^{\infty} n/2^n$, denn für $n \ge 1$ gilt mit Satz 7.14*

$$\frac{(n+1)/2^{n+1}}{n/2^n} = \frac{n+1}{2n} = \frac{1}{2}\left(1 + \frac{1}{n}\right) \qquad \textit{und} \qquad \lim_{n\to\infty} \frac{1}{2}\left(1 + \frac{1}{n}\right) = \frac{1}{2} < 1. \qquad \triangle$$

Beispiel 7.53. *Für die harmonische Reihe $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} \frac{1}{n}$ aus Beispiel 7.37 und die Reihe $\sum_{n=1}^{\infty} a_n$ mit $a_n = \frac{1}{n(n+1)}$ aus Beispiel 7.38 ergeben sich die Quotienten*

$$\frac{b_{n+1}}{b_n} = \frac{1/(n+1)}{1/n} = \frac{n}{n+1}, \qquad \frac{a_{n+1}}{a_n} = \frac{1/((n+1)(n+2))}{1/(n(n+1))} = \frac{n}{n+2}.$$

Mit den Grenzwertregeln aus Satz 7.14 folgt

$$\lim_{n\to\infty} \frac{b_{n+1}}{b_n} = \lim_{n\to\infty} \frac{1}{1+1/n} = 1, \qquad \lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to\infty} \frac{1}{1+2/n} = 1.$$

Beide Quotientenfolgen konvergieren also gegen denselben Grenzwert, obwohl die harmonische Reihe divergiert und die Teleskopreihe konvergiert. Das Quotientenkriterium kann bei Grenzwert 1 *also nicht entscheiden.* △

Bei der geometrischen Reihe aus Beispiel 7.36 gilt für $q \geq 0$ und $n \geq 1$ stets $\sqrt[n]{|q^n|} = q$. Das Wurzelkriterium überträgt die Unterscheidung zwischen Konvergenz und Divergenz auf Reihen, bei denen die n-ten Wurzeln der Summandenbeträge gegen q konvergieren.

Satz 7.54 (Wurzelkriterium)**.** Sei $(a_n)_{n\in\mathbb{N}}$ eine reelle Folge mit $\lim_{n\to\infty} \sqrt[n]{|a_n|} = q \in \mathbb{R}$. Ist $q < 1$, so konvergiert $\sum_{n=0}^{\infty} a_n$ absolut. Ist $q > 1$, so divergiert die Reihe.

Beweis. Übung. □

Wie beim Quotientenkriterium ist auch beim Wurzelkriterium im Fall $q = 1$ keine allgemeine Aussage über Konvergenz oder Divergenz möglich.

Beispiel 7.55. *Für* $a_n = (\frac{n+1}{3n})^n$ *mit* $n \geq 1$ *ist*

$$\sqrt[n]{|a_n|} = \frac{n+1}{3n} = \frac{1}{3}\left(1 + \frac{1}{n}\right).$$

Nach Satz 7.14 konvergiert dieser Ausdruck gegen $1/3 < 1$*. Satz 7.54 liefert die absolute Konvergenz von* $\sum_{n=1}^{\infty} (\frac{n+1}{3n})^n$*.* △

Beim Multiplizieren zweier Reihen entstehen Produkte $a_k b_\ell$. Wir fassen jeweils diejenigen mit $k + \ell = n$ zu einem Summanden c_n zusammen. Der folgende Satz zeigt, dass diese Anordnung bei absolut konvergenten Reihen das Produkt der Reihenwerte liefert.

Satz 7.56 (Produktsatz)**.** Seien $\sum_{n=0}^{\infty} a_n$ und $\sum_{n=0}^{\infty} b_n$ absolut konvergente Reihen. Wir setzen

$$c_n := \sum_{k=0}^{n} a_k b_{n-k} \qquad \text{für } n \in \mathbb{N}.$$

Dann konvergiert auch $\sum_{n=0}^{\infty} c_n$ absolut, und es gilt

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right).$$

Beweis. Wir setzen

$$s_m \coloneqq \sum_{n=0}^{m} a_n, \quad t_m \coloneqq \sum_{n=0}^{m} b_n, \quad \bar{s}_m \coloneqq \sum_{n=0}^{m} |a_n|, \quad \bar{t}_m \coloneqq \sum_{n=0}^{m} |b_n|$$

und bezeichnen die zugehörigen Grenzwerte mit $s, t, \bar{s}, \bar{t}$. Für alle $m \in \mathbb{N}$ gilt

$$\begin{aligned}\sum_{n=0}^{m} |c_n| = \sum_{n=0}^{m} \left| \sum_{k=0}^{n} a_k b_{n-k} \right| &\overset{3.93}{\leq} \sum_{n=0}^{m} \sum_{k=0}^{n} |a_k|\,|b_{n-k}| \\ &\leq \sum_{k=0}^{m} \sum_{\ell=0}^{m} |a_k|\,|b_\ell| = \bar{s}_m \bar{t}_m \leq \bar{s}\bar{t}.\end{aligned}$$

Die zweite Ungleichung ergänzt die nichtnegativen Beiträge mit $k + \ell > m$. Die Partialsummen der Beträge wachsen monoton und sind beschränkt; nach Satz 7.12 konvergiert daher $\sum_{n=0}^{\infty} |c_n|$. Nach Satz 7.45 konvergieren damit auch die Partialsummen $u_m \coloneqq \sum_{n=0}^{m} c_n$ gegen einen Wert u.

Im Produkt $s_{2m}t_{2m}$ stehen gegenüber u_{2m} zusätzlich genau die Beiträge mit $k + \ell > 2m$. Für diese gilt $k > m$ oder $\ell > m$, also

$$\begin{aligned}|s_{2m}t_{2m} - u_{2m}| = \left| \sum_{k=1}^{2m} \sum_{\ell=2m-k+1}^{2m} a_k b_\ell \right| &\overset{3.93}{\leq} \sum_{k=1}^{2m} \sum_{\ell=2m-k+1}^{2m} |a_k|\,|b_\ell| \\ &\leq \sum_{k=m+1}^{2m} \sum_{\ell=0}^{2m} |a_k|\,|b_\ell| + \sum_{k=0}^{2m} \sum_{\ell=m+1}^{2m} |a_k|\,|b_\ell| \\ &= (\bar{s}_{2m} - \bar{s}_m)\bar{t}_{2m} + \bar{s}_{2m}(\bar{t}_{2m} - \bar{t}_m) \\ &\leq (\bar{s} - \bar{s}_m)\bar{t} + \bar{s}(\bar{t} - \bar{t}_m).\end{aligned}$$

Doppelt gezählte Beiträge vergrößern nur die obere Schranke. Nach Beobachtung 7.29 (i) und den Grenzwertregeln aus Satz 7.14 konvergiert die linke Seite gegen $|st-u|$ und die rechte gegen null. Der Grenzwertvergleich aus Satz 7.20 liefert $|st - u| = 0$, also $u = st$. □

Beispiel 7.57. *Gewöhnliche Konvergenz beider Reihen reicht für Satz 7.56 nicht aus. Wir betrachten zweimal dieselbe Reihe mit Summanden*

$$a_n = b_n \coloneqq \frac{(-1)^n}{\sqrt{n+1}} \qquad \textit{für } n \in \mathbb{N}.$$

Die Beträge fallen monoton gegen null: Für jedes $\varepsilon > 0$ gilt $1/\sqrt{n+1} < \varepsilon$, sobald $n + 1 > 1/\varepsilon^2$. Nach Satz 7.49 konvergieren daher beide Reihen.

Für die im Produktsatz angegebenen Summanden erhalten wir

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

Alle Beiträge haben dasselbe Vorzeichen. Für $0 \leq k \leq n$ sind außerdem beide Faktoren unter der Wurzel höchstens $n + 1$. Deshalb gilt

$$|c_n| = \sum_{k=0}^{n} \frac{1}{\sqrt{(k+1)(n-k+1)}} \geq \sum_{k=0}^{n} \frac{1}{n+1} = 1 \qquad \textit{für alle } n \in \mathbb{N}.$$

Die Folge $(c_n)_{n\in\mathbb{N}}$ ist keine Nullfolge. Nach Beobachtung 7.41 divergiert die Reihe $\sum_{n=0}^{\infty} c_n$, obwohl beide Ausgangsreihen konvergieren. △

7.3 Asymptotik

Bisher haben wir untersucht, ob Folgen konvergieren und welchen Grenzwert sie besitzen. Nun untersuchen wir, wie schnell Folgen wachsen oder sich ihrem Grenzwert nähern. Derartige Fragen sind zum Beispiel in der Algorithmik wichtig, wo uns interessiert, wie schnell der Rechenaufwand eines Verfahrens mit der Datenmenge wächst, wie effizient also das Verfahren ist, oder wie schnell sich ein Approximationsverfahren der Lösung nähert.

Betrachten wir beispielsweise das Gram–Schmidt-Verfahren aus Satz 4.98, das aus einer Basis mit $n \geq 1$ Vektoren eine Orthonormalbasis bestimmt. Für den j-ten Vektor ziehen wir die $j-1$ Projektionsanteile nacheinander vom Eingabevektor ab. Dies erfordert $j-1$ Vektoradditionen, wobei wir eine Subtraktion als Addition des negativen Vektors zählen. Die Gesamtzahl der Vektoradditionen ist durch die Partialsumme

$$t(n) \coloneqq \sum_{i=0}^{n-1} i \overset{2.35}{=} \frac{n(n-1)}{2}$$

gegeben. Das Wachstum dieser Anzahl wird also durch die Partialsummen der bestimmt divergenten Reihe $\sum_{n=0}^{\infty} n$ beschrieben.

Wie können wir dieses Wachstum messen? Wenn sich die Anzahl der Basisvektoren verdoppelt, erhöht sich die Anzahl der Vektoradditionen für $n \geq 2$ um den Faktor

$$\frac{t(2n)}{t(n)} = \frac{2(2n-1)}{n-1} = \frac{4n-4}{n-1} + \frac{2}{n-1} = 4 + \frac{2}{n-1}.$$

Da $\lim_{n\to\infty} \frac{1}{n-1} = 0$ gilt, gilt also $\lim_{n\to\infty} \frac{t(2n)}{t(n)} = 4$. Also verhält sich die Anzahl Vektoradditionen für große n wie eine Funktion $f\colon \mathbb{N} \to \mathbb{R}$ mit $f(2n) = 4f(n)$, also wie zum Beispiel $f(n) = n^2$.

Tatsächlich ist

$$\lim_{n\to\infty} \frac{t(n)}{n^2} = \lim_{n\to\infty} \frac{1}{2}\left(1 - \frac{1}{n}\right) = \frac{1}{2},$$

was bestätigt, dass die Anzahl Vektoradditionen bis auf einen Faktor 2 so wächst wie n^2. Im Gegensatz dazu ist $\lim_{n\to\infty} t(n)/n = \infty$ und $\lim_{n\to\infty} t(n)/n^3 = 0$, also wächst $t(n)$ in diesem Sinne schneller als n und langsamer als n^3.

Wir fassen Folgen in diesem Abschnitt als Funktionen $f\colon \mathbb{N} \to \mathbb{R}$ auf. Um auch (unbestimmt) divergente Folgen einzufangen, vergleichen wir das Wachstum zweier Folgen in der Definition unten nicht über Grenzwerte. Da Restfehler auch negativ sein können, vergleichen wir außerdem die Beträge der Folgenglieder.

Definition 7.58. Für $g\colon \mathbb{N} \to \mathbb{R}$ definieren wir die Menge der Folgen, die asymptotisch

- *nicht schneller wachsen als* g durch $\mathcal{O}(g) \coloneqq \{f\colon \mathbb{N} \to \mathbb{R} \mid \exists c > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| \leq c|g(n)|\}$;
- *nicht langsamer wachsen als* g durch $\Omega(g) \coloneqq \{f\colon \mathbb{N} \to \mathbb{R} \mid g \in \mathcal{O}(f)\}$;
- *genauso schnell wachsen wie* g durch $\Theta(g) \coloneqq \mathcal{O}(g) \cap \Omega(g)$.

Die Mengen $\Omega(g)$ und $\Theta(g)$ lassen sich auch analog zu $\mathcal{O}(g)$ definieren.

Beobachtung 7.59. Für $g\colon \mathbb{N} \to \mathbb{R}$ gilt

$$\Omega(g) = \{f\colon \mathbb{N} \to \mathbb{R} \mid \exists c > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| \geq c|g(n)|\},$$
$$\Theta(g) = \{f\colon \mathbb{N} \to \mathbb{R} \mid \exists c, d > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon c|g(n)| \leq |f(n)| \leq d|g(n)|\}.$$

Wir beobachten zunächst den Zusammenhang von Definition 7.58 mit unseren anfänglichen Überlegungen zur Effizienz des Gram-Schmidt-Verfahrens. Dazu merken wir an, dass $f \in \mathcal{O}(g)$ gleichbedeutend damit ist, dass die Folge $(f(n)/g(n))_{n\in\mathbb{N}}$ beschränkt ist (vgl. Definition 7.4), sofern der Quotient für alle Indizes definiert ist. Also ist insbesondere $f \in \mathcal{O}(g)$, wenn diese Folge konvergiert (Satz 7.9 (ii)). Dabei genügt es, die Quotienten ab einem beliebigen festen Index zu betrachten.

Beobachtung 7.60. Für $f, g\colon \mathbb{N} \to \mathbb{R}$ gilt:

- Ist $\lim_{n\to\infty} f(n)/g(n) = d \in \mathbb{R}$, dann folgt $f \in \mathcal{O}(g)$.
- Ist $\lim_{n\to\infty} g(n)/f(n) = d \in \mathbb{R}$, dann folgt $f \in \Omega(g)$.
- Ist $\lim_{n\to\infty} f(n)/g(n) = d \in \mathbb{R} \setminus \{0\}$, dann folgt $f \in \Theta(g)$.

Wie schon für Reihen und Grenzwerte schreiben wir $f(n) \in \mathcal{O}(g(n))$ für $f \in \mathcal{O}(g)$, beispielsweise also $n \in \mathcal{O}(n^2)$, ohne die Funktionen explizit zu benennen; dieselbe Konvention verwenden wir für die anderen Symbole. Endlich viele Anfangsglieder beeinflussen die Definitionen nicht. Deshalb dürfen auch hier Ausdrücke vorkommen, die erst ab einem bestimmten Index definiert sind, beispielsweise $1/n \in \Omega(1/n^2)$; wir ergänzen die fehlenden Werte wie in Definition 7.11 beliebig.

Beispiel 7.61. *Für die Anzahl der Vektoradditionen im Gram-Schmidt-Verfahren gilt $t(n) \in \Theta(n^2)$ nach Beobachtung 7.59, denn*

$$\frac{1}{4}|n^2| \leq |t(n)| = \left|\frac{n(n-1)}{2}\right| \leq \frac{1}{2}|n^2| \qquad \textit{für alle } n \geq 2. \qquad \triangle$$

Beispiel 7.62. *Bei der gewöhnlichen Multiplikation zweier Matrizen $A, B \in \mathbb{R}^{n\times n}$ mit $n \geq 1$ berechnen wir nach Definition 5.46 jeden Eintrag durch*

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \qquad \textit{für } 1 \leq i, j \leq n.$$

Für jeden der n^2 Einträge addieren wir n Produkte und benötigen dafür jeweils $n-1$ Additionen. Insgesamt ist die Anzahl der benötigten Additionen also

$$n^2(n-1) = n^3 - n^2 \in \Theta(n^3).$$

Es sind inzwischen Verfahren bekannt, bei denen die Gesamtzahl der Additionen und Multiplikationen in $\mathcal{O}(n^{2{,}371177})$ liegt.[1] $\triangle$

[1] Stand September 2026: E. Dupont et al., Improving the matrix multiplication exponent with modern optimization and AlphaEvolve (2026), arXiv:2608.16884.

Beispiel 7.63. *Wir wenden das Gauß-Verfahren aus Abschnitt 6.2 auf eine invertierbare Matrix $A \in \mathbb{R}^{n\times n}$ mit $n \geq 1$ an und zählen die Additionen reeller Zahlen. Beim i-ten Pivotelement addieren wir in der Vorwärtsphase ein Vielfaches der Pivotzeile zu $n-i$ anderen Zeilen und in der Rückwärtsphase zu $i-1$ anderen Zeilen. Zusammen sind dies $(n-i)+(i-1) = n-1 \in \Theta(n)$ Zeilenadditionen pro Pivotelement. Bei komponentenweiser Ausführung benötigt jede Zeilenaddition n Additionen reeller Zahlen, sodass pro Pivotelement $n(n-1) \in \Theta(n^2)$ Additionen reeller Zahlen anfallen. Da die invertierbare Matrix A genau n Pivotelemente besitzt, beträgt die Gesamtzahl der Additionen*

$$n \cdot (n-1) \cdot n = n^3 - n^2 \in \Theta(n^3).$$

Auch für eine nichtinvertierbare Matrix $A \in \mathbb{R}^{n\times n}$ gibt es höchstens n Pivotelemente. Die Anzahl der Additionen liegt daher allgemein in $\mathcal{O}(n^3)$. △

Beispiel 7.64. *Es gilt $f \in \mathcal{O}(1)$ genau dann, wenn f beschränkt ist.* △

Beispiel 7.65. *Auch der Fehler einer Reihe, also wie weit die Partialsumme vom Wert der Reihe entfernt ist, lässt sich mit der $\mathcal{O}$-Notation beschreiben. Für die Reihe aus Beispiel 7.38 und die geometrische Reihe aus Beispiel 7.36 gilt für $m \geq 1$*

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} - \sum_{k=1}^{m} \frac{1}{k(k+1)} = 1 - \sum_{k=1}^{m} \frac{1}{k(k+1)} \overset{7.38}{=} \frac{1}{m+1} \in \mathcal{O}(1/m),$$
$$\sum_{k=1}^{\infty} \frac{1}{2^k} - \sum_{k=1}^{m} \frac{1}{2^k} = 1 - \sum_{k=1}^{m} \frac{1}{2^k} \overset{7.36}{=} \frac{1}{2^m} \in \mathcal{O}(2^{-m}).$$

△

Für einen strengeren Vergleich soll jeder noch so kleine positive Faktor als obere Schranke genügen. Dabei darf der Anfangsindex vom Faktor abhängen.

Definition 7.66. Für $g\colon \mathbb{N} \to \mathbb{R}$ definieren wir die Menge der Folgen, die asymptotisch

- *echt langsamer wachsen als g* durch $o(g) \coloneqq \{f\colon \mathbb{N} \to \mathbb{R} \mid \forall \varepsilon > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| \leq \varepsilon |g(n)|\}$,
- *echt schneller wachsen als g* durch $\omega(g) \coloneqq \{f\colon \mathbb{N} \to \mathbb{R} \mid \forall c > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| \geq c|g(n)|\}$.

Insbesondere ist $f \in \omega(g)$ genau dann, wenn $g \in o(f)$.

Die letzte Äquivalenz gilt dabei, weil wir $\varepsilon = 1/c$ identifizieren können.

Bemerkung 7.67. Die Zeichen $\mathcal{O}, \Omega, \Theta, o$ und ω heißen *(Bachmann-)Landau-Symbole*.

Beispiel 7.68. *Es ist $f \in o(1)$ genau dann, wenn f eine Nullfolge ist, denn $\forall \varepsilon > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| \leq \varepsilon$ ist äquivalent zu $\forall \varepsilon' > 0\colon \exists n_0 \in \mathbb{N}\colon \forall n \geq n_0\colon |f(n)| < \varepsilon'$: In der Hinrichtung setzen wir $\varepsilon = \varepsilon'/2$ und in der Rückrichtung $\varepsilon' = \varepsilon$.* △

Analog zu Beobachtung 7.60 können wir die Symbole aus Definition 7.66 mit Grenzwerten in Verbindung bringen. Im Gegensatz zu Beobachtung 7.60 sind die Grenzwertbedingungen hier nicht nur hinreichend, sondern auch notwendig.

Beobachtung 7.69. Seien $f, g\colon \mathbb{N} \to \mathbb{R}$ und $n_0 \in \mathbb{N}$ mit $f(n) \geq 0$ und $g(n) > 0$ für alle $n \geq n_0$. Dann gilt:

- Es ist $\lim_{n\to\infty} f(n)/g(n) = 0$ genau dann, wenn $f \in o(g)$.
- Es ist $\lim_{n\to\infty} f(n)/g(n) = \infty$ genau dann, wenn $f \in \omega(g)$.

Beweis. Zur Aussage über $f \in o(g)$: Division durch $g(n) > 0$ macht aus $f(n) \leq \varepsilon g(n)$ die Abschätzung $0 \leq f(n)/g(n) \leq \varepsilon$. Aus $f \in o(g)$ erhalten wir zu einer vorgegebenen Fehlertoleranz $\varepsilon > 0$ die strengere Schranke zum Faktor $\varepsilon/2$, also Konvergenz gegen null. Umgekehrt liefert die Konvergenz die verlangte Schranke für jedes $\varepsilon > 0$.

Zur Aussage über $f \in \omega(g)$: Die untere Schranke wird zu $f(n)/g(n) \geq c$. Aus $f \in \omega(g)$ folgt für jede Schranke $s > 0$ mit $c = 2s$, dass der Quotient ab einem Index größer als s ist. Umgekehrt liefert bestimmte Divergenz gegen ∞ für jeden Faktor $c > 0$ sogar $f(n)/g(n) > c$. □

Eine Warnung: Es liegt nahe zu vermuten, dass die Menge $o(g)$ der echt langsamer als g wachsenden Folgen genau die Folgen in $\mathcal{O}(g) \setminus \Theta(g)$ enthält, also die, die nicht schneller und nicht genauso schnell wie g wachsen. Das ist allerdings im Allgemeinen nicht der Fall (siehe Übung). Wir zeigen, dass zumindest eine Inklusion gilt.

Beobachtung 7.70. Ist $g(n) > 0$ für alle n ab einem Index, so gilt

$$o(g) \subseteq \mathcal{O}(g) \setminus \Theta(g), \qquad \omega(g) \subseteq \Omega(g) \setminus \Theta(g).$$

Beweis. Aus $f \in o(g)$ folgt mit $\varepsilon = 1$ zunächst $f \in \mathcal{O}(g)$. Gäbe es zugleich eine untere Schranke $|f(n)| \geq cg(n)$ mit $c > 0$, so widerspräche die obere Schranke zu $\varepsilon = c/2$ dieser für alle hinreichend großen Indizes. Also ist $f \notin \Theta(g)$. Für $f \in \omega(g)$ folgt mit dem Faktor 1 die Zugehörigkeit zu $\Omega(g)$; jede obere Schranke $|f(n)| \leq dg(n)$ wird durch die untere Schranke zum Faktor $2d$ ausgeschlossen. □

Rechnen mit Landau-Symbolen

Um längere Ausdrücke zu vergleichen, brauchen wir Regeln, mit denen wir bereits bekannte Abschätzungen kombinieren können. Summen und Produkte von Folgen verstehen wir dabei punktweise, also $(f + h)(n) \coloneqq f(n) + h(n)$ und $(fg)(n) \coloneqq f(n) \cdot g(n)$.

Satz 7.71. Sei $\Xi \in \{\mathcal{O}, \Omega, \Theta, o, \omega\}$. Für reelle Folgen gelten die folgenden Regeln:

(i) Aus $f \in \Xi(g)$ und $g \in \Xi(h)$ folgt $f \in \Xi(h)$.

(ii) Aus $f, h \in \Xi(g)$ mit $f, h \geq 0$ folgt $\alpha f + \beta h \in \Xi(\gamma g)$ für alle $\alpha, \beta, \gamma > 0$.

(iii) Aus $f_1 \in \Xi(g_1)$ und $f_2 \in \Xi(g_2)$ folgt $f_1 f_2 \in \Xi(g_1 g_2)$.

Beweis. Übung. □

Für den Vergleich von Wachstum und Fehlerabnahme benötigen wir typische Vergleichsfolgen. Der folgende Satz ordnet sie nach ihrer asymptotischen Größe.

Satz 7.72. Wir schreiben $f \prec g$ für $f \in o(g)$. Für $a, b \in \mathbb{Q}$ mit $0 < a < b$ und $\alpha, \beta \in \mathbb{R}$ mit $1 < \alpha < \beta$ gilt

$$\beta^{-n} \prec \alpha^{-n} \prec n^{-b} \prec n^{-a} \prec 1 \prec n^a \prec n^b \prec \alpha^n \prec \beta^n \prec n! \prec n^n.$$

Beweis. Übung. □

Die Größenordnungen $\Theta(n)$, $\Theta(n^2)$ und $\Theta(n^3)$ nennen wir *linear*, *quadratisch* beziehungsweise *kubisch*. Eine obere Schranke $\mathcal{O}(n^k)$ mit festem $k \in \mathbb{N}$ heißt *polynomiell*. Eine Größenordnung $\Theta(c^n)$ mit $c > 1$ heißt *exponentiell*. Der letzte Satz zeigt, dass jede feste Potenz gegenüber jeder solchen Exponentialfolge asymptotisch verschwindet.

Beispiel 7.73. *Die Vergleiche beschreiben auch die Abnahme von Fehlern. Beim Lastenausgleich aus Beispiel 7.26 gilt $|u_n - v_n| = |u_0 - v_0|\, 3^{-n} \in \mathcal{O}(3^{-n})$. Für ungleiche Anfangslasten liegt dieser Fehler sogar in $\Theta(3^{-n})$. Nach Satz 7.72 ist $3^{-n} \in o(2^{-n})$: Der Fehler klingt also schneller ab als die Vergleichsfolge $(2^{-n})_{n\in\mathbb{N}}$.* △

Für Umformungen mit mehreren Schritten ist es üblich, mit Landau-Symbolen zu rechnen, wobei Verknüpfungen von Folgenmengen als elementweise Verknüpfungen zwischen den enthaltenen Folgen verstanden werden. Dabei wird das Symbol „=“ als (unsymmetrischer) Mengenvergleich genutzt. Beispielsweise bedeutet

$$n^2 + 4n = n^2 + \mathcal{O}(n) = n^2 + o(n^2)$$

die Aussage

$$n^2 + 4n \in n^2 + \mathcal{O}(n) \subseteq n^2 + o(n^2).$$

Wir halten diese Schreibweise als Definition fest.

Definition 7.74. Seien $\mathcal{A}, \mathcal{B}$ Mengen reeller Folgen, $f\colon \mathbb{N} \to \mathbb{R}$ eine reelle Folge und $\star\colon D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^2$ eine Verknüpfung. Wir wenden die Verknüpfung punktweise auf Folgen an, also $(u \star v)(n) := u(n) \star v(n)$. Dabei berücksichtigen wir nur Kombinationen, für die der Ausdruck ab einem Index definiert ist, und ergänzen fehlende Anfangsglieder beliebig. Wir schreiben

$$\begin{aligned} \mathcal{A} \star \mathcal{B} &:= \{u \star v \mid u \in \mathcal{A},\ v \in \mathcal{B}\}, \\ f \star \mathcal{A} &:= \{f \star u \mid u \in \mathcal{A}\}, \\ \mathcal{A} \star f &:= \{u \star f \mid u \in \mathcal{A}\}. \end{aligned}$$

Insbesondere können $\mathcal{A}$ und $\mathcal{B}$ durch Landau-Symbole oder bereits zusammengesetzte Ausdrücke gegeben sein. Für Ausdrücke mit Landau-Symbolen vereinbaren wir die Schreibweisen

$$\begin{aligned} f = \mathcal{A} \quad &\text{bedeutet} \quad f \in \mathcal{A}, \\ \mathcal{A} = \mathcal{B} \quad &\text{bedeutet} \quad \mathcal{A} \subseteq \mathcal{B}. \end{aligned}$$

Beispiel 7.75. *Für die Anzahl der Vektoradditionen des Gram-Schmidt-Verfahrens können wir nun schreiben (siehe Beispiel 7.61)*

$$t(n) = \frac{1}{2}n^2 - \frac{1}{2}n = \frac{1}{2}n^2 + \mathcal{O}(n) = \frac{1}{2}n^2 + o(n^2).$$

Hier ist der konkrete Rest $-n/2$ für alle $n \geq 1$ negativ. Wegen $|-n/2| \leq n$ für alle $n \in \mathbb{N}$ gehört er zu $\mathcal{O}(n)$. Die Beträge in der Definition erfassen die Größe des Restes unabhängig von seinem Vorzeichen. △

Bemerkung 7.76. Gelegentlich schreibt man $f \sim g$ für $f = g + o(g)$. Diese Beziehung ist eine Äquivalenzrelation. Insbesondere gilt $f \sim g$, wenn $\lim_{n\to\infty} f(n)/g(n) = 1$.